Hidden kinetic traps in multidomain folding highlight the presence of a misfolded but functionally competent intermediate

Candice Gautier, Francesca Troilo, Florence Cordier, Francesca Malagrinò, Angelo Toto, Lorenzo Visconti, Yanlei Zhu, Maurizio Brunori, Nicolas Wolff, Stefano Gianni

Hidden kinetic traps in multidomain folding highlight the presence of a misfolded, but functionally competent, intermediate

Candice Gautier[1,*], Francesca Troilo[1,*], Florence Cordier[2], Francesca Malagrinò[1], Angelo Toto[1], Lorenzo Visconti[1], Yanlei Zhu[3], Maurizio Brunori[1], Nicolas Wolff[3] and Stefano Gianni[1,§]

*These authors equally contributed to this work

[1] Istituto Pasteur - Fondazione Cenci Bolognetti, Dipartimento di Scienze Biochimiche "A. Rossi Fanelli" and Istituto di Biologia e Patologia Molecolari del CNR, Sapienza Università di Roma, 00185, Rome, Italy

[2] Structural Bioinformatics Unit, Department of Structural Biology and Chemistry, Institut Pasteur, CNRS UMR3528, Paris, France.

[3] Channel-Receptors Unit, Institut Pasteur, CNRS UMR3571, Paris, France.

§Corresponding author: Stefano.gianni@uniroma1.it

CLASSIFICATION

Biological Sciences - Biochemistry

KEYWORDS

Folding, Kinetics, Misfolding

AUTHOR CONTRIBUTIONS

‡These authors contributed equally.

All authors have given approval to the final version of the manuscript.

ABSTRACT

Despite over 75% of the proteome is composed of multidomain proteins, current knowledge of protein folding is primarily based on studies on isolated domains. In this work, we provide the description of the folding mechanism of a multi-domain tandem construct, comprising two distinct covalently bound PDZ domains, belonging to a protein called Whirlin, a scaffolding protein of the hearing apparatus. In particular, by using a synergy between NMR and kinetic experiments, we demonstrate the presence of a misfolded intermediate that competes with productive folding. In agreement with the view that tandem domain swapping is a potential source of transient misfolding, we demonstrate that such kinetic trap retains a native-like functional activity, as probed by the preserved capability to bind its physiological ligand. Thus, despite the general knowledge of protein misfolding is intimately associated to dysfunction and diseases, we provide a direct example of a functionally competent misfolded state. Remarkably, a bioinformatic analysis of the amino acidic sequence of Whirlin from different species suggests that the tendency to perform tandem domain swapping between PDZ1 and PDZ2 is highly conserved, as mirrored by their unexpectedly high sequence identity. On the basis of these observations, we discuss on a possible physiological role of such misfolded intermediate.

Significance Statement

Much of our current knowledge on protein folding is based on work focussed on isolated domains. In this work, by using a combination of NMR and kinetic experiments, we successfully depict the folding pathway of a multidomain construct comprising two PDZ domains in tandem, belonging to the protein Whirlin. We demonstrate the presence of a misfolded intermediate that competes with productive folding. Interestingly, we show that this misfolded state unexpectedly retains the native-like functional ability to bind its physiological ligand, representing a clear example of a functionally competent misfolded state. On the basis of these results and a comparative analysis of the amino acidic sequences of Whirlin from different species, we propose a possible physiological role of the misfolded intermediate.

INTRODUCTION

Despite over 75% of the eukaryotic proteome is composed by multi-domain proteins (1), much of our current knowledge on protein folding relies on studies on single domain proteins. In fact, protein domains are generally assumed to be able to fold independently, as proven by the possibility to express them in isolation. However, it cannot be excluded that the interaction between these structural subunits may play a critical role in the formation of the native structure, as well as in dictating misfolding events (2, 3). Consequently, recent years have seen a considerable growing effort in developing methods to understand the folding of more complex multi-domain systems (2, 4–12).

One of the simplest mechanisms whereby proteins increase their complexity lies in the duplication of their structural subunits (13). Thus, some proteins may be characterized by repetition of contiguous homologous domains that are often found in tandem. Whilst tandem repeats are frequently found in the proteome, a clear relationship between the presence of these repetitions and their role in the mechanical and functional properties of proteins is still relatively elusive (14–16).

By considering their folding mechanism, it has been shown that the presence of tandem repeats may complicate folding substantially (4, 8, 11, 12). In fact, the concurrent unfolding of contiguous domains can represent an additional potential source of misfolding. In this scenario, misfolded intermediates may be stabilized by interactions involving two adjacent domains, thus competing with productive folding and resulting in an increased complexity of the observed pathway.

Whirlin is a large scaffolding protein involved in the hearing apparatus where it participates to the transduction of sound-induced vibrations in the cochlea into electric potentials, which are then transmitted to the brain (17, 18). Whirlin is located within the actin filled stereocilia of the hair cells and its physiological role is exerted by mechanically transmitting, in complex with other proteins, the concerted deflection of these cilia as induced by sound waves, opening

mechanotransduction channels of the hair bundle. The genes encoding these proteins are affected by mutations responsible for a hereditary sensory disease called Usher syndrome, that associate deafness and progressive blindness (19). The C-terminal PDZ binding motif of Sans, a scaffolding protein mediating several protein-protein interactions and also containing ankyrin repeats and a SAM domain, interacts with the N-terminal region of Whirlin comprising two PDZ domains, namely PDZ1 and PDZ2 (20, 21). One mutation responsible for Usher syndrome results in a Whirlin with truncated PDZ domains (17, 19). Although the presence of two PDZ domains is almost universally conserved in Whirlin in different species, only PDZ1 interacts with Sans and other partners and no specific physiological partners for PDZ2 could be found. The structure of the PDZ1-PDZ2 tandem of Whirlin (P1-P2) has been recently addressed (22). By employing a combination of solution NMR and SAXS, it was observed that P1-P2 behaves as a highly dynamic supramodule where the two individual domains transiently interact each other and populate a well defined closed formation, as well as a heterogeneous ensemble of more open conformations. This dynamic arrangement of the PDZ1-PDZ2 supramodule regulates its binding activity.

To shed light on multistate folding, we resorted to investigate the pathway of folding of PDZ1-PDZ2. As detailed below, the fortuitous differences in stabilities between PDZ1 and PDZ2 in the tandem P1-P2 makes this construct as an ideal system to study multidomain folding, allowing to selectively unfold either one domain (PDZ1, being the least stable) or both with increasing denaturant concentration. We demonstrate that, whilst the (un)folding of PDZ1 is essentially unaffected when PDZ2 is held in its native conformation, the concurrent denaturation of both domains leads to the accumulation of a misfolded kinetic trap that competes with native folding and substantially slows down the productive pathway. In agreement with the view that tandem domain swapping is a potential source of transient misfolding (4, 8), we demonstrate that the kinetic trap retains a native-like functional activity, as probed by the preserved capability to bind a peptide mimicking Sans. Thus, while protein misfolding is classically associated with highly debilitating medical conditions (23–26), we provide the a direct example of a functionally competent misfolded

state. Moreover, a bioinformatic analysis of the amino acidic sequence of Whirlin from different species suggests that the tendency to perform tandem domain swapping between PDZ1 and PDZ2 is highly conserved, as mirrored by their unexpectedly high sequence identity. On the basis of these observations, we discuss on a possible physiological role of such misfolded intermediate.

RESULTS

Given that the binding of Whirlin to its physiological ligand Sans is mediated by its PDZ1 domain, in order to make our folding and binding studies possible, we resorted to engineer a fluorescent variant of PDZ1, which could be employed for both folding and binding experiments. In fact, as detailed below, whilst the wild type construct, previously used for structural characterization by NMR (22), contains two Trp residues in position 237 and 375, located in the hairpin connecting PDZ1 with PDZ2 and in a C-terminal tail of the protein respectively, these residues do not return a detectable fluorescence change upon unfolding, preventing a kinetic characterization. Therefore, based on our previous experiments on the (un)folding of PDZ domains (27–31), position Tyr 168 in PDZ1 was mutated into Trp. Trp168 proved to be an excellent probe for both folding and binding studies. The variant Y168W, denoted below as pP1-P2, was therefore employed as a fluorescent pseudo-wild-type protein in all our experiments.

Equilibrium denaturation of P1-P2. Evidence for sequential independent unfolding of the two individual domains.

The GdnHCl induced equilibrium denaturation of pP1-P2 monitored by fluorescence and CD is depicted in Figure 1. A qualitative comparison between the equilibrium transitions obtained using the two different spectroscopic techniques highlights a clear difference, with the data obtained

by CD returning a broader transition. A two-state analysis of the fluorescence data displays a denaturation midpoint of 0.96 ± 0.02 M. The calculated $m_{D\text{-}N}$ value, which reflects the change in surface area becoming accessible to the solvent upon unfolding (32), is 2.53 ± 0.07 kcal mol^{-1} M^{-1}, consistent with the value expected for an individual PDZ domain. Conversely, the apparent $m_{D\text{-}N}$ value obtained from CD experiments was 0.74 ± 0.03 kcal mol^{-1} M^{-1}, indicating that the transition observed by CD is characterized by an inherent complexity, possibly detecting the unfolding of the two individual PDZ domains of the construct. Importantly, whilst no detectable change could be observed upon unfolding on wild-type P1-P2, the CD transition of pP1-P2 and P1-P2 were super-posable (Figure 1), indicating that mutation of Tyr 168 to Trp had a negligible effect on the stability of the construct.

In order to interpret the data obtained by CD, we resorted to compare the denaturation curves obtained for pP1-P2 with those obtained for PDZ1 and PDZ2, expressed and purified in isolation (Figure 1C). Inspection of Figure 1 indicates that in both cases, equilibrium unfolding was consistent with an apparent two state transition, returning similar values of $m_{D\text{-}N}$, being 2.33 ± 0.09 and 2.61 ± 0.12 kcal mol^{-1} M^{-1} for PDZ1 and PDZ2 respectively. Furthermore, a global fit of the equilibrium unfolding obtained for the tandem pP1-P2, with shared thermodynamic parameters, indicates that denaturation transition of this construct followed by CD are perfectly consistent with the values expected from the sum of the individual curves obtained for PDZ1 and PDZ2 in isolation. This finding suggests that the two domains of the construct pP1-P2 unfold as independent units, and there is no evident interactions among them; the two behaving as the same domains in isolation.

From the data reported in Figure 1 it is possible to identify a GdnHCl concentration at which only PDZ1 is denatured, whereas PDZ2 should maintain its native conformation, i.e. 2.2 M. To further support this observation, we monitored the chemical denaturation of pP1-P2 by NMR. $^{1}H–^{15}N$ heteronuclear single quantum correlation (HSQC) spectra were recorded on ^{15}N-labeled

pP1-P2 at increasing GdnHCl concentration, ranging from 0 to 3.6 M (Figure 2A). The addition of GdnHCl induces overall changes of the pP1-P2 spectrum, with the progressive disappearance of the well-dispersed folded resonances of pP1-P2 (black spectrum) and the appearance of a fully unfolded state with ^{1}H chemical shifts around 8-8.5 ppm (light blue). Due to the non-trivial dependence of chemical shifts on denaturant concentration (33), we chose to monitor pP1-P2 unfolding by studying the changes in intensities of the non-ambiguous and isolated peaks. As illustrated in Figure 2B, peak intensities in PDZ1 decrease at a lower GdnHCl concentration than in PDZ2, indicating earlier unfolding of PDZ1. From the decay of average peak intensities in each PDZ, PDZ1 is totally unfolded at 2.2 M GdnHCl and PDZ2 at 3.5 M (Figure 2C), consistent with fluorescence and CD measurements. Of additional interest, it is worth noticing that, whilst the denaturation of PDZ1 conforms to a two state transition when monitored by NMR, in the case of PDZ2 we could observe an additional complexity. In particular, the calculated average of PDZ2 displays a broader transition. This effect may arise from the unfolding of the previously described supertertiary structure of the tandem (22).

The folding mechanism of PDZ1 in tandem to native PDZ2 is marginally perturbed.

As highlighted above, the difference in stability between the two domains in the pP1-P2 construct allows to selectively unfold PDZ1 or both domains, at different denaturant concentrations. Therefore, pP1-P2 offers the unique opportunity to compare the folding of PDZ1 in the presence and in the absence of a flanking folded or unfolded domain. To this purpose, we first characterized the folding and unfolding of PDZ1 in isolation and then compared it to that observed in the pP1-P2 construct. Unfolding and refolding kinetics of PDZ1 were monitored by stopped-flow kinetics, by rapidly diluting the protein into solutions containing different GdnHCl concentrations. In all experiments, the time course of the emission fluorescence was satisfactorily fitted to a single-exponential equation. The measured chevron plot of PDZ1 is reported in Figure 3A. Analysis of the

data is consistent with a simple two-state V-shaped chevron, with no clear deviation from linearity in the folding and unfolding branches. Furthermore, both the $m_{D\text{-}N}$ value and the calculated midpoint are consistent with the same parameters obtained from equilibrium experiments, confirming the two-state nature of the reaction (34). To measure the folding kinetics of PDZ1 in the pP1-P2 construct in the presence of native PDZ2, we first denatured pP1-P2 in 2.2 M GdnHCl and then triggered refolding by rapid mixing with buffer. Analogously, unfolding was initiated by mixing the native protein with solutions containing GdnHCl at different concentrations. In analogy to what observed with PDZ1 in isolation, folding and unfolding time courses were both fitted satisfactorily to a single exponential decay at any final denaturant concentration. A comparison between the chevron plots of PDZ1 in isolation and in the tandem is reported in Figure 3A. It is evident that, when PDZ2 is held in its native conformation, the folding of PDZ1 is essentially unaffected by the presence of a neighboring domain, with both folding and unfolding rate constants being unaltered.

A hidden kinetic trap in the folding of pP1-P2.

The experiments in Figures 1, 2 and 3 demonstrate that, when pP1-P2 is denatured at mild denaturant concentrations, PDZ2 retains its native conformation and its presence has a marginal effect on the folding and stability of PDZ1. To investigate further the effect of the presence of a neighboring domain on PDZ1, we conducted additional refolding experiments in which pP1-P2 was denatured at high denaturant concentrations, i.e. where both PDZ domains were denatured. The refolding kinetics of pP1-P2 measured in these experiments are reported in Figure 3B. It is evident that, when both PDZ domains are denatured, productive folding at low denaturant concentration is remarkably slowed down, as mirrored by the presence of a pronounced roll-over (35). Moreover, the refolding rate constants increase with increasing denaturant concentrations, indicating that the intermediate is more structured than the transition state and, therefore, it has to unfold for

productive folding to occur (36, 37). On the basis of these observations we conclude that the concurrent denaturation of both PDZ domains causes the rapid accumulation of a kinetic trap, which competes with productive folding. Of additional interest, it may be noticed that the apparent slope of the refolding rate constant, at low denaturant concentrations, appears steeper that that of the unfolding branch by nearly 2-fold. This finding indicates that the misfolded intermediate may contain more folded structure that of a single PDZ domain. This hypothesis is further described below.

To further highlight the presence of an intermediate in the folding of pP1-P2 we repeated folding and unfolding experiments in the presence of a stabilizing agent, i.e. 0.5 M sodium sulfate. Figure 4A compares the folding rate constants obtained in stabilizing conditions when pP1-P2 is denatured in mild or high denaturant concentrations. In agreement with our previous observations on other PDZ domains (31, 36), the presence of sodium sulfate emphasizes the presence of a misfolded kinetic trap and, at low denaturant concentration, folding is slowed down by over two orders of magnitudes. To exclude the presence of intermolecular interactions, refolding experiments were also conducted at different protein concentrations, varying from 250 nM to 15 μM. As expected from a monomocular folding reaction, rate constants were found to be insensitive on protein concentration (Figure 4A). Inspection of data in Figure 4 also reveals that the folding kinetics starting from 2.2 M GndHCl in the presence of sodium sulfate display a pronounced roll-over, being therefore inconsistent with a two-state mechanism even when misfolding is avoided. A likely interpretation of this effect is that the salt stabilizes a marginally stable folding intermediate, which is barely detectable at low salt, but well populated in the presence of sulfate.

The misfolded intermediate of pP1-P2 retains the ability to bind the physiological partner of PDZ1

To address the functional properties of the intermediate observed in pP1-P2 folding, we designed a double jump stopped-flow experiment. In detail, the fully denatured pP1-P2 (i.e. in the

presence of 5.37 M GdnHCl) was first mixed with the refolding buffer in the presence of 0.5 M sodium sulfate, to populate the misfolded intermediate. Then, in a second mix, the solution was challenged with different concentrations of a peptide mimicking the C-terminal sequence of Sans, the physiological partner of Whirlin (21). The delay time between the first and the second mixing was held below 2 s. In fact, since at these experimental conditions the formation of the native state of pP1-P2 displays a rate constant of 0.04 ± 0.002 s^{-1}, any faster binding events to the Sans peptide might be assigned to the misfolded intermediate. Importantly, it was previously shown that only PDZ1 binds Sans and no physiological partners has been identified for PDZ2 (22).

In analogy to previous work on other PDZ domains (38–40), binding was followed by monitoring FRET between the tryptophan, acting as a donor, and a dansyl group covalently attached to the N-terminus of the peptide, acting as acceptor. Under all the investigated conditions, binding time courses were consistent with a single exponential decay. The pseudo-first order plot of the binding experiment obtained by double jump is reported in Figure 4B. Surprisingly, we observe that the intermediate binds Sans with a similar, but not identical, affinity to that of native pP1-P2, with an apparent k_{on} of 0.030 ± 0.004 s^{-1} μM^{-1} and a k_{off} of 23.6 ± 0.7 s^{-1} versus the values of 0.07 ± 0.01 s^{-1} μM^{-1} and 24 ± 1 s^{-1} measured for native pP1-P2 at the same experimental conditions. On the basis of these observations, we conclude that, whilst misfolded, the intermediate in pP1-P2 retains its ability to bind the physiological ligand of PDZ1.

The double jump experiment reported above relies on the assumption that only PDZ1, and not PDZ2, is capable of binding to the peptide. Therefore, we tested the binding of isolated PDZ2 against the peptide and could not observe any transition. Likewise, no binding could be detected when pP1P2 was challenged with the peptide at mildly denaturing conditions, i.e. 2.2 M GdnHCl where PDZ1 is essentially denatured whereas PDZ2 is still folded. To provide an additional control in this set of experiments, we also performed an additional double jump experiment in which pP1P2, kept at mildly denaturing conditions, was mixed with refolding buffer to the same the final

denaturant concentration of the double jump experiment reported above; then, after a controlled delay time, the mixture was mixed with Sans peptide. Importantly, at a peptide concentration of 200 μM and long delay times, we observed a rate constant of 45 s^{-1}, consistent with the value measured for the native state. Conversely, at very short delay times, we observed a binding rate constant of 10 s^{-1}, which is in agreement with the folding rate constant observed in the refolding experiments starting from the mildly denaturing conditions. Importantly, also the dependence of the observed amplitude of the fast binding phase as a function of the delay time between the first and the second mix returned a rate constant of about 10 s^{-1} (Figure 4C). This finding indicates that, when the misfolded intermediate does not accumulate, only the fully folded pP1P2 can bind the peptide.

An unexpectedly high sequence identity between PDZ1 and PDZ2

A plausible scenario whereby tandem repeats may form metastable kinetic traps involves the formation of domain swapped intermediates (4, 8, 11). Domain swapping occurs when two neighboring domains form intertwined interactions, thereby reciprocally exchanging structural subunits. In these conditions, inter-domain rather then intra-domain interactions may satisfy the structural architecture of the native domain topology. It has been previously shown that tandem domain swapping is promoted when the sequence identity of continuous domains is high (12) . Furthermore, a large bioinformatic analysis has demonstrated that consecutive homologous domains almost exclusively have sequence identities of less than 40%, such that misfolding is minimized (12).

We analyzed the sequence identity of the PDZ domains in Whirlin from 70 species, as obtained from UniProt database. In particular, we compared the pairwise sequence identities between the contiguous PDZ1 and PDZ2, found in the pP1-P2 construct, to those of a third PDZ domain, PDZ3, which is located downstream in the sequence of Whirlin. Contrary to expectations,

we found that, the contiguous PDZ1 and PDZ2 have an unusually high level of sequence identity, ranging from 34 to 53%, whilst PDZ1 displays a lower sequence identity to PDZ3 (Figure 5).

Remarkably, nearly all pairwise sequence identities between PDZ1 and PDZ2 in Whirlin were found to be higher than 40% and about 80% of the sequences display a pairwise identity higher than 45%. This finding suggests that these two domains might be under an evolutionary pressure to maintain their unusual similarities.

DISCUSSION

The general knowledge on protein misfolding is intimately linked to dysfunction and disease. In fact, since decades it is well established how the aberrant incorrect structures are the recurrent trigger of protein aggregation leading to several diseases spanning from neurological disorders, such Alzheimer's or Parkinson's diseases, to systemic amyloidosis (23–25). Consequently, Nature has evolved several systems to solve incorrect protein folding or to degrade irreversibly misfolded polypeptides.

A peculiar protein misfolding event is represented by tandem domain swapping (4, 8) . In this cases, if the increasing complexity of protein architecture demands the presence of contiguous homologuous domains, the concurrent unfolding of two neighbouring subunits may induce intertwined interactions. This phenomenon causes two adjacent domains to form a 'central domain', equivalent to a circular permutation, formed from the central region of the two-domain sequence, and a 'terminal domain' comprising the remaining parts of the sequence. Notably, by following this scenario, a misfolded intermediate may appear to contain more folded structure than that expected when studying a single domain, as in the case of pP1-P2.

Studies on titin and immunoglobulin domains have shown that evolution tends to minimize domain swapped misfolding by lowering the sequence identity of adjacent domains, which is almost

universally less than 40% in tandem repeat proteins (8, 11, 12). In this context, it is of interest to notice that all the sequences of Whirlin in the UniProt database display unusually high sequence identities between their adjacent PDZ1 and PDZ2. Therefore, also by considering that PDZ domains are naturally prone to circular permutation (36), domain swapping (41) and that previous protein engineering experiments suggested that PDZ domains may be successfully expressed as circularly permutated variants (42,43), it is somewhat surprising to find that the tendency to perform domain swapping might be evolutionary conserved in Whirlin. Can the misfolded intermediate play some physiological role in the function of Whirlin?

Mechanical proteins are exposed to tensile forces that might unfold them and lead to the accumulation of long-lived states displaying multiple unfolded domains (12). In these conditions, there could be a tendency to form aberrant structures leading to a loss of function. In the case of Whirlin, the anchoring to Sans protein via its PDZ1 domain, guarantees the mechanical communication from the surface of the stereocilia to the actin filaments, ensuring the hearing of sound waves. It is intriguing that the presence of two PDZ domains is almost universally conserved in Whirlin (17, 21), and yet no physiological partners could be found for PDZ2 (22, 44). Thus, on the basis of our findings, we speculate that the second PDZ domain, enhancing the possibility to accumulate a functionally competent domain swapped state, may represent a structural buffer to increase the repertoire of structures to conserve the anchoring to Sans, even in the presence of mechanical unfolding.

In summary, our work highlights an example of a functionally competent misfolded intermediate. In spite of expectations, the sequence patters in natural proteins highlight that the tendency to form this metastable state appears to be conserved. Future work on other PDZ containing tandem repeats will shed light on the generality of these findings.

MATERIALS AND METHODS

Site directed mutagenesis

The construct P1-P2 is identical to that previously described (22). The variant pP1-P2 was obtained by substituting the tyrosine residue in position 168 with a tryptophan by using Quickchange Lightning mutagenesis kit. The DNA encoding for the pseudo-wild type fluorescent PDZ1 Y168W (pPDZ1) domain was obtained using the Quickchange Lightning mutagenesis kit (Agilent technologies) to insert a STOP codon in the pP1-P2 sequence downstream the region encoding for PDZ1 domain (residues 140-224). The sequence was confirmed by DNA sequencing. PDZ2 was produced from a DNA library at NZYTech, Lisbon, Portugal. The primers oligos were purchased from Eurofins Genomics. .

Protein expression and purification

DNA encoding for pP1-P2 and pPDZ1 and PDZ2 was transformed into Escherichia coli BL21 (DE3) cells for protein expression. The protein expression was induced with IPTG to a final concentration 100 μg/mL when the optical density at 600 nm reached 0.8, and the cells were incubated for 3 hours at 30 °C. The cells were then cultured after overnight incubation at 20 °C. Each construct and mutant was purified from the soluble fraction in 50 mM NaPho at pH 7.2 with 0.3 mM NaCl with a HiTrap Chelating High Performance column from GE Healthcare, and eluted with a gradient to 1 M imidazole. The imidazole was removed using a HiPrep™ 26/10 Desalting column from GE Healthcare. The purity of the protein sample was confirmed by SDS-PAGE. The peptide mimicking the C-terminal sequence of Sans was purchased from JPT Peptide Technologies.

Equilibrium Experiments

Fluorescence equilibrium (un)folding experiment were performed on a standard spectrofluorometer Fluoromax single photon counting spectrofluorometer (Jobin-Yvon, NJ, USA). The protein pP1-P2 was excited at 280 nm and emission spectra were recorded between 300 and 400 nm, at increasing denaturant concentrations. The experiment was performed at 25°C using a quartz cuvette with a path length of 1 cm. The far-UV CD denaturation spectra were measured using a Jasco 810 dichrograph, flushed with N2 and equipped with a Peltier thermoregulation system. One-mm thick quartz cuvettes were used. Spectra were measured between 200 and 250 nm at 25°C. The scanning speed was 50 nm/min, with data pitch of 0.5 nm. CD ellipticities were selected at a single wavelength (225 nm) and plotted against denaturant concentration. The spectra of single PDZ1 and PDZ2 were analyzed quantitatively following a two-state model. The spectra of pP1P2 was fitted with the sum of two two-stated model.

NMR experiments

^{15}N labelled pP1-P2 was expressed and purified as previously described [42]. The ^{15}N-labeled NMR sample of pP1-P2 was prepared at an initial concentration of 266 μM in 240 μl of a buffer containing 50 mM Sodium phosphate, 150 mM NaCl and 2 % D_2O, pH 3.7. Chemical denaturation of pP1-P2 was followed by GdnHCl titration. Aliquots of a 6 M GdnHCl stock solution were successively added to the pP1-P2 sample. For each GdnHCl point (0.5, 1, 1.5, 2, 2.2, 2.6, 3.1 and 3.6 M), a 2D ^{1}H–^{15}N HSQC spectrum was recorded at 25°C on an 800 MHz Bruker Avance NEO spectrometer equipped with a TCI cryoprobe. From the series of HSQC, the resonance peak intensities were measured for each residue as a function of GdnHCl concentration, using CcpNmr Analysis 2.4 software (45). The intensities were corrected to account for the dilution effect at each titration point and normalized to the intensities without GdnHCl. Assignment of most of the isolated resonances could be deduced from the one at pH 7.5, 6.5 and 5.9.

Kinetic experiments

Kinetic Folding Experiments

Rapid mixing for kinetics folding and unfolding experiments were carried out on a stopped-flow device (Pi-star, Applied Photophysics, Leatherhead, UK) and fluorescence emission of pP1P2 and pPDZ1 was measured with a 320 nm band-pass filter using an excitation wavelength of 280 nm. The kinetic folding experiments were made with 1.5 μM final concentration of protein and denaturant concentrations varying from 0.2 to 5.27M. Refolding experiments of pP1P2 were performed in two different starting conditions: the protein was diluted in mild denaturant concentration (i.e. 2.2M GndHCl) or in strong denaturant concentration (i.e. 5.37 M GndHCl). The temperature was set at 25 °C. Buffers used were: 50 mM tris-HCl pH 7.5, 0.3 M NaCl; and 50 mM tris-HCl pH 7.5, 0.5 M Na_2SO_4. Kinetic traces were fitted with a single exponential decay using Applied Photophysics software to obtain the observed rate constant k_{obs}. The logarithmic values of k_{obs} were plotted versus GndHCl concentration (chevron plot). The traces that did not show a roll-over effect were fitted to a standard two-state chevron.

Due to the complexity of the observed kinetics, observed data were fitted as follows. For the chevron arising from the unfolding and the refolding data starting from mildly denaturing conditions, data were fitted to a three state mechanism involving the accumulation of a transient intermediate (35). Subsequently, the chevron arising from the data at strong denaturing conditions was fitted by applying the analytical solution to an off-pathway scenario, obtained from the eigenvalues of a three-state matrix, as previously described (46). Since only the slow λ2 phase could be observed, for the fitting procedure to converge, we forced the parameters for productive folding to those obtained from the fit at mildly denaturing conditions.

Kinetic binding experiment

Kinetic experiments of binding between pP1-P2 and the peptide mimicking the C-terminal region of Sans was measured by monitoring FRET between the tryptophan, acting as a donor, and a dansyl group covalently attached to the N-terminus of the peptide, acting as aceptor. The

experiment was performed on a single-mixing SX-18 stopped-flow instrument (Applied Photophysics), recording the change of fluorescence emission. The excitation wavelength used was 280 nm while the fluorescence emission was collected using a 320-nm cut-off glass filter. The binding experiments were carried out at pseudo-first order conditions by mixing a constant concentration of pP1-P2 (4 μM) versus increasing concentrations of Sans (ranging from 40 to 200 μM). The buffer used was 50 mM TrisHCl, 0.5 M Na_2SO_4, pH 7.5. The observed rate constants (k_{obs}) were calculated from the average of 3–6 single traces and by fitting of the time-course for binding using a single exponential equation.

Double jump kinetic experiment

Refolding and rapid binding experiments were carried out on an Applied Photophysics stopped-flow instrument with double jump facility (Leatherhead, UK); the excitation wavelength was 280 nm and the fluorescence emission was measured using a 320 nm cut-off glass filter. Refolding and binding were initiated by a symmetric mixing of the denatured and SANS peptide with the appropriate buffer. *First mix.* Unfolded pP1P2 (4 μM) was first mixed against a refolding buffer (0.54 M GdnHCl at 50 mM sodium phosphate pH 7.5, 0.5 M sodium sulfate). Unfolded pP1P2 was obtained by incubation in GdnHCl 5.37 M at pH 7.5. *Second mix.* After 2 s of delay time, binding was initiated by transferring the solution into final conditions of 0.54 M GdnHCl at 50 mM sodium phosphate pH 7.5, 0.5 M sodium sulfate, in presence of varying concentrations of SANS (ranging from 40 to 300 μM). A control experiment, reported as an inset panel in Figure 4B, was also run by using the same conditions described above, but starting the experiment at mildly denaturing conditions, i.e. 2.2 M GdnHCl at 50 mM sodium phosphate pH 7.5, and measuring binding at different delay times between the first and the second mixing event.

Data Availability Statement

The data utilized in this manuscript to justify the results and conclusions of this work are entirely presented within the manuscript.

ACNKOWLEDGEMENTS

The project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement PDZnet No 675341. Work partly supported by grants from the Italian Ministero dell'Istruzione dell'Università e della Ricerca (Progetto di Interesse 'Invecchiamento' to S.G.), Sapienza University of Rome (B52F16003410005, RP11715C34AEAC9B and RM1181641C2C24B9 to S.G), the Associazione Italiana per la Ricerca sul Cancro (Individual Grant - MFAG 2016, 18701 to S.G.) the Istituto Pasteur Italia (Teresa Ariaudo Research Project 2018, to A.T.). F.M. was supported by a fellowship from the Associazione Italiana per la Ricerca sul Cancro. The authors also acknowledge the Conseil Régional d'Ile de France for financial support through the SESAME 2014 NMRCHR program No. 14014526 (800 MHz spectrometer).

REFERENCES

1. S. Batey, J. Clarke, Apparent cooperativity in the folding of multidomain proteins depends on the relative rates of folding of the constituent domains. *Proc. Natl. Acad. Sci. USA* **103**, 18113–18118 (2006).

2. P. Tian, R. B. Best, Structural Determinants of Misfolding in Multidomain Proteins. *PLoS Comput. Biol.* **12**, e1004933 (2016).

3. J. H. Han, S. Batey, A. A. Nickson, S. A. Teichmann, J. Clarke, The folding and evolution of multidomain proteins. *Nat. Rev. Mol. Cell. Biol.* **8**, 319–330 (2007).

4. A. Lafita, P. Tian, R. B. Best, A. Bateman, Tandem domain swapping: determinants of multidomain protein misfolding. *Curr. Opin. Struct. Biol.* **58**, 97–104 (2019).

5. R. Kantaev, *et al.*, Manipulating the Folding Landscape of a Multidomain Protein. *J. Phys. Chem. B* **122**, 11030–11038 (2018).

6. V. Kumar, T. K. Chaudhuri, Spontaneous refolding of the large multidomain protein malate synthase G proceeds through misfolding traps. *J. Biol. Chem.* **293**, 13270–13283 (2018).

7. D. T. Gruszka, *et al.*, Disorder drives cooperative folding in a multidomain protein. *Proc. Natl. Acad. Sci. USA* **113**, 11841–11846 (2016).

8. A. Borgia, *et al.*, Transient misfolding dominates multidomain protein folding. *Nat. Commun.* **6**, 8861 (2015).

9. J. Valle-Orero, *et al.*, The elastic free energy of a tandem modular protein under force. *Biochem. Biophys. Res. Commun.* **460**, 434–438 (2015).

10. T. Inanami, T. P. Terada, M. Sasai, Folding pathway of a multidomain protein depends on its topology of domain connectivity. *Proc. Natl. Acad. Sci. USA* **111**, 15969–15974 (2014).

11. M. B. Borgia, *et al.*, Single-molecule fluorescence reveals sequence-specific misfolding in multidomain proteins. *Nature* **474**, 662–665 (2011).

12. C. F. Wright, S. A. Teichmann, J. Clarke, C. M. Dobson, The importance of sequence diversity in the aggregation and evolution of proteins. *Nature* **438** (2005).

13. A. D. McLachlan, Repeating sequences and gene duplication in proteins. *J. Mol. Biol.* **64**, 417–437 (1972).

14. A. Pena-Francesch, M. C. Demirel, Squid-Inspired Tandem Repeat Proteins: Functional Fibers and Films. *Front. Chem.* **7**, 69 (2019).

15. A. Perez-Riba, M. Synakewicz, L. S. Itzhaki, Folding cooperativity and allosteric function in the tandem-repeat protein class. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **373**, 1749 (2018).

16. B. Kobe, A. V. Kajava, When protein folding is simplified to protein coiling: the continuum of solenoid protein structures. *Trends Biochem. Sci.* **25**, 509–515 (2000).

17. I. Ebermann, *et al.*, A novel gene for Usher syndrome type 2: mutations in the long isoform of whirlin are associated with retinitis pigmentosa and sensorineural hearing loss. *Hum. Genet.* **121**, 203–211 (2007).

18. T. Maerker, *et al.*, A novel Usher protein network at the periciliary reloading point between molecular transport machineries in vertebrate photoreceptor cells. *Hum. Mol. Genet.* **17**, 71–86 (2008).

19. P. D. Mathur, J. Yang, Usher syndrome and non-syndromic deafness: Functions of different whirlin isoforms in the cochlea, vestibular organs, and retina. *Hear Res.* **375**, 14–24 (2019).

20. N. Sorusch, K. Wunderlich, K. Bauss, K. Nagel-Wolfrum, U. Wolfrum, Usher syndrome protein network functions in the retina and their relation to other retinal ciliopathies. *Adv. Exp. Med. Biol.* **801**, 527–533 (2014).

21. N. Sorusch, *et al.*, Characterization of the ternary Usher syndrome SANS/ush2a/whirlin protein complex. *Hum. Mol. Genet.* **26**, 1157–1172 (2017).

22. F. Delhommel, *et al.*, Structural Characterization of Whirlin Reveals an Unexpected and Dynamic Supramodule Conformation of Its PDZ Tandem. *Structure* **25**, 1645–1656 (2017).

23. F. Chiti, C. M. Dobson, Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade. *Annu. Rev. Biochem.* **86**, 27–68 (2017).

24. C. M. Dobson, T. P. Knowles, M. Vendruscolo, The Amyloid Phenomenon and Its Significance in Biology and Medicine. *Cold Spring Harb. Perspect. Biol.*, Apr 1. pii: a033878. doi: 10.1101/cshperspect.a033878. [Epub ahead of print] (2019).

25. T. P. Knowles, M. Vendruscolo, C. M. Dobson, The amyloid state and its association with protein misfolding diseases. *Nat. Rev. Mol. Cell. Biol.* **15**, 384–396 (2014).

26. E. P. O'Brien, J. Christodoulou, M. Vendruscolo, C. M. Dobson, New scenarios of protein folding can occur on the ribosome. *J. Am. Chem. Soc.* **133**, 513–526 (2011).

27. N. Calosci, *et al.*, Comparison of successive transition states for folding reveals alternative early folding pathways of two homologous proteins. *Proc. Natl. Acad. Sci. U S A* **105**, 19241–19246 (2008).

28. C. N. Chi, *et al.*, A conserved folding mechanism for PDZ domains. *FEBS Lett.* **581**, 1109–1113 (2007).

29. G. Hultqvist, *et al.*, An expanded view of the protein folding landscape of PDZ domains. *Biochem. Biophys. Res. Commum.* **421**, 550–553 (2012).

30. Y. Ivarsson, *et al.*, An On-pathway Intermediate in the Folding of a PDZ Domain. *J. Biol. Chem.* **282**, 8568–8572 (2007).

31. Y. Ivarsson, C. Travaglini-Allocatelli, M. Brunori, S. Gianni, Folding and misfolding in a naturally occurring circularly permuted PDZ domain. *J. Biol. Chem.* **283**, 8954–8960 (2008).

32. J. K. Myers, Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci.* **4**, 2138–2148 (1995).

33. K. W. Plaxco, *et al.*, he effects of guanidine hydrochloride on the "random coil" conformations and NMR chemical shifts of the peptide series GGXGG. *J. Biomol. NMR* **10**, 221–230 (1997).

34. S. E. Jackson, Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry* **30**, 10428–10435 (1991).

35. M. J. Parker, J. Spencer, A. R. Clarke, An integrated kinetic analysis of intermediates and transition states in protein folding reactions. *J Mol Biol* **253**, 771–86 (1995).

36. S. Gianni, *et al.*, Structural characterization of a misfolded intermediate populated during the folding process of a PDZ domain. *Nat. Struct. Mol. Biol.* **17**, 1431–1437 (2010).

37. D. E. Otzen, M. Oliveberg, Salt-induced detour through compact regions of the protein folding landscape. *Proc. Natl. Acad. Sci. U S A* **96**, 11746–11751 (1999).

38. C. N. Chi, A. Bach, K. Strømgaard, S. Gianni, P. Jemth, Ligand binding by PDZ domains. *Biofactors* **38**, 338–348 (2012).

39. C. N. Chi, A. Engstrom, S. Gianni, M. Larsson, P. Jemth, Two conserved residues govern the salt and pH dependencies of the binding reac- tion of a PDZ domain. *J. Biol. Chem.* **281**, 36811–36818 (2006).

40. S. Gianni, *et al.*, The kinetics of PDZ domain-ligand interactions and implications for the binding mechanism. *J. Biol. Chem.* **280**, 34805–34812 (2005).

41. A. S. Fanning, M. F. Lye, J. M. Anderson, A. Lavie, Domain swapping within PDZ2 is responsible for dimerization of ZO proteins. *J. Biol. Chem.* **282**, 37710–37716 (2007).

42. Y. Ivarsson, C. Travaglini-Allocatelli, M. Brunori, S. Gianni Engineered Symmetric Connectivity of Secondary Structure Elements Highlights Malleability of Protein Folding Pathways *J. Am. Chem. Soc.* **131**, 11727-11733 (2009)

43. G. Hultqvist, *et al.*, Energetic Pathway Sampling in a Protein Interaction Domain *Structure,* **21,** 1193-1202 (2013)

44. F. Delhommel, N. Wolff, F. Cordier, (1)H, (13)C and (15)N backbone resonance assignments and dynamic properties of the PDZ tandem of Whirlin. *Biomol. NMR Assign.* **10**, 361–365 (2016).

45. W. F. Vranken, *et al.*, The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59**, 687–696 (2005).

46. P. Jemth, et al., Demonstration of a Low-Energy On-Pathway Intermediate in a Fast-Folding Protein by Kinetics, Protein Engineering, and Simulation. *Proc. Natl. Acad. Sci. U S A.* **101**, 6450-6455 (2004)

FIGURE LEGENDS

Figure1. Equilibrium denaturation experiments. Panel A: GdnHCl induced equilibrium denaturation of wild type P1-P2 (open diamonds) and pP1-P2 (black circles) monitored by fluorescence. The change of the intrinsic fluorescence of the tryptophan residue versus GdnHCl concentrations is consistent to a two-state transition. No detectable change could be observed in the case of wild type P1-P2. Panel B: GdnHCl induced equilibrium denaturation of wild type P1-P2 (open diamonds) and pP1-P2 (black circles) pP1-P2 monitored by CD. As discussed in the text, the denaturation obtained by CD experiments does not fit well to a two-state transition, revealing an additional complexity. Panel C: GdnHCl induced equilibrium denaturation of pP1-P2 (black circles), pPDZ1 (empty circles) and PDZ2 (grey circles) monitored by CD. The curves obtained for pPDZ1 and PDZ2 are consistent with a two-state transition. The pP1-P2 denaturation curve was satisfactorily fitted by the sum of the individual curves obtained for pPDZ1 and PDZ2 in isolation..

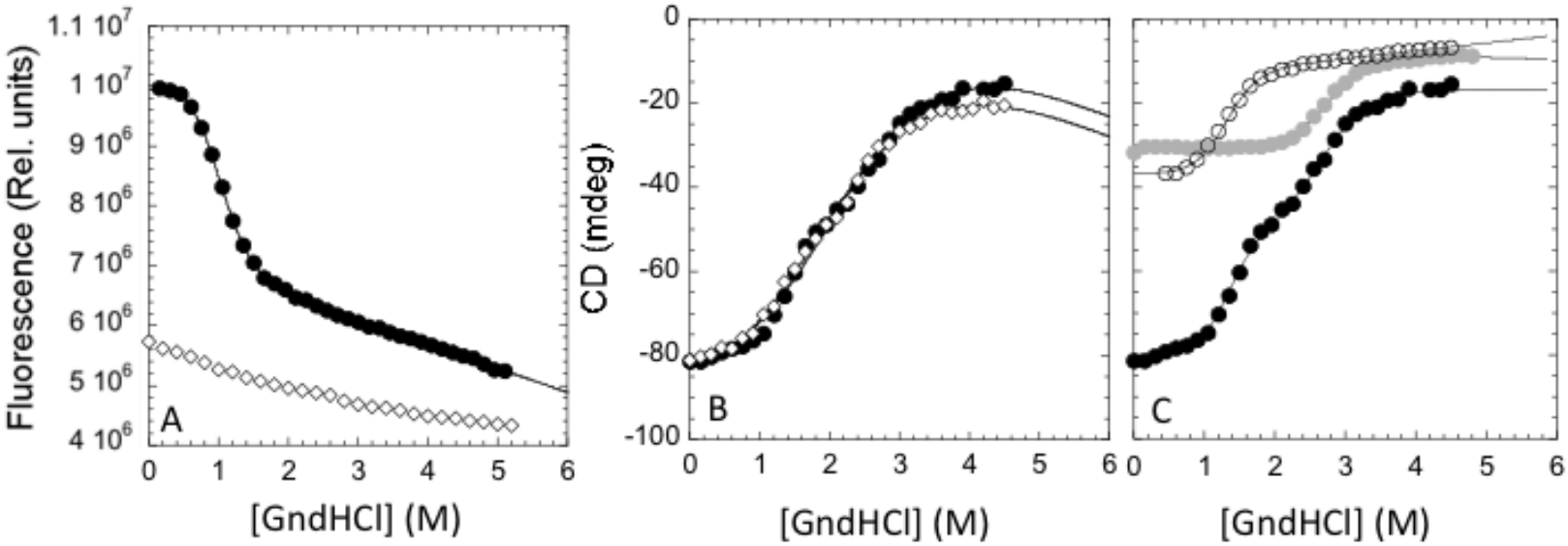
Figure 2: Chemical denaturation of pP1-P2 followed by NMR. (A) Region of ^{1}H-^{15}N HSQC spectra recorded on ^{15}N-labeled pP1-P2 at increasing GdnHCl concentration from 0 (black) to 3.6 M (light blue). Unambiguous assignment of isolated peaks is indicated in red for PDZ1 and blue for PDZ2. (B) Normalized peak intensities for each unambiguously assigned residue plotted at 1, 2, 2.6 and 3.6 M GdnHCl from dark to light blue. Domains of pP1-P2 and secondary structure elements are indicated at the top. (C) Average of the normalized peak intensities plotted as a function of GdnHCl concentration for PDZ1 (left) and PDZ2 (right) residues, highlighting the denaturation midpoints measured for PDZ1 and PDZ2, 1.3 M and 2.7 M GdnHCl, respectively.

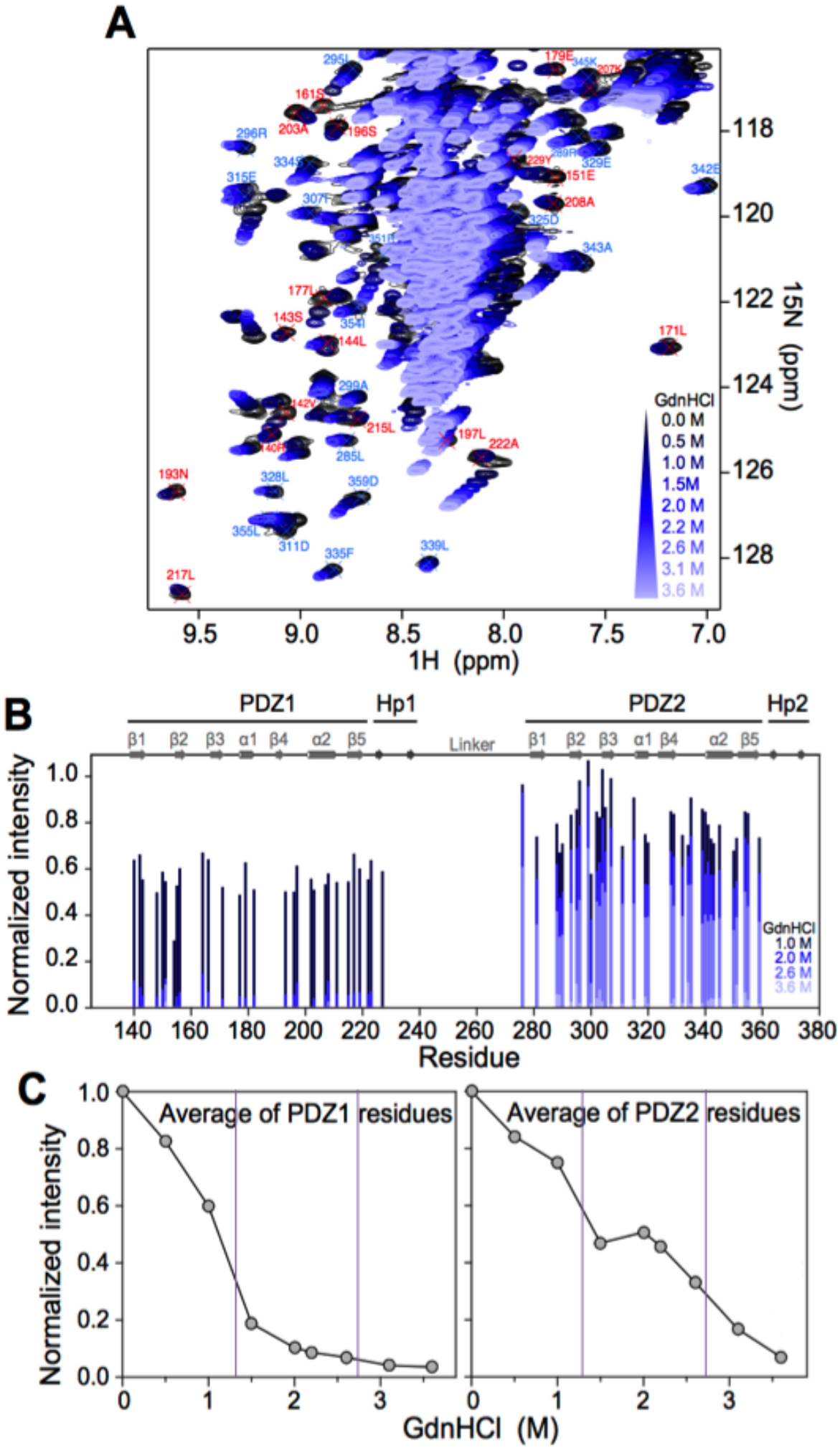
Figure 3. Unfolding and refolding kinetics of pP1-P2 and pPDZ1. Panel A: chevron plot of pPDZ1 (empty circles) and pP1-P2 (black circles) obtained starting the refolding experiment diluting the protein with mild denaturant concentrations (i.e. 2.2 M GdnHCl). The two chevron plots are perfectly superimposable, indicating that, in this condition, the folding of PDZ1 is not affected by the presence of a flanking domain, when PDZ2 is held in its native conformation (see the text for details). Panel B: chevron plot of pPDZ1 (empty circles) and pP1-P2, obtained starting the refolding diluting the protein with mild denaturant concentrations (i.e. 2.2M GdnHCl, black circles) and high denaturant concentrations (i.e. 5.37 M GdnHCl, gray circles). As explained in details in the text, when the refolding starts at higher GdnHCl concentrations, both PDZ1 and PDZ2 domains are denatured leading to the accumulation of a kinetic trap (represented by the presence of a pronounced roll-over in the refolding arm) that competes with the productive folding of the protein. The buffer used for both experiments was 50 mM TrisHCl pH 7.5, 0.3 M NaCl.

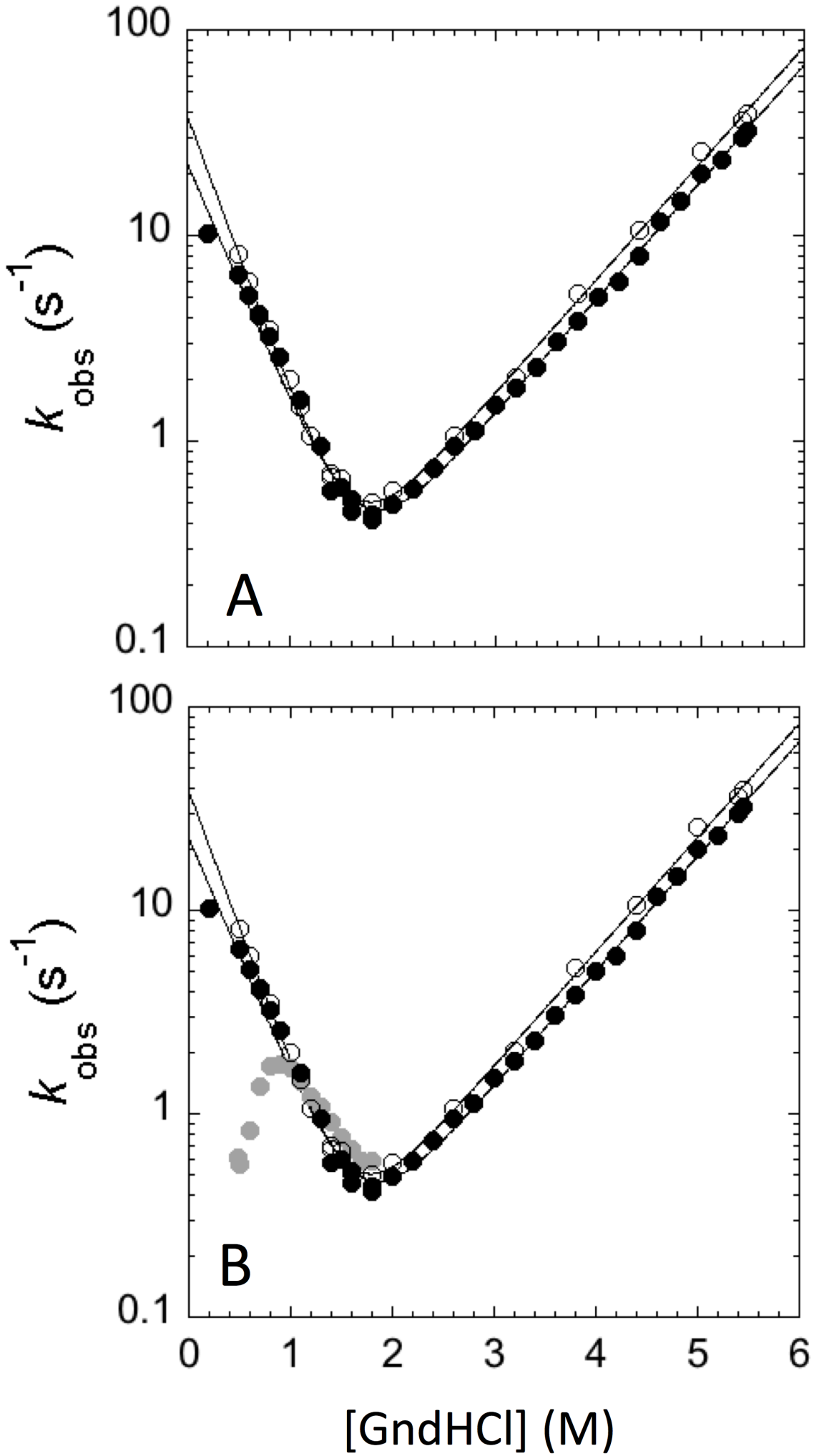
Figure 4. Panel A: Unfolding and refolding kinetics of pP1-P2 in presence of a stabilizing agent (0.5M Na_2SO_4). pP1-P2 was denatured in mild (2.2 M, black circles) or high (5.37M, grey circles) GdnHCl concentrations. The refolding of pP1-P2 in the latter condition is slowed down by over two orders of magnitudes by the presence of Na_2SO_4. Data were fitted as described in the Material and Methods Section. Inset panel. Refolding kinetics of pP1-P2 measured at 0.5 M [GdnHCl], starting from high denaturant concentration, at different protein concentrations, varying from 250 nM to 15 μM. As expected from a monomocular folding reaction, rate constants were found to be insensitive on protein concentration Panel B: Binding experiments measured by FRET between the Trp (donor) and dansyl (acceptor) attached to the Sans peptide. Empty circles represent the pseudo-first order plot of the binding experiment between pP1-P2 and the peptide, performed at 25°C in buffer 50mM TrisHCl 0.5 M Na_2SO_4 0.54 M GdnHCl. The binding obtained between the peptide and the misfolded intermediate of pP1-P2 was obtained by a double jump stopped-flow experiment (black circles). Inset panel. Dependence of the relative amplitude of native-like binding competent species as a function of delay time between the first (refolding) and second (binding) mix. In this case, pP1-

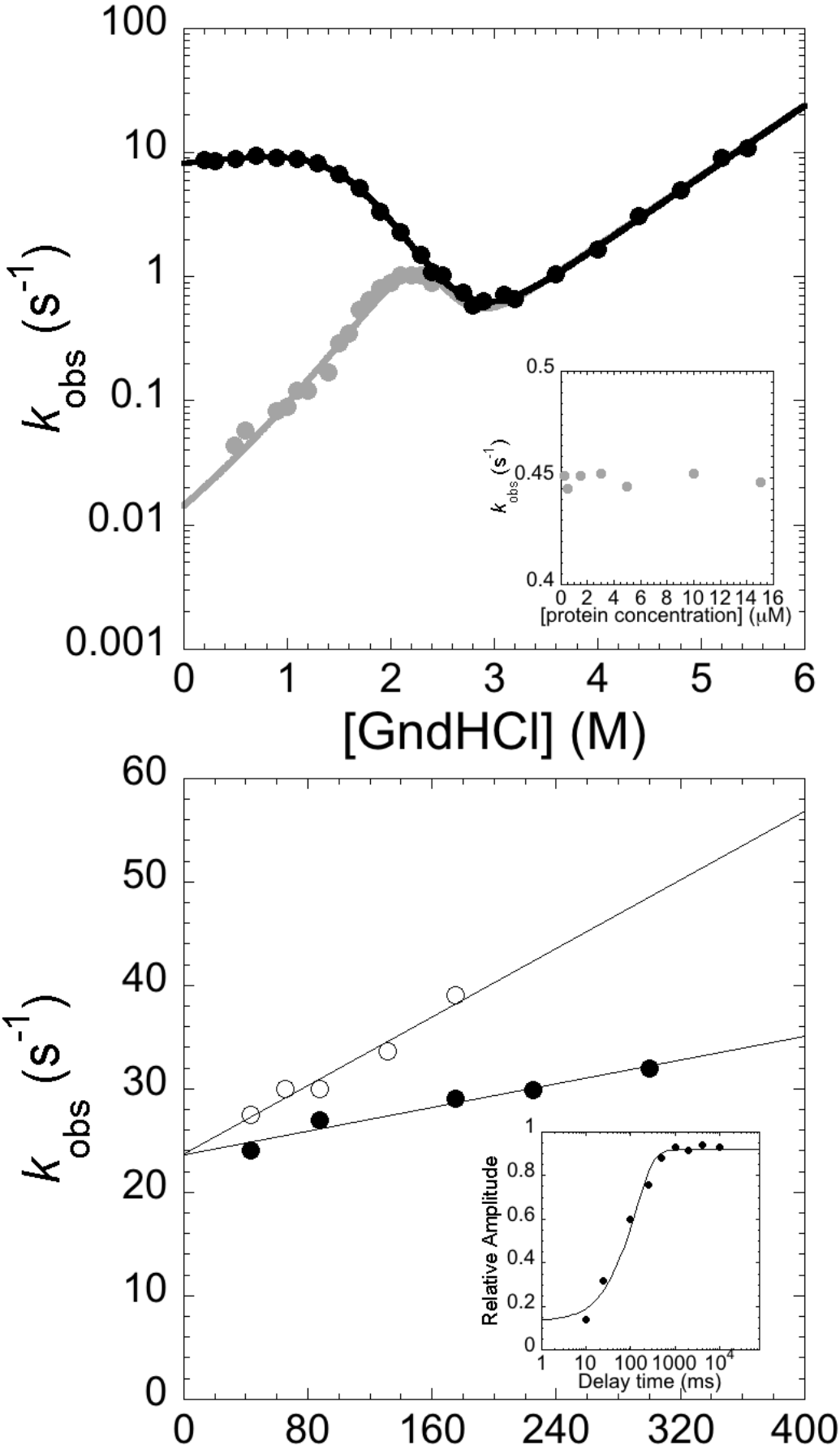
P2 denatured in 2.2 M GdnHCl was mixed with refolding buffer and then challenged with 200 μM peptide.

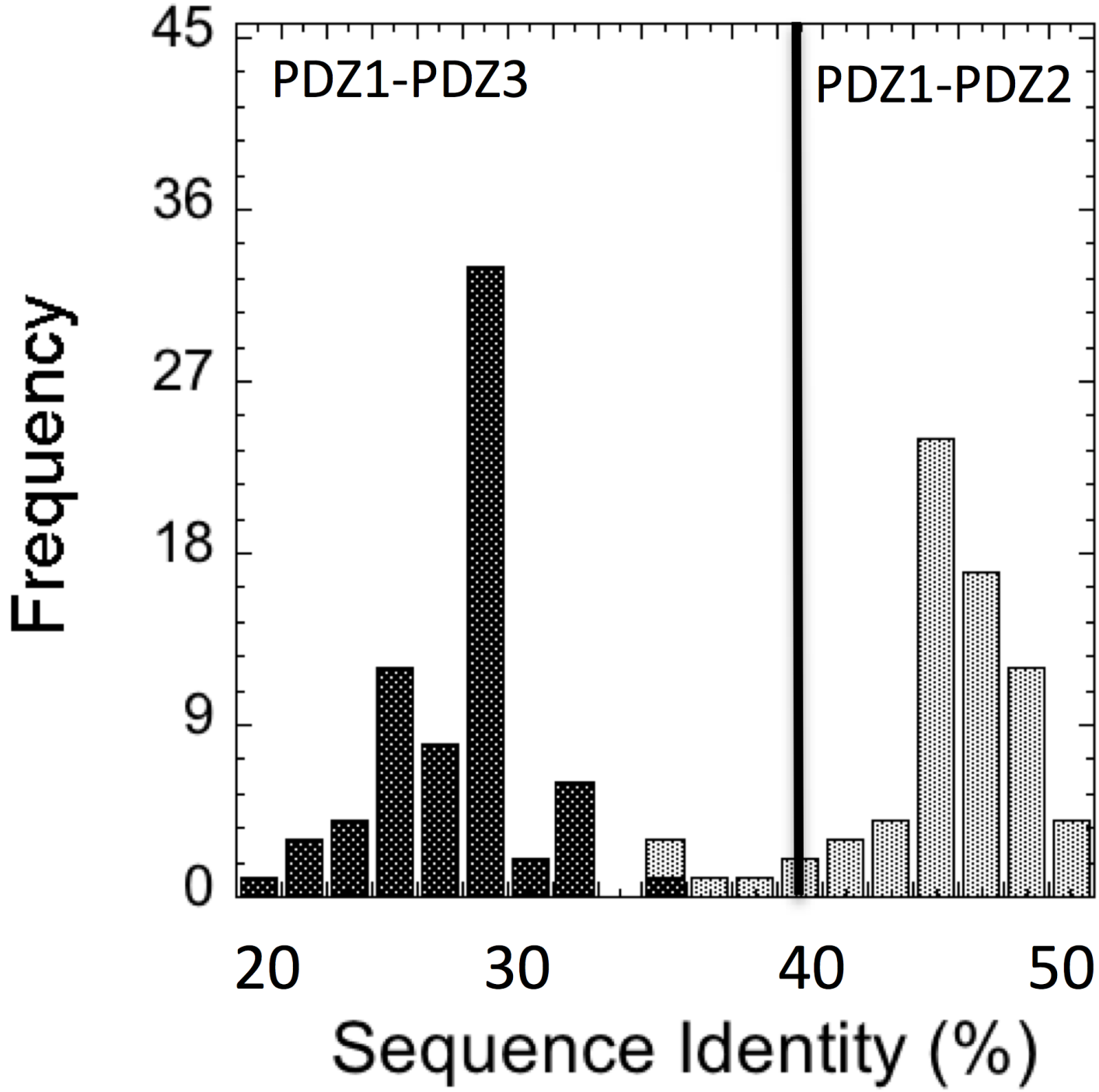
Figure 5. Analysis of the sequence identity (%) between PDZ1 and PDZ2 (grey bars) and PDZ1 and PDZ3 (black bars) of Whirlin from 70 species obtained by Uniprot database.


Fluorescence (Rel. units)
1.1 10^7
1 10^7
9 10^6
8 10^6
7 10^6
6 10^6
5 10^6
4 10^6
A
0 1 2 3 4 5 6
[GndHCl] (M)
CD (mdeg)
0
-20
-40
-60
-80
-100
B
0 1 2 3 4 5 6
[GndHCl] (M)
C
0 1 2 3 4 5 6
[GndHCl] (M)


A
15N (ppm)
1H (ppm)
GdnHCl
0.0 M
0.5 M
1.0 M
1.5M
2.0 M
2.2 M
2.6 M
3.1 M
3.6 M
B
PDZ1
Hp1
PDZ2
Hp2
Linker
Normalized intensity
Residue
GdnHCl
1.0 M
2.0 M
2.6 M
3.6 M
C
Average of PDZ1 residues
Average of PDZ2 residues
Normalized intensity
GdnHCl (M)


100
10
1
0.1
k_{obs} (s^{-1})
A
100
10
1
0.1
k_{obs} (s^{-1})
B
0
1
2
3
4
5
6
[GndHCl] (M)


100
10
1
0.1
0.01
0.001
k_{obs} (s^{-1})
0
1
2
3
4
5
6
[GndHCl] (M)
0.5
0.45
0.4
k_{obs} (s^{-1})
0 2 4 6 8 10 12 14 16
[protein concentration] (μM)
60
50
40
30
20
10
0
k_{obs} (s^{-1})
0
80
160
240
320
400
1
0.8
0.6
0.4
0.2
0
Relative Amplitude
1 10 100 1000 10^4
Delay time (ms)


45
36
27
18
9
0
Frequency
PDZ1-PDZ3
PDZ1-PDZ2
20
30
40
50
Sequence Identity (%)