Dissecting Inter-domain Cooperativity in the Folding of a Multi Domain Protein

Louise Laursen [1], Stefano Gianni [2*] and Per Jemth [1*]

*1 - **Department of Medical Biochemistry and Microbiology,** Uppsala University, BMC Box 582, SE-75123 Uppsala, Sweden*
*2 - **Istituto Pasteur-Fondazione Cenci Bolognetti and Istituto di Biologia e Patologia Molecolari del CNR,** Dipartimento di Scienze Biochimiche "A. Rossi Fanelli," Sapienza Università di Roma, 00185 Rome, Italy*

***Correspondence to Stefano Gianni and Per Jemth:** Stefano.Gianni@uniroma1.it (S. Gianni), Per.Jemth@imbim.uu.se (P. Jemth)*
https://doi.org/10.1016/j.jmb.2021.167148
Edited by Daniel Otzen

Abstract

Correct protein folding underlies all cellular functions. While there are detailed descriptions and a good understanding of protein folding pathways for single globular domains there is a paucity of quantitative data regarding folding of multidomain proteins. We have here investigated the folding of a three-domain supramodule from the protein PSD-95, consisting of one PDZ domain, one SH3 domain and one guanylate kinase-like (GK) domain. This supramodule has previously been shown to work as one functional unit with regard to ligand binding. We used equilibrium and kinetic folding experiments to demonstrate that the PDZ domain folds faster and independently from the SH3-GK tandem, which folds as one cooperative unit. However, concurrent folding of the PDZ domain slows down folding of SH3-GK by non-native interactions, resulting in an off-pathway folding intermediate. Our data contribute to an emerging description of multidomain protein folding in which individual domains cannot *a priori* be viewed as separate folding units.

Introduction

Folding of polypeptides into well-defined tertiary structures, defined by the amino acid sequence, has been an area of intense studies ever since the ground-breaking work of Anfinsen in the early 1960′s. Since then, a combination of experiment and computational chemistry has provided a good understanding of folding of protein domains. However, whilst our knowledge about the folding of single protein domains is well established,[1–3] many proteins consist of multiple domains and folding studies of such multi-domain proteins are surprisingly scarce. It has been shown that domains fold one by one as they emerge from the ribosome and it appears as if codon usage has been selected to slow down translation of linker regions, to allow nascent domains to fold before the next one is synthesized.[4] Yet, many small protein domains (50–150 residues) are marginally stable (a few kcal mol^{-1})[5,6] and may therefore unfold and refold continuously after translation. For example, a typical stability of 3 kcal mol^{-1} would correspond approximately to 0.5% population of the denatured state.

A protein domain is generally defined as a stable sub-structure that is capable to fold independently of the remainder of the protein.[7] Consequently, there is the implicit assumption that, within the same protein, domains affect their respective folding only marginally. Thus, the folding of protein domains in isolation is generally expected to resemble closely what is observed in the context of a multidomain system. Nevertheless, experience suggests that not all protein domains may be successfully

expressed in isolation nor necessarily behave as independent cooperative folding units, posing the definition of protein domain as somewhat more complex. Only a few studies have directly addressed these questions and protein domains seem to display different behavior.[8–13] In fact, while in some cases adjacent domains do not affect the folding pathway much,[14] in other cases, transient misfolding has been observed,[15–19] which can be avoided by co-translational folding,[20,21] but is still an intriguing finding, which stimulates additional research.

In the present work we investigated the folding of a three-domain supramodule from the protein PSD-95 using equilibrium and kinetic experiments. The supramodule, which displays interdomain effects on ligand binding[22,23] contains one PDZ, one SH3 and one guanylate kinase like (GK) domain. While equilibrium data showed that the PDZ domain folds independently, the SH3-GK tandem behaves as a single cooperative unit. This is noteworthy, since the SH3 domain is an archetypical protein domain, which has been shown to fold cooperatively in multiple studies.[24–28] Furthermore, we demonstrate that the unfolded PDZ domain can slow down SH3GK folding by stabilizing a transient intermediate, forming a misfolded kinetic trap. Thus, our work pinpoints the non-trivial cooperative folding interactions that may be established by protein domains, as exemplified by the SH3GK supramodule, while also addressing possible misfolding events occurring when several contiguous domains are denatured. On the basis of our data, we discuss some considerations pertaining to folding of multidomain proteins in the physiological cellular environment.

Results

We used four different protein constructs to dissect the folding of the PSG supramodule: wild-type PSG (denoted PSG WT), PSG WT with a Phe → Trp substitution at position 337 (PSG), the SH3-GK tandem (SH3GK) and PDZ3 with the engineered Trp337 (PDZ3) (Figure 1). The SH3GK tandem contains five Trp residues, two in the SH3 domain, two in the linker between SH3 and GK and one in the GK domain. These Trp residues together with Trp337 in PDZ3 were used to monitor folding using fluorescence. In parallel we used circular dichroism (CD) to monitor secondary structure. The combination of CD and fluorescence spectroscopy (with and without the Phe337 → Trp substitution) allowed us to deduce the folding mechanism for the PSG supramodule.

Equilibrium experiments demonstrate that PDZ3 and SH3GK fold independently in the context of PSG. First, we performed guanidinium chloride (GdnCl) induced unfolding experiments with PSG at equilibrium and at different pH values (Figure 2(A)) and different initial ionic strength (Figure 2(B)), respectively. CD-monitored experiments between pH 2.0–9.5 displayed two distinct unfolding transitions with GdnCl midpoints of approximately 1.7 M and 2.9 M, respectively, at neutral and basic pH values. This behavior was consistent over several replicates at different NaCl concentrations (Figure 2(B)). At low pH (3.5) the unfolding displayed two transitions but the profile was different from that at higher pH values (Figure 2(A)). In order to assign the respective transition to the domains of PSG, we performed similar GdnCl-induced unfolding experiments for PSG WT, SH3GK and PDZ3 (Figure 3(A)). The data of PSG WT were identical to those of PSG, showing that Phe337Trp does not change the equilibrium unfolding.

Next, we compared the folding of PSG with SH3GK and PDZ3. The GdnCl midpoints as well as the $m_{D\text{-}N}$ values suggested that the transition at 1.7 M is due to SH3GK (un)folding and the

Figure 1. Structure and topology of the PDZ3-SH3-GK (PSG) supramodule from PSD-95. (A) Structures of the three domains from PSD-95 PSG. The structure of PDZ3 was solved in isolation whereas SH3GK is a tandem construct; from left to right: PDZ3 (red), SH3 (cyan) and GK (grey). Protein data bank (PDB) accession numbers: 314 W (PDZ3) and 5YPR (SH3GK). (B) Schematic topology diagram of the four constructs: PSG, PSG WT, SH3GK, and PDZ3 used in our unfolding and refolding experiments. Number and positions of Trp (W) residues are highlighted in spheres in the structure. These residues contribute to the fluorescence signal in the equilibrium and kinetic experiments.

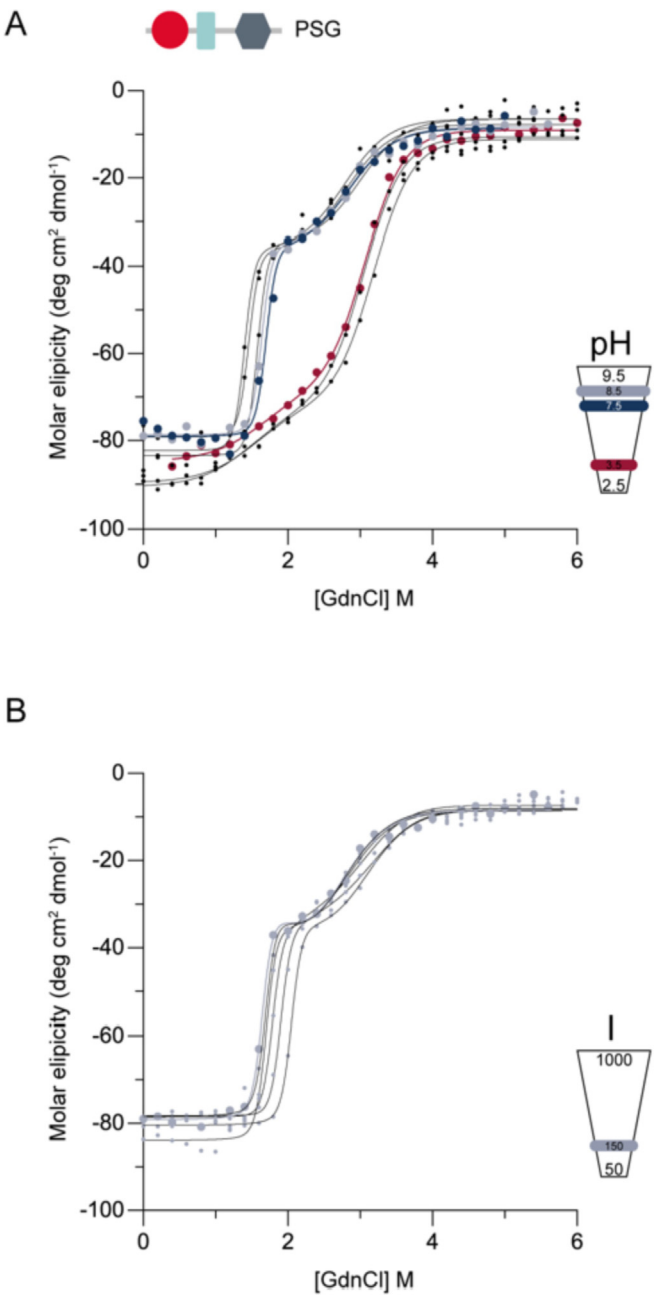

Figure 2. Equilibrium (un)folding of PSG is pH dependent, but independent of NaCl concentration. The equilibrium unfolding of PSG was monitored by circular dichroism at 222 nm as a function of GdnCl concentration. Experiments were performed at different pH values at a constant ionic strength (A) or at different NaCl concentrations at a constant pH (B). Experiments were performed at 25 °C. (A) The following buffers (50 mM) were used at the respective pH value: pH 2.0 (sodium phosphate), pH 3.0 (sodium formate), pH 3.5 (sodium formate), pH 5.5 (sodium acetate), pH 6.0 (Bis-Tris), pH 7.5 (sodium phosphate), pH 8.5 (Tris) and pH 9.5 (CHES). The ionic strength in absence of GdnCl was adjusted to 150 mM using NaCl. (B) Experiments were performed in 50 mM Tris, pH 8.5, and the ionic strength in absence of GdnCl was adjusted to 50, 150, 250, 450, 750 and 1000 mM using NaCl. 1 mM TCEP was added to all samples.

transition at 2.9 M corresponds to (un)folding of PDZ3. The $m_{D\text{-}N}$ value is obtained from the slope of the transition and is related to the change in solvent accessible surface area upon unfolding. Thus, it correlates well with size for globular domains. Accordingly, the transition at 1.7 M should be associated with the higher $m_{D\text{-}N}$ value of the SH3GK tandem as compared to the smaller $m_{D\text{-}N}$ value of the PDZ3 domain, which unfolds around 2.9 M. Quantitative analysis of a two-step

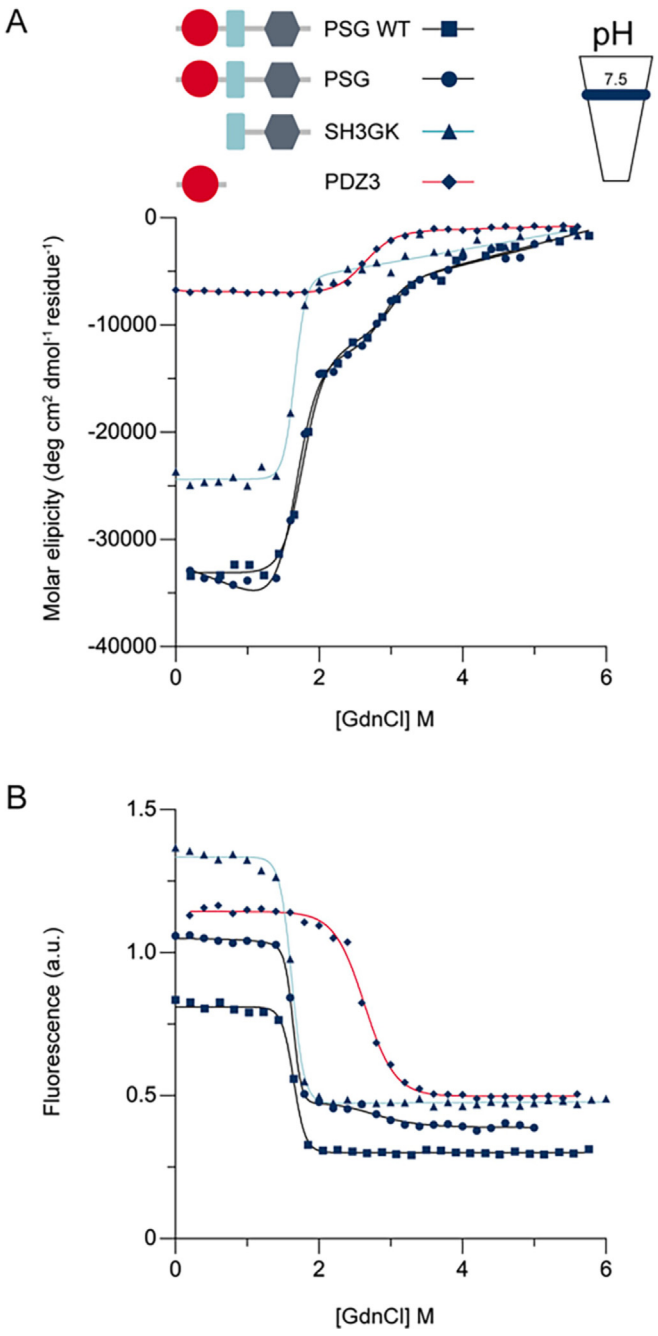

Figure 3. Sequential unfolding of SH3-GK and PDZ3 domains in PSG at neutral pH. GdnCl-induced unfolding in 50 mM sodium phosphate, pH 7.5, of PSG WT, PSG, SH3GK and PDZ3, respectively. The unfolding was monitored by (A) circular dichroism at 222 nm and (B) fluorescence emission at 340 nm (excitation at 280 nm). Parameters from the curve fitting ($m_{D\text{-}N}$, $[GdnCl]_{50\%}$ and $\Delta G_{D\text{-}N}$) are given in Supplementary Table S1.

unfolding is challenging, and the $m_{D\text{-}N}$ values for the two transitions in PSG unfolding are determined with large errors (Supplementary Table S1). Nonetheless, the fitted parameters were overall consistent with our interpretation: $m_{D\text{-}N}$ values for SH3GK in PSG and PSG WT (7.7 kcal mol^{-1} M^{-1}) agreed well with that from the SH3GK tandem (7.6 kcal mol^{-1} M^{-1}). The $m_{D\text{-}N}$ value for PDZ3 in PSG and PSG WT was 1.3 kcal mol^{-1} M^{-1}, whereas PDZ3 displayed an $m_{D\text{-}N}$ value of 2.9 kcal mol^{-1} M^{-1}. This discrepancy is reasonable given the relatively small transition associated with PDZ3 unfolding in the context of PSG. These numbers are also consistent with the respective size of PDZ3 (93 residues, 2.8 kcal mol^{-1} M^{-1}) and SH3GK (305 residues, 10 kcal mol^{-1} M^{-1}) as calculated by an empirical equation, which is based on experimental data[29]

and the earlier notion that $m_{D\text{-}N}$ values correlate with change in solvent-accessible surface area upon (un)folding.[30] Furthermore, since CD measures secondary structure content, we can directly compare the amplitudes of the transitions for the SH3GK tandem and PDZ3 domain with the corresponding transitions in PSG. Indeed, these amplitudes correlate well (Figure 3(A)) and are thus consistent with the conclusion based on GdnCl midpoints and $m_{D\text{-}N}$ values.

We then monitored the equilibrium unfolding using fluorescence, which usually reports on loss of tertiary structure. While the large number of Trps, five in SH3GK and PSG WT, and six in PSG, could complicate the analysis, the overall fluorescence intensity decreased upon unfolding, over the entire GdnCl range (Figure S1). The fluorescence intensity at 340 nm, where a large change in fluorescence upon denaturation was observed, was plotted versus GdnCl concentration (Figure 3(B)). Midpoints and $m_{D\text{-}N}$ values for the transitions were consistent with the CD data (Supplementary Table S1). Note that the amplitudes of the fluorescence data cannot be directly compared between the four constructs. Importantly, the transition around 2.9 M was present for PSG (with Trp337 in PDZ3) but not for PSG WT (with Phe337) corroborating that this transition is associated with PDZ3 (un)folding.

To further investigate the equilibrium unfolding we performed CD-monitored GdnCl denaturation experiments at lower pH values (2.5–3.5). While the total difference in secondary structure content upon unfolding was similar at low and high pH, the profile of the GdnCl denaturation changed (Figure 2(B)). Based on a set of experiments at low pH, similar to that performed at high pH (Figure 2(B) and Figure 4) we found that PDZ3 is destabilized by low pH (in agreement with earlier results) but that SH3GK is stabilized. This resulted in overlapping transitions for PDZ3 and SH3GK in the context of PSG, making a quantitative analysis challenging. However, a cooperative unfolding of the entire PSG would result in a sharp transition and a large $m_{D\text{-}N}$ value. Thus, the relatively extended unfolding transition at low pH is consistent with two (or more) independent and partially overlapping unfolding transitions, suggesting that PDZ3 and SH3GK fold as independent units at low pH as well as at high pH.

Kinetic experiments reveal the presence of an off-pathway intermediate in SH3GK folding. The equilibrium experiments demonstrated independent folding of PDZ3 and SH3GK but such data cannot deduce any further details of the mechanism. In an attempt to learn more about the mechanism we performed kinetic folding and unfolding experiments. In agreement with our previous studies, the folding kinetics of PDZ3 followed a two-state mechanism at neutral pH, corresponding to an equilibrium between the denatured and the folded states without any detectable intermediates. Thus, kinetic folding transients were well described by single exponential kinetics and the GdnCl dependence of the observed rate constant (k_{obs}) followed a V-shaped so-called chevron plot (Figure S2, Figure 5(A)). The left arm of the chevron plot corresponds to the folding rate constant (k_f), which decreases linearly with increasing GdnCl concentration, whereas the right arm corresponds to the unfolding rate constant (k_u), which increases linearly with increasing GdnCl concentration, consistent with two-state folding. The SH3GK tandem, on the other hand, displayed more complex kinetics. While unfolding kinetic transients were single exponential, the refolding kinetics were clearly bi-exponential (Figure S2) up to approximately 1.1 M GdnCl, where the two phases merge. Due to the complex kinetics, we cannot establish unequivocally what the respective phase corresponds to, but the slow phase might be due to cis-trans proline isomerization. Thus, the faster of the two phases is considered the main folding phase. The k_{obs} values for SH3GK followed a chevron plot with deviation from linearity ("rollovers") in both the refolding and unfolding arm (Figure 5(B)). The rollover in the unfolding arm (at high [GdnCl]) can be interpreted in terms of a high energy intermediate and a change in rate limiting step involving the two transition states surrounding the intermediate. An alternative explanation is a broad transition state barrier, with movement of the top of the barrier according to Hammond behavior.[31–33] Experimentally, it is practically impossible to distinguish these two models, since they predict very similar dependence of the observed rate constants on denaturant concentration. Turning to the rollover in the refolding arm (at low [GdnCl]), this also suggests the presence of an intermediate. The formation of an intermediate was corroborated by analysis of amplitudes and endpoints of the kinetic transients (Figure S3). While the amplitudes from unfolding kinetic transients corresponded well to the expected ones, the amplitudes from refolding kinetics were smaller than expected based on the observed change in endpoint fluorescence in the experiment. Such "loss of amplitude" is consistent with fast formation of an intermediate occurring in the dead time of the experiment, i.e., faster than the mixing of solutions in the stopped-flow instrument. However, whether this is a productive on-pathway intermediate between the denatured and the native state, or an off-pathway species could not be deduced from the kinetic data.

Next, we performed similar kinetic folding experiments with PSG WT. The kinetic transients for PSG WT were relatively well described by a single exponential function, although the residuals show a trend (Figure S2). However, the difference in magnitude of the two rate constants obtained from fitting a double exponential to the transients

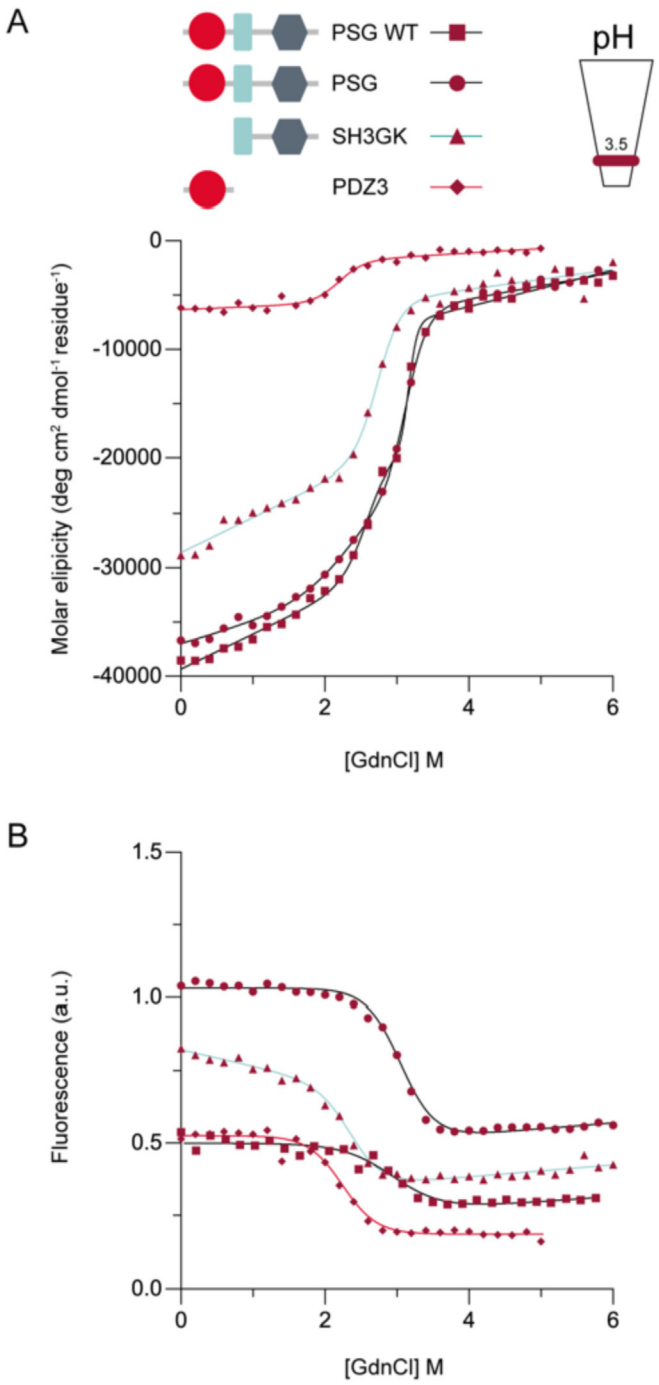

Figure 4. Equilibrium unfolding at low pH. GdnCl-induced unfolding in 50 mM sodium formate, pH 3.5, of PSG WT, PSG, SH3GK and PDZ3, respectively. The unfolding was monitored by (A) circular dichroism at 222 nm and (B) fluorescence emission at 340 nm (excitation at 280 nm).

was only 2–3 fold, which is not enough for a reliable curve fitting. The observed rate constants for PSG WT were identical to those of SH3GK in the unfolding arm (Figure 5(B)). However, the k_{obs} values in the refolding arm (from a single exponential fit) follow a distinct pattern. In fact, between 0–1 M GdnCl the k_{obs} values for PSG WT increase with increasing GdnCl concentration, a kinetic signature of an off-pathway intermediate. Like for SH3GK, there is a loss of amplitude in the kinetic transients consistent with formation of a folding intermediate.

Finally, we performed kinetic refolding and unfolding experiments with PSG, i.e., with a Trp337 in PDZ3. Now, the kinetic transients were clearly biphasic (Figure S2) thus reporting directly

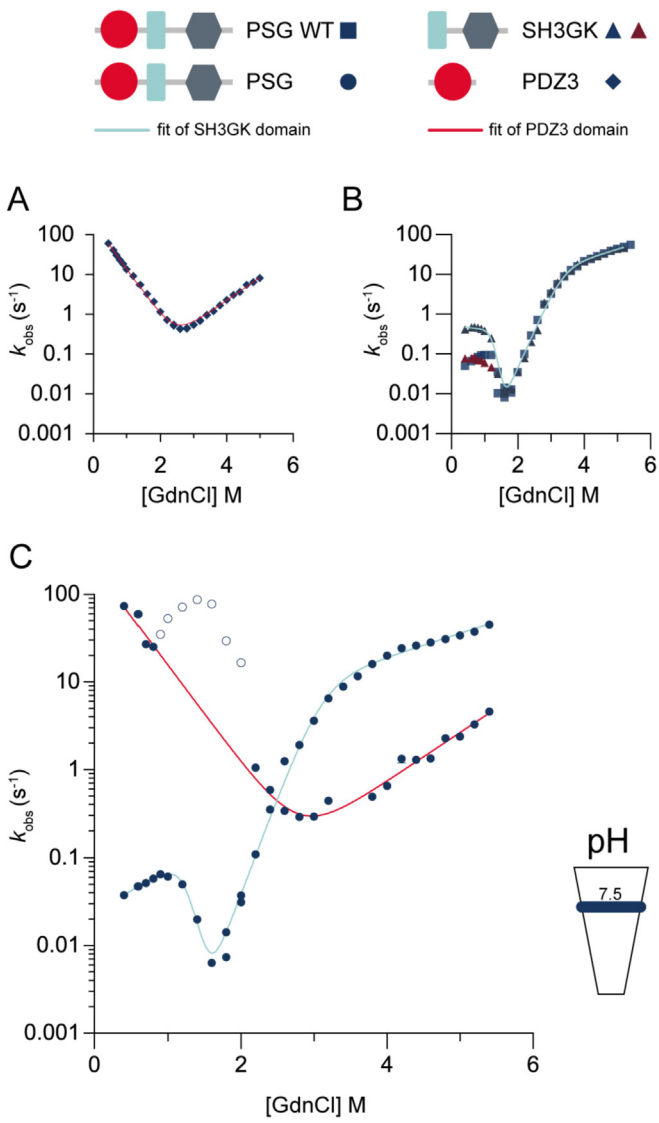

Figure 5. Unfolding and refolding kinetics of PDZ3, SH3GK PSG WT and PSG. Refolding was performed by mixing protein from an initial GdnCl concentration of 4.5 M into lower buffer-GdnCl concentrations. Unfolding was performed by mixing protein in buffer with buffers containing GdnCl at different concentrations. (A) PDZ3, (B) SH3GK and PSG WT (no Trp in PDZ3) and (C) PSG (with Trp337 in PDZ3). The experiments were performed at 25 °C in 50 mM sodium phosphate, pH 7.5, 1 mM TCEP. Data for PDZ3 were fitted to Eq. (3) (parameters are shown in Supplementary Table S2) and data for SH3GK, PSG WT and PSG to Eq. (4).

on a two-step (three state) folding reaction. The observed rate constants yielded two chevrons, one similar to that of PSG WT and one similar to that of PDZ3 (Figure 5(C)). The data points at 1–2 M GdnCl for the fast phase displayed a peculiar behavior that we cannot explain, and which is likely due to an artefact from low kinetic amplitudes (Figure S3(D)). The observed rate constants from the slow phase displayed a similar behavior to that of PSG WT, with a clearly positive slope of k_{obs} versus GdnCl concentration, and with a loss of amplitude in the kinetic transients, suggesting accumulation of an off-pathway intermediate during refolding under native-like conditions.

As a control, we performed refolding experiments at different protein concentrations for both SH3GK and PSG to rule out that the rollover in the refolding arm was due to transient aggregation (Figure S4).

Refolding rate constants suggest that unfolded PDZ3 traps an intermediate of the SH3GK tandem. The data presented in Figure 5 suggested that an intermediate accumulates during refolding of SH3GK, and that the formation of this intermediate is affected by the concurrent

refolding of PDZ3. To corroborate this hypothesis, we took advantage of the fact that PDZ3 displays a higher GdnCl unfolding midpoint than SH3GK. At [GdnCl] = 4.5 M, both PDZ3 and SH3GK are unfolded. However, at [GdnCl] = 2 M, PDZ3 remains folded but SH3GK is partially unfolded. Thus, by performing refolding experiments from 2 M GdnCl we were able to monitor folding of SH3GK in the presence of a folded PDZ3 domain. It was clear from these experiments that the refolding of PSG adjacent to a folded PDZ3 domain shows a similar behavior as refolding of SH3GK without the PDZ3 domain (Figure 6). We can therefore conclude that the concurrent refolding of PDZ3 slows down the refolding of SH3GK, via an inter-domain interaction between non-native states.

A folding pathway consistent with all data involves rapid formation of an intermediate from SH3GK (Figure 7). This intermediate folds into the native structure with an observed rate constant following the slow kinetic phase (Figure 5(B)–(D)). In the presence of unfolded PDZ3, which occurs when PSG is refolded from 4.5 M GdnCl, the denatured state of PDZ3 will interact with and trap the SH3GK intermediate to form an off-pathway intermediate and slow down refolding to the native state. Note that we have depicted the SH3GK intermediate as an off-pathway species, but, based on the kinetic data, we cannot distinguish whether it is on- or off-pathway.

Discussion

There is a general assumption based on the inherent definition of a protein domain that, within a protein, individual domains represent self-stabilizing units, capable to fold cooperatively whether or not other domains are present. Nevertheless, due to the complexity of the thermodynamics and kinetics of multidomain systems, the experimental evidence supporting this postulation is still surprisingly limited. Among multidomain systems, it is of particular interest to explore those cases of supertertiary structure in which different protein domains are closely linked in function, displaying features that have been classically associated to so-called supramodules and where independent folding of the individual domains is not obvious. By following these premises, the experimental work reported in this study are particularly instructive. In fact, whilst the

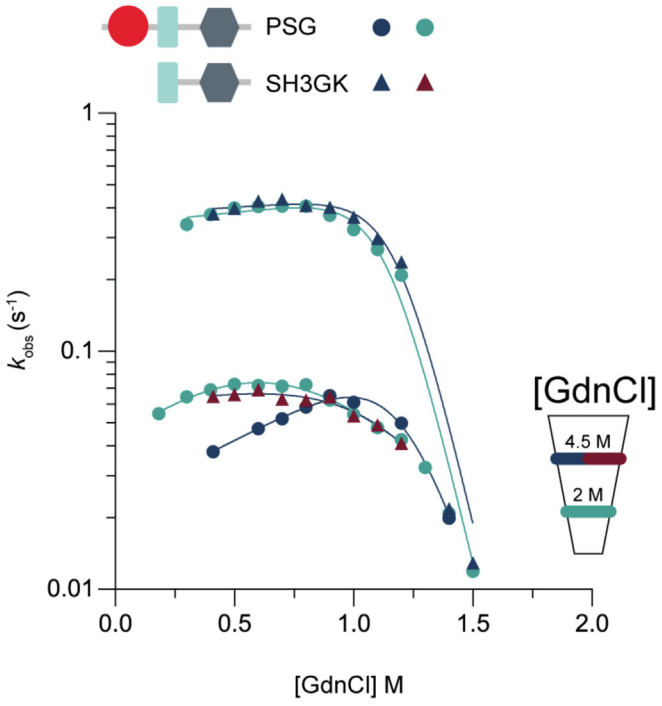

Figure 6. Concurrent PDZ3 folding slows down refolding of SH3GK. Refolding of PSG was started from different initial conditions, 4.5 M and 2 M GdnCl, respectively. Refolding from the fully denatured state (4.5 M) resulted in lower refolding rate constants for PSG and a clear positive slope at GdnCl concentrations below 1 M. However, refolding from 2 M GdnCl where PDZ3 is in its native state yielded rate constants, which behaved like those of the SH3GK tandem domain without PDZ3. The latter scenario includes two kinetic phases of which the slower one is possibly due to cis-trans Pro isomerization. The experiments were performed at 25 °C in 50 mM sodium phosphate, pH 7.5, 1 mM TCEP.

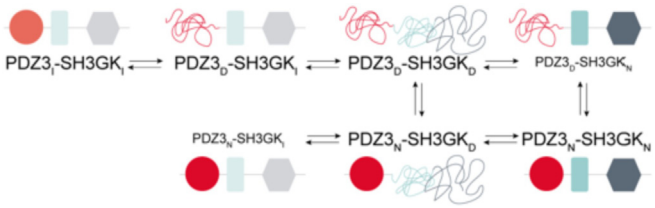

Figure 7. A model for folding of PSG. The supramodule PSG from the protein PSD-95 consists of three domains: PDZ3, SH3 and GK. PSG contains two units, which fold independently: PDZ3 and the SH3GK tandem. The folding of SH3GK involves a fast forming "burst phase" intermediate $SH3GK_I$, which may be on- or off-pathway to the native state. Unfolded $PDZ3_D$ interacts intramolecularly with $SH3GK_I$ to form an off-pathway intermediate $PDZ3_I$-$SH3GK_I$.

PDZ domain appears to fold independently from the reminder of the protein, the SH3 and GK domains fold and unfold as a single cooperative unit as shown by the high $m_{D\text{-}N}$ value (7–8 kcal mol^{-1} M^{-1}). In the case of the whole PSG supramodule, this behavior is clearly very robust and is maintained at different experimental conditions such as low and high pH.

As mentioned in the introduction, a problem of interest lies in understanding the capability of protein domains to fold and function efficiently in isolation. In fact, whilst countless examples demonstrate that many domains behave as robust functional and structural units, it is also very common to observe different behaviors with other domains expressing poorly and/or displaying a very low thermodynamic stability. In these cases, it may be assumed that the architecture of the multidomain protein is essential to stabilize each distinct domain. For example, the folding pathways of different PDZ and SH3 domains have been previously successfully described, showing that in several cases both these type of domains may behave as independent folding units. In this context, the results provided in this work allow the description of an additional behavior. In fact, in the case of PSG folding, the SH3 and GK are intimately linked. Remarkably, these two domains fold and unfold as a single cooperative unit, as mirrored by the high $m_{D\text{-}N}$ value, which is consistent with the size of the SH3GK tandem and robust to changes in experimental conditions. On the basis of these observations, it may be concluded that, in the case of PSG, the SG supramodule behaves as a single protein domain. A similar behavior was previously observed for a speficic type of multidomain protein, namely tandem repeats.[34] Our findings on the PSG supramodule together with previous data suggest that general rules assuming protein domains as independent folding units are unlikely and reinforces the importance to study directly the folding of multidomain proteins.

It is of interest to analyze the kinetic role of the different intermediates observed in the folding of PSG. The simplest scenario consistent with the observed equilibrium and kinetic data is depicted in Figure 7. In summary, focusing on the SH3-GK tandem, it may be noted that its folding displays two intermediates: one high energy intermediate occurring late on the folding pathway and one lower energy intermediate occurring earlier. Interestingly, the data reveal that the low energy intermediate is prone to interact with the unfolded, but not folded, PDZ3 domain. Interaction with unfolded PDZ3 results in stabilization of the SH3GK intermediate into a new off-pathway misfolded species (Figure 7). To get to the native state, this relatively stable off-pathway intermediate must unfold, resulting in an overall slowing down of the folding of native PSG.

The cooperative transitions typically observed in the folding of single protein domains strongly suggest that proteins have evolved to conform to an all-or-none type of folding reaction, avoiding the accumulation of intermediates, which could lead to aggregation and dysfunction. In addition, there is likely a trade-off between folding efficiency and function such that a perfectly smooth energy landscape is rarely seen.[35–37] While the modular nature of proteins allows for stepwise folding of domains as they emerge from the ribosome, the individual domains are often marginally stable such that they unfold and fold reversibly. It is hypothesized that sequential folding of domains as they emerge from the ribosome allows for expression of multidomain proteins while avoiding misfolding and aggregation. Based on these considerations, it is worth noticing how the inherent stability of the PDZ3 domain is particularly high compared to other characterized PDZ domains.[38,39] In fact, by considering that transient denaturation of PDZ3 may lead to a misfolding event of the SH3GK module, it is tempting to speculate that the high thermodynamic stability of this PDZ domain is under positive selection to circumvent dysfunction and aggregation. Future work on different multidomain proteins will further test this speculation.

Methods

Protein expression and purification. Wild-type PSD-95 PDZ3-SH3-GK (residues 309–724 of PSD-95, denoted PSG WT), wild-type SH3GK tandem domain (residues 419–724), pseudo wild-type PSG (with a Phe337 → Trp substitution in the

PDZ3 domain, denoted PSG) and pseudo wild-type PDZ3 with the Phe337Trp mutation (residues, 309–401, denoted PDZ3) were expressed and purified as described previously.[22] Briefly, a pRSET A plasmid encoding the proteins were transformed into *Escherichia coli* BL21(DE3) pLys cells (Invitrogen). Cells were first grown in LB medium at 37 °C in a rotary shaker. Over-expression of protein was induced with 1 mM isopropyl-β-D-1-thiogalactopyranoside at OD_{600} of 0.6–0.8, and the cells were incubated overnight at 18 °C. Each protein was purified from the soluble fraction on a Nickel Sepharose Fast Flow column (GE Healthcare) equilibrated with 50 mM Tris pH 7.8, 100 mM NaCl, 10% glycerol, 20 mM Imidazole and 0.5 mM DTT. Bound proteins were eluted by increasing the imidazole concentration to 250 mM. Proteins were dialyzed into 50 mM Tris pH 7.8, 100 mM NaCl, 10 % glycerol and 2 mM DTT, concentrated and further purified using size exclusion chromatography (S-100, GE Healthcare). Protein purity was quantified by SDS-PAGE and identity by MALDI-TOF mass spectrometry.

Equilibrium Experiments. Equilibrium unfolding experiments were performed with 4–20 μM protein in 50 mM sodium phosphate pH 7.45, 21 mM NaCl (I = 150), 1 mM TCEP at 25 °C. Circular dichroism was measured on a JASCO 1500 spectropolarimeter in a 1 mm quartz cuvette using the average of 5 scans, and ellipticity at 222 nm was plotted against GdnCl concentration (0 to 5 M, with 0.2 M steps). Fluorescence equilibrium unfolding experiments were performed with the same unfolded protein sample (0 to 5 M GdnCl) in a quartz cuvette with 10 mm path length. Samples were incubated for at least 30 min before the first measurement to ensure that the (un)folding reaction had reached equilibrium. The protein was excited at 280 nm and emission spectra were recorded from 300 to 450 nm at the respective denaturant concentrations. Fluorescence emission at 340 nm was plotted against GdnCl concentration. CD and fluorescence data were analyzed quantitatively either by two state or three state models (Eqs. (1) and (2)).

$$F = (((\alpha_N + \beta_N \cdot [GdnCl]) + (\alpha_D + \beta_D \cdot [GdnCl]) \cdot \exp((m_{D-N} \cdot ([GdnCl] - [GdnCl]_{50\%}))/RT))/((1 + \exp(m_{D-N} \cdot ([GdnCl] - [GdnCl]_{50\%}))/RT)) \quad (1)$$

$$F = (((\alpha_N + \beta_N \cdot [GdnCl]) + (\alpha_D + \beta_D \cdot [GdnCl]) \cdot \exp((m_{D-N} \cdot ([GdnCl] - [GdnCl]_{50\%}))/RT))/((1 + \exp(m_{D-N} \cdot ([GdnCl] - [GdnCl]_{50\%}))/RT)) + ((\alpha_D^2 + \beta_D^2 \cdot [GdnCl]) \cdot \exp((m_{D-N}^2 \cdot ([GdnCl] - [GdnCl]_{50\%}^2))/RT))/((1 + \exp(m_{D-N}^2 \cdot ([GdnCl] - [GdnCl]_{50\%}^2))/RT)) \quad (2)$$

The same experimental setup and equations as described above were used for analysis of the pH dependence of the folding (pH 2.5 to 9.5, *I* = 150 mM, adjusted with NaCl) and for the (initial) ionic strength dependence (*I* = 50 to 1000 mM, adjusted with NaCl in 50 mM Tris pH 8.5, 1 mM TCEP). Buffer used for the pH dependence were: pH 2.0, sodium phosphate; pH 3.0, sodium formate; pH 3.5, sodium formate; pH 5.5, sodium acetate; pH 6.0, Bis-Tris; pH 7.5, sodium phosphate; pH 8.5, Tris; and pH 9.5, CHES.

Kinetic experiments. Rapid mixing for kinetic folding and unfolding experiments were performed using an upgraded SX-17 MV stopped-flow spectrophotometer (Applied Photophysics, Leatherhead, UK) at 25 °C. All proteins used in the kinetic experiments have one or several Trp residues, so fluorescence emission was measured with a 320 nm bandpass filter using an excitation wavelength at 280 nm. PDZ3, SH3-GK, PSG WT or PSG with or without 4.5 M denaturant were rapidly mixed with buffer-GdnCl solutions to final protein concentrations of 0.05–2 μM and GdnCl concentrations of 0.41 to 5 M to induce refolding or unfolding. Kinetic experiments were performed at neutral pH (50 mM sodium phosphate, pH 7.45, 21 mM NaCl (I = 150), 1 mM TCEP). Kinetic traces were fitted to a single or double exponential decay using Applied Photophysic's software to obtain observed rate constants (k_{obs}). Chevron plots were obtained by plotting k_{obs} versus GdnCl concentration.

Data analysis of kinetic experiments. All data were analyzed assuming a linear dependence of the logarithm of each of the microscopic rate constants on denaturant concentration.[40] Kinetic data from PDZ3 followed a two-state scenario where $k_{obs} = k_u + k_f$ and was therefore fitted to Eq. (3).

$$k_{obs} = k_u^{H_2O} \cdot \exp(m_{\ddagger-N}[GdnCl]/RT) + k_f^{H_2O} \cdot \exp(-m_{D-\ddagger}[GdnCl]/RT) \quad (3)$$

SH3-GK, PSG and PSG WT showed more complex kinetics, implying a multistate folding pathway. The data were fitted to Eq. (4) describing an equilibrium between intermediates and ground states. While the kinetics are qualitatively consistent with equilibrium data, the high number of parameters in Eq. (4) leads to large errors in several parameters, which precludes a meaningful quantitative comparison between kinetics and equilibrium.

$$k_{obs} = k_f^{H_2O} \cdot exp(-m_f[GdnCl]/RT)/(1 + K_I \cdot exp(m_I \cdot [GdnCl]/RT)) + k_u^{H_2O} \cdot exp(m_u[GdnCl]/RT)/(1 + K_p \cdot exp(m_p[GdnCl]/RT)) \quad (4)$$

All data were analyzed with Prism 9 (GraphPad Software, Inc.)

Control experiments to ensure that denatured PSG refolds to its native state. PSG was denatured in 4.5 M GdnCl and refolded overnight against the experimental buffer. This refolded PSG sample was then used in five different experiments to demonstrate that refolding returns

the same native state as obtained after expression and purification (Figure S5). CD spectra (Figure S5(A)) as well as GdnCl denaturation experiments (Figure S5(B) and (C)) were identical for the two samples. Furthermore, kinetic experiments involving manual mixing refolding experiments in a fluorimeter (JASCO 1500 spectropolarimeter) showed that the time window used in stopped flow experiments captured the entire refolding reaction (Figure S5(D)). Manual mixing in the JASCO instrument was necessary since long recordings in the stopped flow resulted in signal drift. Finally, peptide-binding experiments were performed as described previously using a dansylated CRIPT peptide.[22] The observed rate constants for binding were almost identical for the two PSG samples, showing that a binding-competent native state was attained after refolding (Figure S5(E)).

CRediT authorship contribution statement

Louise Laursen: Methodology, Validation, Formal analysis, Investigation, Writing - review & editing, Visualization, Project administration. **Stefano Gianni:** Conceptualization, Methodology, Writing - original draft, Writing - review & editing, Funding acquisition. **Per Jemth:** Conceptualization, Methodology, Formal analysis, Resources, Writing - original draft, Writing - review & editing, Supervision, Project administration, Funding acquisition.

Acknowledgements

The project has received funding from the European Union´s Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 67341 (to SG and PJ) and from the Swedish Research Council (grant no. 2020-04395) (to PJ). The work was partly supported by grants from Sapienza University of Rome (B52F16003410005, RP11715C34AEAC9B, RM1181641C2C24B9 and RG12017297FA7223 to S.G), and the Associazione Italiana per la Ricerca sul Cancro (individual grant, IG 2020, 24551, to S.G.).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmb.2021.167148.

Received 19 March 2021;
Accepted 5 July 2021;
Available online 8 July 2021

Keywords:
protein folding;
multi-domain protein;
PSD-95;
kinetics;
folding mechanism

References

1. Lapidus, L.J., (2020). The road less traveled in protein folding: evidence for multiple pathways. *Curr. Opin. Struct. Biol.*, **66**, 83–88.
2. Gianni, S., Jemth, P., (2016). Protein folding: Vexing debates on a fundamental problem. *Biophys. Chem.*, **212**, 17–21.
3. Jackson, S.E., (1998). How do small single-domain proteins fold?. *Fold Des.*, **3**, R81–R91.
4. Kramer, G., Shiber, A., Bukau, B., (2019). Mechanisms of cotranslational maturation of newly synthesized proteins. *Annu. Rev. Biochem.*, **88**, 337–364.
5. Goldenzweig, A., Fleishman, S.J., (2018). Principles of protein stability and their application in computational design. *Annu. Rev. Biochem.*, **87**, 105–129.
6. Sanchez-Ruiz, J.M., (2011). Probing free-energy surfaces with differential scanning calorimetry. *Annu. Rev. Phys. Chem.*, **62**, 231–255.
7. Wetlaufer, D.B., (1973). Nucleation, rapid folding, and globular intrachain regions in proteins. *Proc. Natl. Acad. Sci. U. S. A.*, **70**, 697–701.
8. Batey, S., Scott, K.A., Clarke, J., (2006). Complex folding kinetics of a multidomain protein. *Biophys. J .*, **90**, 2120–2130.
9. Arora, P., Hammes, G.G., Oas, T.G., (2006). Folding mechanism of a multiple independently-folding domain protein: double B domain of protein A. *Biochemistry*, **45**, 12312–12324.
10. Kumar, V., Chaudhuri, T.K., (2018). Spontaneous refolding of the large multidomain protein malate synthase G proceeds through misfolding traps. *J. Biol. Chem.*, **293**, 13270–13283.
11. Batey, S., Clarke, J., (2006). Apparent cooperativity in the folding of multidomain proteins depends on the relative rates of folding of the constituent domains. *Proc. Natl. Acad. Sci. U. S. A.*, **103**, 18113–18118.
12. Han, J.H., Batey, S., Nickson, A.A., Teichmann, S.A., Clarke, J., (2007). The folding and evolution of multidomain proteins. *Nat. Rev. Mol. Cell Biol.*, **8**, 319–330.
13. Ramsay, G., Freire, E., (1990). Linked thermal and solute perturbation analysis of cooperative domain interactions in proteins. Structural stability of diphtheria toxin. *Biochemistry*, **29**, 8677–8683.

14. Batey, S., Clarke, J., (2008). The folding pathway of a single domain in a multidomain protein is not affected by its neighbouring domain. *J. Mol. Biol.*, **378**, 297–301.
15. Borgia, A., Kemplen, K.R., Borgia, M.B., Soranno, A., Shammas, S., Wunderlich, B., et al., (2015). Transient misfolding dominates multidomain protein folding. *Nat. Commun.*, **6**, 8861.
16. Gautier, C., Troilo, F., Cordier, F., Malagrino, F., Toto, A., Visconti, L., et al., (2020). Hidden kinetic traps in multidomain folding highlight the presence of a misfolded but functionally competent intermediate. *Proc. Natl. Acad. Sci. U. S. A.*, **117**, 19963–19969.
17. Lafita, A., Tian, P., Best, R.B., Bateman, A., (2019). Tandem domain swapping: determinants of multidomain protein misfolding. *Curr. Opin. Struct. Biol.*, **58**, 97–104.
18. Borgia, M.B., Borgia, A., Best, R.B., Steward, A., Nettels, D., Wunderlich, B., et al., (2011). Single-molecule fluorescence reveals sequence-specific misfolding in multidomain proteins. *Nature*, **474**, 662–665.
19. Visconti, L., Malagrino, F., Troilo, F., Pagano, L., Toto, A., Gianni, S., (2021). Folding and misfolding of a PDZ tandem repeat. *J. Mol. Biol.*, **433**, 166862
20. Kim, S.J., Yoon, J.S., Shishido, H., Yang, Z., Rooney, L.A., Barral, J.M., et al., (2015). Protein folding. Translational tuning optimizes nascent protein folding in cells. *Science*, **348**, 444–448.
21. Samelson, A.J., Bolin, E., Costello, S.M., Sharma, A.K., O'Brien, E.P., Marqusee, S., (2018). Kinetic and structural comparison of a protein's cotranslational folding and refolding pathways. *Sci. Adv.*, **4** (eaas9098)
22. Laursen, L., Kliche, J., Gianni, S., Jemth, P., (2020). Supertertiary protein structure affects an allosteric network. *Proc. Natl. Acad. Sci. U. S. A.*, **117**, 24294–24304.
23. Zeng, M., Shang, Y., Araki, Y., Guo, T., Huganir, R.L., Zhang, M., (2016). Phase transition in postsynaptic densities underlies formation of synaptic complexes and synaptic plasticity. *Cell*, **166**, (1163–1175) e1112
24. Troilo, F., Bonetti, D., Camilloni, C., Toto, A., Longhi, S., Brunori, M., et al., (2018). Folding mechanism of the SH3 domain from Grb2. *J. Phys. Chem. B*, **122**, 11166–11173.
25. Guerois, R., Serrano, L., (2000). The SH3-fold family: experimental evidence and prediction of variations in the folding pathways. *J. Mol. Biol.*, **304**, 967–982.
26. Martínez, J.C., Serrano, L., (1999). The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved. *Nat. Struct. Biol.*, **6**, 1010–1016.
27. Martinez, J.C., Pisabarro, M.T., Serrano, L., (1998). Obligatory steps in protein folding and the conformational diversity of the transition state. *Nat. Struct. Biol.*, **5**, 721–729.
28. Grantcharova, V.P., Riddle, D.S., Santiago, J.V., Baker, D., (1998). Important role of hydrogen bonds in the structurally polarized transition state for folding of the src SH3 domain. *Nat. Struct. Biol.*, **5**, 714–720.
29. Geierhaas, C.D., Nickson, A.A., Lindorff-Larsen, K., Clarke, J., Vendruscolo, M., (2007). BPPred: a Web-based computational tool for predicting biophysical parameters of proteins. *Protein Sci.*, **16**, 125–134.
30. Myers, J.K., Pace, C.N., Scholtz, J.M., (1995). Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci.*, **4**, 2138–2148.
31. Matouschek, A., Fersht, A.R., (1993). Application of physical organic chemistry to engineered mutants of proteins: Hammond postulate behavior in the transition state of protein folding. *Proc. Natl. Acad. Sci. U. S. A.*, **90**, 7814–7818.
32. Schatzle, M., Kiefhaber, T., (2006). Shape of the free energy barriers for protein folding probed by multiple perturbation analysis. *J. Mol. Biol.*, **357**, 655–664.
33. Sanchez, I.E., Kiefhaber, T., (2003). Hammond behavior versus ground state effects in protein folding: evidence for narrow free energy barriers and residual structure in unfolded states. *J. Mol. Biol.*, **327**, 867–884.
34. Perez-Riba, A., Synakewicz, M., Itzhaki, L.S., (2018). Folding cooperativity and allosteric function in the tandem-repeat protein class. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, **373**
35. Gianni, S., Freiberger, M.I., Jemth, P., Ferreiro, D.U., Wolynes, P.G., Fuxreiter, M., (2021). Fuzziness and frustration in the energy landscape of protein folding, function, and assembly. *Acc. Chem. Res.*,.
36. Ferreiro, D.U., Komives, E.A., Wolynes, P.G., (2018). Frustration, function and folding. *Curr. Opin. Struct. Biol.*, **48**, 68–73.
37. Gianni, S., Camilloni, C., Giri, R., Toto, A., Bonetti, D., Morrone, A., et al., (2014). Understanding the frustration arising from the competition between function, misfolding, and aggregation in a globular protein. *Proc. Natl. Acad. Sci. U. S. A.*, **111**, 14141–14146.
38. Chi, C.N., Gianni, S., Calosci, N., Travaglini-Allocatelli, C., Engström, Å., Jemth, P., (2007). A conserved folding mechanism for PDZ domains. *FEBS Lett.*, **581**, 1109–1113.
39. Calosci, N., Chi, C.N., Richter, B., Camilloni, C., Engström, Å., Eklund, L., et al., (2008). 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.
40. A. Fersht, U.A. Fersht, W.H. Freeman, Company, Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding: W. H. Freeman, 1999.