Archives of Biochemistry and Biophysics 731 (2022) 109444

Contents lists available at ScienceDirect

Archives of Biochemistry and Biophysics

journal homepage: www.elsevier.com/locate/yabbi

Exploring the effect of tethered domains on the folding of Grb2 protein

Livia Pagano, Valeria Pennacchietti, Awa Diop, Daniele Santorelli, Paola Pietrangeli, Lucia Marcocci, Caterina Nardella, Francesca Malagrinò, Angelo Toto **, Stefano Gianni *

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

ARTICLE INFO

Keywords:
Kinetics
Multidomain folding
Misfolding
SH2
SH3

ABSTRACT

Two thirds of eukaryotic proteins have evolved as multidomain constructs, and *in vivo*, domains fold within a polypeptide chain, with inter-domain interactions possibly crucial for correct folding. However, to date, most of the experimental folding studies are based on domains in isolation. In an effort to better understand multidomain folding, in this work we analyzed, through equilibrium and kinetic folding experiments, the folding properties of the Growth factor receptor-bound protein 2 (Grb2), composed by one SRC homology 2 domain flanked by two SRC homology 3 domains. In particular we compared the kinetic features of the multidomain construct with the domains expressed in isolation. By performing single and double mixing folding experiments, we demonstrated that the folding of the SH2 domain is kinetically trapped in a misfolded intermediate when tethered to the C-SH3. Importantly, within the multidomain construct, misfolding occurred independently if refolding is started with C-SH3 in its unfolded or native state. Interestingly, our data reported a peculiar scenario, in which SH2 and C-SH3 domain reciprocally and transiently interact during folding. Altogether, the analysis of kinetic folding data provided a quantitative description of the multidomain folding of Grb2 protein, discussed under the light of previous works on multidomain folding.

1. Introduction

The multidomain nature of proteins is at the heart of the numerous activities and interactions essential for cell development and homeostasis. This plethora of vital and complex functions is ensured by a limited pool of domain families that are shuffled and in some cases duplicated within the protein architecture [1]. However, even though the analysis of sequenced genomes have demonstrated that a significant proportion of proteins consist in more than one domain [2–5], much of theoretical and experimental work in the protein folding field have been focused on globular and isolated domains. Given the multi-modular structure of proteins, a complete comprehension of the protein-folding problem may not be achieved without considering the supramodular context to which single domains are exposed [6–10].

In the last decades the growing attention on the multidomain protein folding have paralleled the ongoing medical research on pathologies induced by protein misfolding; at cellular level, co-translational folding seems to avoid misfolding events by gradually fold the nascent polypeptide in a "domain-by-domain" manner [11]. However, for large multidomain protein, with a sufficient long lifetime, (un)folding processes may occur several times [12]. In particular, while a single and independent domain can fold by following its minimally frustrated energy landscape [13], in a multidomain system, the inter-domain interactions can vary the frustration pattern, resulting in an increase of the probability of misfolded states [9,10,12,14–17]. Determining the molecular bases of such inter-domain interactions can provide information on how domains affect, interact and communicate within a more complex multidomain system.

In an effort to better understand multidomain folding we investigated the folding pathway of the Growth factor receptor-bound protein 2 (Grb2). Grb2 is an adapter protein composed by three interaction modules: one SRC homology 2 (SH2) flanked by two terminal SRC homology 3 domains (N-SH3 and C-SH3). This structural arrangement enables Grb2 to interact with different cellular partners, playing an essential role in the communication between the cell surface growth factor receptors and the RAS signaling pathway [18–21]. While the allosteric regulation of the binding of the full-length protein with its physiological partners (e.g. SOS1 and HER2) has been investigated using

* Corresponding author.
** Corresponding author.
E-mail addresses: angelo.toto@uniroma1.it (A. Toto), stefano.gianni@uniroma1.it (S. Gianni).

https://doi.org/10.1016/j.abb.2022.109444
Received 30 July 2022; Received in revised form 27 September 2022; Accepted 14 October 2022
Available online 18 October 2022

different experimental techniques [22–25], how the full-length protein folds in respect to the isolated domains has not been studied so far.

Here, we reported an analysis of the folding profile of Grb2 protein, comparing the kinetic features of the multidomain construct with the domains expressed in isolation. In particular our data are compatible with a complex folding mechanism implying the accumulation of a misfolded kinetic trap in the refolding of the SH2 domain only when linked to its neighboring C-SH3 domain, suggesting the presence of a specie that compete with productive folding. Interestingly, our data points to the evidence that SH2 populates the misfolded kinetic trap independently to the folding state of the C-SH3 domain. Moreover, the refolding rate constant of C-SH3 appears to be remarkably slowed down in experimental conditions in which the SH2 domain is highly destabilized. Under the light of previous and emerging studies on multidomain protein folding we argue that Grb2 is an "out of line" protein, giving an additional glimpse in the complexity of multidomain folding.

2. Materials and methods

2.1. Protein expression and purification

DNA encoding for N-SH3, C-SH3, SH2, C-SH3SH2 tandem and full-length Grb2 were transformed into *Escherichia coli* BL21 (DE3) cells for protein expression. Bacterial cells were grown in LB medium, containing 30 μg/ml of kanamycin, at 37 °C until $OD_{600} = 0.7–0.8$, and then protein expression was induced with 0.5 mM IPTG. After induction, cells were grown at 25 °C overnight and then collected by centrifugation. Each construct (His-tagged) was purified from the soluble fraction in 50 mM Tris-HCl at pH 7.5 with 0.3 mM NaCl with a HiTrap Chelating High-Performance column (GE Healthcare) and then eluted with a gradient to 1 M imidazole. The imidazole was removed using a HiPrep 26/10 Desalting column (GE Healthcare). The purity of the protein sample was confirmed by sodium dodecyl sulfate polyacrylamide gel electrophoresis.

2.2. Equilibrium experiments

Fluorescence equilibrium (un)folding experiments were performed on a standard spectrofluorometer (FluoroMax single photon counting spectrofluorometer; Horiba). The proteins were excited at 280 nm, and emission spectra were recorded between 300 and 400 nm at increasing Guanidinium chloride (GdnHCl) concentrations. Experiments were performed with all the constructs at constant concentration of 2 μM, in Hepes 50 mM at pH 7.5, at 25 °C, using a quartz cuvette with a path length of 1 cm. Data were fitted using Equation (1):

$$Y_{obs} = Y_N + Y_D \frac{e^{\left(m_{D-N}\left([GdnHCl]-[GdnHCl]_{1/2}\right)\right)}}{1 + e^{\left(m_{D-N}\left([GdnHCl]-[GdnHCl]_{1/2}\right)\right)}} \quad (1)$$

where Y_N and Y_D represent the baselines for the 100% native and 100% denatured protein populations, and $m_{D\text{-}N}$ is the calculated m-value for the unfolding reaction (see Results for details).

2.3. Stopped-flow experiments

Single mixing folding experiments were performed on an Applied Photophysics Pi-star 180 stopped-flow apparatus, monitoring the change of fluorescence emission, exciting the sample at 280 nm and recording the fluorescence emission by using a 360 nm cutoff glass filter. The experiments were performed at 25 °C, by using GdnHCl as denaturant agent. The final protein concentration was 2 μM for all the proteins analyzed. For each experiment, an average calculated from at least 5 independent traces was satisfactorily fitted with a single exponential equation. Data regarding the N-SH3 and the C-SH3 domains were fitted using a two-state equation [26]:

$$k_{obs} = k_f^0 \exp\left(-m_f[GdnHCl]\right) + k_u^0 \exp\left(m_u[GdnHCl]\right) \quad (2)$$

Where k_f^0 and k_u^0 are the microscopic folding and unfolding rate constants in the absence of denaturant; m_f and m_u are the associated m-values (see Results section for details).

Grb2 full-length and SH2 kinetic folding data were fitted using the following equation [27]:

$$k_{obs} = \frac{k_{IN}^0 \exp\left(-m_{IN}[GdnHCl]\right)}{\left(1 + K_{DI}\, exp(m_{DI}[GdnHCl]\right)} + k_{NI}^0 \exp\left(m_{NI}[GdnHCl]\right) \quad (3)$$

Where k_{IN}^0 and k_{NI}^0 are the microscopic folding and unfolding rate constants in the absence of denaturant of the intermediate and native states, with the associated m_{IN} and m_{NI} values. K_{DI} is the equilibrium rate constant between the denatured and intermediate state.

The buffers used for the pH dependence were: 50 mM Acetate pH 5, 50 mM Acetate pH 5.5, 50 mM Bis-tris pH 6, 50 mM Bis-tris pH 6.5, 50 mM Hepes pH 7, 50 mM Hepes pH 7.5, 50 mM Tris-HCl pH 8, 50 mM Tris-HCl pH 9. The pH-jump experiment was performed as follows: in one syringe the full-length protein Grb2 was diluted in buffer 50 mM Hepes at pH 7.5 with 2 M GdnHCl (mild denaturant condition) that was rapidly mixed in a 1:10 dilution with the refolding buffer 50 mM Acetate at pH 5. The refolding traces obtained were satisfactorily fitted with a double exponential.

Double-jump interrupted unfolding experiments were carried out on an Applied Photophysics stopped-flow instrument with double-jump capability at an excitation wavelength of 280 nm. Fluorescence emissions were measured using a 360 nm cutoff glass filter. Unfolding was initiated by symmetric mixing (1:1 diluition) of the native Grb2 with Hepes 50 mM buffer, pH 7.5, 4 M GdnHCl, leading to a final 2 M denaturant concentration. In the second mix, refolding was induced by challenging the solution (1:1 diluition) with Hepes 50 mM buffer, pH 7.5, with final 1 M denaturant concentration. The calculated amplitudes were plotted as a function of delay times and fitted to a double exponential decay.

3. Results

3.1. Equilibrium unfolding of Grb2 and its constituent domains

The stability of the four constructs (N-SH3, C-SH3, SH2 isolated domains and full-length Grb2) has been studied through a GdnHCl-induced equilibrium denaturation, following the intrinsic fluorescence of the tryptophan residues W36, W60, W121, W193 and W194 (Fig. 1).

In order to avoid multimerization, cysteine residues in positions 32 and 198 were mutated into serine. The decrease of emission fluorescence of tryptophan residues was observed upon denaturation (as reported in Fig. 2), and all the transitions could be satisfactorily fitted with a two-state model equation (Equation (1)). Quantitative analysis of the (un) folding profiles returned the $m_{D\text{-}N}$ values summarized in Table 1.

The $m_{D\text{-}N}$ parameter represents the dependence of the free energy of unfolding on denaturant concentration (defined as $\partial\Delta G/\partial$[denaturant]) and for a two-state system correlates with the change in accessible surface area upon unfolding [28]. While for the isolated N-SH3 domain we could not reliably measure thermodynamic unfolding parameters, C-SH3 and SH2 reported m-values that were consistent with the size of the domains. Interestingly, the full-length protein showed a $m_{D\text{-}N}$ value of 1.50 ± 0.10 kcal $mol^{-1}M^{-1}$, lower than what is expected from a protein of 215 residues [28]. It is worth noticing that the analysis of equilibrium unfolding of the isolated domains show very similar midpoints of the denaturation curves. This suggests that the apparent two-state transition reported in Fig. 2 might hide the independent unfolding processes of individual domains in the context of the full length protein, leading to a miscalculation of the $m_{D\text{-}N}$ value [29].

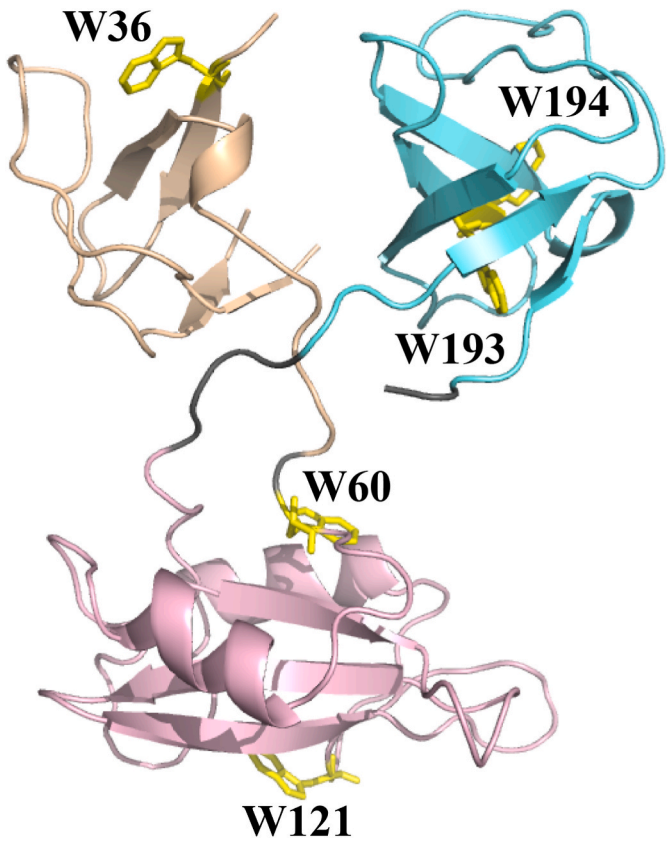

Fig. 1. Cartoon representation of the Grb2 protein (PDB code: 1GRI). The three domains N-SH3, SH2 and C-SH3 are colored in orange, pink and light blue respectively. Naturally occurring tryptophan residues (W36, W60, W121, W193 and W194) are highlighted in the structure in yellow sticks. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

3.2. Folding kinetics of the N-SH3, SH2 and C-SH3 in isolation and within multidomain systems

As described above, the main goal of this work is to characterize the differences in the folding features of the individual domains of Grb2 in isolation and in the context of multidomain constructs. The (un)folding kinetics were monitored by using stopped-flow technique, following the time courses of the refolding and unfolding reactions at different GdnHCl concentrations. The semilogarithmic plot of the observed rate constants versus denaturant concentrations (chevron plot) is reported in Fig. 2. Interestingly, while for isolated domains all traces were satisfactorily fitted with a single exponential equation, traces acquired for the full-length construct clearly followed a double exponential decay (data shown in Fig. 1 in the S.I.). The superimposition of the chevron plots of the isolated domains with the full-length protein allows to ascribe each kinetic phase to the (un)folding of different isolated domains. In fact, while the N-SH3 domain in the context of the full-length protein resulted optically silent (probably because of a quenching effect exerted by the neighboring tryptophan residues on W36), the two k_{obs} obtained for the full length Grb2 matched almost perfectly the k_{obs} calculated for the isolated SH2 and C-SH3 isolated domains.

Chevron plots obtained for the isolated N-SH3 and C-SH3 domains were fitted with a two-state equation, suggesting the absence of any detectable intermediate [26]. Notably, the chevron plot of the isolated N-SH3 domain displays a poorly characterized refolding arm, suggesting a low thermodynamic stability of the N-SH3 domain as reported by equilibrium unfolding experiments. On the other hand, significant differences have been addressed in the SH2 refolding branch showing a clear deviation from linearity, a typical signature of the presence of intermediate(s) along the folding pathway [27,30–34]. SH2 kinetic data were fitted with a three-state equation implying the presence of an intermediate accumulating along the reaction pathway (Equation (3)) (kinetic parameters are reported in Table 2). Interestingly, a comparison of kinetic data obtained for the SH2 domain in isolation and in the context of the full-length Grb2 reveals a remarkably slowed down refolding in the latter case, differing by one order of magnitude (77 ± 30 s^{-1} and 5.3 ± 1.3 s^{-1} respectively). Moreover, it appears evident that, in

Table 1
Equilibrium folding parameters of the Grb2 full length, N-SH3, SH2 and C-SH3 domains. Midpoints and $m_{D\text{-}N}$ values were obtained from the fitting process of the equilibrium experiments. ΔG_{H2O} were calculated as $m_{D\text{-}N}{}^{*}[d]_{50\%}$. (/= Thermodynamic parameters regarding N-SH3 domain could not be reliably calculated).

Equilibrium parameters	n° of amino acids	ΔG_{H2O} (kcal mol^{-1})	$[d]_{50\%}$ (M)	$m_{D\text{-}N}$ (kcal mol^{-1} M^{-1})
Grb2 FL	215	3.05 ± 0.09	2.01 ± 0.07	1.51 ± 0.12
N-SH3	58	/	/	/
SH2	93	3.43 ± 0.07	1.73 ± 0.04	1.98 ± 0.14
C-SH3	59	2.05 ± 0.01	1.15 ± 0.01	1.77 ± 0.02

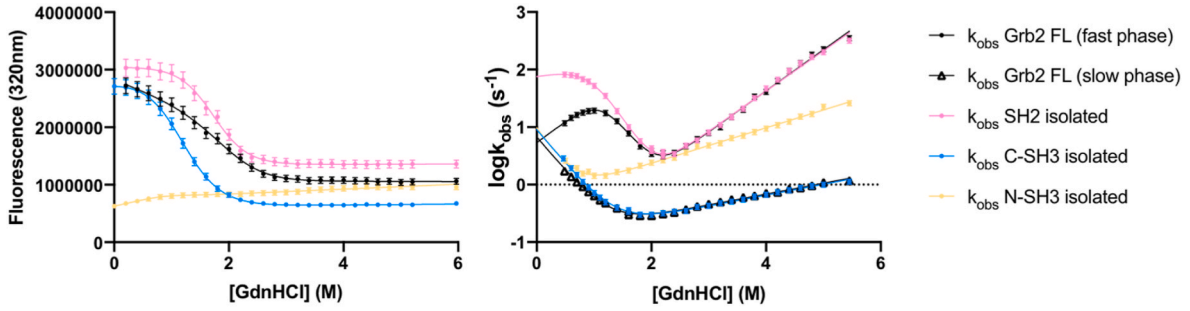

Fig. 2. Left panel - GdnHCl-induced equilibrium denaturation of Grb2 (black), SH2 domain (pink), N-SH3 (orange) and C-SH3 domain (light blue). All the denaturation curves were fitted with a two-state model equation described in Materials and Methods section. Thermodynamic parameters obtained from the fit are reported in Table 1. Right panel -– Superimposition of the chevron plots of the Grb2 full-length and the three individual domains. The full-length protein is represented in black circles (faster phase) and empty triangles (slower phase) whereas the single domains follow the same color code of the equilibrium denaturation. Chevron plots of N-SH3 and C-SH3 domains were fitted with a two-state equation (see materials and methods section). Kinetic data of the SH2 domain in the full-length and in isolation were fitted with an equation describing a three-state folding mechanism (see text for details and Materials and Methods section). All the experiments were performed in buffer Hepes 50 mM pH 7.5, at 25 °C. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 2
Kinetic (un)folding data were calculated by fitting chevron plots with a two-state equation for isolated N-SH3 and C-SH3 domains and for the C-SH3 domain in the context of the full length Grb2 construct, and a three-state equation for isolated SH2 domain and the SH2 domain in the context of the full-length Grb2 (equations are reported in the Materials and Methods section). As detailed in the text, for the full-length construct we could monitor two phases ascribable to the (un)folding of SH2 and C-SH3 domains.

	k_f (s^{-1})	k_u (s^{-1})	m_f (kcal mol^{-1} M^{-1})	m_u (kcal mol^{-1} M^{-1})	$\Delta G_{D\text{-}N}$ (kcal mol^{-1})
Grb2FL (C-SH3)	5.6 ± 0.4	0.11 ± 0.01	1.50 ± 0.05	0.26 ± 0.01	2.3 ± 0.07
C-SH3 isolated	10.3 ± 0.5	0.12 ± 0.01	1.7 ± 0.1	0.25 ± 0.01	2.6 ± 0.06
N-SH3 isolated	9.0 ± 2.0	0.56 ± 0.02	1.9 ± 0.2	0.41 ± 0.01	1.6 ± 0.13

	k_{IN} (s^{-1})	k_{NI} (s^{-1})	K_{DI}	m_{IN} (kcal mol^{-1} M^{-1})	m_{NI} (kcal mol^{-1} M^{-1})	m_{DI} (kcal mol^{-1} M^{-1})	$\Delta G_{D\text{-}N}$ (kcal mol^{-1})
Grb2FL (SH2)	5.3 ± 1.3	0.06 ± 0.01	2.0 ± 0.8	−1.1 ± 0.2	0.95 ± 0.01	2.9 ± 0.2	3.1 ± 0.3
SH2 isolated	77 ± 30	0.06 ± 0.01	3.0 ± 1.0	0.11 ± 0.01	0.93 ± 0.07	2.0 ± 0.1	4.9 ± 0.3

the full-length protein, the k_{obs} obtained in refolding experiments increased at increasing denaturant concentrations. This evidence is compatible with a rapid accumulation of a kinetic trap populated during refolding, i.e. a misfolded intermediate state that must unfold in order to reach the native state [9]. To exclude this effect to be caused by aggregates formation, we performed concentration-dependent refolding experiments (reported in Fig. S2). Importantly, protein concentration did not affect the calculation of refolding rate constants, confirming the monomolecular nature of the folding reaction.

3.3. The folding state of C-SH3 domain does not affect the folding of SH2

To better interpret the effect of the presence of contiguous domains in the folding of the full-length Grb2 we resorted to conduct kinetic and equilibrium experiment on a construct composed by the SH2 and the C-SH3 tandem (Grb2ΔN-SH3). In particular, by following this strategy, we could pinpoint possible inter-domain interactions, occurring between the SH2 and the C-SH3 domain, responsible for the formation of the misfolded specie. The equilibrium denaturation of Grb2ΔN-SH3 is reported in the supplementary information (Fig. S3). A comparison of the chevron plots of Grb2 and Grb2ΔNSH3 is reported in Fig. 3. Clearly, kinetic data obtained from the two constructs are comparable, suggesting that the kinetic trap populated by the SH2 domain may be due to transient interactions with the covalently linked C-SH3 during refolding.

As previously shown for the PDZ1-PDZ2 tandem of Whirlin protein [9], misfolding events may arise depending on the folding state of contiguous domains; in fact, in the case of Whirlin, whenever the PDZ2 domain retained its native conformation, the folding of PDZ1, in the context of the tandem, seemed to be equivalent to PDZ1 domain in isolation, while a misfolded kinetic trap was populated when refolding started from both PDZ1 and PDZ2 denatured. On the basis of this evidence, and following an analogous strategy, we compared the folding of the SH2 domain in the presence of the C-SH3 helding its native conformation and in its unfolded state. To do so, we performed kinetic double-mixing interrupted unfolding experiments. In the first mix, native Grb2ΔNSH3 was rapidly mixed (1:1 dilution) with denaturing buffer (Hepes 50 mM pH 7.5, 4 M GdnHCl), leading to a final 2 M denaturant concentration. Then, after a controlled delay time, the solution was challenged with a renaturing buffer, inducing the consequent refolding at 1 M denaturant final concentration. The fluorescence exponential traces, collected at different delay times (ranging from 0.02s to 40s), revealed a double exponential decay.

The dependence of the amplitudes of the refolding traces versus the delay time was fitted to a double exponential equation (Fig. 3). While at long delay time (i.e. allowing the C-SH3 domain to unfold before the second mix) the amplitude analysis showed a slow phase of 0.11 ± 0.07 s^{-1}, probably relative to prolines *cis-trans* isomerization, at shorter delay times (retaining the C-SH3 domain in its native conformation) we could calculate a fast phase of 3.70 ± 0.70 s^{-1}. This latter is comparable to the refolding rate constant obtained for the SH2 in the full length Grb2 at 2 M GdnHCl, being 3.40 ± 0.02 s^{-1}. These data suggest that, whether the C-SH3 retains its native conformation, or it is unfolded, no significant differences can be measured in the refolding pathway of the SH2 within the multidomain construct.

3.4. The stability of SH2 domain influences the folding of C-SH3

A largely used method to obtain information about the folding mechanism of a protein is to perturb the stability of the protein by varying the experimental conditions (pH, temperature, salt concentration) and monitor how the system responds to the perturbations in terms of kinetic parameters [35]. We therefore investigated the folding of Grb2 at different pH (from pH 4 to pH 9) and the data obtained are shown in Fig. 4. In all conditions the kinetic traces displayed two phases related to the two optically active domains (SH2 and C-SH3). The analysis of the chevron plots revealed that while the stability of the C-SH3 domain appears to be slightly affected by the change of pH conditions, the SH2 domain is strongly destabilized at acidic pH, with a dramatic increase in the microscopic unfolding rate constant and low resolution of refolding arm of chevron plots at acidic pH. In particular, at pH 5.0 no refolding

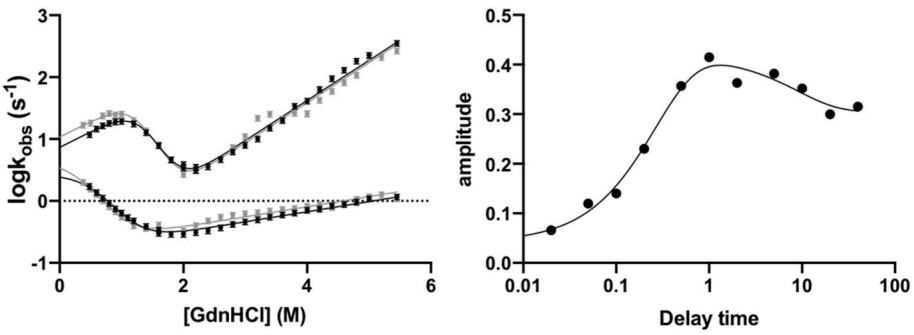

Fig. 3. Left panel - Chevron plot of the Grb2 full-length (black circles) and C–SH3 – SH2 tandem (grey circles). Kinetic data from the two constructs appear to be very similar, suggesting that the N-terminal domain does not affect the formation of the kinetic trap. Chevron plots obtained from the dependence of fast and slow (un)folding phases of C–SH3 – SH2 tandem and full-lenght Grb2 were fitted respectively by sharing kinetic m-values. Right panel - Interrupted unfolding double jump experiment. Amplitudes of refolding traces were plotted versus delay times. Line represents the best fit to a double exponential equation. The fit of the first exponential phase reports a very similar k_{obs} (3.7 ± 0.7 s^{-1}) to the one obtained for the SH2 domain in the context of the full-length protein at 2 M denaturant concentration (k_{obs} = 3.40 ± 0.02 s^{-1}). The slower k_{obs} (0.11 ± 0.07 s^{-1}) may be ascribable to possible *cis-trans* proline isomeration.

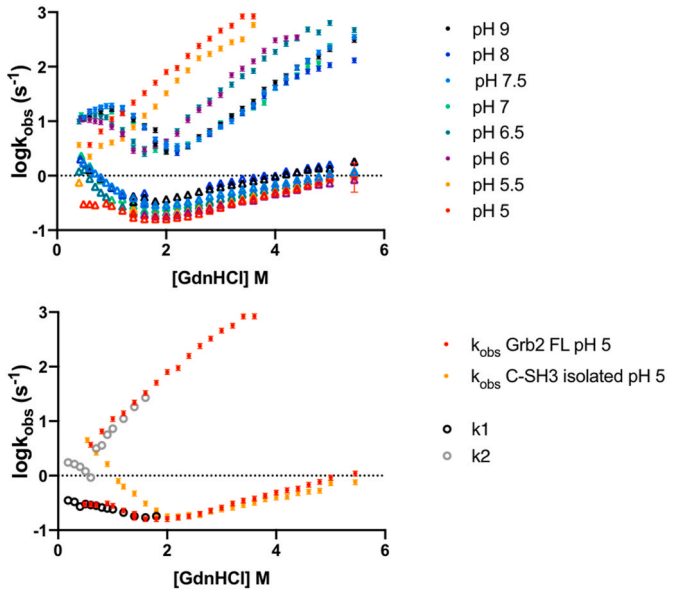

Fig. 4. Upper panel - Kinetic (un)folding profile of full-length Grb2 at different pH conditions, from pH 9 to pH 5 (the experimental setting is provided in Materials and Method sections). Lower panel - Chevron plots of Grb2 (red circles) and C-SH3 in isolation (orange circles) at pH 5. . As detailed in the text, the experiment was performed by inducing the refolding of Grb2 (partially unfolded at 50 mM Hepes 2 M GdnHCl pH 7.5) with a renaturing buffer 50 mM Acetate at pH 5. The traces recorded in the refolding experiment showed two phases. From the fit of the double exponential we could obtain two observed rate constants represented in empty black and grey circles Whilst black empty circles (k1) are compatible with the slow phase of the C-SH3 domain in the same final conditions (pH 5), we could measure an additional faster phase k2 (grey empty circles) that matches the k_{obs} related to the C-SH3 at pH 7.5, condition in which the SH2 is folded. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

traces could be measured for SH2, the domain being incapable to reversibly fold in those conditions. On the other hand, a clear roll-over effect in the C-SH3 refolding branch became more pronounced at lower pH values, with a progressive decrease in the microscopic folding rate constant. Interestingly, the chevron plot of the C-SH3 domain in isolation at pH 5.0 is compatible with a two-state folding mechanism, showing linear unfolding and refolding arms.

To further investigate whether the decreased stability of SH2 domain at acidic pH influenced the folding behavior of C-SH3 we designed a refolding pH-jump experiments. This strategy allowed us to monitor the refolding kinetics of C-SH3 at pH 5.0 starting from experimental conditions in which the SH2 domain was held in its native conformation. In particular, to observe the folding of C-SH3 when linked to the natively folded SH2, we diluted the protein in mild denaturant concentration (2 M GdnHCl, pH 7.5) and the solution was rapidly mixed (in a 1:10 dilution) with the refolding buffer to reach final conditions of Acetate 50 mM pH 5.0. The comparison of chevron plot at pH 5 and the k_{obs} obtained from pH-jump experiment (empty circles) is reported in the right panel of Fig. 4. The traces collected in the pH-jump experiment were consistent with a double exponential decay. Whilst lower k_{obs} are very similar to the ones obtained for the Grb2 construct at pH 5.0, we could monitor a faster phase (empty grey circles) matching the k_{obs} ascribable to the C-SH3 at pH 7.5, condition in which SH2 is folded. Altogether our results suggest that the reduced stability of the SH2 domain at pH 5.0 affects the refolding of the C-SH3 domain in the context of a SH2-CSH3 tandem.

4. Discussion

The human proteome is composed by over 75% of multidomain proteins and yet, our understanding on multidomain folding is mostly limited to smaller individual domain constructs. The reason of such downsizing is not only due to the challenging interpretations of the experimental data, but also to the assumption that the folding features of a single domain can be applied to a more general system [36–39].

Expanding the folding field to more complex topologies requires the analysis of inter-domain interactions. The biophysical work of Wolynes and coworkers on the *in silico* prediction of the frustration patterns of multidomain protein [40] emphasized how the frustration originated from interdomain interactions can easily lead to misfolding. On the other hand, protein evolved by lowering the sequence identity of neighboring domains as a strategy to decrease the probability of misfolding events to occur [12].

In this scenario, our data on Grb2 folding confirm and, at the same time, disprove some of the major assumptions of the multidomain folding. It is, in fact, of interest to compare Grb2 folding with what has been previously seen for the PDZ1-PDZ2 tandem from Whirlin [9]. The remarkable difference in thermodynamic stability between the PDZ1 and PDZ2 allowed to perform refolding experiments by selectively unfold both domains or helding PDZ2 native. We demonstrated that, while the folding kinetics of PDZ1 were identical for the domain in isolation or in the presence of native PDZ2, a misfolded kinetic trap was populated when refolding started from high denaturant concentrations, condition in which both domains were unfolded. Peculiarly, the misfolded intermediate showed native-like functional properties, being able to bind a physiological ligand of Whirlin, providing the evidence of a functionally competent misfolded state.

In the case of Grb2, as reported by folding kinetic data on the full-length protein, the folding of the SH2 domain is kinetically trapped only when it is tethered to the C-SH3, even though they have very different topologies and share just 11.86% of sequence identity (calculated using Clustalw). Moreover, kinetic data obtained for the Grb2ΔNSH3 construct and their comparison with isolated domains revealed a peculiar scenario, in which the SH2 domain folding is trapped in a misfolded intermediate independently to the folding state of C-SH3. In addition, the folding of C-SH3 domain is also affected by the presence of SH2, in particular in experimental conditions in which the SH2 is strongly destabilized. This evidence supports the hypothesis of reciprocal transient interactions occurring during the folding process of the two domains. In a recent paper published by our group [41] we characterized the folding and binding properties of a tandem of PDZ domains from the sPDZD2 protein. Our data highlighted the presence of a misfolded intermediate along the folding pathway that displayed an unexpected ability to bind the physiological ligand of sPDZD2 with increased affinity compared to the native state. Under this light, one possible reason of the presence of multiple energetic minima along the folding pathway of multidomain proteins may be represented by the ability to exert different functions, in particular in the case of protein-protein interaction domains, the possibility to interact with multiple partners and/or display different determinants of specificity and affinity.

Altogether our results bring novel and additional information in the study of the multidomain folding, providing the evidence that Grb2 does not fold as a “sum of the parts”, and proving divergences in the folding energy landscape of its domains when isolated and in the multidomain contexts, incurring into complex interdomain interactions and transient misfolding events that characterize a singular folding pathway. Future works based on extensive site-directed mutagenesis will be aimed to address the molecular basis of such interdomain interactions, possibly pinpointing residues with significant effect on the folding pathway of Grb2.

Funding sources

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 (RP11715C34AEA C9B and RM1181641C2C24B9, RM11916B414C897E, RG12017297F A7223 to S.G, AR22117A3CED340A to C.N.), by an ACIP grant (ACIP 485–21) from Institut Pasteur Paris to S.G., the Associazione Italiana per la Ricerca sul Cancro (Individual Grant – IG 24551 to S.G.), the Regione Lazio (Progetti Gruppi di Ricerca LazioInnova A0375-2020-36,559 to S. G.), European Union's Horizon 2020 research and Innovation programme under the Marie Skłodowska-Curie Grant Agreement UBIMOTIF No 860517 (to S.G.) the Istituto Pasteur Italia, "Teresa Ariaudo Research Project" 2018, and "Research Program 2022–2023 Under 45 Call 2020" (to A.T.).

Declaration of competing interest

All Authors declare no conflict of interests.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.abb.2022.109444.

References

[1] 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 (2007) 319–330.
[2] G. Apic, J. Gough, S.A. Teichmann, Domain combinations in archaeal, eubacterial and eukaryotic proteomes, J. Mol. Biol. 310 (2001) 311–325.
[3] D. Ekman, Å.K. Björklund, J. Frey-Skött, A. Elofsson, Multi-domain proteins in the three kingdoms of life: orphan domains and other unassigned regions, J. Mol. Biol. 348 (2005) 231–243.
[4] M. Gerstein, How representative are the known structures of the proteins in a complete genome? A comprehensive structural census, Folding Des. 3 (1998) 497–512.
[5] S.A. Teichmann, C. Chothia, M. Gerstein, Advances in structural genomics, Curr. Opin. Struct. Biol. 9 (1999) 390–399.
[6] 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 (2006) 18113–18118.
[7] S. Batey, K.A. Scott, J. Clarke, Complex folding kinetics of a multidomain protein, Biophys. J. 90 (2006) 2120–2130.
[8] L. Pagano, F. Malagrinò, L. Visconti, F. Troilo, V. Pennacchietti, C. Nardella, A. Toto, S. Gianni, Probing the effects of local frustration in the folding of a multidomain protein, J. Mol. Biol. 433 (2021), 167087.
[9] C. Gautier, F. Troilo, F. Cordier, F. Malagrinò, A. Toto, L. Visconti, Y. Zhu, M. Brunori, N. Wolff, S. Gianni, Hidden kinetic traps in multidomain folding highlight the presence of a misfolded but functionally competent intermediate, Proc. Natl. Acad. Sci. U. S. A 117 (2020) 19963–19969.
[10] P. Tian, R.B. Best, Structural determinants of misfolding in multidomain proteins, PLoS Comput. Biol. 12 (2016), e1004933.
[11] C.A. Waudby, C.M. Dobson, J. Christodoulou, Nature and regulation of protein folding on the ribosome, Trends Biochem. Sci. 44 (2019) 914–926.
[12] M.B. Borgia, A. Borgia, R.B. Best, A. Steward, D. Nettels, B. Wunderlich, B. Schuler, J. Clarke, Single-molecule fluorescence reveals sequence-specific misfolding in multidomain proteins, Nature 474 (2011) 662–665.
[13] S. Gianni, M.I. Freiberger, P. Jemth, D.U. Ferreiro, P.G. Wolynes, M. Fuxreiter, Fuzziness and frustration in the energy landscape of protein folding, function, and assembly, Acc. Chem. Res. 54 (2021) 1251–1259.
[14] A. Borgia, K.R. Kemplen, M.B. Borgia, A. Soranno, S. Shammas, B. Wunderlich, D. Nettels, R.B. Best, J. Clarke, B. Schuler, Transient misfolding dominates multidomain protein folding, Nat. Commun. 6 (2015) 8861.
[15] V. Kumar, T.K. Chaudhuri, Spontaneous refolding of the large multidomain protein malate synthase G proceeds through misfolding traps, J. Biol. Chem. 293 (2018) 13270–13283.
[16] L. Visconti, F. Malagrinò, F. Troilo, L. Pagano, A. Toto, S. Gianni, Folding and misfolding of a PDZ tandem repeat, J. Mol. Biol. 433 (2021), 166862.
[17] A. Lafita, P. Tian, R.B. Best, A. Bateman, Tandem domain swapping: determinants of multidomain protein misfolding, Curr. Opin. Struct. Biol. 58 (2019) 97–104.
[18] A. Giubellino, T.R. Burke, D.P. Bottaro, Grb2 signaling in cell motility and cancer, Expert Opin. Ther. Targets 12 (2008) 1021–1033.
[19] P.G. Dharmawardana, B. Peruzzi, A. Giubellino, T.R. Burke, D.P. Bottaro, Molecular targeting of growth factor receptor-bound 2 (Grb2) as an anti-cancer strategy, Anti Cancer Drugs 17 (2006) 13–20.
[20] K. Neumann, T. Oellerich, H. Urlaub, J. Wienands, The B-lymphoid Grb2 interaction code, Immunol. Rev. 232 (2009) 135–149.
[21] J.A. Simon, S.L. Schreiber, Grb2 SH3 binding to peptides from Sos: evaluation of a general model for SH3-ligand interactions, Chem. Biol. 2 (1995) 53–60.
[22] N.S. Kazemein Jasemi, C. Herrmann, E. Magdalena Estirado, L. Gremer, D. Willbold, L. Brunsveld, R. Dvorsky, M.R. Ahmadian, The intramolecular allostery of GRB2 governing its interaction with SOS1 is modulated by phosphotyrosine ligands, Biochem. J. 478 (2021) 2793–2809.
[23] J. Schlessinger, M.A. Lemmon, SH2 and PTB domains in tyrosine kinase signaling, Sci. STKE (2003) 2003.
[24] N. Bisson, D.A. James, G. Ivosev, S.A. Tate, R. Bonner, L. Taylor, T. Pawson, Selected reaction monitoring mass spectrometry reveals the dynamics of signaling through the GRB2 adaptor, Nat. Biotechnol. 29 (2011) 653–658.
[25] C.B. McDonald, K.L. Seldeen, B.J. Deegan, V. Bhat, A. Farooq, Assembly of the Sos1-Grb2-Gab1 ternary signaling complex is under allosteric control, Arch. Biochem. Biophys. 494 (2010) 216–225.
[26] S.E. Jackson, A.R. Fersht, Folding of chymotrypsin inhibitor 2 1 Evidence for a two-state transition, Biochemistry 30 (1991) 10428–10435.
[27] 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 (1995) 771–786.
[28] J.K. Myers, C. Nick Pace, J. Martin Scholtz, Denaturant *m* values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding, Protein Sci. 4 (1995) 2138–2148.
[29] S. Batey, A.A. Nickson, J. Clarke, Studying the folding of multidomain proteins, HFSP J. 2 (2008) 365–377.
[30] A. Matouschek, J.T. Kellis, L. Serrano, M. Bycroft, A.R. Fersht, Transient folding intermediates characterized by protein engineering, Nature 346 (1990) 440–445.
[31] S. Khorasanizadeh, I.D. Peters, H. Roder, Evidence for a three-state model of protein folding from kinetic analysis of ubiquitin variants with altered core residues, Nat. Struct. Biol. 3 (1996) 193–205.
[32] D. Bonetti, C. Camilloni, L. Visconti, S. Longhi, M. Brunori, M. Vendruscolo, S. Gianni, Identification and structural characterization of an intermediate in the folding of the measles virus X domain, J. Biol. Chem. 291 (2016) 10886–10892.
[33] A. Bachmann, T. Kiefhaber, Apparent two-state tendamistat folding is a sequential process along a defined route11Edited by A R Fersht, J. Mol. Biol. 306 (2001) 375–386.
[34] S. Gianni, N. Calosci, J.M.A. Aelen, G.W. Vuister, M. Brunori, C. Travaglini-Allocatelli, Kinetic folding mechanism of PDZ2 from PTP-BL, Protein Eng. Des. Sel. 18 (2005) 389–395.
[35] A. Fersht, Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding, WORLD SCIENTIFIC, 2017.
[36] S.E. Jackson, How do small single-domain proteins fold? Folding Des. 3 (1998) R81–R91.
[37] C.M. Dobson, Protein folding and misfolding, Nature 426 (2003) 884–890.
[38] E.J. Guinn, B. Jagannathan, S. Marqusee, Single-molecule chemo-mechanical unfolding reveals multiple transition state barriers in a small single-domain protein, Nat. Commun. 6 (2015) 6861.
[39] E.J. Guinn, P. Tian, M. Shin, R.B. Best, S. Marqusee, A small single-domain protein folds through the same pathway on and off the ribosome, Proc. Natl. Acad. Sci. USA 115 (2018) 12206–12211.
[40] W. Zheng, N.P. Schafer, P.G. Wolynes, Frustration in the energy landscapes of multidomain protein misfolding, Proc. Natl. Acad. Sci. USA 110 (2013) 1680–1685.
[41] F. Malagrinò, G. Fusco, V. Pennacchietti, A. Toto, C. Nardella, L. Pagano, A. de Simone, S. Gianni, Cryptic binding properties of a transient folding intermediate in a PDZ tandem repeat, Protein Sci. Publ. Protein Soc. 31 (2022), e4396.