



**HAL**  
open science

## PDZ Sample Quality Assessment by Biochemical and Biophysical Characterizations

Célia Caillet-Saguy, Sébastien Brûlé, Nicolas Wolff, Bertrand Raynal

► **To cite this version:**

Célia Caillet-Saguy, Sébastien Brûlé, Nicolas Wolff, Bertrand Raynal. PDZ Sample Quality Assessment by Biochemical and Biophysical Characterizations. Jean-Paul Borg. PDZ Mediated Interactions: methods and protocols, 2256, Humana Press, pp.89-124, 2021, Methods in molecular biology, 978-1-0716-1166-1 (ebook). 10.1007/978-1-0716-1166-1\_6 . hal-03234405

**HAL Id: hal-03234405**

**<https://hal.science/hal-03234405>**

Submitted on 25 May 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# PDZ sample quality assessment by biochemical and biophysical characterizations

Célia Caillet-Saguy<sup>1\*</sup>, Sébastien Brûlé<sup>2\*</sup>, Nicolas Wolff<sup>1§</sup>, Bertrand Raynal<sup>2</sup>

1. Institut Pasteur, Unité Récepteurs-Canaux, Institut Pasteur, CNRS UMR 3571, Paris, France.

2. Institut Pasteur, Plate-forme de Biophysique Moléculaire, CNRS UMR 3528, Paris, France.

\* These authors equally contributed to this work

§ Corresponding author: nicolas.wolff@pasteur.fr

## ABSTRACT

PDZ domains are small globular domains involved in protein-protein interactions. They participate to a wide range of critical cellular processes. These domains, very abundant in the human proteome, are widely studied by high-throughput interactomics approaches and by biophysical and structural methods. However, the quality of the results are strongly related to the optimal folding and solubility of the domains. We provided here a detailed description of protocols for a strict quality assessment of the PDZ constructs. We described appropriate experimental approaches that have been selected to overcome the small size of such domains to check the purity, the identity, the homogeneity, the stability and the folding of samples.

Running title : PDZ sample quality assessment

Key Words : PDZ domain; folding; protein sample; quality control ; biophysical techniques ; stability ; purity ; identity ; structure ; quantification

## 1. Introduction

PDZ domains are one of the most abundant protein-protein interaction domain found in metazoan. The human proteome contains 266 identified PDZ domains, termed the PDZome, spread over 152 proteins [1]. PDZ domains are mainly found in large scaffold proteins, existing with a wide variety of modular domains. The PDZ-mediated interactions participate in numerous critical cellular processes and are often targeted in pathologies and by numerous viral proteins during infection [2].

PDZs are globular domains of ~90 amino-acids residues, which recognize PDZ Binding Motifs (PBM), mainly at the extreme C-terminus of their partner proteins. They are recognized as well-folded and compact domains. The boundaries of most experimental PDZ constructs that have been used for structural studies or interaction assays, generally extend beyond the strict boundaries of the core PDZ structure. If PDZ domains are mainly robust, soluble domain, inappropriate delimitations can lead to unstable protein. Indeed, many studies have shown that extensions can influence the dynamics, stability and solubility of the PDZ domains [3] for often unknown reasons. The extension of N- and C-terminus can entropically stabilize the domain affecting its internal dynamics. Thus, the thermostability and the folding of PDZ domain are dependent in many cases on short disordered extensions at their two termini but also of adjacent modules forming homotypic and heterotypic PDZ supramodules [4]. These N- and C-terminal extensions could form functional and structural units playing a crucial role of activity modulation of the adjacent PDZ domain. Indeed, sequence context impacts the stability and solubility of constructs, and can deeply influence binding affinity and specificity, and structured and disordered extensions may affect the structure and function of the core PDZ domain [5].

Optimal folding and solubility of PDZ domains are needed to ensure accurate affinity measurement and structure determination. Conformational exchange can also be an intrinsic

property of PDZ domains; for example with the tandem of PDZ domains in whirlin [6] or of GRIP1 [7], and also the auto-association capacity of the unique PDZ domain of MAST2 (microtubule-associated serine and threonine kinase 2; [8]) or of the second PDZ domain of ZO2 [9]. Notably, the frequency of PDZ-PDZ interactions in eukaryotic proteins has been documented and dimerization occurs *in vitro* in 30% of the 157 PDZ domains tested [10].

Altogether, these examples illustrate the need for a strict quality assessment of the PDZ domains before deeply characterizing their biochemical, biophysical and functional properties. Appropriate experimental approaches should be selected to overcome the small size of such domains. Here, we report the contribution of a series of biochemical and biophysical approaches applied to PDZ domains constructs for sample quality evaluation. These analyses were used to assay the purity, the identity, the homogeneity, the stability and the folding of PDZ domain construct samples. We used as example the PDZ domain of the kinase protein MAST2 and detailed every step of this workflow, the principles, the techniques used and some protocols in the chapter below. The biochemical/biophysical characterizations are separated in four parts: purity, homogeneity, identity and conformational stability/folding state.

## **1.1 Purity**

### **1.1.1 SDS-PAGE electrophoresis**

SDS-PAGE electrophoresis (polyacrylamide gel electrophoresis containing sodium dodecyl sulfate) is a technique that separates proteins in a gel [11]. The migration is controlled by an electric field, enabling their separation according to their molecular weight.

The polyacrylamide gel is made by the co-polymerization of acrylamide and bis-acrylamide, in the presence of polymerization agents (TEMED, ammonium persulfate, for example). The concentration of acrylamide used controls the size of the pores; the higher the percentage, the

smaller the pore. SDS (sodium dodecyl sulfate) is a strong detergent with a long hydrophobic hydrocarbon tail and a negatively charged extremity. It interacts with proteins by binding their hydrophobic regions through its hydrocarbon part. By binding to the protein, the SDS prevents its folding. The native structure of the protein is therefore denatured, and an apparent negative charge is then conferred to the protein. In the presence of SDS, all the proteins will, therefore, have an apparent negative charge proportional to their polypeptide chain length. This means that only the molecular weight of the proteins will be the factor of their separation. Using the determined molecular mass, the presence of a given known protein will be evaluated.

### **1.1.2 Capillary gel electrophoresis**

SDS-capillary gel electrophoresis (SDS-CGE), also called capillary gel electrophoresis (CGE), is another technique to easily determine the purity of a PDZ construct. This technique is typically used to separate proteins according to their size in a capillary which is filled with polyacrylamide gel and SDS. The presence of SDS aids the electrophoretic mobility of proteins, as it coats their surface proportional to their size. Consequently, the molecular structure will have little influence on mobility, so macromolecules will migrate according to their molecular mass very similarly to the SDS-PAGE technique.

CGE has many advantages over classical SDS-PAGE, including on-capillary detection, high separation efficiency, the capability of accurate protein quantification and molecular weight determination and adaptable to high-throughput method with high reproducibility.

### **1.1.3 UV visible spectroscopy between 200 nm and 340 nm**

Absorption spectrophotometry is the measure of the attenuation of the light passing through a medium in order to obtain the concentrations of absorbent substances (chromophores).

PDZs can be analyzed by absorption spectrophotometry. These biomolecules have two main absorption bands in the near UV-region and do not absorb in the visible region (400 to 800 nm) of the electromagnetic spectrum. This near-UV absorption allows to quantify PDZs and provides information about their purity.

In the absorption spectrum of protein, a maximal absorption band can be found around 280 nm (less intense) and between 200 and 220 nm (more intense). The bands at 280 nm and 205 nm correspond to transitions  $\pi \rightarrow \pi^*$  in aromatic  $\alpha$ -amino acids (tyrosine, phenylalanine and tryptophan) and in amide groups (peptide bonds), respectively. At 280 nm, tyrosine, cystines and tryptophan have specific molar absorptivities that allow to calculate a molar absorption coefficient for a given protein sequence (<http://web.expasy.org/protparam/>; [12]). The main contribution at 280 nm comes from the tryptophan and tyrosine with an absorbance of  $5580 \text{ M}^{-1} \cdot \text{cm}^{-1}$  and  $1480 \text{ M}^{-1} \cdot \text{cm}^{-1}$ , respectively. The quantification at 280 nm is therefore suitable in presence of at least 1 tryptophan or 1 tyrosine in the PDZ sequence.

The molar absorption relates the measured absorbance to the concentration of the protein through the Beer-Lambert law:  $A(\lambda) = \epsilon \cdot l \cdot c$  where  $A$  is the absorbance at a given wavelength  $\lambda$  (280 nm for a protein),  $\epsilon$  (in  $\text{M}^{-1} \text{ cm}^{-1}$ ) is the molar absorption coefficient (also called the extinction coefficient) for a given PDZ and  $c$  is the concentration of PDZ of interest (in M)(see Note 1). These properties can be used in protein analysis, either to identify protein-containing fractions or to determine the concentration of protein in a purified sample. Quantification of PDZs in a complex mixture (eg isolated fractions of cells) by UV visible absorption is difficult since protein compositions and their absorption coefficients are not known. Absorption in the 205 nm region (peptide bonds) can also be used for the spectrophotometric assay of proteins [13], this is of particular importance as some PDZ do not contain tryptophan or tyrosine. For the majority of proteins, UV-vis absorption allows to measure a mass concentration with concentration as low as  $100 \mu\text{g/mL}$ .

In addition to the measure of the concentration using the molar absorption coefficient of the PDZ, a complete spectrum will also inform about the general quality of the preparation. It should be emphasized that the exact PDZ buffer should be used as a blank to avoid misinterpretation. A strong signal at 260 nm is usually a sign of nucleic acid contamination or small compounds such as DTT, detergents, imidazole, ATP, etc [13–15]. In the presence of a 260 nm contamination, the quantification at 280 nm can easily be overestimated. In addition, a regularly increasing absorbance between 340 nm and 300 nm is generally indicative of scattering due to aggregation, also resulting in quantification overestimation.

## **1.2 Identity**

### **1.2.1 Intact mass spectrometry**

Mass spectrometry is a sensitive analytical method that can be easily implemented to determine the intact molecular mass of PDZ domains. The combination of the small size of the PDZs and the high accuracy in  $m/z$  of mass spectrometry devices allows, in addition to ensuring the molecular mass, to determine potential post-translational modifications (PTMs) of the domains. Proteolytic processing due to the loss of one or several amino acids is crucial to monitor and nearly impossible to detect by gel electrophoresis. In addition, the fragmentation of PDZs into small peptides allows sequencing and thus confirms the identity of the PDZ of interest (see in-source decay (ISD) part). Many mass spectrometer configurations have been developed, each with its strengths and limitations. In this chapter, only the MALDI-TOF configuration will be explained.

In 1988, Karas and Hillenkamp developed a way to produce ions from proteins with a molecular mass greater than 10 kDa [16] by laser desorption using a matrix. As a result of this work, MALDI (Matrix-Assisted Laser Desorption Ionization) was made possible and opened new perspectives in proteomic analysis. MALDI is a soft ionization which is crucial

to preserve intact mass of fragile and non-volatile compounds such as proteins. The principle of this ionization mode is based on the irradiation by laser pulses of a crystalline phase containing the co-crystallized PDZ and an organic matrix. This irradiation is carried out using a laser pulsed usually at a UV wavelength, which causes the desorption of the PDZ/matrix pair that is charged by the transfer of protons from the matrix to the analyte. Importantly, the sample must be able to release or capture protons (transfer with the matrix), depending if the positive or negative ionization mode is used to ionize the sample. For proteins, a linear and positive mode is usually applied and the matrix must absorb within the laser emission range to allow the sample to vaporize. The ions are then accelerated by the application of an electrical pulse and pass through a separation tube, called Time of Flight (TOF), maintained under vacuum. As the difference of potential is constant for all ions, ions with a smaller  $m/z$  value (lighter ions) move faster until they reach the detector. Therefore, the time of flight is different depending on the mass to charge ratio ( $m/z$ ). The signal at the detector is recorded and amplified and then processed electronically to generate a  $m/z$  spectrum. In MALDI-TOF, proteins mainly result in singly-charged or doubly-charged ion species.

One of the main advantages is that the sample droplets can be prepared rapidly on the metal target, allowing the analysis of a large number of proteins in a short time. Also, nowadays, most of the scientific centers have access to this technology and research teams can easily be trained to complement their SDS-PAGE electrophoresis.

### **1.2.2 Top Down PDZ sequencing**

Top-down sequencing (TDS) with MALDI is a complementary application of trypsin digestion. The generated trypsin digestion peptides are analyzed and compared to queries databases to determine a protein identity from SDS gel bands or liquid samples. The objective of the TDS strategy is to ensure that the protein sample is the expected one by



comparison of both the N- and C- termini sequences with the generated sequencing, assuming no degradation or enzymatic cleavage. This technique allows researchers to work directly on intact and undigested PDZs. TDS is truly suitable for the analysis of small domains such as PDZs. The fragmentation is done directly in the source (in-source decay or ISD) using a laser power much higher than the one used for intact mass. In practice, ISD fragmentation, compared to collision-induced fragmentation (CID), triggers a much more independent fragmentation of the protein sequence. Depending on the matrix, ISD fragments are mainly b-, c-, y- and z- ions. The 2,5-Dihydroxybenzoic acid (DHB) and super DHB matrices contribute to the N-terminal validation, while the 1,5-Diaminonaphtalene (DAN) matrix helps in determining the C-terminal residues. Compare with the Edman degradation, ISD has the advantage to detect both termini and in presence of modifications. This method can confirm 10-50 residues from both sides of the intact protein. Due to the matrix contribution below 800 Da, the first residues cannot be determined. This weakness can be overcome by using the T3-sequencing approach or Pseudo-MS3, where selected ISD fragments are further fragmented (MALDI-TOF/TOF) [17]. Nowadays, the all procedure intact mass and T3-sequencing can be performed in less than 1 hour.

### **1.3 Homogeneity**

By PDZ domain homogeneity measurements, we intend to measure if the domain is monomeric or oligomeric, the state of assemblies, and if the sample contains soluble high-order assemblies and aggregates. There is no single technique that will answer all these issues [18], as one of the challenges is the concentration required to make the measurement given the small molecular mass of PDZ domains.

#### **1.3.1 Dynamic light scattering**

Dynamic light scattering (DLS), because of its speed and low sample consumption, is a very convenient method to determine the mono-dispersity of the species of interest and the presence of soluble high-order assemblies and aggregates, simultaneously [19]. However, for PDZ domain measurement, a concentration of at least 1mg/ml is needed to get a reliable calculation of the hydrodynamic radius ( $R_h$ ) of the domain. It is important to emphasize that DLS does not measure molecular mass, but Brownian motion, which is related to the  $R_h$  of the particles. With the translational diffusion deduced from this motion, the hydrodynamic radius can be calculated i.e., the radius of the sphere that would diffuse with the same rate as the molecule of interest [20]. One should note that light scattering intensity is related to the cube of the radius, and consequently, DLS is the method of choice to detect small quantities of large aggregates in a PDZ sample [18]. This property makes it difficult to estimate the relative amount of the detected specie with the added difficulty that DLS has a weak resolving power and cannot distinguish between protein monomer and dimer in such molecular weight range, i.e. 10-20kDa (Fig. 1).

### **1.3.2 Analytical size exclusion chromatography**

To further analyze the oligomeric states of the PDZ domains, DLS measurement can be completed by analytical size exclusion chromatography (SEC). SEC is currently the standard separation technique to quantify protein oligomers. It separates molecules according to their Stokes or hydrodynamic radius [21], with larger molecular species eluting before smaller ones. Aggregates, contaminants and potentially different molecular arrangements of the protein of interest can be readily separated and quantified, with an online concentration detection (UV or refractive index (RI)). However, the precise quantification of the molecular mass of each eluted species requires an online static light scattering (SLS) detector [22], in combination with the UV or RI detector (Fig. 2). Several types of SLS systems are available

on the market with multi-angle detectors or MALS (Multi-angle light scattering), and low angle detectors. Their main differences are centered around the way they extrapolate light scattering intensity at  $0^\circ$  to measure molecular mass [20]. In the case of PDZ domains, both methods will give the same result as the domain is small and the instrument will not detect any angular dependence. However, the protein sample will be diluted during the SEC experiment by as much as a 10-fold factor that might alter the equilibrium between oligomeric species. Furthermore, “inert” may the gel filtration resins be, some proteins do interact with them, rendering SEC impossible. Finally, for small domains such as PDZs, the gel filtration column may not be resolute enough to separate different oligomeric states that differ by only 10 kDa. In all these cases, the technique of choice to measure molecular mass will be analytical ultracentrifugation.

### **1.3.3 Analytical ultracentrifugation**

During analytical ultracentrifugation experiment, there are no dissociating forces acting on the sample and the concentration is constant during the all-time. However, it is more sample- and time-consuming than SEC SLS and will require a specialist help. An analytical ultracentrifuge allows monitoring the sedimentation process of a sample while moving in the centrifugal field [23]. It will separate PDZ domains according to their sedimentation features which are linked to their size and molecular mass. Sedimentation velocity experiments are done at high speed to analyze the movement of the sedimenting species until all sample collects at the bottom of the cell. Then, the experiments are analyzed using a specific algorithm such as the one used in SEDFIT and UltraScan [24, 25] to determine, for each species present, a sedimentation coefficient, a frictional ratio and consequently a molecular mass, i.e an oligomeric state (Fig. 3). For each species, we will get macromolecular shape information as well as the oligomeric state. The resolving power of the experiment will allow

to separate contaminants and aggregates and the different oligomeric states of the PDZ. Nowadays, the detectors are really versatile and can detect sample absorption at any wavelengths between 190 nm and 600 nm, or refractive index changes allowing measurement of PDZ samples at concentration as low as 10  $\mu\text{g/ml}$ . We will not detail the technique in this chapter. However, a good starting point for new users is the technical paper of Andrea Balbo and al. [26]. The techniques will use between 100 and 400  $\mu\text{L}$  of sample at a concentration ranging from 0.1 to 3  $\text{mg/ml}$ . Data can be analyzed with SEDFIT or Ultrascan algorithms [25, 27] using the most adapted model such the continuous size distribution  $c(S)$  model presented in the figure for MAST2-PDZ (Fig. 3).

## **1.4 Conformational stability/folding state**

To verify that the same folding signature can be seen for different batches of PDZ, one of the classical technique is circular dichroism (CD).

### **1.4.1 Circular Dichroism (CD)**

Circular dichroism (CD) relies on the ability of a sample having a chiral chromophore or placed in an asymmetric environment, to absorb differently circularly polarized right light and circularly polarized left light. The spectrum of circular dichroism corresponds to the difference of absorbance between these two polarizations of light, for each of the wavelengths. Studying macromolecules by CD allows to obtain information about their folding. For proteins, far UV (180-260 nm) and near UV (250-330 nm) circular dichroism measurements give insight respectively into their secondary structure content and their tertiary organization, respectively. The far-UV CD spectroscopy of proteins and peptides is predominantly based on the excitation of electronic transitions in amide groups. Two types of electron transitions are responsible for the CD signals in this wavelength region, an  $n \rightarrow \pi^*$

transition at around 222 nm, and  $\pi \rightarrow \pi^*$  transitions (both parallel and perpendicular orientations) at  $\sim 208$  and 190 nm. The peptide backbone forms characteristic secondary structures such as  $\alpha$ -helices,  $\beta$ -sheets, turns, and disordered sections with specific  $\Phi$ ,  $\Psi$  dihedral angles and H-bond patterns affecting the CD spectrum. Numerous algorithms have been developed for the quantitative estimation of the secondary structure composition from the CD spectra with a good overall prediction [28]. For PDZ domains, the CD spectrum is a linear combination of the spectra of its secondary structural elements, i.e. mainly 1 to 2  $\alpha$ -helices, a  $\beta$ -sheet (5 to 6 strands) and turns (Fig. 4). In the near-UV wavelength region, the major chromophores generating CD features of proteins are the aromatic side chains of tryptophan, tyrosine and phenylalanine together with the disulfide bonds. Only qualitative tertiary structural information can be obtained from the environmental-dependent CD spectra of protein aromatic residues in the near UV wavelength range.

CD has been used for protein folding assays, intermolecular interactions and in the investigations of protein disorder. It is one of the easiest and sensitive methods to detect conformational changes that might result from changes in pH, salt concentration, ionic strength, added solvents or mutations in native protein, to assess the thermal or chemical stability by following molecule unfolding, and to analyze protein-ligand interactions.

#### **1.4.2 Differential Scanning Calorimetry (DSC) and Differential Scanning Fluorimetry (DSF)**

Delimitations of PDZ domain limits are tricky to define. Stability, folding and dynamics of PDZ constructs need to be optimized. High thermostability is an important feature to improve PDZ behavior during crystallization [29–31] or other structural studies [32]. Several techniques can be used to monitor the thermostability of the sample such as: CD coupled to a temperature gradient that monitors the loss of secondary structure as a sign of global unfolding [29], DSF that detects changes in the tryptophan and tyrosine environment, as well

as thermofluor and DSC [33]. All these approaches are based on partial or full unfolding of proteins during heat denaturation. These techniques use different observable data to monitor protein unfolding that could lead to a sharp modification of the signal over a short temperature range. The melting temperature ( $T_m$ ) that corresponds to the temperature where the protein is 50% unfolded (Fig. 5) can be derived by calculating the peak of the first derivative.

A biomolecule in solution is in equilibrium between its native (folded) and denatured (unfolded) conformations. A higher thermal transition midpoint ( $T_m$ ) corresponds to a more stable molecule. DSC measures the enthalpy ( $\Delta H$ ) of unfolding that results from heat-induced denaturation. It is also used to determine the change in heat capacity ( $\Delta C_p$ ) of denaturation. The combination of  $\Delta H$ ,  $\Delta C_p$  and  $T_m$  helps to define the intrinsic properties of the protein, in our case a PDZ domain. As an example, some proteins may have a low  $T_m$  associated with a large  $\Delta C_p$  which will confer a great stability *in vivo*. In parallel, PDZs with a mass > 5000 Daltons, such as proteins, form well-defined structures that undergo thermally-induced conformational changes [34]. These structural rearrangements result in the absorption of heat caused by the redistribution of non-covalent bonds. Differential scanning calorimeters measure this heat uptake.

Concerning NanoDSF, this technique records the intrinsic fluorescence of Tryptophan (Trp) and Tyrosine (Tyr) residues, which are very sensitive to changes in their local environment. Thermal unfolding is measured by monitoring the intrinsic Trp and Tyr fluorescence intensity, and the position of the emission maximum as a function of temperature. The fluorescence intensity ratio between 330 and 350 nm is defined as an empirical parameter to monitor the evolution of the micro-environment of the aromatic residues during protein denaturation throughout the temperature increases. This ratio sharply increases/decreases during thermal unfolding, allowing to determine a  $T_m$  value [35]. The applicability of

nanoDSF is highly dependent on the presence of Trp and Tyr in the folded core of the PDZ that are exposed upon unfolding.

Moreover, it is necessary to exclude that the observed signal changes are caused by aggregation, as this will also lead to variations in the fluorophore environment. A back-scattering measurement can also be performed to determine if aggregation occurs (before or concomitantly with the denaturation). The characteristics of the temperature gradient are essential as it is related to the activation energy via the Arrhenius equation [6]. Typically, a heating rate of 1°C/min is applied. Overall, comparing the melting temperatures ( $\Delta T_m$ ) in different buffer compositions allows researchers to define the optimal buffer condition as an increase in  $T_m$  corresponds to a better thermal stability and to a reduced conformational flexibility.

### **1.4.3 Nuclear Magnetic Resonance (NMR)**

NMR is based on the measure of the absorption of radiofrequency (RF) radiation by an atomic nucleus located in a strong magnetic field. The principle of NMR is that atomic nuclei, with an odd number of protons ( $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ ,  $^{31}\text{P}$ , ...), neutrons, or both, have an intrinsic nuclear spin. When an atomic nucleus with a nonzero spin is placed in a magnetic field, the nuclear spin aligned in the same direction or in the opposite direction to the field. Different energies characterize these two types of nuclear spin alignment, and the application of a magnetic field facilitates the degeneration of nuclear spins. An atomic nucleus whose spin is aligned with the field will have less energy than when its spin is aligned in the opposite direction of the field.

The energy of an NMR transition depends on the magnetic field strength as well as on the proportionality factor applied to each nucleus called the gyromagnetic ratio. The local environment around a given nucleus in a molecule tends to slightly disturb the local magnetic

field exerted on that nucleus and to affect its transition energy. This dependence of the transition energy on the position of a particular atom in a molecule makes NMR extremely useful for determining the structure of molecules.

The sample placed in an intense magnetic field will be disturbed by radiofrequency pulses. The recording of the return to the equilibrium of the spins makes it possible to have access to the chemical environment of the atoms. This information offers the possibility of structural and dynamic analyzes as well as the study of interactions involving biological macromolecules.

NMR can be used to characterize the degree of folding of a protein by observing the dispersion of the resonance peaks. Indeed, in the one-dimensional (1D)  $^1\text{H}$  spectrum or the two-dimensional (2D)  $^1\text{H}$ - $^{15}\text{N}$  correlation spectra if the protein is  $^{15}\text{N}$ -labeled, the peaks for a well-folded protein are narrow and sharp and distributed over a large range of chemical shifts (good signal dispersion) (Fig. 6).  $^1\text{H}$  resonances can be especially found at values  $< 0.5$  ppm (corresponding to high field-shifted methyl group protons) or  $> 8.5$  ppm (corresponding to down field-shifted amide protons). In contrast, the peaks can be broader and not as widely dispersed in the spectrum of an unfolded or partially folded protein. Moreover, observed linewidths of peaks are related to the molecular weight of the protein, and then may be indicative of auto-association or aggregation. In addition to the evaluation of the folding and the stability of a protein, a 1D  $^1\text{H}$  spectrum provides information about purity. Indeed, impurities with low molecular weight and observable nuclei give rise to sharp signals amongst the broader envelope of the protein resonances.

The most common 2D spectrum to obtain structural information about a protein is the  $^1\text{H}$ - $^{15}\text{N}$  heteronuclear single quantum coherence (HSQC) spectrum. It correlates the nitrogen atom of an amide group with the directly attached amide proton. Since there is only one backbone  $\text{H}^{\text{N}}$  per amino acid, except for Pro, each HSQC signal represents one single amino acid. The



HSQC also contains signals from the NH<sub>2</sub> groups of the side chains of Asparagine and Glutamine (also Lysine and Arginine depending on the pH values) and of the aromatic H<sup>N</sup> protons of Trp and Histidine. The signals may cover a spectral range from 6.0-12 ppm. If a protein is folded, its signals are distributed over the complete spectral range; if not, signals are located between 7.5 and 8.5 ppm in the proton dimension. Signals outside these regions indicate that the protein is folded and in a defined three-dimensional (3D) state. The chemical shift is very sensitive to the overall structure so that even slight conformational changes will affect the signals in a <sup>1</sup>H-<sup>15</sup>N HSQC spectrum, making it a very efficient tool to check conformational stability and folding states.

## **2. Materials**

### **2.1 Purity**

#### **2.1.1 Basic reagents for SDS-PAGE electrophoresis**

1. Electrophoresis tank.
2. Power supply.
3. Polyacrylamide gels. Novex™ 4-12% Tris-Glycine Mini Gels offers a broad range of molecular weight resolution from 6 to 400 kDa. Standard 10 to 12 % polyacrylamide gels can also be used to resolve PDZ domains.
4. 2X Sample buffer: 100mM Tris-HCl pH6.8, 0.2% BromoPhenol Blue (BBP), 4% SDS, 20% Glycerol, 200mM Dithiotreitol (DTT) (see Note 2). Add 1ml of 1M Tris pH6.8, a pinch of BromoPhenol Blue (BPB) and vortex, 4ml of 10% SDS, 2ml of 100% Glycerol and 2ml of 1M DTT.
5. Molecular-weight markers such as PageRuler™ Prestained Protein Ladders from 10 to 180 kDa.

6. Migration buffer such as NuPAGE MES SDS Running Buffer. Add 50 ml of the NuPAGE™ MES SDS Running Buffer (20X) and complete to 1 L with pyrolysis water.
7. Home-made or commercial protein staining solution. InstantBlue is a ready to use Coomassie protein stain for polyacrylamide gels.

### **2.1.2 Capillary Electrophoresis (CE)**

1. Capillary electrophoresis system in this case a LabChip GX II (Perkin Elmer).
2. LabChip HT Protein Express Chip (Perkin Elmer, 760499).
3. Protein Express Assay Reagent Kit (Perkin Elmer, CLS960008).
4. HT Protein 200 Sample Buffer (Perkin Elmer 760518).
5. 384-well PCR plate (Greiner Bio-One).

(see Note 3)

### **2.1.3 UV spectroscopy**

1. UV-Vis spectrometer.
2. Quartz cuvette suitable for UV (190-400 nm).
3. Sample buffer as described in 2.2.4 item 4.
4. Special wipe such as Kimwipes to clean the cuvette before measuring.

## **2.2 Integrity measurements by MALDI-TOF**

### **2.2.1 Total mass**

1. MALDI steel plate.

A freshly prepared matrix solution: 25 mg/ml  $\alpha$ -Cyano-4-hydroxycinnamic acid (HCCA) in 50% (v/v) Acetonitrile, 0.1 % (v/v) Trifluoroacetic acid (TFA) in HPLC water.

2. ZipTip with C4 resin (Merck Millipore).
3. High quality protein standard from Bruker, Laserbio labs or Sciex.

- **ISD and T3 sequencing**

1. MALDI steel plate.
2. Two matrix solutions: 25 mg/ml 1,5-Diaminonaphthalene (DAN) and 25 mg/ml of a mixture composed of 2,5-Dihydroxybenzoic acid and 2-hydroxy-5-methoxybenzoic acid (super-DHB). Both matrices are dissolved in 50% Acetonitrile (v/v), 0.1 % (v/v) TFA in HPLC water.
3. ZipTip with C4 resin (Merck Millipore).
4. 1 mg/ml BSA in PBS as ISD standard.

## **2.3 Homogeneity**

### **2.3.1 Dynamic Light Scattering (DLS)**

1. 0.22  $\mu$ M Syringe filter.
2. Reusable quartz cuvettes (Hellma) or plates depending on the model.
3. Detergent solution : 2% (v/v) Hellmanex.
4. Buffer used for PDZ sample: 1ml.
5. PDZ sample.

### **2.3.2 Size Exclusion Chromatography (SEC) with Multi Angle Light Scattering (SEC-MALS)**

1. Size exclusion column (eg. Superdex 75 increase from GE healthcare).

2. PBS (137 mM [NaCl], 2.7 mM [KCl], 10 mM [Na<sub>2</sub>HPO<sub>4</sub>], 1.8 mM [KH<sub>2</sub>PO<sub>4</sub>]) (1 liter) or other buffer (see Note 4).
3. PDZ sample. Prepare your sample in the chosen buffer with a protein concentration at least of 1mg/ml. This is particularly true for PDZ domain if you want sufficient signal on the light scattering detector as the signal is proportional to the molecular mass of the protein.

## **2.4 Conformational stability/folding state**

### **2.4.1 Circular Dichroism (CD)**

1. Aviv 215 spectropolarimeter.
2. 0.2 mm path-length cylindrical cell (Hellma) (see Note 5).
3. PDZ sample diluted at 0.2 mg/ml with PDZ sample buffer.
4. PDZ sample buffer.
5. 70% ethanol.
6. 2% (v/v) Hellmanex.

### **2.4.2 Differential scanning fluorimetry (nano DSF)**

1. Prometheus NT.48.
2. 70% Ethanol to clean the lens.
3. Quartz capillary.
4. Buffer used for PDZ domain preparation and storage.
5. PDZ sample diluted at 0.2 mg/ml with PDZ sample buffer.
6. Peptide solubilized in PDZ sample buffer (10-fold molar excess relative to PDZ sample).

### **2.4.3 NMR**

1. 600MHz (or higher fields) NMR spectrometer equipped with triple resonance  $^1\text{H}\{^{13}\text{C}/^{15}\text{N}\}$  PFG probe (for heteronuclear NMR detection).
2. 4 mm NMR tube (Shigemi INC).
3. 100  $\mu\text{M}$  PDZ sample (250  $\mu\text{l}$ ).
4. 100%  $\text{D}_2\text{O}$  to add to the PDZ sample at 1-5% final concentration.

### **3. Methods**

We used the PDZ domain of the human kinase MAST2 (UniProtKB Q6P0Q8) as a practical model. This construct of 96 residues (MAST2-PDZ) (Fig. 7A) has a molecular weight of 10435.9 Da.

#### **3.1 Purity**

The objective is to quantify the PDZ construct of interest and assess the presence (and level) of contaminants using techniques that are available to most laboratories.

##### **3.1.1 SDS-PAGE**

A simple protocol to run an SDS-PAGE is given below:

1. Prepare 20  $\mu\text{l}$  of sample by adding 10  $\mu\text{l}$  of 2X denaturing loading buffer to 10  $\mu\text{l}$  of protein and heat sample at 95°C for 5 minutes (see Note 6).
2. Prepare the pre-casted gel by removing the comb and the white tape near the bottom of the gel cassettes and place the gel in the mini gel tank (see Note 7).
3. Fill the chambers with running buffer.
4. Load samples and ladders in the appropriate wells.
5. Run the gel following the manufacturer's instructions (see Note 8).

6. After electrophoresis, remove the gel from the tank and transfer directly into the InstantBlue staining solution. Be sure that the gel moves freely in stain to facilitate diffusion. Typically about 20 ml is needed to cover the gel.
7. Colored proteins bands will start to develop immediately and a suitable intensity is typically achieved after 15 minutes incubation at room temperature with gentle shaking.

The protein sample is considered pure when no other band other than the expected one for your protein of interest is detected by SDS-PAGE using sensitive staining [20].

In the case of MAST2-PDZ, the protein is purified as previously described [36]. The 2-step purification process improved drastically the purity of the sample as assessed by SDS-PAGE in Figure 7B. The GSTrap column allowed to remove a high content of contaminants and to concentrate the MAST2-PDZ construct. The SDS-PAGE gel of the eluted fractions showed a major band of the protein and only a few contaminants extra-bands (Fig. 7B left panel). Fractions containing MAST2-PDZ are pooled and then purified by size-exclusion chromatography and the only band remaining on the corresponding gel SDS-PAGE is the MAST2-PDZ construct. A very weak band around 30kDa slightly overlap with the first fractions of MAST2-PDZ. The pool of fractions of gel filtration is then concentrated and the gel SDS-PAGE revealed one unique band indicative of the high degree of purity by SDS-PAGE technique. This construct has a migration that does not match with its calculated molecular weight as already observed. This might be due to the effect of SDS binding [37] and/or the hydrophobicity of the construct [38]. The limits of this technique are that the detection is restricted to protein contaminants, and the detection is proportional to the sample loading and gel staining and should be confirmed by mass spectroscopy.

### **3.1.2 Capillary Gel Electrophoresis (CGE)**

Capillary gel electrophoresis (CGE) separates proteins according to their molecular mass similarly to SDS-PAGE and this methodology has been previously exploited to validate bacterial expression of a library of 266 known human PDZ constructs developed for the high-throughput "holdup" chromatographic assay for the determination of PDZ-PBM affinities [39]. Below we outline the main steps of this high-throughput protocol.

1. Dilute to 1:8 (v:v), 1:16 (v:v) and 1:32 (v:v) each PDZ supernatant to be quantified.
2. Dispense 5.6  $\mu\text{L}$ /well of HT Protein 200 Sample buffer in as many wells of a 384 well plate as required.
3. Add 4  $\mu\text{L}$ /well of the PDZ dilutions to each well.
4. Boil the samples at 95 °C for 5 min in a dry bath.
5. Add 20.6  $\mu\text{L}$ /well of distilled H<sub>2</sub>O.
6. PDZ samples within the 384 well plate is then analyzed with the High Sensitivity HT Protein Express protocol (10–100 kDa program) with the LabChip GXII device (Perkin Elmer) following supplier's instructions. This technology allows to electrophoretically separate, stain, destain, detect by laser-induced fluorescence and analyze the protein samples. The data analysis provides protein concentration, molecular weight sizing and purity evaluation using ladder and marker calibration standards. With sample acquisition time of about 40 seconds, the instrument takes approximately 4 h to analyze 384 protein samples.
7. The LabChip GX software is then used for data analysis. This allows to visualize results via an electropherogram or virtual gel view (Fig. 8) or in a tabular form to export into a spreadsheet format. At the end of the run, the concentrations of the initial cultures of the soluble PDZ are calculated from the concentrations determined per PDZ in each serial dilution. The PDZ lysate concentrations range from 10  $\mu\text{M}$  for the least concentrated and up to 100  $\mu\text{M}$  or higher. The concentration of lysozyme is

constant and used as an internal reference for quantification. A limitation to the quantification is reached when the construct is poorly expressed and falls in the background of the *E. coli* proteins in the case of a lysate. On the contrary, the electropherogram allows to easily detect contaminants (Fig. 8). This CGE approach is fast, efficient and allows high resolving separations with low solvent consumption and minimal operating cost considering the high number of samples. However, the LabChip GXII system is not common in most laboratories. Its running cost, the short-use warranty of chips and kits, can constitute a limitation. During the 4 h run, a diminution in data quality may happen due to samples drying into the 384 well plate.

### 3.1.3 UV spectroscopy

UV spectroscopy between 220-240nm and 340nm is a good quality test to determine protein concentration (using  $A_{280\text{nm}}$ ) and to detect aggregation and molecule contamination. Indeed, UV-visible spectroscopy can detect the presence of large particles such as aggregates (radius higher than 200nm) in a protein preparation by monitoring the absorbance signal above 320nm, where aggregate-free protein samples are not supposed to absorb light. If the signal increases as the wavelength diminishes between 340nm and 320nm, it can be attributed to the scattering of light by large aggregates present in the sample. In Figure 9, the two UV spectra of a sample of MAST2-PDZ show a difference in the absorbance above 320 nm, indicative of a difference in the presence of aggregates. The aggregation index (AI) can be calculated:  $AI = 100 \times A_{340\text{nm}} / (A_{280\text{nm}} - A_{340\text{nm}})$  with  $A_{280\text{nm}}$  and  $A_{340\text{nm}}$ , the absorbances at 280 and 340nm, respectively.  $AI < 2$  is indicative of a homogeneous sample with no sign of aggregation.

UV-visible spectroscopy is also useful for the detection of non-protein contaminants by monitoring the absorbance over a large range (at least 240 – 340 nm). Contaminating nucleic acids, reducing agents and detergents may show extra absorbance around 260 nm. The



contribution of these contaminants is evaluated with the ratio  $A_{260\text{nm}}/A_{280\text{nm}}$ , which should be lower than 0.6 [14]. We already observed such contaminations with the first PDZ domain of Whirlin, a protein involved in the human auditory system, which presents a cationic patch at the surface of its PDZ1 domain. This basic cluster of three or four Arg or Lys close to the carboxylate binding site is common to a subset of PDZ domains that bind lipids [40]. More generally, we might presume that PDZ domains known to interact with lipids [41] may also interact with nucleic acids through the cluster of basic residues. In line with this, samples with PDZ domains having high isoelectric point (pI) values are prone to nucleic acids contamination.

If the sample shows no sign of contamination or aggregation, the sample concentration can be calculated using  $A_{280\text{nm}}$ , the protein molar absorption coefficient at the working wavelength and the Beer-Lambert law [42]. However, PDZ domains are small domains and some of them lack of tryptophan. In that case, a weak  $A_{280\text{nm}}$  is obtained and might alter the precision of the concentration. In that case, an alternative way is to use  $A_{205\text{nm}}$  to determine concentration [13].

The protocol for UV spectrum measurement comprises a run of the baseline/blank with buffer and ones with the protein solution. The exact same buffer that was used to prepare the protein solution should be used, ideally the buffer from the last step of purification, i.e. the exclusion chromatography in the case of MAST2-PDZ. For concentration measurement, the absorbance at the wavelength used must be below the specified saturation range of the instrument to keep the linear relationship between the absorbance and the concentration. If the absorbance is too high, the sample must be diluted and the measurement should be repeated. PDZ spectra are also checked for aggregation around 320-340nm and for contamination by nucleic acids at 260nm, for example. If any signs of aggregation or contaminants are detected, the  $A_{280\text{nm}}$  and the corresponding extinction coefficient are used to

calculate the concentration of the protein in the cuvette.

A simple protocol to run a UV-visible spectrum is given below:

1. Turn on the UV-Vis spectrometer and allow the lights to warm up for an appropriate period of time (approximately 20 min) to stabilize.
2. Fill a cuvette with the buffer used to generate the protein sample and make sure the outside is clean. This will serve as a blank and help account for slight losses due to diffusion or absorption by the solvent.
3. Place the cuvette in the spectrometer (see Note 9). Make sure to align it correctly.
4. Read for the blank assay in the range of 200 to 500 nm for example for a protein sample (see Note 10). Absorbance should be minimal, but any absorption should be deducted from future samples. Most instruments store empty data and perform subtraction automatically.
5. Fill the cuvette with the sample. To ensure that the transfer is quantitative, rinse the cuvette 3 times with water then 3 times with alcohol and dry with pressured air. Make sure the outside is clean of fingerprints, etc, by cleaning the cuvette with a special wipe.
6. Place the cell in the spectrometer in the correct direction. Close the cap to avoid any ambient light and collect an absorbance spectrum in the range of 200 to 500 nm.

### **3.2 PDZ mass integrity by MALDI-TOF**

There are many ways to prepare deposits on the MALDI target with variations in terms of the matrix concentration, the solvent used, or the crystallization process. The purpose of the droplet preparation on the MALDI plate is to allow the protein-matrix mixture to be close enough to co-crystallize. The dried droplet method is the most commonly used technique

because it is swift and straightforward to implement. There are several alternatives to proceed that are all roughly equivalent. Indeed, it is possible to directly mix the PDZ domain with the matrix on the MALDI plate or to pre-mix it in a tube and deposit the mixture on the target.

### **3.2.1 Total mass measurement**

For total mass measurement of PDZ domain (Fig. 10), the protocol is as follow:

1. Use the dried droplet method: Apply 1 $\mu$ l of PDZ sample on the plate and overlay with 1 $\mu$ l of HCCA. For a PDZ domain of 15kDa, 1 $\mu$ L at 0.1mg/ml is sufficient. For a 30kDa PDZ construct (a PDZ tandem for example), a concentration of 0.3mg/ml is recommended. These concentrations are acceptable for protein samples in PBS-type buffer. For complex buffer or low PDZ quantity see Note 11.
2. Let the co-crystals air-dry for 15min.
3. Acquire the data in linear and positive mode on the MALDI-TOF mass spectrometer. Use the minimum laser power required to get a nice spectrum (see Note 12).
4. Calibrate the experiment with a standard mixture, for instance, the Protein Standard I from BRUKER company.
5. Analyze the data (see Note 13).

### **3.2.2 ISD and T3 sequencing**

The protocol to perform ISD acquisition is:

1. Use the dried droplet method: Apply 1 $\mu$ l of PDZ sample on the plate and overlay with 1 $\mu$ l of super-DHB matrix. Do a second spot with the DAN matrix (see Note 14).
2. Select an ISD method usually between 1 and 10kDa in reflectron and positive mode. In that range of m/z, avoid any contaminations such as detergents or PEG.

3. Increase by 10% the laser power required compare to the intact mass of the PDZ of interest.
4. Acquire at least 10000 shoots.
5. Save your spectrum and calibrate it with BSA ions generated in that range.
6. Open the spectrum in Biotoools or another software and assign it as an ISD-type.  
Load the amino-acid sequence and analyze the data as illustrated in (Fig. 11)
7. If the first amino acid needs to be determined, proceed with a T3 sequencing procedure that uses a LIFT method as described in the following steps. Define an ISD fragment ion around 1000m/z as parent. Copy the value of the parent ion.
8. Open the LIFT method. Paste the peak around 1000 m/z in the LIFT method. Acquire the signal of the parent ion. Acquire signals of the fragment ions on the parent ion spectrum. Gain and laser power are critical issues as the sensitivity is low. Save the data and annotate manually the ions. A search using Mascott software is also possible.

### **3.3 Homogeneity**

The objective is to assess if the sample has a tendency to form aggregates and to consider the potential for oligomerization of the protein sample.

#### **3.3.1 Dynamic Light Scattering (DLS)**

DLS experiments can be performed using nearly any buffer (see Note 15). Special care should be taken during measurement if the buffer contains large scattering particles such as crowding agents, etc. The required protein concentration depends on the molecular mass of the protein and can be estimated by the rule of thumb given in equation 1.

$$c \left( \frac{\text{mg}}{\text{ml}} \right) = \frac{15}{MW} \quad (1)$$

where  $MW$  is the molecular mass of the target protein in kDa.

For most PDZ domains, a concentration around 1mg/ml is well adapted to the measurement. In terms of volume, users should refer to the manufacturer's specification/manual as it varies from instrument to instrument, ranging from only a few  $\mu$ l to a couple of ml.

A simple protocol for a DLS measurement is given below:

1. Switch on the instrument and select the experimental temperature.
2. Wait for temperature stabilization.
3. If needed, clean the reusable cuvettes or plates with detergent (e.g. 2%(v/v) Hellmanex) and water, and dry it with ethanol.
4. Filter buffers.
5. Load buffer and samples into the wells or the cuvettes (see Note 16). Avoid the formation of bubbles.
6. Setup a program containing 5-20 acquisitions with an acquisition time of 2-10 seconds per sample depending on the presence of dust or particles that may disturb the measurement. Measurements should be done in triplicates using automatic laser power and detector attenuation.
7. If the DLS instrument has a camera, check for potential bubble or precipitate.
8. During data analysis, the buffer quality should be first checked.
9. Check the shape of the autocorrelation curve for the protein (Fig. 1B, inset). Check the fit and the error of the measurement and adjust the fitting limit accordingly. The presence of aggregates is indicated if particles with a  $R_h$  of 10-100nm are observed (Fig. 1B).
10. If the protein is not detectable due to the presence of aggregates or artifacts, the sample should be centrifuged using a benchtop centrifuge (15min, at full speed). However, it is important to consider the potential loss of protein during

centrifugation; it is therefore recommended to measure the concentrations before and after centrifugation.

### **3.3.2 Size Exclusion Chromatography (SEC) with Multi Angle Light Scattering (SEC-MALS)**

Test runs with small injection volumes of the protein of interest may be performed to choose a good column setup (see Note 17). In line with the column, the refractometer and the light scattering detectors allow the measurement of the molecular mass of all the peaks of interest.

A simple protocol for a SEC-MALS experiment is given below:

1. Switch on the instrument.
4. Purge the system with water.
5. Connect the column to the SEC-SLS/SEC-MALS system.
6. Choose the flow speed and pressure limit according to the column manufacturer's instruction.
7. Equilibrate the system with at least 2 column volumes of water (see Note 18).
8. Purge the system with the buffer of interest (see Note 19).
9. Equilibrate the system with at least 2 column volumes of buffer.
10. Purge the refractometer extensively until the baseline remains stable for one column run (see Note 20).
11. Centrifuge the sample for 5 minutes in a benchtop centrifuge at full speed.
12. Load the sample into the injection loop. The volume (usually 100µl) that will be injected onto the column should be lower than 1% of the total column volume to ensure a good separation. If the sample concentration needs to be quantified, flush the injection loop to homogenize the sample injection volume.

13. At the end of the run, analyze all peaks of the chromatogram for sample recovery and molecular mass determination according to the manufacturer's manual. If the column type and the buffer conditions are chosen correctly to avoid non-specific binding of the PDZ, you should recover nearly 100% of your sample. The baseline for the detectors should be set up wisely, i.e., in a region where no proteins elute, such as the region prior to the void volume of the column (first third of the column volume). The baseline should remain constant if the sample is in the same buffer as the elution buffer and if the temperature is strictly controlled.

### **3.4 Conformational stability/folding state**

#### ***3.4.1 Circular Dichroism (CD)***

CD spectroscopy is a fast and easy method that is particularly suitable for quickly judging the folding state of a protein; to compare the structure of proteins obtained from different sources; to analyze the impact of point mutations on the structure of a protein; to judge the stability of the structure facing environmental changes (pH, salinity); to determine the impact on the structure in the presence of a ligand.

One important point to improve the quality of the result is to use well adapted buffer; it must not have an absorbance higher than 1 in the range of the spectrum (see Note 21). CD spectrometers typically measure from longer to shorter wavelengths. The combination of accumulations and the time over which the instrument averages each data point needs to be optimized to get the best signal to noise (S/N) ratio. The parameters of a measurement should be carefully chosen [20]. Usually, the CD signal, i.e. the ellipticity  $\theta$ , the high voltage, and the absorbance are recorded. In order to obtain reliable, interpretable, and high-quality CD data, the absorbance of the sample (including buffer) should be between 1 and 2, i.e. on most laboratory instruments the high voltage should be kept below 600V. A CD spectrum of the

buffer is recorded with the same set of parameters that will be applied in the sample measurement to allow a baseline subtraction. A CD spectrum recorded in the 180-260nm range allows to evaluate the secondary structure content and then the folding state of the protein. The resulting spectrum has a characteristic shape and magnitude determined by the secondary structure elements. The fraction of each secondary structure element ( $\alpha$ -helices,  $\beta$ -sheets, turns and disordered structures) is evaluated using different algorithms combining a set of reference spectra [43]. Dichroic signals in the near UV (250-330nm, absorption zone of aromatic amino acids) provide information on the environment of these aromatic amino acids and is a signature of tertiary and quaternary structure. In these wavelengths, the CD signals of proteins are weak; cuvettes with larger pathlengths and higher concentrations of PDZ sample are required.

CD is particularly suited to the study of structural changes induced by environmental variations (pH, temperature, detergents, co-solvents, buffers ...). Among the advantages of CD compared to other techniques, we can mention that one works in dilute solution, with small amounts of product (a few tens of  $\mu\text{g}$ ), without molecular size limit on the protein. The measurements are simple and fast, allowing kinetic measurements at very short times. Moreover, the measurement is not destructive and it is possible to recover the sample after the acquisition of the spectrum.

Figure 4 shows far-UV CD spectra of free MAST2-PDZ, free PDZ-binding motif peptide and the complex MAST2-PDZ with a PDZ-binding motif peptide in blue, red and grey, respectively. The similarity of spectra of MAST2-PDZ free or complexed to a PDZ-ligand, in blue and grey respectively, is indicative of the maintenance of the PDZ folding upon PDZ-ligand binding while the peptide alone (in red) is unfolded.

Through this technique, the thermal stability of a protein can also be evaluated by looking at the CD shape of spectra recorded at different temperatures; this operation can also be done



continuously with a temperature spectrum at a fixed wavelength, at 222nm for example to follow the helical content of PDZ domains. Thermal stability is now often evaluated by DSC or DSF (see the section below).

Procedure for a standart CD acquisition;

1. Switch on the instrument and select the experimental temperature.
2. Wait for temperature stabilization.
3. Clean the cell with water, dry with ethanol (with detergent, e.g. 2%(v/v) Hellmanex, if needed) .
4. Centrifuge the buffer using a benchtop centrifuge (15 min, at full speed) to remove any aggregates.
5. Fill the cell with 40-50  $\mu$ l of buffer. Avoid the formation of bubbles.
6. Record ellipticity values every 0.5 nm, averaged over 2s between 195-260 nm. Three to five successive scans are collected.
7. Clean again the cell with water, dry with ethanol.
8. Fill the cell with 40-50  $\mu$ l of PDZ sample. Avoid the formation of bubbles.
9. Collect the data with the same parameters as the ones used for the buffer spectra.
10. The final spectrum of PDZ sample is obtained by averaging the 3-5 successive scans and subtracting the baseline spectrum of the buffer.

### **3.4.2 Nano differential scanning fluorimetry (nanoDSF)**

A simple protocol to run a nanoDSF experiment with a MAST2-PDZ solution free or in complex with a peptide on the Prometheus NT.48 is given below:

1. Turn on the instrument and allow the system to warm up for at least 15 minutes.

2. Clean the system with 70% ethanol.
3. Prepare in four different PCR tubes: 10 $\mu$ L of buffer, 10 $\mu$ L of PDZ sample, 10 $\mu$ L of PDZ sample with peptide a 10-fold concentration excess and 10 $\mu$ L of a 10-fold more concentrated peptide solution.
4. Insert a capillary into the tubes and position them in order on the rack in the DSF instrument.
5. Adjust the excitation power so that the fluorescence signal is greater than 2000 counts to obtain clear denaturation transitions.
6. Set the temperature range from 20 to 95°C with a desired temperature gradient of (between 0.1 and 5°C).
7. Perform all measurements with the same temperature ramp.
8. Start the measurement.

### **3.4.3 Differential scanning calorimetry (DSC)**

DSC is useful to characterize the stability of a protein by measuring the emission or absorption of heat from a biomolecule during a controlled increase or decrease in temperature. This allows the study of molecules in a native state and is widely applied in protein engineering, for rational drug design and biopharmaceutical production, when the development of stable proteins is a critical goal.

The thermal core of a DSC system consists of two cells, a reference and a sample cell maintained at the same temperature, as they are heated. To perform a DSC measurement, the reference cell is first filled with buffer and the sample cell with the sample solution. These cells are then heated at a constant scan rate. The absorption of heat that occurs when a protein unfolds causes a temperature difference ( $\Delta T$ ) between the cells, resulting in a thermal gradient across the Peltier units. From the  $\Delta T$  values, DSC profiles are generated to provide information on thermal stability by measuring the thermal transition temperature (thermal

transition midpoint or melting temperature,  $T_m$ ) and the energy required to break the stabilization interactions of the tertiary structure (enthalpy) of proteins. Comparisons are made between samples (wild-type and mutants proteins or various conditions of buffer, for example) and differences in the derived values indicate differences in thermal stability and structural conformation. Compared with other methods of evaluating the thermal stability of protein conformations, DSC is cost-effective and only requires few sample preparation steps. We previously performed DSC on the PDZ domain of the tyrosine phosphatase PTPN3 (PTPN3-PDZ) [34] free or in complex with various PBMs known to interact with the PTPN3-PDZ. A single endothermic peak was observed in the DSC thermograms of PTPN3-PDZ free or complexed. The free domain showed a  $T_m$  of 41°C, while  $T_m$  values between 45°C and 52°C were observed for PTPN3-PDZ complexed to different PBMs. Thus, an increase of 4 to 11°C in the  $T_m$  was observed depending on the sequence of the ligands, showing that the PBM binding onto PTPN3-PDZ stabilizes the domain in all complexes (Fig. 5A). A gain in thermal stability is also observed for other PDZ domains such as MAST2-PDZ, whose  $T_m$  increases from 47 °C to 55-62°C upon PBM binding (Fig. 5B). The ligand binding increases the  $T_m$  of the PDZ domain, either by stabilizing the native conformation or by the destabilization of the unfolded state.

#### **3.4.4 NMR**

Compared to other spectroscopic techniques, NMR sensitivity level is low. To record spectra with a sufficient signal to noise (S/N) ratio within rational acquisition time the sample needs to be quite concentrated. Currently, the common samples correspond to a volume of 500µL or 200µL (5mm or 3mm tubes) containing 50 to 1000µM of a medium-sized protein (10-

20kDa). The sample buffer also contains 3-10% (v/v) D<sub>2</sub>O to record a reference signal (lock signal) to compensate for small changes/drifts of the magnetic field. The NMR tube is cleaned with a soft tissue from the outside before it is inserted into the instrument to remove anything spilled on the tube or grease from the fingertips. When the sample is inserted into the spectrometer and the temperature is equilibrated, the lock signal is defined. The inhomogeneities in the magnetic field are then corrected (shim), and the impedance of the probes for each nucleus is precisely adjusted. This is called “tuning the probe”. Finally, the length of the 90° proton pulse is determined. The setup of the experiment can be automated and can be achieved within 15-20 minutes. The quality of an NMR spectrum depends on the S/N ratio and the resolution that are first related to the strength of the static magnetic field. The S/N ratio also depends on the concentration of the sample and the molecular mass and intrinsic dynamics of the protein, which affect the linewidth of peaks in the spectrum. The S/N ratio correlates with the number of FIDs accumulated and is enhanced by using a cryoprobe. The resolution of a NMR spectrum depends on the linewidth and the separation of the resonances, which is directly correlated with the magnetic field strength. Furthermore, the separation of the signals is determined by the number of data points collected in the respective dimension.

A one-dimensional (1D) <sup>1</sup>H NMR spectrum of a protein requires a few amount of molecules and is fast (less than one minute) to perform and can unambiguously indicate if a protein is folded or not. Protein folding is then observed with the dispersion of the resonance peaks, reflecting the structure of the protein (large dispersion = stable folding), on 1D proton spectra or on two-dimensional (2D) <sup>1</sup>H-<sup>15</sup>N correlation spectra if the protein is <sup>15</sup>N-labeled.

In the 1D <sup>1</sup>H spectrum of a well-folded protein, the peaks are narrow and sharp and distributed over a large range of chemical shifts meaning a good signal dispersion; signals can be especially found at <sup>1</sup>H resonance values lower than 0.5ppm or higher than 8.5ppm

corresponding to high field-shifted methyl group protons and downfield-shifted amide protons respectively. In contrast, the peaks are broader and not as widely dispersed in the spectrum of an unfolded or partially folded protein [44].

For example, in a fast 1D NMR spectrum of MAST2-PDZ (Fig. 6A), we observed sharp and narrow peaks that cover a large range of chemical shifts, i.e. peaks in the negative ppm range between 0 and -1ppm corresponding to upfield methyl protons located in the hydrophobic core of the domain, and peaks out near 10ppm corresponding to downfield backbone amide protons. This is a good indication that MAST2-PDZ is folded. This is a rather fast and qualitative technique to assess protein folding. Also, all the protons of the sample are seen in the 1D spectrum. In the past decades, many methods have been developed to reduce the water signal around 4.7ppm; however, signals from protons from the buffer can drastically complicate the analysis of spectra. For example, the 1D NMR spectrum of MAST2-PDZ (Fig. 6A) displays intense Tris peaks around 3ppm, but the peaks of the protein are still visible.

The 2D  $^1\text{H}$ - $^{15}\text{N}$  correlation HSQC spectrum [45] provides the “fingerprint” of a protein as the dispersion of cross-peaks is unique to each protein folding and sequence. On the contrary, unfolded proteins have all similar easily recognizable restrained patterns with a very limited dispersion of cross peaks. In addition, the peak width is related to the molecular mass of the molecule. We expect for MAST2-PDZ quite narrow peaks consistent with the small size of the domain (Fig. 6B). Unexpected large peaks can be observed for auto-associated domains or/and for proteins in conformational exchange. For example, in the 2D spectrum of the free MAST2-PDZ, we do not observe the expected number of peaks and the peaks are broad due to an auto-association and an exchange in the low-to-medium NMR timescales between monomers and dimers while upon binding of the ligand many additional peaks (in red) appear and peaks are sharper when MAST2-PDZ recovers its monomeric form [8, 36]. In

conclusion, NMR spectrum is highly sensitive to the folded state, oligomerization and aggregation state of a protein, and is thus efficient for the quality control of protein production [46].

A simple protocol to run a NMR experiment is given below:

1. Transfer each sample into an NMR tube, add 1-5% D<sub>2</sub>O and insert the sample plus the spinner into the Magnet with the sampler changer.
2. Adjust manually the probe equipped with an ATM module for each nucleus adapted to the experience to ensure the best sensitivity of the probe.
3. Lock the magnet to keep its field constant and stable during the data acquisition and shim the magnet to make it uniform and homogenous for a sample to get a reasonable resolution and line shape.
4. Determine the 90° transmitter pulse.
5. Setup Parameters and Acquire Data.
6. Process Data to convert data into NMR spectra.
7. Conformational stability/folding state.

#### **4. Notes**

1. Absorbance at the characteristic wavelength of 280 nm (i.e. A<sub>280nm</sub>) cannot be used for quantitative analysis without protein sequence since the amino acid composition varies from one protein to another. Indeed, variations in amino acid composition give molar absorptivities ( $\epsilon_{280nm}$ ) that vary from one protein to another.
2. Add glycerol gives the solution better to put down samples in the wells. If the solution is yellow while it should be blue because BBP turns yellow at acidic pH, add a few microliters of NaOH to turn blue the solution.
3. The materials are already fully described in [39].

4. The buffer should contain at least 50mM of salt to avoid interactions with the column matrix and 0.1% (w/v) NaN<sub>3</sub> to prevent microbial growth. All solutions must be freshly filtered and degassed to avoid background scattering and an unstable baseline.
5. Cuvettes with pathlengths ranging from 0.01-10mm are used with a 20-1000μL volume of protein solution at a concentration of 50μg/ml to 1mg/mL.
6. If you load 10μg that corresponds to 10μL at 1mg/ml of protein, you will need sensitivity of about 100ng per band to detect contamination. The detection limit of Coomassie blue staining is approximately 100ng per band, whereas silver stains and fluorescent dye have a detection limit of approximately 1ng of protein per band [20].
7. Using gradient gels with increasing acrylamide concentration from top to bottom usually 4-12% allows the separation and visualization of large proteins (200 kDa) that can still enter the gel at the top and small (10 kDa) polypeptides such as PDZ domains on the same gel as they retain small and fast-moving proteins in the bottom of the gel.
8. Proteins larger than 800 kDa will not enter in the polyacrylamide gel.
9. Check that there is no air bubble in the cuvette.
10. Record a spectrum over a wide wavelength range with a wavelength as a starting point where the solute does not absorb.
11. HCCA, DHB, and sDHB for PDZ samples with molecular masses below 20kDa and Sinapinic acid is usually used for constructs encompassing PDZ domain above 20kDa. However, sinapinic acid produces adducts that may impede the resolution and the precision of the m/z. In this case, the HCCA matrix could be a good alternative for large PDZ-containing constructs.

If the buffer is supplemented with more than 1% glycerol or detergent, the best practice is to concentrate the sample and then dilute it in water to reach the concentration required, to minimize the effect of the detergent or glycerol.

Sometimes, the concentration of the purified PDZ is below what is required to perform the MALDI-TOF analysis. Concentration and desalting of PDZ constructs of molecular masses lower than 15 kDa could be achieved on a ZipTip C18. ZipTip C4 is used for PDZ constructs of higher molecular mass.

12. High laser power triggers the loss of resolution and precision on the m/z.
13. A database of the most common protein post translation modifications is found here: <https://abrf.org/delta-mass>
14. Between 25 and 50 pmol of PDZ are required. In case of quantity limitation, use a ZipTip to desalt and concentrate the protein. Keep in mind that the PDZ solution must be pure to avoid false positive.
15. To remove dust or salt crystals, the buffer must be filtered.
16. For quality control, we do not recommend to filter or centrifuge the sample in order to detect aggregates; if DLS is used for other purposes, it is advisable to filter the sample and centrifuge it for 5min in a benchtop centrifuge at full speed before loading.
17. Care should be taken to choose the column as an elongated 10kDa protein may behave on the column like a globular 35kDa protein. Be aware that SEC columns are often described by their separative power on globular protein.
18. Columns are usually stored in 20% (v/v) ethanol. Therefore, the column must first be washed extensively with water (at least 2 column volumes) as ethanol may precipitate salts from the buffer.
19. The buffer should be stable over time as any change, such as oxidation, will lead to a drift in the baseline of the refractometer. The software baseline correction can compensate negligible drift.



20. High concentrations of chloride ions absorb at wavelengths lower than 215nm and limit the use of the spectrum below that wavelengths. To avoid scattering, buffer and sample should be filtered and degassed.
21. Chloride ions absorb strongly in the far-UV so low concentrations of NaCl and KCl in CD buffer are recommended. The best way would be to prepare a phosphate buffer by mixing di- and mono-sodium phosphates. In addition, DTT, detergents, glycerol or imidazole that absorb in the far-UV should be removed.

## Acknowledgement

The authors acknowledge Florence Cordier for technical support in NMR. The authors thank the molecular biophysics facility at Institut Pasteur for providing cutting edge instruments.

## References

1. Weatheritt RJ, Luck K, Petsalaki E, Davey NE, Gibson TJ (2012) The identification of short linear motif-mediated interfaces within the human interactome. *Bioinformatics* 28:976–982 . <https://doi.org/10.1093/bioinformatics/bts072>
2. Caillet-Saguy C, Maisonneuve P, Delhommel F, Terrien E, Babault N, Lafon M, Cordier F, Wolff N (2015) Strategies to interfere with PDZ-mediated interactions in neurons: What we can learn from the rabies virus. *Prog Biophys Mol Biol* 119:53–59 . <https://doi.org/10.1016/j.pbiomolbio.2015.02.007>
3. Luck K, Charbonnier S, Travé G (2012) The emerging contribution of sequence context to the specificity of protein interactions mediated by PDZ domains. *FEBS Lett* 586:2648–2661 . <https://doi.org/10.1016/j.febslet.2012.03.056>
4. Ye F, Zhang M (2013) Structures and target recognition modes of PDZ domains: recurring themes and emerging pictures. *Biochem J* 455:1–14 . <https://doi.org/10.1042/BJ20130783>
5. Wang CK, Pan L, Chen J, Zhang M (2010) Extensions of PDZ domains as important structural and functional elements. *Protein Cell* 1:737–751 . <https://doi.org/10.1007/s13238-010-0099-6>
6. Delhommel F, Cordier F, Bardiaux B, Bouvier G, Colcombet-Cazenave B, Brier S, Raynal B, Nouaille S, Bahloul A, Chamot-Rooke J, Nilges M, Petit C, Wolff N (2017) Structural Characterization of Whirlin Reveals an Unexpected and Dynamic

- Supramodule Conformation of Its PDZ Tandem. *Structure* 25:1645-1656.e5 .  
<https://doi.org/10.1016/j.str.2017.08.013>
7. Long J, Wei Z, Feng W, Yu C, Zhao Y, Zhang M (2008) Supramodular nature of GRIP1 revealed by the structure of its PDZ12 tandem in complex with the carboxyl tail of Fras1. *J Mol Biol* 375:1457–1468 . <https://doi.org/10.1016/j.jmb.2007.11.088>
  8. Delhommel F, Chaffotte A, Terrien E, Raynal B, Buc H, Delepierre M, Cordier F, Wolff N (2015) Deciphering the unconventional peptide binding to the PDZ domain of MAST2. *Biochem J* 469:159–168 . <https://doi.org/10.1042/BJ20141198>
  9. Wu J, Yang Y, Zhang J, Ji P, Du W, Jiang P, Xie D, Huang H, Wu M, Zhang G, Wu J, Shi Y (2007) Domain-swapped dimerization of the second PDZ domain of ZO2 may provide a structural basis for the polymerization of claudins. *J Biol Chem* 282:35988–35999 . <https://doi.org/10.1074/jbc.M703826200>
  10. Chang BH, Gujral TS, Karp ES, BuKhalid R, Grantcharova VP, MacBeath G (2011) A Systematic Family-wide Investigation Reveals that ~30% of Mammalian PDZ Domains Engage in PDZ-PDZ Interactions. *Chemistry & Biology* 18:1143–1152 .  
<https://doi.org/10.1016/j.chembiol.2011.06.013>
  11. Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227:680–685 . <https://doi.org/10.1038/227680a0>
  12. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins M, Appel RD, Bairoch AM (2005) Protein identification and analysis tools on the ExPASy server. In: *The Proteomics Protocols Handbook*. Humana Press, pp 571–607
  13. Noble JE (2014) Quantification of protein concentration using UV absorbance and Coomassie dyes. *Meth Enzymol* 536:17–26 . <https://doi.org/10.1016/B978-0-12-420070-8.00002-7>
  14. Glasel JA (1995) Validity of nucleic acid purities monitored by 260nm/280nm absorbance ratios. *BioTechniques* 18:62–63
  15. Pace CN, Vajdos F, Fee L, Grimsley G, Gray T (1995) How to measure and predict the molar absorption coefficient of a protein. *Protein Sci* 4:2411–2423 .  
<https://doi.org/10.1002/pro.5560041120>
  16. Karas M, Hillenkamp F (1988) Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem* 60:2299–2301 .  
<https://doi.org/10.1021/ac00171a028>
  17. Suckau D, Resemann A (2003) T3-sequencing: targeted characterization of the N- and C-termini of undigested proteins by mass spectrometry. *Anal Chem* 75:5817–5824 .  
<https://doi.org/10.1021/ac034362b>
  18. Philo JS (2006) Is any measurement method optimal for all aggregate sizes and types? *The AAPS journal* 8:E564–71 . <https://doi.org/10.1208/aapsj080365>
  19. Nobbmann U, Connah M, Fish B, Varley P, Gee C, Mulot S, Chen J, Zhou L, Lu Y, Shen F, Yi J, Harding SE (2007) Dynamic light scattering as a relative tool for assessing the

molecular integrity and stability of monoclonal antibodies. *Biotechnology & genetic engineering reviews* 24:117–28

20. Raynal B, Lenormand P, Baron B, Hoos S, England P (2010) Quality assessment and optimization of purified protein samples: Why and how? *Microbial Cell Factories* 13: . <https://doi.org/10.1186/s12934-014-0180-6>
21. Fekete S, Beck A, Veuthey J-L, Guillarme D (2014) Theory and practice of size exclusion chromatography for the analysis of protein aggregates. *J Pharm Biomed Anal* 101:161–173 . <https://doi.org/10.1016/j.jpba.2014.04.011>
22. Sahin E, Roberts CJ (2012) Therapeutic Proteins. 899:403–423 . <https://doi.org/10.1007/978-1-61779-921-1>
23. Daviter T, Chmel N, Rodger A (2013) Circular and Linear Dichroism Spectroscopy for the Study of Protein-Ligand Interactions. In: Williams M, Daviter T (eds) *Protein-Ligand Interactions: Methods and Applications*. Humana Press, pp 211–241
24. Schuck P (2003) On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation. *Analytical Biochemistry* 320:104–124 . [https://doi.org/10.1016/S0003-2697\(03\)00289-6](https://doi.org/10.1016/S0003-2697(03)00289-6)
25. Demeler B (2010) Methods for the design and analysis of sedimentation velocity and sedimentation equilibrium experiments with proteins. *Current Protocols in Protein Science* 1–24 . <https://doi.org/10.1002/0471140864.ps0713s60>
26. Balbo A, Zhao H, Brown PH, Schuck P (2009) Assembly, loading, and alignment of an analytical ultracentrifuge sample cell. *J Vis Exp* e1530 . <https://doi.org/10.3791/1530>
27. Schuck P (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys J* 78:1606–1619 . [https://doi.org/10.1016/S0006-3495\(00\)76713-0](https://doi.org/10.1016/S0006-3495(00)76713-0)
28. Micsonai A, Wien F, Bulyáki É, Kun J, Moussong É, Lee Y-H, Goto Y, Réfrégiers M, Kardos J (2018) BeStSel: a web server for accurate protein secondary structure prediction and fold recognition from the circular dichroism spectra. *Nucleic Acids Res* 46:W315–W322 . <https://doi.org/10.1093/nar/gky497>
29. Medrano G, Dolan MC, Condori J, Radin DN, Cramer CL (2012) Quality assessment of recombinant proteins produced in plants. *Methods Mol Biol* 824:535–564 . [https://doi.org/10.1007/978-1-61779-433-9\\_29](https://doi.org/10.1007/978-1-61779-433-9_29)
30. Structural Genomics Consortium, China Structural Genomics Consortium, Northeast Structural Genomics Consortium, Gräslund S, Nordlund P, Weigelt J, Hallberg BM, Bray J, Gileadi O, Knapp S, Oppermann U, Arrowsmith C, Hui R, Ming J, dhe-Paganon S, Park H, Savchenko A, Yee A, Edwards A, Vincentelli R, Cambillau C, Kim R, Kim S-H, Rao Z, Shi Y, Terwilliger TC, Kim C-Y, Hung L-W, Waldo GS, Peleg Y, Albeck S, Unger T, Dym O, Prilusky J, Sussman JL, Stevens RC, Lesley SA, Wilson IA, Joachimiak A, Collart F, Dementieva I, Donnelly MI, Eschenfeldt WH, Kim Y, Stols L, Wu R, Zhou M, Burley SK, Emtage JS, Sauder JM, Thompson D, Bain K, Luz J, Gheyi T, Zhang F, Atwell S, Almo SC, Bonanno JB, Fiser A, Swaminathan S, Studier FW, Chance MR, Sali A, Acton TB, Xiao R, Zhao L, Ma LC, Hunt JF, Tong L, Cunningham

- K, Inouye M, Anderson S, Janjua H, Shastry R, Ho CK, Wang D, Wang H, Jiang M, Montelione GT, Stuart DI, Owens RJ, Daenke S, Schütz A, Heinemann U, Yokoyama S, Büsow K, Gunsalus KC (2008) Protein production and purification. *Nat Methods* 5:135–146 . <https://doi.org/10.1038/nmeth.f.202>
31. Dupeux F, Röwer M, Seroul G, Blot D, Márquez JA (2011) A thermal stability assay can help to estimate the crystallization likelihood of biological samples. *Acta Crystallogr D Biol Crystallogr* 67:915–919 . <https://doi.org/10.1107/S0907444911036225>
  32. Bruce D, Cardew E, Freitag-Pohl S, Pohl E (2019) How to Stabilize Protein: Stability Screens for Thermal Shift Assays and Nano Differential Scanning Fluorimetry in the Virus-X Project. *J Vis Exp*. <https://doi.org/10.3791/58666>
  33. Boivin S, Kozak S, Meijers R (2013) Optimization of protein purification and characterization using Thermofluor screens. *Protein Expr Purif* 91:192–206 . <https://doi.org/10.1016/j.pep.2013.08.002>
  34. Genera M, Samson D, Raynal B, Haouz A, Baron B, Simenel C, Guerois R, Wolff N, Caillet-Saguy C (2019) Structural and functional characterization of the PDZ domain of the human phosphatase PTPN3 and its interaction with the human papillomavirus E6 oncoprotein. *Scientific Reports* 9:7438 . <https://doi.org/10.1038/s41598-019-43932-x>
  35. Monsellier E, Bedouelle H (2005) Quantitative measurement of protein stability from unfolding equilibria monitored with the fluorescence maximum wavelength. *Protein Eng Des Sel* 18:445–456 . <https://doi.org/10.1093/protein/gzi046>
  36. Terrien E, Chaffotte A, Lafage M, Khan Z, Préhaud C, Cordier F, Simenel C, Delepierre M, Buc H, Lafon M, Wolff N (2012) Interference with the PTEN-MAST2 interaction by a viral protein leads to cellular relocalization of PTEN. *Sci Signal* 5:ra58 . <https://doi.org/10.1126/scisignal.2002941>
  37. Rath A, Glibowicka M, Nadeau VG, Chen G, Deber CM (2009) Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. *Proc Natl Acad Sci USA* 106:1760–1765 . <https://doi.org/10.1073/pnas.0813167106>
  38. Shirai A, Matsuyama A, Yashiroda Y, Hashimoto A, Kawamura Y, Arai R, Komatsu Y, Horinouchi S, Yoshida M (2008) Global analysis of gel mobility of proteins and its use in target identification. *J Biol Chem* 283:10745–10752 . <https://doi.org/10.1074/jbc.M709211200>
  39. Duhoo Y, Girault V, Turchetto J, Ramond L, Durbesson F, Fourquet P, Nominé Y, Cardoso V, Sequeira AF, Brás JLA, Fontes CMGA, Travé G, Wolff N, Vincentelli R (2019) High-Throughput Production of a New Library of Human Single and Tandem PDZ Domains Allows Quantitative PDZ-Peptide Interaction Screening Through High-Throughput Holdup Assay. *Methods Mol Biol* 2025:439–476 . [https://doi.org/10.1007/978-1-4939-9624-7\\_21](https://doi.org/10.1007/978-1-4939-9624-7_21)
  40. Ivarsson Y, Wawrzyniak AM, Kashyap R, Polanowska J, Betzi S, Lembo F, Vermeiren E, Chiheb D, Lenfant N, Morelli X, Borg J-P, Reboul J, Zimmermann P (2013) Prevalence, specificity and determinants of lipid-interacting PDZ domains from an in-cell

screen and in vitro binding experiments. PLoS ONE 8:e54581 .  
<https://doi.org/10.1371/journal.pone.0054581>

41. Wawrzyniak AM, Kashyap R, Zimmermann P (2013) Phosphoinositides and PDZ domain scaffolds. *Adv Exp Med Biol* 991:41–57 . [https://doi.org/10.1007/978-94-007-6331-9\\_4](https://doi.org/10.1007/978-94-007-6331-9_4)
42. Layne E (1957) [73] Spectrophotometric and turbidimetric methods for measuring proteins. In: *Methods in Enzymology*. Academic Press, pp 447–454
43. Lees JG, Miles AJ, Wien F, Wallace BA (2006) A reference database for circular dichroism spectroscopy covering fold and secondary structure space. *Bioinformatics* 22:1955–1962 . <https://doi.org/10.1093/bioinformatics/btl327>
44. Page R, Peti W, Wilson IA, Stevens RC, Wüthrich K (2005) NMR screening and crystal quality of bacterially expressed prokaryotic and eukaryotic proteins in a structural genomics pipeline. *Proc Natl Acad Sci USA* 102:1901–1905 .  
<https://doi.org/10.1073/pnas.0408490102>
45. Bodenhausen G, Ruben DJ (1980) Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chemical Physics Letters* 69:185–189 .  
[https://doi.org/10.1016/0009-2614\(80\)80041-8](https://doi.org/10.1016/0009-2614(80)80041-8)
46. Kwan AH, Mobli M, Gooley PR, King GF, Mackay JP (2011) Macromolecular NMR spectroscopy for the non-spectroscopist. *FEBS J* 278:687–703 .  
<https://doi.org/10.1111/j.1742-4658.2011.08004.x>

## Figure captions

Fig. 1 Hydrodynamic radius distribution, measured by DLS. Inset autocorrelation function of the scattered intensity.

Fig. 2 Molecular mass and intrinsic viscosity measurement of MAST2 after size exclusion chromatography separation. Measure were performed by light scattering measurement at 90°C and the determined molecular mass showed the presence of monomer with an intrinsic viscosity of 4.2 dl/g.

Fig. 3 Sedimentation coefficient distribution of MAST2. Two species can be noticed with sedimentation compatible with a main peak of monomer and some dimer formation in the second peak.

Fig. 4 Circular dichroism spectra of MAST2-PDZ. Spectra of the free MAST2-PDZ, the free peptide and the complex MAST2-peptide are shown in blue, red and grey respectively.

Fig. 5 Thermal stability followed by DSC and DSF. A) DSC thermograms of PTPN3-PDZ unbound and complexed to PBM peptides. Free PTPN3-PDZ curve is shown as a black line and PTPN3-PDZ complexed to PBMs curves are shown as dashed lines. B) By calculating the maximum of the first derivative of the ratio of fluorescence (F) at 350 nm over 330 nm, the melting temperature (T<sub>m</sub>) can be derived; it corresponds to the temperature where 50% of the proteins are unfolded.

Fig. 6 NMR spectra of MAST2-PDZ. A) <sup>1</sup>H 1D spectrum of 150 μM MAST2-PDZ in complex with a viral peptide at 600 MHz <sup>1</sup>H frequency. B) <sup>1</sup>H-<sup>15</sup>N 2D HSQC spectra of 150 μM MAST2-PDZ alone (in black) and in complex with a viral peptide (in red) at 600 MHz <sup>1</sup>H frequency.

Fig. 7 Sequence and gel SDS-PAGE of MAST2-PDZ. A) Primary sequence of MAST2-PDZ. B) The SDS-PAGE gels of a GSTrap chromatography (left panel) and a size-exclusion chromatography (right panel) of the eluted fractions containing MAST2-PDZ.

Fig. 8 Interface of the data analysis software LabChip GX reviewer. Electropherogram (top) and virtual gel view (bottom) of the serial dilution of the supernatant of lysed cell extract of MBP-tagged MAST2-PDZ construct of the PDZome library. The concentration of lysozyme is constant and used as an internal reference for quantification. The peak corresponding to the PDZ construct is decreasing upon increasing dilution factor.

Fig. 9 UV spectra of two MAST2-PDZ samples after purification.

Fig. 10 MALDI-TOF spectra for two PDZs. In black, singly and doubly charged ions of an intact PDZ; in red, several peaks showing the proteolysis of the PDZ that cannot be easily identified on a SDS-gel.

Fig. 11 Top down sequencing of MAST2-PDZ. More than 10 and 20 amino acids match in C- and N-termini, respectively.