



HAL
open science

An in silico and chemical approach towards small protein production and application in phosphoproteomics

Ana M G C Dias, Olga Iranzo, Ana C A Roque

► To cite this version:

Ana M G C Dias, Olga Iranzo, Ana C A Roque. An in silico and chemical approach towards small protein production and application in phosphoproteomics. RSC Advances, 2015, 5 (25), pp.19743-19751. 10.1039/C4RA16934D . hal-01477091

HAL Id: hal-01477091

<https://hal.science/hal-01477091>

Submitted on 8 Jun 2020

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.



CrossMark
click for updates

Cite this: *RSC Adv.*, 2015, 5, 19743

An *in silico* and chemical approach towards small protein production and application in phosphoproteomics†

Ana M. G. C. Dias,^{ab} Olga Iranzo^{‡*b} and Ana C. A. Roque^{*a}

The human Pin1 WW domain (hPin1_WW) is a 38 residue protein which specifically recognizes ligands rich in proline and phosphorylated in Ser and Thr residues. This work presents a protocol for the improved chemical synthesis and modification of this protein through automated microwave assisted synthesis combined with the incorporation of pseudoproline units in the protein sequence. After purification, the protein was characterized by Mass Spectrometry and Circular Dichroism spectroscopy with results comparable to the same WW domain chemically synthesized by other strategies or biologically expressed. The protein was further immobilized on a matrix and tested for the selective binding and mild elution of phosphorylated sequences at Ser, Thr and Tyr residues. These results suggest that hPin1_WW is a useful protein scaffold for the purification of phosphorylated species in pTyr and pSer, which can be easily produced and modified by chemical methods.

Received 23rd December 2014

Accepted 9th February 2015

DOI: 10.1039/c4ra16934d

www.rsc.org/advances

1. Introduction

Antibodies and derived structures represent the most widespread platform to develop affinity reagents against several targets of medical and biotechnological relevance. Still, the need for stable and robust protein scaffolds amenable of randomization and easy production, has driven several researchers in the quest of alternatives to antibodies.^{1,2} Such scaffolds include protein domains as for example affibodies (derived from the protein A domain), fibronectin, lipocalins, DARPins and affitins.^{2,3} WW domains possess 38–40 residues in length, with two conserved tryptophan (W) residues spaced by 20–22 amino acids (which gives the name to the domains family)⁴ and are assembled on a three β -sheet structure. These motifs bind to proline rich sequences. WW domains are found in at least 200 multidomain proteins and are usually localized in the recognition region known to mediate protein–protein interactions.⁵ The characteristic structure of the proline residue allows a specific recognition, through weak interactions, which triggers quick responses to the changes in the cellular environment, further translated in signal pathways changes.⁶ WW

domains have been characterized and classified in five different groups based on their recognition sequences.⁴

One of the most widely studied WW domain is found in the human Pin1 protein (hPin1). This protein is a mitotic peptidyl-prolyl isomerase (PPIase) which has been identified as a mitosis cell cycle regulator. Its WW domain is representative of Group IV and specifically recognizes ligands rich in prolines and phosphorylated in Ser and Thr residues.⁷ Furthermore, hPin1 has been identified in several cellular pathways, since it regulates the isomerisation of phosphorylated residues, and its deregulation has been correlated with cancer, Alzheimer and autoimmune diseases.⁸ For that reason human Pin1WW (hPin1_WW) domain has been highly studied and its structure is well characterized, as well as its binding mode to phosphorylated peptides.⁹

Despite their small size, the biological production of WW domains usually requires co-expression with other proteins.^{10–12} In addition, their chemical synthesis is also challenging because WW domains possess a three β -sheet secondary structure usually regarded as difficult for Solid Phase Peptide Synthesis (SPPS). This is due to the establishment of hydrogen bonds between chains during peptide elongation, resulting in the formation of hydrophobic clusters and peptide aggregation. As a result, the solvation and coupling efficiencies decrease drastically, resulting in truncated peptides and deleted sequences. This turns out in long and difficult purification processes that lower to a large extent the reaction yields.¹³ However, the chemical synthesis allows the ease incorporation of mutations in the peptide sequence generating variability in the peptides. In addition, when employing automated synthesis, peptides can be obtained in a short period of time.

^aUCIBIO, REQUIMTE, Departamento de Química, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Campus Caparica, 2829-516 Caparica, Portugal. E-mail: cecilia.roque@fct.unl.pt; Fax: +351 21 294 8550

^bAix Marseille Université, Centrale Marseille, CNRS, iSm2 UMR 7313, 13397 Marseille, France. E-mail: olga.iranzo@univ-amu.fr

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c4ra16934d

‡ Previous address: Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, 2780-157 Oeiras, Portugal.

This is a very flexible strategy for peptide production especially useful for peptides shorter than 70 residues, as for higher peptide sizes, the synthesis gets complex, drastically decreasing the product purity and yield.¹³

Considering the potential of WW domains as scaffold proteins, we optimized the automated microwave assisted SPPS of hPin1_WW through the incorporation of pseudoproline dipeptide units. The stability of the produced structure was analyzed *in silico* and experimentally. The hPin1_WW was immobilized onto agarose beads to assess the potential of WW scaffolds for the affinity purification of target peptides and proteins.

2. Results and discussion

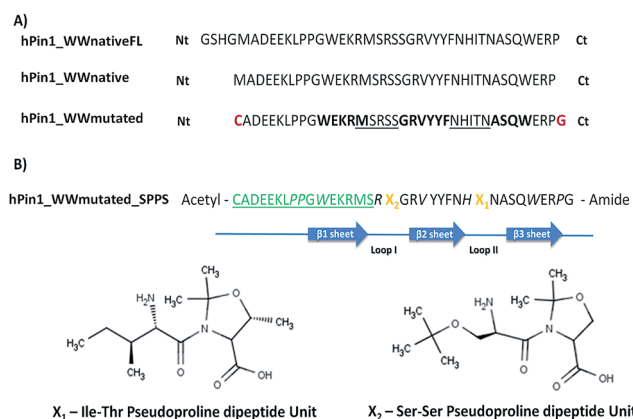
2.1 *In silico* studies – stability and affinity of human Pin1 WW domain

The human Pin1 WW domain (hPin1_WWnativeFL, PDB code: 1F8A) was selected as a model WW protein. The amino acid sequences responsible for the folding of the domain and recognition of the targets were maintained,²⁵ but two mutated structures were generated: (i) a smaller native version (hPin1_WWnative) containing only the amino acids important for folding;⁴ and (ii) a mutated version (hPin1_WWmutated) containing a Cys in the N-terminal (Nt) for subsequent peptide immobilization on a solid support. Additionally, hPin1_WWmutated has a Gly in the C-terminal (Ct) to facilitate the first coupling reaction in SPPS (Scheme 1A). The stability of the

hPin1_WW structures was assessed by Molecular Dynamics (MD) simulations (Fig. S1A†). Different parameters were considered, including grid size, and structural features that required adaptations to leave the Cys group outside the grid box – mimicking the immobilization onto a matrix. MD simulations analysis is presented for hPin1_WW in Fig. 1 and S1.† In general, in the Root Mean Square Deviation (RMSD) analysis all the structures of hPin1_WW domain presented similar behavior during the time of simulation. All the structures tend to equilibrate after 20–30 ns and the behavior is retained until the end of the simulation. The Root Mean Square Fluctuation (RMSF) analysis, which informs about the flexibility of each residue in the sequence over the time of the simulation, shows that the flexibility of the smaller versions decreases at the N terminal when compared with the WWnativeFL domain structure. Specifically, for hPin1_WWmutated, there is an increase in rigidity in β -sheet I, Loop II and β -sheet III compared with hPin1_WWnative structure (Fig. 1A).

The β -sheets from WW domains are less flexible, as they establish hydrogen bond interactions between each other, as observed in the H-bond interactions analysis (Fig. S1B and C†). In particular, the interactions between the backbone of amino acids residues Asn27–Ala32, Gln34–Tyr25, Glu13–Phe26, and Tyr24–Arg15 are important for the characteristic tertiary structure and WW domain stability.¹¹ The most flexible region is Loop I, since it is involved in the molecular recognition of ligands in the hPin1_WWnativeFL structure,^{4,26} but we still observed H-bond interactions between Ser17–Arg22 that are important for loop maintenance.^{11,27}

For the affinity studies the WW domain was used as a receptor and different peptides as ligands. The ligands tested were the native phosphorylated ligand, the non-phosphorylated version and a proline-rich peptide, and their structures were docked at the surface of WW domain using a blind docking strategy. As the three ligands tested possess proline residues, the recognition of the peptides through the consensus “X-P-groove region” formed by Tyr-27 and Trp-38 residues from WW domain was observed (Fig. S2†). The proline rich peptide presented the highest affinity towards the WW domain (K_a 2.49×10^7 M⁻¹). In this case, apart from the “X-P-groove region” the other proline residues are also recognized through hydrophobic interactions established with other WW domain residues. Regarding the phosphorylated peptide, the best binding conformation and the estimated affinity constant were similar to those described in the literature (K_a 2.28×10^5 M⁻¹ versus K_a 2.04×10^5 M⁻¹ (ref. 9)) (Fig. 1C and Table S1†). For the non-phosphorylated peptide version, the binding conformation is similar to the phosphorylated peptide but there is a decrease in affinity of one order of magnitude, due to the lack of hydrogen bond formation between Arg-21 and the phosphate group. These results can be explained due to the conserved recognition in WW domains of the Pro residue by a Tyr and Trp residues in the β -sheet. This recognition event occurs first, and restricts the liberty of the other residues in the peptide ligand.²⁸ Additionally, in the case of hPin1_WW, when peptides contain pSer/pThr residues the binding affinity increases as the phosphate group interaction with the loop I residues will contribute positively for



Scheme 1 Nomenclature of the sequences of hPin1 WW domain used in this work. (A) Sequences used in the *in silico* studies derived from PDB code 1F8A file, the mutants were generated using PyMol software. The hPin1_WWnativeFL is the native full length hPin1 WW domain (42 residues); hPin1_WWnative, corresponds to a sequence of 37 residues; and hPin1_WWmutated, corresponds to hPin1_WWnative mutated in the N terminal with a Cys and an introduction of a Gly in the C terminal (see Experimental section for details); (B) strategy for SPPS using pseudoproline units (see Experimental section for details), where X₁ = pseudoproline unit Fmoc-Ile-Thr(psiMe, Mepro)-OH and X₂ = pseudoproline unit Fmoc-Ser(tBu)-Ser(psiMe, Mepro)-OH; amino acids in italic needed special couplings, and for amino acids in green and underlined, a capping step was added. The structures of the pseudoproline units used are sketched. The chemical structures sketches were designed using Marvin Beans version 6.3.0 (ChemAxon).

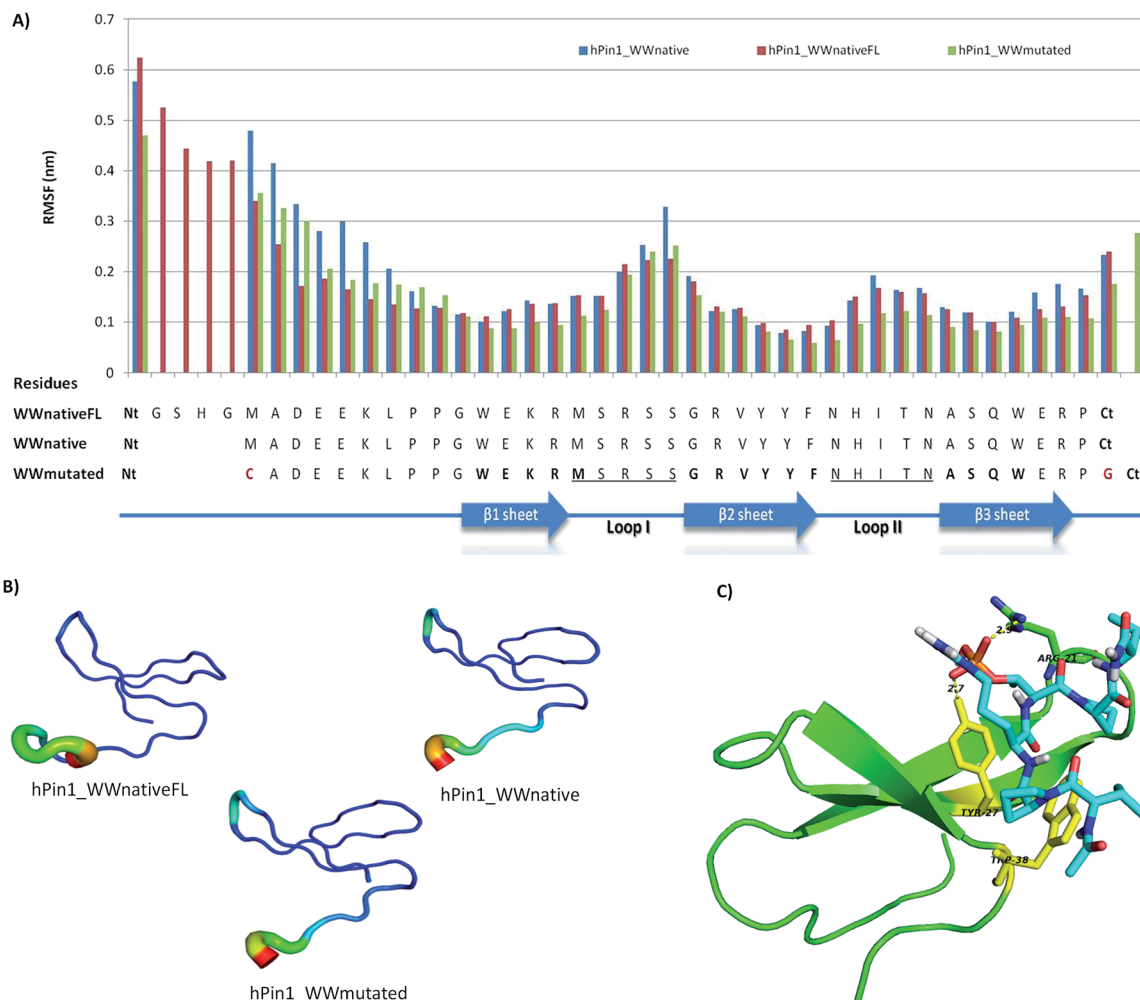


Fig. 1 *In silico* studies results. (A) Root Mean Square Fluctuation (RMSF) that analyses the flexibility of each residue in the structure over the time of simulation; (B) B-factor analysis that identifies the most flexible residues during simulation compared with the initial structure that was simulated; the results of the analysis are displayed in a color mode (most flexible residues in red and less flexible in dark blue). (C) Best docking solution between hPin1 and peptide VP*RpTPV. The H-bond interactions between the phosphorylated group in the Thr and Arg-21 (loop I) and Tyr-27 (β -sheet II) are identified. In yellow are represented the residues Tyr-27 and the Trp-38 that are described in the literature as part of the "X-P-groove" where the Pro* is recognized.

the binding. From these *in silico* studies was possible to conclude that the hPin1_WWmutated structure maintains its stability and molecular recognition sites against the native ligands of hPin1_WW.

2.2 Peptide solid-phase chemical synthesis using pseudoproline units

After studying *in silico* the stability of hPin1_WWmutated, this structure was chemically synthesized by Fmoc-based SPPS in an automated microwave system. Firstly, the hPin1_WWmutated peptide was produced using a standard protocol for SPPS (Table SII†). Using this approach, several co-products were obtained as shown in the HPLC chromatograms (Fig. S3A†) diffculting the purification step and therefore the overall yield of the synthesis.

Besides microwave technology, that promotes better coupling reaction between amino acids due to the kinetic power input,^{13,29} different strategies have been described in the

literature to overcome problems with difficult chemical synthesis. These include resins with an hydrophobic character (*e.g.* poli(ethyleneglicol) (PEG) based resins);³⁰ different solvent mixtures;³¹ chaotropic salts;¹³ pseudoproline units which are oxazolidine dipeptides known to disrupt secondary structure;³² and fragments condensation using native chemical ligation.^{33,34}

We explored the use of a PEG based resin and the incorporation of pseudoproline dipeptide units during synthesis.^{32,35} This was possible because hPin1_WWmutated domain possess Ser and Thr residues, and their location in the sequence is appropriate for the insertion of pseudoproline units³² (see Scheme 1B for details). Additionally, we introduced capping steps after the 23rd amino acid. The synthesized peptide (hPin1_WWmutated) was purified by reverse-phase preparative HPLC (Fig. S3B†), and its purity determined by reverse-phase analytical HPLC (Fig. 2C). ESI-MS analysis (Fig. S4†) confirmed the identity of the peptide. For hPin1_WWmutated,

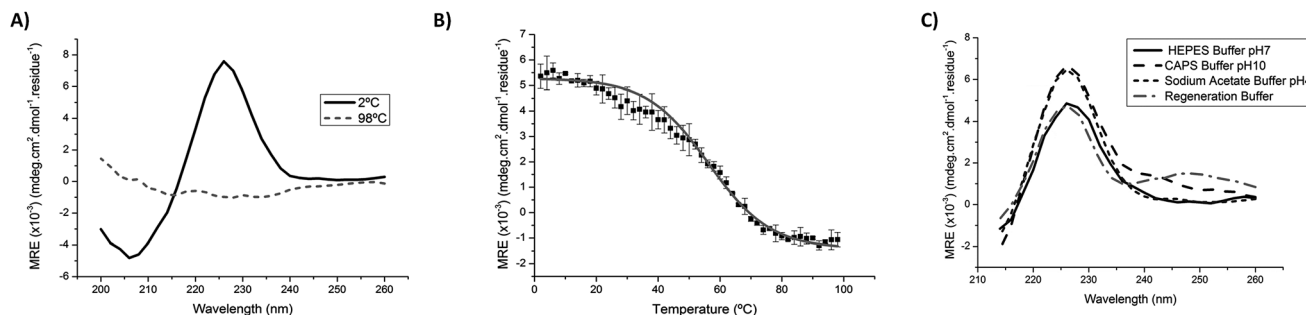


Fig. 2 Circular dichroism studies: (A) far-UV CD spectra of hPin1_WWmutated (50 μ M) at 2 $^{\circ}$ C and 98 $^{\circ}$ C in 20 mM MOPS pH 7.2, 100 mM NaCl; (B) fitting to the two state-model curve for determination of melting temperature (T_m); (C) far-UV CD spectra obtained at 4 $^{\circ}$ C using different buffers: 10 mM HEPES buffer pH 7, 10 mM sodium acetate pH 4, and 10 mM CAPS buffer pH 10 (all with 100 mM NaCl) and the regeneration buffer used for the solid support: 0.1 M NaOH in 30% isopropanol.

the production yield was 30% and the purity 95%. The hPin1_WWmutated was obtained through SPPS in less than 24 h and this strategy was repeated four times with high reproducibility. In each production cycle, 150–200 mg of peptide was obtained which demonstrates the robustness of the methodology.

Our yield results are difficult to compare with the literature since few authors report the actual yields obtained during chemical and biological production of WW domains. A similar yield (\sim 25%) was reported for the manual synthesis of another WW domain with 37 residues, FBP28 (Formin-binding protein), using also pseudoproline units, but with considerable longer synthesis time (\sim 60 h *versus* \sim 24 h in this work).³² These data highlights the benefits of using microwave technology (efficient kinetic power input) combined with automated systems.^{35,13}

2.3 Peptide characterization through circular dichroism spectroscopy

Circular Dichroism (CD) spectroscopic studies were performed to determine the secondary structure and stability of the chemically synthesized hPin1_WWmutated. The spectra were recorded in the Far-UV region to check for the characteristic signals of the WW domains. These proteins have a three β -sheet structure showing a CD spectrum with a maximum positive ellipticity at \sim 230 nm and a maximum negative ellipticity at 206 nm.¹² Fig. 2A shows the CD spectra obtained for hPin1_WWmutated at 2 $^{\circ}$ C in MOPS buffer pH 7.2, which shows a maximum positive ellipticity at 226 nm, the specific signal for hPin1_WW.²³ In addition, temperature denaturation experiments were performed to determine peptide stability. The change in ellipticity at 230 nm was monitored as temperature increased from 2 $^{\circ}$ C to 98 $^{\circ}$ C (Fig. 2B). The CD unfolding curve was fitted to a two state-model to determine the T_m of hPin1_WWmutated, using the equation described by Koepf *et al.*¹² The T_m determined using these conditions was 56.3 ± 0.5 $^{\circ}$ C ($R^2 = 0.998$) (Fig. 2B). This result is in agreement with the T_m value of 55 $^{\circ}$ C determined by Kraemer-Pecore *et al.* for hPin1_WW (34 residues) under the same buffer conditions.²³ Different T_m values have been reported for native hPin1_WW at pH 7 using different buffer conditions. Kaul *et al.*³⁶ reported a

T_m of 58 $^{\circ}$ C (36 residues length, 10 mM sodium phosphate buffer) and Deechongkit *et al.*³⁷ reported T_m values of 59 $^{\circ}$ C (34 residues length, 20 mM sodium phosphate buffer, 1 mM DTT). Interestingly, the same authors observed a 7 $^{\circ}$ C decrease in T_m values when the same structure was capped in the N and C-terminal with acetyl and amide group, respectively. This information suggests that the charges in the terminal regions of the peptide provide stability to the overall structure.³⁷

hPin1_WWmutated will be immobilized onto a solid support to explore its capability to capture target peptides and proteins. For that purpose different buffer conditions were tested (10 mM HEPES buffer pH 7; 10 mM sodium acetate buffer pH 4; 10 mM CAPS buffer pH 10 with 100 mM NaCl; and finally, a regeneration buffer: 0.1 M NaOH in 30% isopropanol) and their influence on the structure of the hPin1_WWmutated analyzed by CD spectroscopy. The results presented in Fig. 3C indicate that the signal at 226 nm of hPin1_WWmutated is maintained despite the nature and pH of the buffer. When temperature was increased to 23 $^{\circ}$ C, no major differences in the signal were observed.

2.4 Immobilization of hPin1_WWmutated onto solid support for affinity purification

Phosphorylation is one of the most important post translational modification in proteins, and it has been implicated in several steps in cellular regulation.³⁸ As this modification can be temporary and reversible in proteins, the recognition of phosphorylated and non-phosphorylated proteins can work as a molecular switch. Kinases and phosphatases are responsible for phosphorylation state modification. For that reason, perturbations in the equilibrium of the natural protein modifications have been correlated with several diseases.⁸ These observations have increased the interest in the phosphoproteome, which relates all the knowledge obtained about these modifications through different methods: phospho-enrichment methods and mass spectrometry (MS).³⁹ Since hPin1_WW domain it is a natural binding motif of phosphorylated peptides, we aim to immobilize it in solid support to demonstrate its application in the enrichment/purification of phosphorylated species.

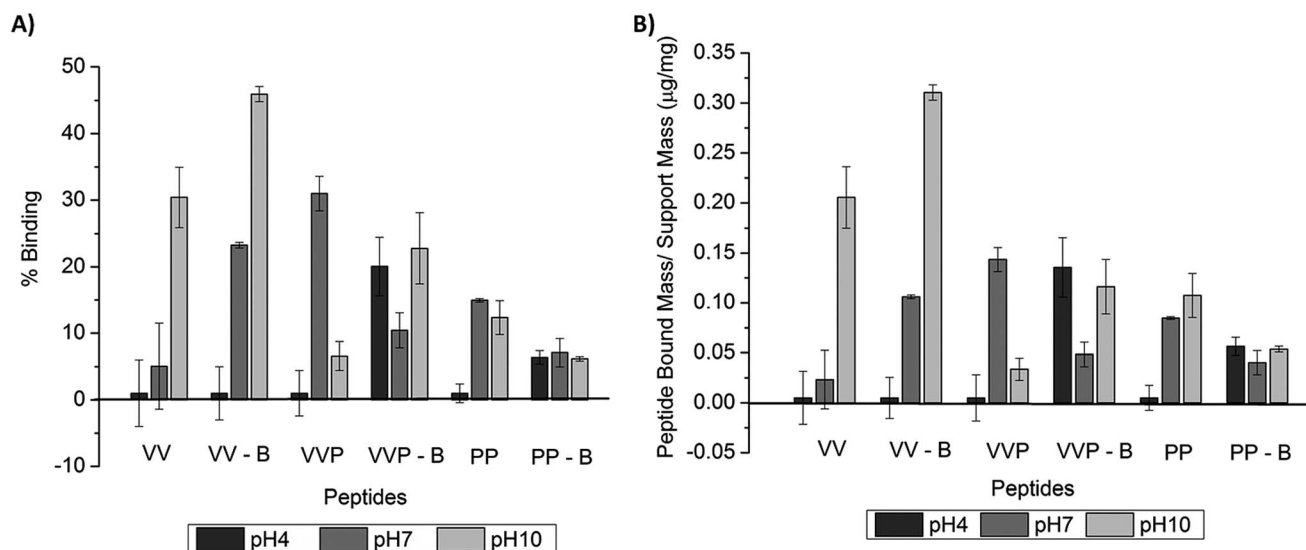


Fig. 3 Binding studies at different pH values for phosphorylated and non-phosphorylated peptides in batch mode using 100 mg of hPin1WW_Ag. (A) Binding percentage and (B) peptide bound mass/support mass. VV, non-phosphorylated native peptide – VPRpTPV; VVP, phosphorylated native peptide – VPRpTPV; PP, pro-rich peptide – PPPYPYPAW; VV-B, agarose unmodified tested with VV; VVP-B, agarose unmodified tested with VVP and PP-B, agarose unmodified tested with PP. Binding percentage is calculated using $(\text{mass peptide loaded} - \text{mass peptide output}) \times 100 / \text{mass peptide loaded}$ and peptide bound mass/support mass ($\mu\text{g mg}^{-1}$) is calculated using $(\text{mass peptide loaded} - \text{mass peptide output}) / \text{mass support}$, (number of repeats $n = 2$).

For the immobilization of the hPin1_WWmutated domain, the chemistry of sulfo-SMCC (sulfo-succinimidyl-4-[N-maleimidomethyl]cyclohexane-1-carboxylate) was chosen using cross-linked agarose as solid-support (hPin1_WWAg). The immobilization yield was 60% (2.45×10^{-3} μmol peptide per mg support) and the strategy was very reproducible. After immobilization the adsorbent hPin1_WWAg was tested with pure peptide solutions for binding, namely the native ligand peptide (VPRpTPV, VVP), the non-phosphorylated version (VPRpTPV, VV), and also the proline-rich peptide (PP). Since hPin1_WW is a natural phosphate binding domain, we aimed first to maximize the conditions to recognize the phosphate group, which possesses different charges at different pHs. To determine the best pH value we tested three different buffers and conditions: pH 4 ($-\text{OPO}_3\text{H}^{-1}$); pH 7 and pH 10 ($-\text{OPO}_3\text{H}^{-1/-2}$),⁴⁰ the buffers were chosen to avoid phosphate groups which act as competitors. The buffers selected were 25 mM sodium acetate buffer at pH 4; 25 mM HEPES buffer at pH 7; 25 mM CAPS buffer at pH 10, with 150 mM NaCl.

The results for a small scale of adsorbent (100 mg) are shown in Fig. 3, where agarose with no modification was used as a control. At pH4 the peptides presented very low binding values, as at this buffer condition the support is positively charged and the peptides too, so there is low electrostatic attraction, *i.e.* low binding. The exception is observed for VVP that has a negative charge ($-\text{OPO}_3\text{H}^{-1}$) and for that reason can present some binding to the unmodified support. At pH10, hPin1_WWAg is negatively charged (-2) and peptide VV is positively charged, this will promote the electrostatic attraction to the support and high binding. However, the peptide VVP presents a phosphate group (negative charge) and thus, its overall positive charge is

decreased which results in lower electrostatic attractions, *i.e.* low binding. The same is observed for the PP peptide that at this pH is negatively charged. Considering the unmodified support, agarose at pH 10 presents a negative charge and consequently the VV will present the highest binding as opposed to the other tested peptides.

The small differences observed between Fig. 3A and B are due to the fact that in Fig. 3A the data is normalized by the mass of peptide loaded whereas in Fig. 3B it is normalized by the mass of support used, as there are slight differences between each assay.

From these results, we selected the buffer condition where the phosphorylated peptide had the highest binding and the non-phosphorylated version had the lowest binding, thus, HEPES buffer at pH 7. Under this condition, the proline-rich peptide was also recognized by the support, however in less extent than the native phosphorylated ligand. The binding percentage for the VPRpTPV is 30% (0.14 μg peptide bound per mg support). We repeated the screening of the phosphorylated peptide with $3 \times$ the solid support mass (300 mg), and the binding percentage increased for 85% (0.16 μg peptide bound per mg support).

After this study, the affinity of hPin1_WWAg was tested for a model protein, Bovine Serum Albumin (BSA) chemically modified with phosphate groups: phosphothreonine (BSA-pT), phosphoserine (BSA-pS) and phosphotyrosine (BSA-pY), which are the most frequent modifications observed in proteins. As a control, we tested unmodified BSA (Fig. 4A and B). The binding percentage was higher for the BSA-pS and BSA-pY than for BSA-pT and ranged from 35% (0.059 μg protein bound per mg support) to 40% (0.065 μg protein bound per mg support). The binding for the BSA-pT

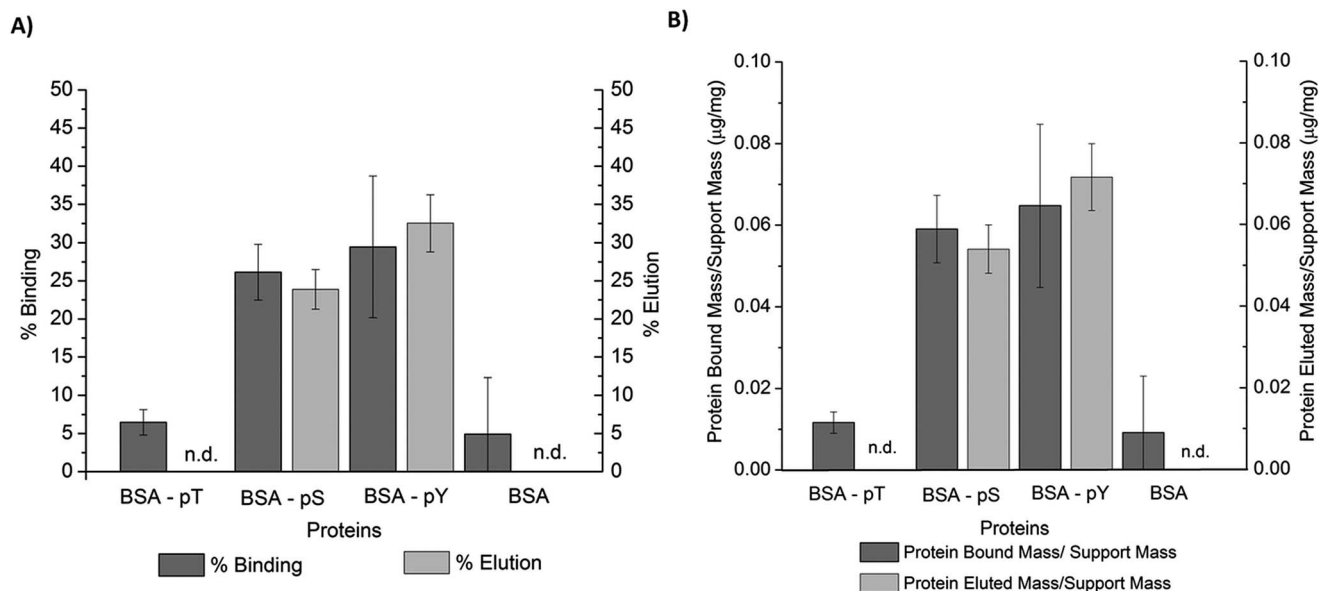


Fig. 4 (A) Binding studies using 25 mM HEPES buffer + 150 mM NaCl at pH 7 for proteins BSA-pT, BSA-pS, BSA-pY and BSA and (B) elution with 50 mM Phosphate buffer + 50 mM NaCl, pH 7. (n.d. – non determined). Elution percentages are calculated as (mass protein eluted) \times 100/mass protein loaded (number of repeats $n = 2$).

and unmodified BSA was approximately 5%. These results indicate that the recognition is made through the phosphorylated group, but also by interactions with other residues in the protein that contribute for the binding. One elution condition was tested for the BSA-pS and BSA-pY, using a competitive strategy with phosphate groups and mild conditions (PBS buffer). The elution was successful for BSA-pS (0.054 μg protein eluted per mg support) and BSA-pY (0.072 μg protein eluted per mg support), which corresponds to protein recoveries of 92% and 90% respectively for BSA-pS and BSA-pY (Fig. 4A and B).

3. Conclusions

The synthetic approach presented here shows the production of the hPin1_WW domain with the characteristic folding of a WW domain and a T_m value in agreement with those reported in the literature. The pseudoproline units strategy using automated microwave technology demonstrated to be a robust, less time consuming and an extremely efficient route for the production and modification of this small protein with good yield.

It was also shown that hPin1_WWmutated immobilized in a solid support maintains the affinity for the native phosphorylated ligand VPRpTPV. From the binding studies with BSA and their modified versions with pSer and pTyr, we conclude that it has affinity for these modified proteins and thus recognition is mediated by the phosphorylated residues. In addition, it is possible to completely recover the protein bound to the solid support under mild conditions (competitive elution with phosphate buffer at neutral pH). In summary, the data presented here show that the chemical production of hPin1_WW at high yields is possible as well as its immobilization onto solid supports. Therefore, we demonstrate a novel platform for the chemical production and modification of a small protein

scaffold and its application for the purification of phosphorylated proteins and peptides.

4. Experimental

4.1 *In silico* studies

4.1.1 Molecular modeling studies – structures preparation.

The structure of hPin1_WW used in this report (hPin1_WW-nativeFL, 1 to 41 residue, 42 residues long) was retrieved from Protein Data Bank (PDB) code 1F8A. This structure is the native and full length hPin1_WW structure that was decreased to have only the amino acids important for folding in WW domains, therefore only the 5 to 41 residue were considered (hPin1_WWnative, 37 residues long). This sequence was then mutated in the N-terminal, where a Met was substituted for a Cys and in the C-terminal where a Gly was added (hPin1_WW-mutated, 38 residues long). These modifications are important, since Cys will be used in the peptide attachment to the solid support, and the Gly in the C-terminal will facilitate the initiation of the SPPS. The N- and C-terminal of all sequences were capped by an acetyl group and an amide group, respectively (Scheme 1A).

4.1.2 Molecular dynamics – stability studies.

Molecular dynamics were carried out on the GROMACS 4.5.5 simulation package¹⁴ running in parallel on the in-house Sun Grid Engine (SGE) high performance computer cluster. The structures of hPin1_WWdomain were used as starting structures for dynamics. Each peptide was placed in an octahedral box with a cut off distance between the peptide and the box edges of 10 Å and filled with an explicit SPCE water model. Chloride counter ions were added in order to neutralize the global charge of the system. The united-atom force field GROMOS 43a1 (ref. 15) was applied. All simulations were run under isothermal-isobaric

(NPT) ensemble conditions, coupled to the Berendsen barostat with a reference pressure of 1.0 bar and a coupling time constant of 0.6 ps,¹⁶ as well as to the V-rescale thermostat with a reference temperature 300 K and coupling time constant of 0.1 ps.¹⁷ The simulation time step was 2 fs and all H-bonds were constrained by using the LINKS algorithm,¹⁸ as implemented in GROMACS software. After an energy minimization of 2000 steps with the steepest descent, followed by 1000 steps with the conjugated gradient algorithm, the system was equilibrated in three subsequent steps of 100 ps each. A position restraint was applied with a decreasing force constant of 1000, 100, and 10 kJ mol⁻¹ on all bonds, respectively. Finally, constraints were removed and each system was simulated up to 50 ns. The visualization software PyMol 1.3 (ref. 19) and VMD 1.9 (ref. 20) were used to identify the main interactions established between the residues of the peptide based on a cut-off distance criteria of 3.2 Å for H-bond interactions between donor and acceptor atoms.

4.2 Docking studies – affinity studies

The affinity studies were conducted between hPin1_WWmutated structure (receptor) and several peptides (ligands): VPRpTPV (native ligands of hPin1_WWnative domain),⁹ the non-phosphorylated version of the same peptide and PPPYP (a proline-rich peptide derived from the natural ligand of YAP65 WW domain).¹² The peptides were designed using Pymol 1.3 software and saved in a pdb file format. The coordinates of receptor and ligand were converted into pdbqt file format. The grid parameter files were setup using the AutoDock 4.0.7 tool package.²¹ A Cubic grid was defined using grid size [100 100 80]; spacing of 0.375 Å and a Grid center coordinates of [-5.826 8.505 54.406]. Since the aim of the work was to immobilize hPin1_WW into a solid support, the box was defined such that the Cys residue was maintained out of the box. The ligands had degree of freedom and were allowed to explore the surface of the receptor. The Lamarckian genetic algorithm was used and a maximum number of 256 solutions were tested. In the solutions analysis was requested that the structures were grouped with a RMS tolerance of 4 Å. After docking, the solutions were analyzed based on an energy criteria – best scored clusters, with lower estimated binding free energy, and also based on number of elements criteria – from the best scored clusters only the ones with highest number of elements were further analyzed. Thus we aimed to select the best binding conformation, based in the energy score and the frequency of this conformation. The results were visualized using PyMol 1.3 software.

4.3 Peptide chemical synthesis and characterization

4.3.1 Solid-phase peptide chemical synthesis. *N,N*-Dimethylformamide (DMF), acetonitrile, dichloromethane and *N*-methylpyrrolidone (NMP) were purchased from Fisher Scientific (Loures, Portugal), *N,N*-diisopropylethylamine (DIEA), piperidine, trifluoroacetic acid (TFA), thioanisole, 1,2-ethanedithiol and anisole were purchased from Sigma-Aldrich (Loures, Portugal). All Fmoc protected amino acids, 2-(1*H*-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) resins and pseudoproline units, were purchased from EMD Biosciences/

Merck Biosciences (Darmstadt, Germany). The hydroxybenzotriazole (HOBt) was purchase from CEM Corporation (North Carolina, USA).

3-(*N*-Morpholino)propanesulfonic acid (MOPS); 4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid (HEPES); 3-(cyclohexylamino)-1-propane sulfonic acid (CAPS), ethylenediaminetetraacetic acid (EDTA) were used in buffers preparation and were obtained from Sigma-Aldrich (Loures, Portugal).

The reagents used in these procedures were always high grade.

The first synthesis of hPin1_WWmutated was carried out in a CEM Liberty microwave peptide synthesizer (CEM Corporation, North Carolina, USA) using the standard protocol provided by the manufacturer (Table SII†). A Rink amide MBHA resin 100–200 mesh with a substitution of 0.59 mmol g⁻¹ was employed (scale 0.25 mmol). All the amino acid reaction couplings were undertaken at 75 °C, except for Asp, His and Cys that were carried out at 50 °C. The deprotection and cleavage of the peptide was carried out using the standard protocol, namely TFA/thioanisole/1,2-ethanedithiol/anisole (%v/v = 90 : 5 : 3 : 2) for 2 h at room temperature. The resin was filtered out and the TFA solution was reduced under a stream of nitrogen. Cold diethyl ether was added to this solution to precipitate the peptide, which was then filtered and washed again with diethyl ether. The crude peptide was dissolved in water and lyophilized. The reverse-phase HPLC chromatogram of the crude peptide showed a mixture of products (Fig. S3A†).

Pseudoproline units were used for the second synthesis of hPin1_WWmutated. This strategy was feasible because this domain contains Ser and Thr residues in the middle of the sequence (Scheme 1B). Consequently, two dipeptide units, Fmoc-Ser(*t*Bu)-Ser(*psi*Me, Mepro)-OH and Fmoc-Ile-Thr(*psi*Me, Mepro)-OH were introduced during synthesis (see Scheme 1B). hPin1_WWmutated was assembled in a NovaPEG Rink amide resin (substitution 0.64 mmol g⁻¹) previously swelled for 30 min in DMF and directly loaded in the reaction vessel. All the amino acid reaction couplings were undertaken at 75 °C, except for Asp, His and Cys that were carried out at 50 °C (Table SII†). Double coupling was used for the pseudoproline units, the amino acids following these units and the amino acids Arg, Pro, Trp, Thr and Val. After the loading of the 23rd amino acid, a capping reaction step was introduced following each amino acid coupling to prevent the formation of truncated sequences and the coupling and deprotection reaction times were slightly increased. The peptide was acetylated at the N-terminal. Cleavage and removal of the protecting groups was simultaneously performed by treatment with the mixture TFA/thioanisole/1,2-ethanedithiol/anisole (%v/v = 90 : 5 : 3 : 2) for 3 hours at room temperature and under nitrogen. The solution was reduced under a nitrogen stream and cold diethyl ether was added to precipitate the crude peptide which was dissolved in water and lyophilized.

hPin1_WWmutated was purified by preparative reverse-phase HPLC (Phenomenex Jupiter Proteo column, 250 mm × 21.2 mm, 4 μ, 90 Å) using solvent A (water/TFA, 99.9 : 0.1 v/v) and solvent B (acetonitrile/water/TFA, 90 : 9.9 : 0.1 v/v). The peptide

produced using the standard method was eluted with a linear gradient of solvent B (10% to 50% in 25 min) at a flow rate of 10 mL min⁻¹ (Fig. S3 A, † $R_t = 18.8$ min). The peptide produced using pseudoproline units was eluted with a linear gradient of solvent B (20% to 44% in 20 min) at a flow rate of 10 mL min⁻¹ (Fig. S3B, † $R_t = 15.8$ min). The purity of the peptide was checked by analytical reverse-phase HPLC (Phenomenex Jupiter Proteo column, 250 mm × 4.6 mm, 4 μ, 90 Å) using a linear gradient of solvent B (10–50% in 20 min) at a flow rate of 0.6 mL min⁻¹ (Fig. S3C, † $R_t = 18.2$ min) and it was greater than 95%. The pure hPin1_WWmutated was characterized by ESI-MS (Fig. S4†).

4.3.2 Circular dichroism spectroscopy. All the CD spectra were collected under a constant flow of nitrogen on a Jasco J-715 spectropolarimeter equipped with a thermostated cell holder and a Peltier bath using 1 mm path length quartz cells. The stock peptide solutions were prepared in Milli-Q water, previously purged with nitrogen, and the concentration was determined by UV-vis spectroscopy using the extinction coefficients of tryptophan (5690 M⁻¹ cm⁻¹) and tyrosine (1280 M⁻¹ cm⁻¹) at 280 nm in 6 M GdnHCl.²² For hPin1_WWmutated the following conditions were used: all the samples contained 50 μM peptide, 10 mM MOPS buffer pH 7.2 and 100 mM NaCl. The spectra of the peptide were recorded in the wavelength range of 200–300 nm at 2 °C and 98 °C. The temperature denaturation experiments were carried between 2 and 98 °C, with an increase step of 2 °C and 1.5 min equilibration time between temperatures. In the stability studies with different buffers, the following conditions were used: all samples contained 50 μM peptide in presence of 10 mM HEPES buffer pH 7, 10 mM sodium acetate buffer pH 4 or 10 mM CAPS buffer pH 10 with 100 mM NaCl and regeneration buffer: 0.1 M NaOH in 30% isopropanol (v/v). The spectra of the peptide were recorded in the wavelength range of 216–300 nm at 4 °C and 23 °C. The following parameters were used during acquisition: 2 nm scale, continuous scanning, 200 nm min⁻¹ velocity and accumulation of 4 to 8 runs. The molar residue ellipticities (MRE, deg cm² dmol⁻¹) were calculated using the following equation:

$$\text{MRE} = \frac{\theta_{\text{obs}}}{C \times l \times \text{residue} \times 10}$$

where θ_{obs} is the observed ellipticity in millidegrees, C is the molar concentration, l is the path length of the cell in cm, residue is the number of residues in the sequence and 10 is a conversion factor to dmol. T_m values were determined using the protocol described by Kraemer-Pecore *et al.*²³

4.4 Immobilization of hPin1_WWmutated and affinity studies

4.4.1 Immobilization in solid support. For immobilization of hPin1_WWmutated in the solid support the chemistry of Sulfo-SMCC was chosen (Pierce, ThermoScientific). This is a cross-linking agent that requires an amine group in the support and a Cys in the peptide to be immobilized. The selected solid support was Sepharose CL-6B (GE Healthcare Life Sciences), which was epoxy-activated using the method described in the literature.²⁴ Subsequently, the epoxy-activated agarose beads (25 μmol of epoxy groups per g of wet gel) were aminated by using a

5× excess conditions of 1.4-diaminobutane (Sigma-Aldrich). The slurry was incubated overnight at 40 °C with orbital agitation. After this, the aminated agarose beads were thoroughly washed (about 10× gel volume) with distilled water and the amine groups content was determined using Kaiser test (14 μmol of amines per g of wet gel). The resin was stored in 20% ethanol at 4 °C.

To proceed with immobilization of hPin1_WWmutated in the aminated support, the following solutions were used: the immobilization buffer, 50 mM HEPES, 150 mM NaCl, 1 mM EDTA pH 7.2 and to dissolve sulfosuccinimidyl 4-[N-maleimidomethyl]cyclohexane-1-carboxylate (Sulfo-SMCC), a 50 mM HEPES pH 7 buffer. These buffers were de-aerated using a nitrogen stream before use. Sulfo-SMCC was used in an excess of 5× with ratio of 0.5 μmol peptide: 1 μmol of aminated support. The support was washed with 5 × 500 μL immobilization buffer. The solution of Sulfo-SMCC was prepared and added to the support and incubated with rotative agitation at 30 rpm for 30 min at room temperature (23 °C). During this time, a solution of hPin1_WWmutated (10 mg mL⁻¹) was prepared in the immobilization buffer and sonicated for 5 min to facilitate the dissolution. To avoid any oxidized Cys in the peptide, Bond-Breaker TCEP Solution (Pierce, ThermoScientific) in a ratio of 1 : 100 was added. After incubation the solution was removed by filtration and the support was washed 3× with aerated buffer to eliminate unreacted Sulfo-SMCC. After this, the solution of hPin1_WWmutated was added to the activated support and incubated for 1 h. The support was washed with immobilization buffer for 4× to remove unreacted peptide. To block the unreacted sites in the support, a solution of L-cystein (prepared in the same proportion as Sulfo-SMCC solution) was incubated with the support for 30 min. Following, the support was washed 4×. Finally, to determine the amount of hPin1_WWmutated that was immobilized in the support, the solution of hPin1_WWmutated that was loaded and all the washes after immobilization were quantified by Abs_{280 nm}. The support (hPin1_WWAg) was saved at 4 °C in immobilization buffer.

4.4.2 Affinity studies. The affinity studies using the support hPin1_WWAg were conducted with different peptides and proteins in batch mode at room temperature (23 °C). The binding buffers used were: 25 mM sodium acetate, 150 mM NaCl pH 4; 25 mM HEPES, 150 mM NaCl pH 7; or 25 mM CAPS, 150 mM NaCl pH10; for proteins only the pH 7 buffer was used as binding condition. The elution buffer used was: 50 mM Sodium Phosphate buffer, 50 mM NaCl pH7. The support hPin1_WWAg was regenerated using a regeneration buffer (0.1 M NaOH in 30% isopropanol) at 4 °C and water. The regeneration was repeated 3× alternating the regeneration buffer and water. Afterwards, each support was equilibrated 5× with appropriate binding buffer and until the absorbance of the equilibration samples reach Abs_{280 nm} < 0.005.

The peptides VPRpTPV and VPRTPV (purity >95% from GeneCust-Europe), and PPPYPYPAW (purity >97% from CASLO, ApS, Lyngby, Denmark) with no modifications in the N and C terminal (free terminals). Peptides solutions were added in 50 μg to 100 mg hPin1_WWAg and the studies were performed using solutions of 300 μL. The proteins Bovine Serum Albumin (BSA), phosphothreonine-BSA (BSA-pT), phosphoserine-BSA

(BSA-pS) and phosphotyrosine-BSA (BSA-pY) were from Sigma-Aldrich (Germany). Solutions of these proteins were prepared with a mass range of 50 μg in a 500 μL total volume. These solutions were incubated with 300 mg of hPin1_WWAg. Each peptide and protein was tested individually with the hPin1_WWAg. The flow-through, 5 \times washes and 5 \times elution samples were analyzed at Abs_{230 nm} for the VPRpTPV and VPRTPV and at Abs_{280 nm} for PPPYPAW and BSA proteins. The Binding Percentage was calculated using (mass peptide loaded – mass peptide output) \times 100/mass peptide loaded. The peptide bound mass/support mass ($\mu\text{g mg}^{-1}$) was calculated using (mass peptide loaded – mass peptide output)/mass support, elution percentages for proteins were calculated as (mass protein eluted) \times 100/mass protein loaded.

Acknowledgements

A.M.G.C.D. would like to thank Dr Iris Batalha and Dr Ricardo Branco for the helpful discussions and technical support in the immobilization and in the *in silico* studies, respectively. A.M.G.C.D. is grateful to Fundação para a Ciência e Tecnologia (FCT) – Portugal for funding through PhD grant SFRH/BD/72664/2010. This work is funded by National Funds through FCT under the project PEst-C/EQB/LA0006/2011, Pest-OE/EQB/LA0004/2011, the National Network of Mass Spectrometry REDE/1504/REM/2005, project PTDC/EBB-BIO/102163/2008, PTDC/EBB-BIO/118317/2010 and ERA-IB-2/0001/2013. Mass Spectrometry data were provided by the Mass Spectrometry Laboratory, Analytical Services Unit, Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa.

References

- 1 V. Marx, *Nat. Methods*, 2013, **10**, 829–833.
- 2 H. K. Binz, P. Amstutz and A. Plückthun, *Nat. Biotechnol.*, 2005, **23**, 1257–1268.
- 3 N. Sawyer, E. B. Speltz and L. Regan, *Biochem. Soc. Trans.*, 2013, **41**, 1131–1136.
- 4 M. J. Macias, V. Gervais, C. Civera and H. Oschkinat, *Nat. Struct. Biol.*, 2000, **7**, 375–379.
- 5 M. Sudol and T. Hunter, *Cell*, 2000, **103**, 1001–1004.
- 6 A. Zarrinpar, R. P. Bhattacharyya and W. A. Lim, *The Structure and Function of Proline Recognition Domains*, 2003.
- 7 P. J. Lu, X. Z. Zhou, M. Shen and K. P. Lu, *Science*, 1999, **283**, 1325–1328.
- 8 Y.-C. Liou, X. Zhen Zhou and K. Ping Lu, *Trends Biochem. Sci.*, 2011, **36**, 501–514.
- 9 M. A. Verdecia, M. E. Bowman, K. P. Lu, T. Hunter and J. P. Noel, *Nat. Struct. Mol. Biol.*, 2000, **7**, 639–643.
- 10 H. Yanagida, T. Matsuura, Y. Kazuta and T. Yomo, *ChemBioChem*, 2011, **12**, 962–969.
- 11 M. Jäger, H. Nguyen, J. C. Crane, J. W. Kelly and M. Gruebele, *J. Mol. Biol.*, 2001, **311**, 373–393.
- 12 E. K. Koepf, H. M. Petrassi, M. Sudol and J. W. Kelly, *Protein Sci.*, 1999, **8**, 841–853.
- 13 S. L. Pedersen, A. P. Tofteng, L. Malik and K. J. Jensen, *Chem. Soc. Rev.*, 2012, **41**, 1826–1844.
- 14 B. Hess, C. Kutzner, D. v. d. Spoel and E. Lindahl, *J. Chem. Theory Comput.*, 2008, **4**, 435–447.
- 15 X. Daura, A. E. Mark and W. F. V. Gunsteren, *J. Comput. Chem.*, 1998, **19**, 535–547.
- 16 H. J. C. Berendsen, J. P. M. Postma, W. F. Van Gunsteren, A. DiNola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684–3690.
- 17 G. Bussi, D. Donadio and M. Parrinello, *J. Chem. Phys.*, 2007, **126**, 1–7.
- 18 B. Hess, H. Bekker, H. J. C. Berendsen and J. G. E. M. Fraaije, *J. Comput. Chem.*, 1997, **18**, 1463–1472.
- 19 W. L. DeLano, *The PyMOL Molecular Graphics System*, DeLano Scientific LLC, San Carlos, CA, USA, 2002.
- 20 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graphics*, 1996, **14**, 33–38.
- 21 G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew and A. J. Olson, *J. Comput. Chem.*, 1998, **19**, 1639–1662.
- 22 *Protein Structure: a Practical approach*, ed. T. E. Creighton, IRL Press at Oxford University, Oxford, 2nd edn, 1997.
- 23 C. M. Kraemer-Pecore, J. T. J. Lecomte and J. R. Desjarlais, *Protein Sci.*, 2003, **12**, 2194–2205.
- 24 J. M. Haigh, A. Hussain, M. L. Mimmack and C. R. Lowe, *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.*, 2009, **877**, 1440–1452.
- 25 H. I. Chen, A. Einbond, S.-J. Kwak, H. Linn, E. Koepf, S. Peterson, J. W. Kelly and M. Sudol, *J. Biol. Chem.*, 1997, **272**, 17070–17077.
- 26 T. Sharpe, A. L. Jonsson, T. J. Rutherford, V. Daggett and A. R. Fersht, *Protein Sci.*, 2007, **16**, 2233–2239.
- 27 T. Peng, J. S. Zintsmaster, A. T. Namanja and J. W. Peng, *Nat. Struct. Mol. Biol.*, 2007, **14**, 325–331.
- 28 C.-e. A. Chang and Y.-m. M. Huang, in *Annual Reports in Computational Chemistry*, ed. A. W. Ralph, Elsevier, 2013, vol. 9, pp. 61–84.
- 29 B. Bacsá, K. Horváti, S. Bosze, F. Andreae and C. O. Kappe, *J. Org. Chem.*, 2008, **73**, 7532–7542.
- 30 B. de la Torre, A. Jakab and D. Andreu, *Int. J. Pept. Res. Ther.*, 2007, **13**, 265–270.
- 31 A. K. Tickler, A. B. Clippingdale and J. D. Wade, *Protein Pept. Lett.*, 2004, **11**, 377–384.
- 32 I. Coin, M. Beyermann and M. Bienert, *Nat. Protoc.*, 2007, **2**, 3247–3256.
- 33 Z. Fidan, A. Younis, P. Schmieder and R. Volkmer, *J. Pept. Sci.*, 2011, **17**, 644–649.
- 34 P. Siman, S. V. Karthikeyan and A. Brik, *Org. Lett.*, 2012, **14**, 1520–1523.
- 35 P. Marek, A. M. Woys, K. Sutton, M. T. Zanni and D. P. Raleigh, *Org. Lett.*, 2010, **12**, 4848–4851.
- 36 R. Kaul, A. R. Angeles, M. Jäger, E. T. Powers and J. W. Kelly, *J. Am. Chem. Soc.*, 2001, **123**, 5206–5212.
- 37 S. Deechongkit and J. W. Kelly, *J. Am. Chem. Soc.*, 2002, **124**, 4980–4986.
- 38 F. Delom and E. Chevet, *Proteome Sci.*, 2006, **4**, 1–12.
- 39 I. L. Batalha, C. R. Lowe and A. C. A. Roque, *Trends Biotechnol.*, 2012, **30**, 100–110.
- 40 L. Ashton, C. Johannessen and R. Goodacre, *Anal. Chem.*, 2011, **83**, 7978–7983.