

# Design and evolution of a synthetic small protein scaffold based on the WW domain

Ana Margarida Gonçalves Carvalho Dias<sup>1,2</sup>  | Gonçalo Duarte Gomes Teixeira<sup>1,2</sup>  |  
 Arménio Jorge Moura Barbosa<sup>1,2</sup>  | Joao Goncalves<sup>3</sup>  | Olga Iranzo<sup>4</sup>  |  
 Ana Cecília Afonso Roque<sup>1,2</sup> 

<sup>1</sup>Associate Laboratory i4HB—Institute for Health and Bioeconomy, Chemistry Department, NOVA School of Science and Technology, Caparica, Portugal

<sup>2</sup>UCIBIO—Applied Molecular Biosciences Unit, Department of Chemistry, NOVA School of Science and Technology, NOVA University Lisbon, Caparica, Portugal

<sup>3</sup>Imed—Research Institute for Medicines, Faculty of Pharmacy, Universidade de Lisboa, Lisbon, Portugal

<sup>4</sup>Department of Chemistry, DR2 CNRS, Institut des Sciences Moléculaires de Marseille, Biosciences, CNRS UMR 7313, Marseille, France

## Correspondence

Ana Cecília Afonso Roque, Associate Laboratory i4HB—Institute for Health and Bioeconomy, Chemistry Department, NOVA School of Science and Technology, Campus Caparica, 2829-516 Caparica, Portugal; UCIBIO—Applied Molecular Biosciences Unit, Department of Chemistry, NOVA School of Science and Technology, NOVA University Lisbon, 2829-516 Caparica, Portugal. Email: [cecilia.roque@fct.unl.pt](mailto:cecilia.roque@fct.unl.pt)

## Funding information

Horizon 2020 Framework Programme, Grant/Award Number: 899732; Fundação para a Ciência e a Tecnologia, Grant/Award Numbers: COVID/BD/152648/2022, LISBOA-01-0145-FEDER-028878, PD/BD/139800/2018, PTDC/BII-BIO/28878/2017, PTDC/CTM-CTM/3389/2021, UIDB/04138/2020, UIDB/04378/2020, UIDP/04138/2020, UIDP/04378/2020

**Review Editor:** Aitziber L. Cortajarena

## Abstract

Protein engineering is a powerful tool in biotechnology and the basis to create unprecedented sequences, structures, and functions. The WW domains are a family of naturally occurring proteins involved in the molecular recognition of proline-rich and phosphorylated peptide sequences with relevance in cellular processes involved in human diseases. Due to their small size, WW domains represent appealing small protein domains for protein engineering and to generate novel functions as binders to non-cognate targets. Here, we designed a synthetic protein scaffold library based on the WW prototype sequence in which the loops were extended and randomized while maintaining structural stability. Using in vitro evolution by phage display against human serum albumin (HSA), we found a lead candidate that was produced by biological and chemical means and further characterized using experimental and computational tools. As a potential application for the lead binder, it was immobilized on a matrix and used to capture the target HSA. Overall, this work shows the versatility of WW domains as peptide scaffolds amenable for in vitro evolution against non-cognate targets.

## KEYWORDS

affinity purification, human serum albumin, phage display, protein engineering, WW domains

## 1 | INTRODUCTION

Protein–protein interactions, namely affinity-based molecular recognition events, are key for biological processes in Nature. Rational and de novo design, as well as in vitro evolution, are tools to engineer existent or new protein ligands to improve binding efficiency toward cognate targets or to recognize non-native

targets (Gray et al., 2020; Linsky et al., 2020; Wang et al., 1979). Typically, small protein domains that are efficiently expressed in bacterial hosts (e.g., *Escherichia coli*) or easily produced by chemical synthesis, that possess a robust folding—ideally independent from disulfide bonds—while still allowing diversity at the sequence level in determined regions, and that present high physical and chemical stability, are

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2025 The Author(s). *Protein Science* published by Wiley Periodicals LLC on behalf of The Protein Society.

ideal for protein engineering and ligand development. A surplus feature is the possibility to evolve these protein sequences by different *in vitro* strategies, namely phage, ribosome, or CIS display (Dias, 2020; Gebauer & Skerra, 2019). Apart from protein sequences derived from antibodies and antibody fragments, there are also several examples of successfully engineered protein scaffolds which are non-antibody derived and that were further commercially exploited, namely centyrin (Aro Biotherapeutics), DARPin (Molecular Partners AG), Nanofitins (Affilogic) and Affibodies (Affibody AB). This demonstrates the importance of protein scaffolds engineering in different biotechnology fields (Dias & Roque, 2017). Recently, there has been also an increasing interest in peptide scaffolds—sequences with molecular weight between 500 and 5000 Da typically containing 40 or fewer amino acids—as biomedical ligands due to their increased cell penetration and selectivity *in vivo* (Cooper et al., 2021).

WW domains are an interesting family of small protein domains in the frontier between proteins and peptides (Figure 1). These possess 38–40 residues in length and a molecular weight below 5 KDa, with two tryptophan residues spaced by 20–22 amino acids, which assemble on a three  $\beta$ -sheet structure connected by two loops (Sudol & Hunter, 2000). These motifs are found in at least 200 multidomain proteins, are localized in recognition regions that mediate protein–protein interactions (Macias et al., 2002), and are known to bind proline-rich and phosphorylated sequences in peptides or proteins. The protein–protein interactions mediated by WW domains are associated with signaling pathways that are important in cell cycle regulation of several diseases, such as cancer (Martínez-Lumbreras et al., 2024; Salah et al., 2012).

Due to their small size, native and mutated WW domains have been used as prototype sequences for theoretical and experimental biophysical studies (Jäger et al., 2006; Piana et al., 2011), and they were produced by recombinant expression or chemically synthesized in high yields (Dias et al., 2015, 2016; Patel et al., 2013). Two of the most studied WW domains are the human YAP65 (hYAP65\_WW) and the human Pin1 (hPin1\_WW). Both sequences have been used for rational design to either improve affinity against their natural targets (Dalby et al., 2000; Espanel et al., 2003; Yanagida et al., 2008) or to generate new sequences that can be further used for *in vitro* evolution or rational design (Patel et al., 2013). Several modifications have been tested in loop I, such as the incorporation of non-natural and glycosylated amino acids (Kaul et al., 2001), or the variation of the loop size with 4–6 amino acids (Price et al., 2011). Loop II was also mutated, but with little impact in WW domain conformation stability (Jäger et al., 2006). Also, the incorporation of one pair of cysteine residues close to the terminal regions for disulfide bond formation increased structure

stability and proteolysis resistance (Patel et al., 2013). In addition, *in vitro* evolution by phage (Dalby et al., 2000; Espanel et al., 2003), ribosome (Yanagida et al., 2008) and CIS display (Patel et al., 2013) generated binders toward native targets (GTPPPPYTVG or PPXY, where X is any amino acid and PPLP (Dalby et al., 2000; Espanel et al., 2003; Yanagida et al., 2008)) or non-cognate targets (VEGFR-2 (Patel et al., 2013)). Overall, these studies demonstrate the high robustness and versatility of the WW domain scaffold for protein engineering, as both framework and loop regions accept mutations in up to 25% of the total sequence, allowing randomization to any amino acid (Patel et al., 2013).

An alignment of WW domain sequences performed by Macias et al. (2000) revealed that loop sequences are highly diverse, but a consensus sequence in the  $\beta$ -sheets regions could be generated. Thus, the authors suggested a WW prototype (WWp) sequence (PDB code: 1E0M, 4.4 KDa) (Figure 1a). The WWp was biologically produced (fused with glutathione-S-transferase (GST)) and characterized by nuclear magnetic resonance (NMR) spectroscopy and circular dichroism (CD), showing folding and thermal stability ( $T_m = 44.2^\circ\text{C}$ ) comparable to other WW domains (Macias et al., 2000). It retained the framework residues responsible for folding, but the loops were reduced to two amino acids, thus eliminating binding specificity.

Loops engineering in the complementarity-determining regions (CDRs), including chemical diversity and loops extension from up to 18 residues, is critical in antibody development. For example, the nanobodies—antibody fragments derived from libraries based on the variable domain of the heavy chain (VHH, Ablynx, Figure 1)—are a successful example of library design based on the extension of the CDR3 loops (Muyldermans, 2021). Furthermore, longer loops grant various protein scaffolds greater shape diversity, thereby targeting binding sites with specific geometry or difficult targets such as membrane proteins (Jiang et al., 2024; Zimmermann et al., 2020). Large loops require specific structural features and optimization of the loop backbone to promote a well-balanced number of hydrophilic and hydrophobic interactions. Previous WW domain libraries addressed partial randomization of the loops and modification of the  $\beta$ -sheet framework to develop affinity reagents by *in vitro* evolution against cognate (Yanagida et al., 2008) and non-cognate targets (Patel et al., 2013). Therefore, we took the challenge of exploring loop diversity in WW domains by total randomization and extension of the loops.

Here, we report the design of a new WW scaffold sequence based on the WWp in which the size of the loops is increased without affecting stability (named as WWp5\_4) (Figure 1). This new scaffold has 42 amino acids and up to 5 KDa and it is one of the smallest scaffolds described in the literature or commercially

available. We generated a naïve library for phage display by randomizing the designed loops and further selected binders against human serum albumin (HSA).

HSA (66.5 kDa) was used as a model protein due to its abundance in the plasma and relevance in the biomedical field (Sugio et al., 1999a). The lead ligand was

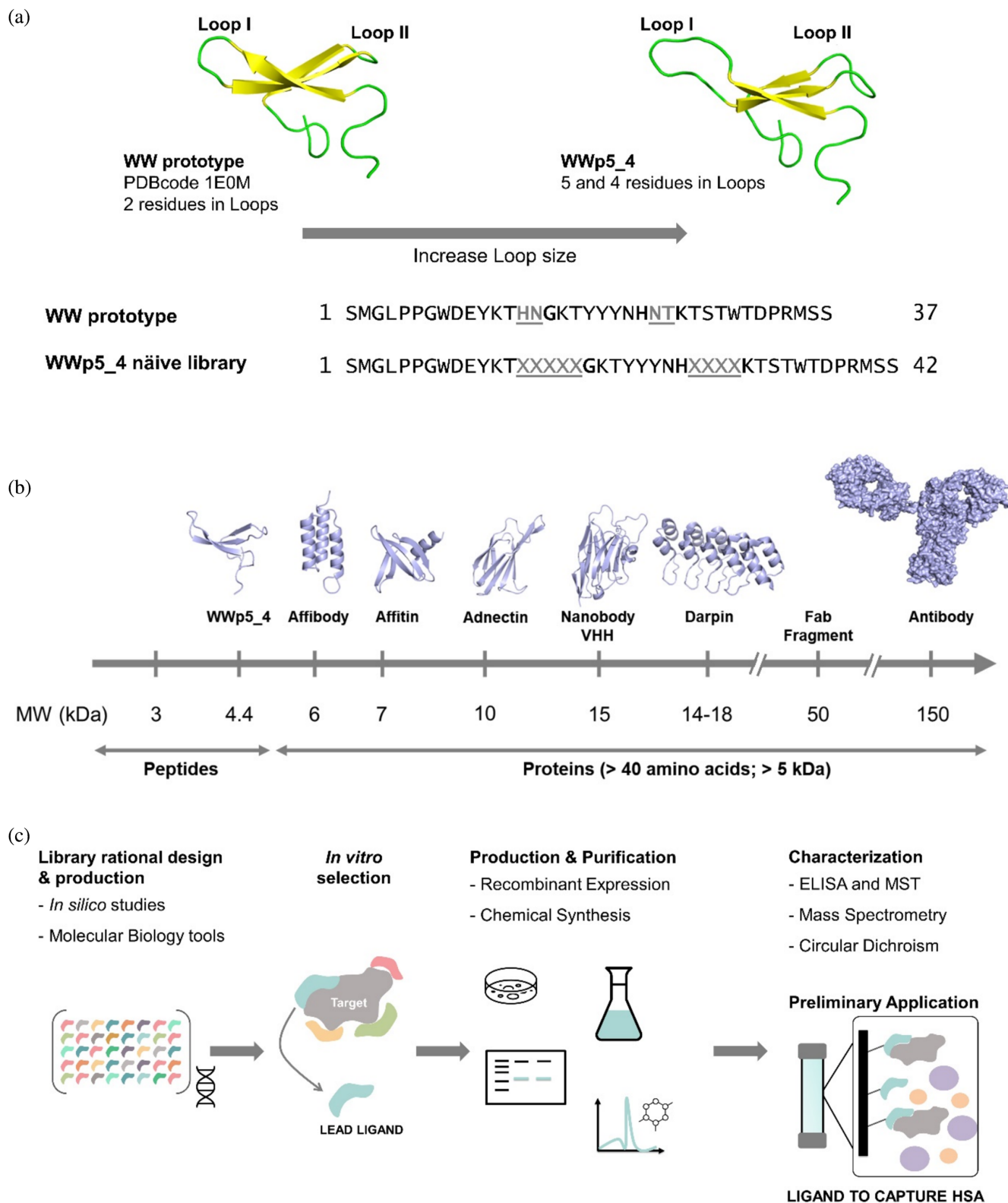


FIGURE 1 Legend on next page.

biologically and chemically produced and then characterized. Finally, as a preliminary application, we employed the best ligand for the selective capture of HSA. This work presents a newly designed synthetic peptide scaffold that can be used to generate ligands against non-cognate targets.

## 2 | RESULTS AND DISCUSSION

### 2.1 | WWp5\_4 synthetic peptide scaffold design and predicted stability

The peptide scaffold developed in this work derives from the sequence reported in the literature as the WWp (PDB code: 1E0M) (Macias et al., 2000) (Figure 1a). This sequence was obtained after analyzing the most frequent amino acids in 40 positions of several WW domains known in the literature. The WWp sequence has 37 amino acids in length, and the loops are composed of non-conserved residues (loop I—HN and loop II—NT) (Jiang et al., 2024). To design the WWp-based scaffold library, we envisioned that the loop regions would be fully randomized, whereas the  $\beta$ -sheet regions would remain constant.

We divided the design strategy into two steps: (i) selecting loop length; and (ii) assessing the predicted stability of mutants with different loop sequences, used as surrogate models for ligands obtained after *in vitro* evolution.

First, we computationally designed the optimum loop length that could yield maximum diversity without compromising structure stability. In the literature, WW domains present different loop sizes providing shape diversity to address the cognate targets. For loop I, the maximum size is found in hPin1\_WW with 6 residues in loop I, but 5 residues are the most frequent length. Loop I is known to take part in recognition of Pro-rich peptide sequences (Jäger et al., 2006). Regarding loop II, it typically has between 3 and 4 residues. We generated a set of 8 mutants with varying number of Ala residues in the loops (between 5–13 and 4–12 for loops I and II, respectively) (Table 1, Figures 2 and S2). The E0MMut\_NC structure was generated by homology modeling (Swiss-model) and by AlphaFold3. The alignment between the two structures overlap, with small backbone RMSD of 1.3 Å (Figure S2), thus the

homology modeling design was used to further generate all the structures used in the *in silico* studies. Molecular dynamic (MD) simulations for 50 ns showed that the most suitable length for loops I and II would be 5 and 4, respectively (E0MMut\_NC), as the root mean square deviation (RMSD) and root mean square fluctuation (RMSF) is similar to the E0M\_control (Figures S3–S5), also represented in the b-factor analysis (Figure S6). The dictionary of secondary structure in proteins (DSSP) analysis indicates a similar folding between the E0M\_control and the E0MMut\_NC variant (Figures 2, and S7). Thus, the WW peptide scaffold to carry on the work was selected with 5 and 4 amino acid residues in loops I and II, and named as WWp5\_4 scaffold.

Second, we assessed the resilience of the WWp5\_4 scaffold to accept distinct amino acid combinations in the loops without compromising structural stability. Therefore, we generated *in silico* a set of mutant sequences: (i) two control sequences, WWp and WWp5\_4 with loops substituted with Ala residues; (ii) six mutants from WWp5\_4 with loops randomly substituted by combinations of four amino acids (YADS—Tyr, Ala, Asp and Ser); (iii) four mutants from WWp5\_4 with loops substituted by a random selection of amino acids with distinct properties (Table 2). The rationale for including combinations of the four amino acids Tyr, Ala, Asp, and Ser arises from their high frequency in antigen–antibody recognition sites (Aires et al., 2008; Gunasekaran et al., 1997). After running MD simulations for 50 ns with the WW variant models, we observed that despite the chemical diversity in the loops, similar RMSD values were obtained (variation up to 0.2, Figures S8, S9A, S9B, S10, S11A, S11B) when compared with the E0M\_control, corresponding to a very stable state. Interestingly, the incorporation of Ala residues in the WWp loops did not disturb the stability of the E0M\_control structure but increased the flexibility of the residues in the region of loop I. The RMSF values of the residues in loop I are variable (variation <0.3 nm); however, the RMSF values in loop II are very similar between the mutants and the E0M\_control (variation <0.2 nm) (Figures S9C, S9D; S11C, S11D) as is possible to visualize in the b-factor analysis (Figure S12). Finally, the DSSP analysis shows the maintenance of the characteristic folding into  $\beta$ -sheets for the different mutants during the time of the

**FIGURE 1** Structure of the designed WW scaffold and overall research strategy. (a) Structures used for design of the phage display library. The structure used as starting point for the design of the synthetic and naïve library was the WW prototype sequence (PDB: 1E0M) developed by Macias et al. (2000). This structure has 2 residues in the loops. Through rational design it was generated a WWp5\_4, which has 5 and 4 residues respectively in loop I and loop II and through *in silico* studies we evaluated the effect of different loop compositions. The amino acid sequence of the WW prototype is represented with the residues in the limits of the loops shown in bold. The randomized positions in loop I and loop II of the WWp5\_4 naïve library are represented as an underlined X in gray. (b) Comparison between protein scaffolds sizes commonly used in biotechnology applications. Antibody (PDB code: 1HZH); Darpin (PDB code: 2QYJ); Nanobody (PDB code: 1MEL); Adnectin (PDB code: 7L0G); Affitin (PDB code: 4CJ2 chain C); Affibody (PDB code: 1Q2N); and WWp5\_4 designed in this work. (c) The research strategy adopted in this work. Protein models were produced in Pymol visualization software.

**TABLE 1** Mutant structures generated from the WW prototype sequence used in molecular dynamics studies with loops of varied lengths composed of alanine residues.

| Structure name                  | WW prototype<br>(WWp, E0M_control) | E0MMut_NC | E0ML |     |     |     |      |       |       |
|---------------------------------|------------------------------------|-----------|------|-----|-----|-----|------|-------|-------|
| Size LoopI_LoopII               | 2_2                                | 5_4       | 6_5  | 7_6 | 8_7 | 9_8 | 10_9 | 10_10 | 13_12 |
| Number of Ala residues in loops | -                                  | 9         | 11   | 13  | 15  | 17  | 19   | 20    | 25    |
| Protein Length                  | 37                                 | 42        | 44   | 46  | 48  | 50  | 52   | 53    | 58    |

simulations (Figures 2, S13, S14). Overall, the in silico results showed that the WWp5\_4 scaffold has a high structural resilience toward chemical diversity in loops I and II.

## 2.2 | Phage display library construction and selection

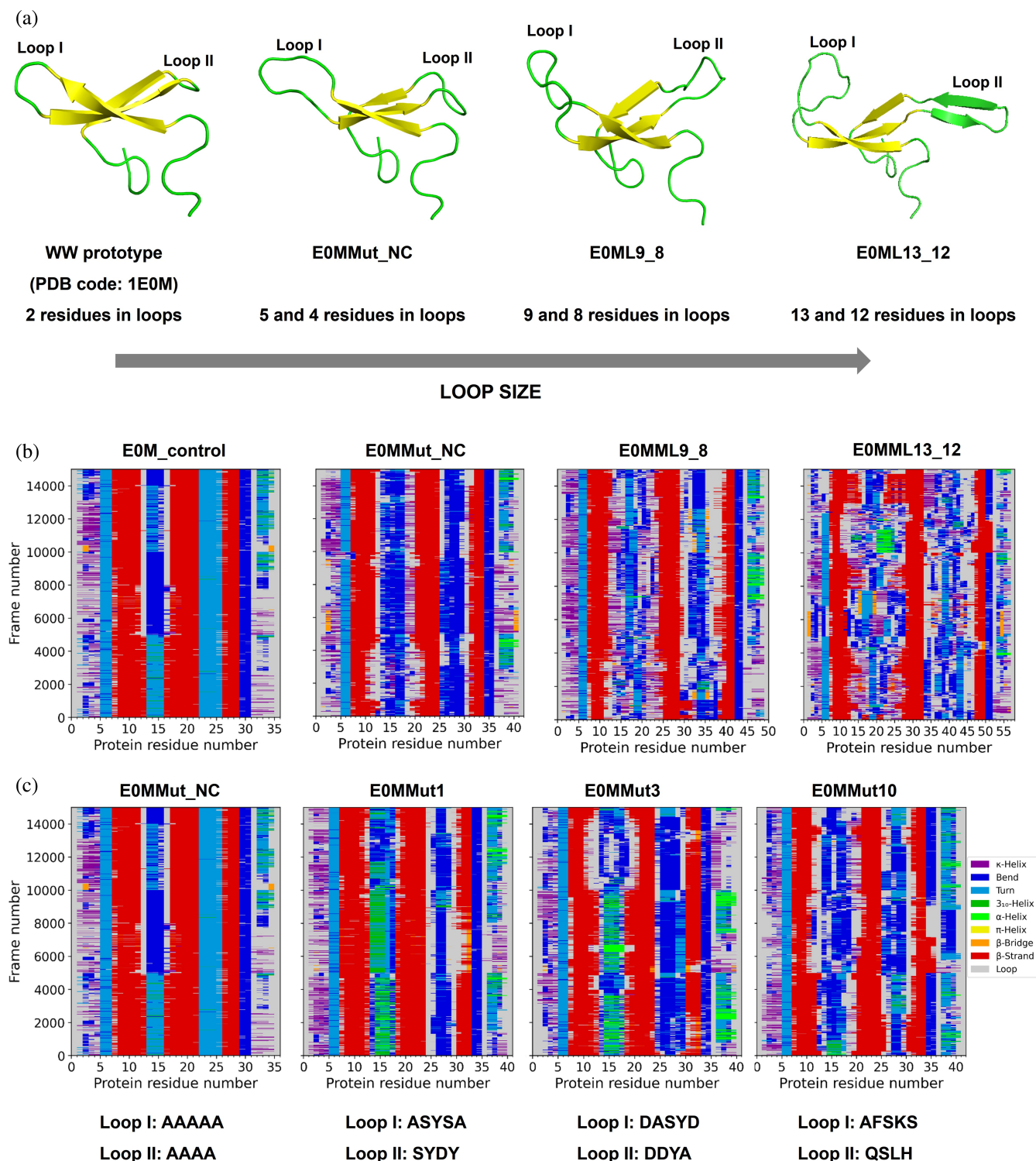
In the design of the WWp5\_4 library, we used the WWp sequence with 5 and 4 residues in loops I and II, respectively, in a total of 9 randomized positions. The loop regions in the gene sequence were randomized using degenerated codons NNN (N = A, T, G or C) (Figure S15). The library was assembled and cloned successfully into pComb3× phagemid vector, which includes a capsid pIII protein. To show the potential of the new WWp5\_4 library to find binders against non-cognate targets, we selected HSA as a first target example. HSA plays important roles for biomedical and biotechnological applications due to its abundance in the plasma (80%) (Sugio et al., 1999a), its natural carrier function, and its long half-life in plasma (19 days in humans) (Elsadek & Kratz, 2012). Thus, HSA is used as a therapeutic or therapeutic adjuvant. Ligands targeting HSA can be useful to capture or detect HSA from complex mixtures (Millioni et al., 2011) or to develop strategies to increase the half-life of small therapeutics (Zorzi et al., 2019).

Three rounds of selection were used in the panning, and in each step the stringency conditions were increased by: (i) reducing the amount of immobilized target (1, 0.5–0.25 µg/ well HSA) and (ii) increasing the number of washes (5, 10–15 times). After three rounds, we proceed for the identification of binders from the selection against HSA. The phages from the third panning elution were used for infection of *E. coli* TOP10F', as recommended in Barbas III et al., (2001). In total, 186 clones were expressed in pComb3× phagemid vector, and the soluble fraction (SF) was isolated. The SF was used in an ELISA assay to verify binding against the target HSA and a negative control, as well as to assess expression levels. The 186 clones presented different binding and expression profiles (Figure S16) and we selected 12 clones with high binding signal (comparing to the negative control) and high expression signal to carry on.

The 12 clones were renamed 1–12, and were recombinantly expressed at three different induction temperatures—25, 30, and 37°C—using 1 mM IPTG. After cell lysis, the SF of the crude extract was used to repeat the ELISA assay (Figure S17). Figure 3a shows the ELISA screening of the SF of the extracts after production at 25°C. The 12 clones show distinct expression levels and some unspecific binding against the negative control (blocking solution, 5% v/v Soya Milk in PBS1×). Nevertheless, clones 3 and 9 present the highest expression level. The binding signal is the highest for clones 3, 5, and 9, well above the signal of clone NC (SF of *E. coli* Top10F' extract and negative control for the expression of WW domains).

The 12 clones plus other 8 randomly selected (renamed from 13 to 20) were Sanger sequenced to evaluate the diversity of the library after the third round of panning. Figure 3b shows as an example the alignment of 10 sequences identified during Sanger sequencing. The alignment of the sequences demonstrated high chemical diversity in loops I and II without repeated sequences (Figure 3b, c; Table S2). Clone 5 appeared as a double WW domain (tandem sequence), probably due to a recombination process during library assembly, presenting in total four loops with different compositions. Although, clone 5 presented the highest binding signal toward HSA, it was poorly expressed, most probably due to its high hydrophobic character. We analyzed the loop sequences in terms of amino acid relative abundance (Figure 3c). In loop I the most predominant amino acids are hydrophobic aliphatic, particularly Leu, that appears in all positions in loop I. Loop II is predominantly hydrophilic and can be polar and uncharged (Ser, Thr, and Asn) or basic (Lys and Arg). WW domains present high variability in the loops as demonstrated by Macias et al. (2000). However, the most common amino acids tend to be hydrophilic and polar, such as Ser, Thr, Asn and hydrophilic with an acidic character (Glu or Asp), followed by hydrophobic and aliphatic as Pro. In our selection the recovered loop sequences are highly diverse and have a strong hydrophilic and polar character.

We decided to select clone 3 (CW3, full sequence in Figure 4a and Table S2) to carry on the work, since it presented one of the highest binding signals and good expression at different temperatures and in several replicates. Nevertheless, clone 9 could be an interesting



**FIGURE 2** Design of a WW prototype-based peptide scaffold (WWp5\_4). (a) Assessing the size of loop I and II through the substitution with alanine residues. (b) DSSP analysis during the simulations for the sequences with different loop size, showing the variability in secondary structure during simulation as the loops are extended. The most stable secondary structure is the E0MMut\_NC with 5 residues in loop I and 4 residues in loop II substituted by Ala residues. (c) DSSP analysis during the all-atom simulations for mutants with 5 residues in loop I and 4 residues in loop II, modified with different combination of amino acids. E0MMut\_NC is the structure with alanine in the loops and E0MMut1 and E0MMut3 are structures with “YADS” combinations in the loops, while the E0MMut10 is an example of a mutant with totally randomized loops. The mutants have similar secondary structures composed with  $\beta$ -sheets (regions identified in red), which are stable during the simulation. Indicating that the characteristic folding of the WW domain is maintained despite the mutated loops with different composition of amino acids.

**TABLE 2** Mutant sequences generated from the WWp5\_4 scaffold sequence and used in molecular dynamic studies to assess resilience to chemical diversity in the loops.

| Structure name                     | Sequence                                    | Loops  |
|------------------------------------|---|--|
| WWprototype,<br>(WWp, E0M_control) | SMGLPPGWDEYKTH---NGKTYYYNHN--TKTSTWTDPRMSS  | Designed   |
| WW naïve library (WWp5_4)          | SMGLPPGWDEYKTXXXXXGKTYYYNHXXXXKTSTWTDPRMSS  | n.a.   |
| E0MMut_NC                          | SMGLPPGWDEYKTA AAAAGKTYYYNHAAAAKTSTWTDPRMSS | Alanine  |
| E0MMut1                            | SMGLPPGWDEYKTASYSAGKTYYYNHSDYKTSTWTDPRMSS   | Combination of four amino acids (YADS—Tyr, Ala, Asp and Ser) |
| E0MMut2                            | SMGLPPGWDEYKTAYSYSAGKTYYYNHAYASKTSTWTDPRMSS |  |
| E0MMut3                            | SMGLPPGWDEYKTASYDGKTYYYNHDDYAKTSTWTDPRMSS   |  |
| E0MMut4                            | SMGLPPGWDEYKTYAYSYGKTYYYNHSSAKTSTWTDPRMSS   |  |
| E0MMut5                            | SMGLPPGWDEYKTYAYS DGKTYYYNHDSYSKTSTWTDPRMSS |  |
| E0MMut6                            | SMGLPPGWDEYKTYDYDAGKTYYYNH DYAKTSTWTDPRMSS  |  |
| E0MMut7                            | SMGLPPGWDEYKTLKSN GKTYYYNHRWAKTSTWTDPRMSS   | Combination of a selection of random amino acids             |
| E0MMut8                            | SMGLPPGWDEYKTERND SGKTYYYNHKLSKTSTWTDPRMSS  |  |
| E0MMut9                            | SMGLPPGWDEYKTDEVKHGKTYYYNHMSTRKTSTWTDPRMSS  |  |
| E0MMut10                           | SMGLPPGWDEYKTAFSKSGKTYYYNHQLHKTSTWTDPRMSS   |  |

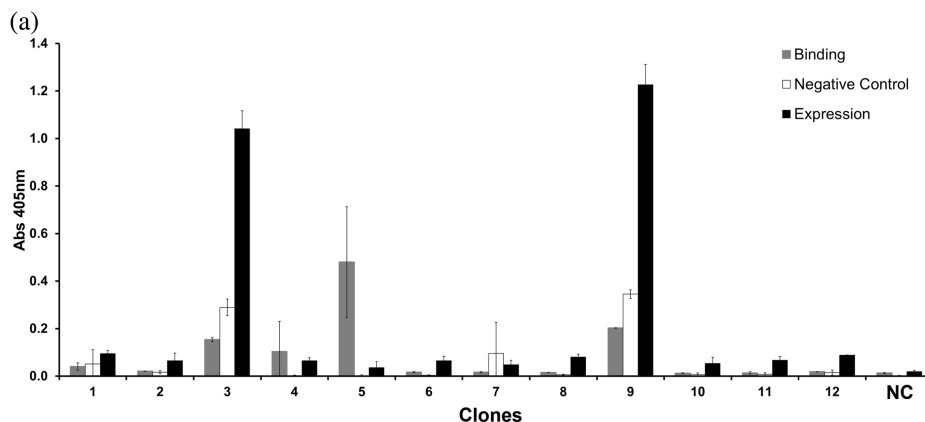
ligand to pursue as a second lead candidate, and clone 5 could be used as an inspiration for future designs.

## 2.3 | Biological production of CW3S protein and characterization of HSA-binding

The production of WW domains has been previously demonstrated by recombinant expression in *E. coli* as fusion protein with glutathione S-transferase (GST) tag or green fluorescent protein (GFP) (Macias et al., 2000; Villegas-Méndez et al., 2012). In this work, we produced CW3 by recombinant expression alone and as a GFP fusion. Sanger sequencing results revealed that CW3 has a Cys residue in the second loop (position 29), which can cause dimerization. We mutated this amino acid for a Ser residue, naming the new sequence as CW3S. This is a residue that appears with high frequency in selections and several reports in literature change Cys for Ser residues since their side chain is structurally similar (Mouratou et al., 2007). Thus, we attempted the expression of the CW3S alone in small scale using *E. coli* Rosetta (DE3) cells. We also produced the WWp described by Macias et al. (2000) under the same conditions to be used as a control in ELISA. After cell lysis the SF of the crude extracts was characterized by SDS-PAGE showing the expression of CW3S and WWp with the expected molecular weight (~6 kDa) in the SF (Figure S18). We then analyzed the SF of the crude extract by ELISA (Figure S19). The ELISA was performed using an anti-His tag antibody conjugate HRP. Results show Abs<sub>405nm</sub> in the same range as described for the CW3 clone when expressed in *E. coli* TOF10F', and this signal is 3.5

times higher than WWp and the clone NC (SF of *E. coli* Rosetta extract). Hence, CW3S is binding HSA and the framework of the WWp5\_4 library that corresponds to the sequence of the WWp sequence does not bind to the target (Figure S19). These results strengthen the interest in the CW3S clone, thus we continued with the production in a different system to improve the solubility and protein yield, as well as to study the versatility of the novel affinity ligand.

For that, we recombinantly expressed the fusion protein CW3S with GFP (CW3S\_GFP). The crude extracts were analyzed by SDS-PAGE gel and we observed a band correspondent to the expected size (~33 kDa) and further confirmed by Western-blot, where a signal for the His-tag in the construct was identified (Figure S20). The amount of total protein and the fusion protein in the SF of the crude extracts were quantified by BCA assay and gel densitometry, respectively. In the best expression condition (temperature 30 °C) we estimated 164 mg of CW3S\_GFP per Liter of culture. This yield is lower comparing with the same expression condition for the fusion hPin1\_GFP (1840 mg/L), which may be due to the lower water solubility of CW3S, as predicted by [PepCalc.com](#). The protein produced at 30 °C in the SF was purified by affinity chromatography using a resin developed in house to capture GFP (Pina et al., 2015). After purification, all fractions were analyzed in terms of Total Protein and GFP concentration (Figure S20). The elutions were analyzed by SDS-PAGE gel and the fractions with higher fluorescence intensity (higher concentration of GFP, E9 to E12) were pooled, dialyzed, and concentrated. At the end of the purification process, we obtained CW3S\_GFP with a concentration of 97 µg/mL (recovery 70% ± 7%, >70% purity). This protein was used for further binding studies by ELISA assays

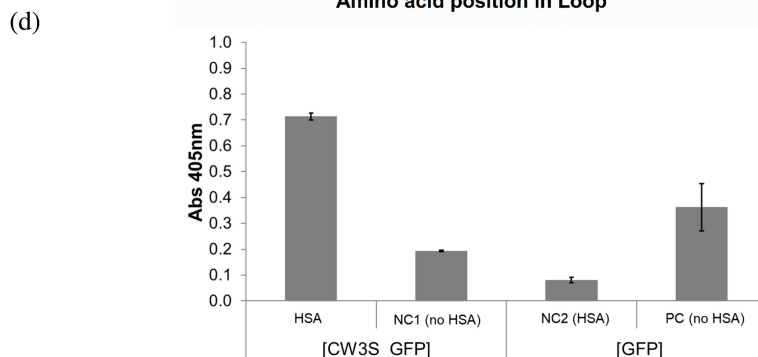
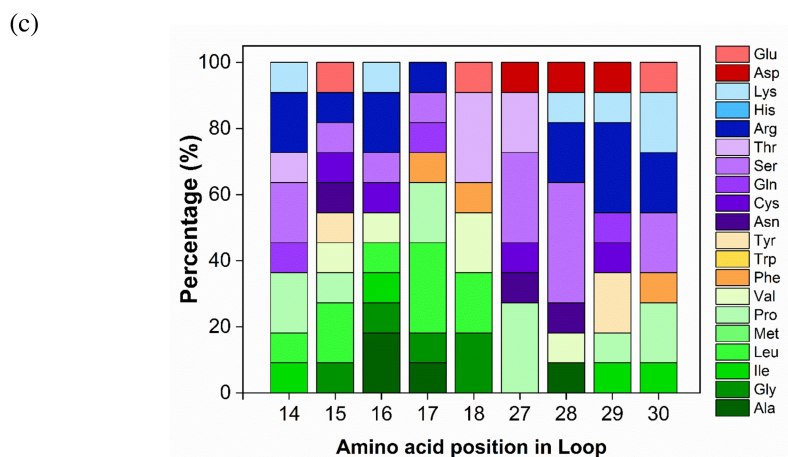


**FIGURE 3** Selection of HSA binders. (a) ELISA screening of top 12 clones selected after in vitro selection. These 12 clones were expressed in *E. coli* TOP10F<sup>+</sup> at 25°C and 1 mM IPTG. The SF of the crude extracts was used for ELISA against HSA: (i) Binding, well with immobilized target; (ii) negative control, well without target with only blocking solution (5% (v/v) soya milk in PBS 1×); (iii) expression, well without target and the crude extract was directly immobilized in the microplate. As negative control of the expression (clone NC) was used an extract from *E. coli* TOP10F<sup>+</sup>. The plotted data corresponds to the average of duplicate assays and error bars represent the standard deviation. (b) DNA Sequencing results with protein sequences selected by phage display for HSA compared with WWp sequence. (c) Amino acid relative abundance in the loops of the different clones. Loop I—amino acid position 14–18 and Loop II—amino acid position 27–30. (d) Binding between CW3S\_GFP and HSA assessed by ELISA, using 1 µg/well of each protein, a well without HSA–NC1 (no HSA) was used as negative control. GFP (1 µg/well) was used as a positive control since a primary antibody anti-GFP was used for detection. Thus, two control were used—(i) NC2 (HSA), HSA was immobilized in the well; (ii) PC (no HSA), well without target, the GFP solution was immobilized in the well and the well was blocked (5% (v/v) Soya Milk in PBS1×).

|             |    |               |       |          |      |               |    |
|-------------|----|---------------|-------|----------|------|---------------|----|
| wwprototype | 1  | SMGLPPGWDEYKT | H---N | GKTYYYNH | N--T | KTSTWTDPRMSS  | 37 |
| Clone1      | 1  | SMGLPPGWDEYKT | IRASL | GKTYYYNH | TSYR | KTSTWTDPRMSS  | 42 |
| Clone3      | 1  | SMGLPPGWDEYKS | SNLFL | GKTYYYNH | DNCP | KTSTWTDPRMSS  | 42 |
| Clone4      | 1  | LMGLSPGWDEYKT | SYKLG | GKTYYYNH | TVRS | KTSTWTDPRMSS  | 42 |
| Clone5      | 1  | -MGLPPGWQYQT  | QLRGT | GKTYYYHH | PSRK | KTSTWTDPRMSS  | 41 |
|             | 42 | -GLAPPGWQYQT  | PSILT | GKTYYYNH | SKRS | KTSTWTDPRMSSG | 83 |
| Clone6      | 1  | SMGLPPGWDEYKT | PLAAE | GKTYYYNH | PSQF | KTSTWTDPRMSS  | 42 |
| Clone9      | 1  | SMGLPPGWDEYKT | TCGPV | GKTYYYNH | PDKI | KTSTWTDPRMSS  | 42 |
| Clone11     | 1  | SMGLPPGWDEYKT | RGSQT | GKTYYYNH | NRYE | KTSTWTDPRMSS  | 42 |
| Clone13     | 1  | SMGLPPGWDEYKT | LPCPV | GKTYYYNH | CRDR | KTSTWTDPRMSS  | 42 |
| Clone14     | 1  | SMGLPPGWDEYKT | KVRLG | GKTYHYNH | SSPP | KTSTWTDPRMSS  | 42 |
| Clone15     | 1  | SMGLPPGWDEYKT | REVRF | GKTYYYNH | SAIK | KTSTWTDPRMSS  | 42 |

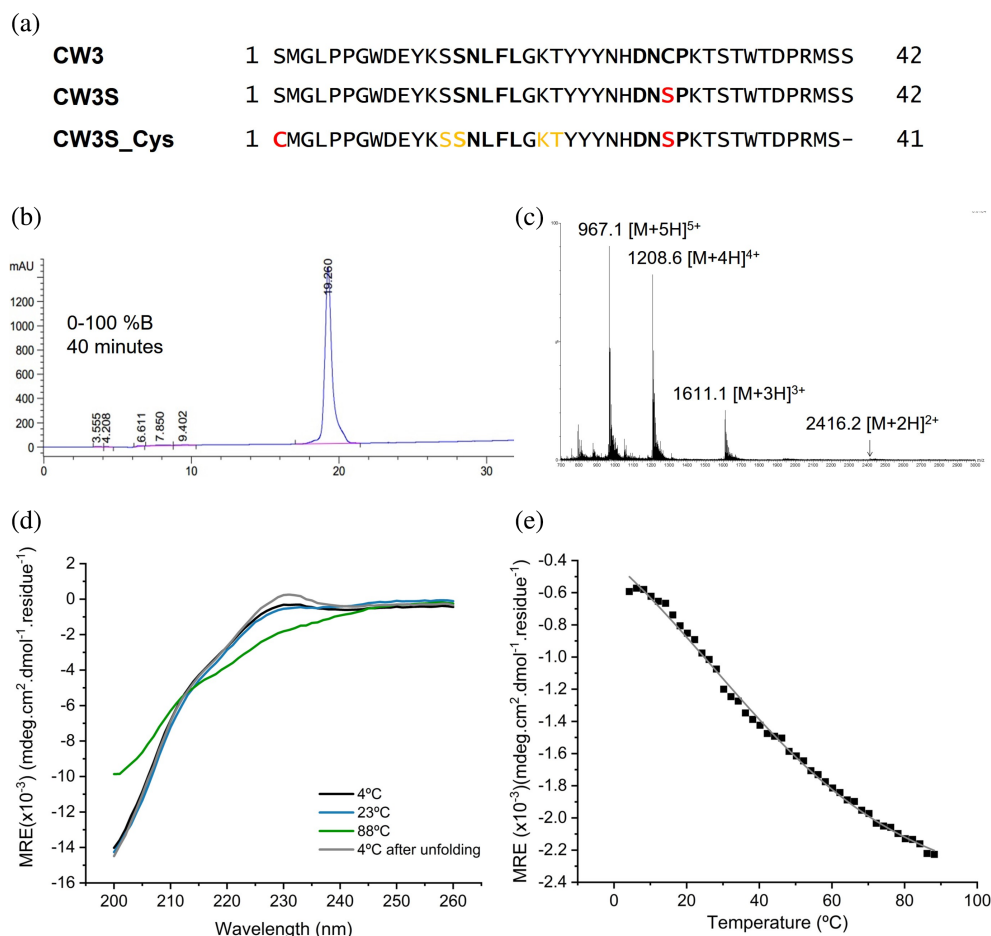
LoopI

LoopII



showing the high selectivity for binding HSA (Figures 3d, S1). It was also observed that the binding signal improved when the WW domain is in fusion with

GFP. We postulate that this effect is due to the stability of the WW domain structure when it is in fusion with another molecule. In nature WW domains are often in



**FIGURE 4** Purification and characterization of the CW3S sequence produced by solid-phase peptide synthesis. (a) CW3 has a Cys in loop I and in CW3S the Cys was mutated to a Ser (in red). The sequence used for chemical synthesis (CW3S\_Cys) was the same as CW3S with incorporation of a Cys in the N-terminal for further immobilization in a solid-support. For the chemical synthesis two pseudoproline dipeptide units were incorporated in the amino acid positions highlighted in yellow. (b) CW3S\_Cys was purified using reverse-phase HPLC, the purified peptide was eluted with a linear gradient of solvent B (0%–100% in 40 min) at a flow rate of 0.6 mL/min ( $R_t = 19.26$  min). (c) Mass spectrum of the collected peak obtained by ESI-MS.  $[M + 2H]^{2+}$  2416.2 (calculated)/2416.2 (determined),  $[M + 3H]^{3+}$  1611.1 (calculated)/1611.1 (determined),  $[M + 4H]^{4+}$  1208.6 (calculated)/1208.6 (determined),  $[M + 5H]^{5+}$  967.1 (calculated)/967.1 (determined). (d) Far-UV Circular Dichroism spectra of CW3S\_Cys solutions at 4, 23, and 88°C, and (e) fitting to the two state-model curve for determination of melting temperature ( $T_m$ ) between 4 and 88°C, monitoring the change in signal at 230 nm. The plotted data corresponds to the average of triplicate assays. The data were fitted with Origin software.

fusion with large proteins and enzymes or in tandem with other WW domains (Martínez-Lumbreras et al., 2024).

## 2.4 | Chemical synthesis of the lead peptide and its characterization

Given the small size of WW domains, they can also be produced by solid-phase peptide synthesis (Dias et al., 2015, 2016; Patel et al., 2013). Thus, we attempted CW3S production by solid-phase peptide synthesis. For the chemical synthesis of CW3S, we used the sequence in Figure 4a named as CW3S\_Cys. In this sequence, the Ser at position 42 was removed since two serine amino acids in the C-terminal can be difficult to

start the synthesis, and the N-terminal Ser was mutated for a Cys. Previous works with WW domains strongly suggest that the Ser deletion will not interfere with the stability or function of the peptide produced since it is in the C-terminal of the peptide (Macias et al., 2000) and the introduction of cysteine at the N-terminal allows further immobilization of the peptide in a solid support, as we already demonstrated for other WW domains (e.g., hPin1\_WW (Dias et al., 2015) or hYAP65\_WW (Dias et al., 2016)). For CW3S\_Cys chemical production, we used the strategy with pseudoproline units already implemented before with success (Dias et al., 2015). Briefly, this synthesis strategy is interesting for peptides and small proteins that possess hydrophobic sequences and tend to aggregate during chemical synthesis. The pseudoproline units resemble a proline

amino acid structure known to promote specific conformational changes in the peptide backbone (Morgan & Rubenstein, 2013) by extending the peptide sequence being produced and increasing the synthesis yield. Therefore, two pseudoproline units—Fmoc-Lys(Boc)-Thr( $\psi^{\text{Me,Me}}$ pro)-OH and Fmoc-Ser(tBu)-Ser( $\psi^{\text{Me,Me}}$ pro)-OH—were incorporated at specific locations identified as KT and SS (in yellow, Figure 4a). After synthesis, the crude peptide was recovered, purified by preparative reversed-phase HPLC, and characterized by mass spectrometry (ESI-MS). The product had a final purity of 99% as determined by analytical HPLC (Figure 4b), and ESI-MS confirmed the identity of the peptide (Figure 4c).

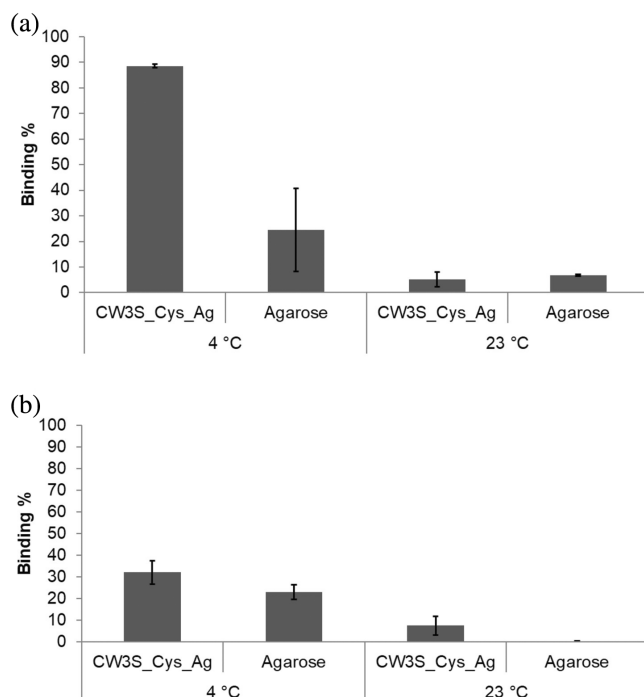
After synthesis and purification of CW3S\_Cys, we proceeded for further characterization by CD to determine the folding and thermal stability of the peptide. Peptide characterization by CD was performed in similar conditions used for hPin1\_WW (Dias et al., 2015) and hYAP65\_WW (Dias et al., 2016). The spectra were recorded in the Far-UV region to search for the characteristic signal of the WW domain—a positive ellipticity at 230 nm and a maximum negative ellipticity at 206 nm (Koepf et al., 1999). The spectra were acquired at 4, 23, and 88°C (Figure 4d). At 4 and 23°C, the results show a positive ellipticity signal at 230 nm and a maximum negative ellipticity close to 200 nm. The signal at 230 nm is not so intense as for the other WW domains reported by our research group (Dias et al., 2015, 2016), the major difference was the buffer composition and pH—for hPin1\_WW, we used a HEPES buffer, and for hYAP65\_WW, we used a phosphate buffer with pH 6 and a different equipment. Besides, at 260 nm, we observe that the signal is not completely zero, suggesting the protein is not totally solubilized in these buffer conditions (10 mM sodium phosphate buffer pH 7.2, 100 mM NaCl). Therefore, we believe the buffer composition could be adjusted to improve the signal at 230 nm and reduce the signal at 260 nm. At 88°C, the signal at 230 nm disappeared, indicating the denaturation of the small protein. After denaturation, the solution was submitted to 4°C, and the signal is recuperated, indicating the refolding of the small protein. The temperature denaturation studies were performed between 4 and 88°C and monitoring the change in signal at 230 nm (Figure 4e). The melting curve was fitted to a two-state model, and a melting temperature ( $T_m$ ) value of  $27.3 \pm 3.2^\circ\text{C}$  ( $R^2 = 0.99$ ) was obtained. This  $T_m$  value is lower than the value calculated for the WWp sequence,  $44.2 \pm 0.2^\circ\text{C}$  ( $R^2 = 0.99$ ) (Macias et al., 2000). The CD results indicate that CW3S\_Cys presents the folding of a WW domain, despite the increase in loop size compared with the WWp sequence. In addition, the  $T_m$  can indicate that this novel structure (42 residues in length) is less stable than the WWp, but the  $T_m$  value is comparable to other WW domains previously studied under similar buffer conditions (with 38 residues in length,

hPin1\_WW— $56.3 \pm 0.5^\circ\text{C}$  ( $R^2 = 0.99$ ) (Dias et al., 2015) or hYAP65\_WW— $31.4 \pm 2.9^\circ\text{C}$  ( $R^2 = 0.99$ ) and  $28.5 \pm 2.6^\circ\text{C}$  ( $R^2 = 0.99$ ) (Dias et al., 2016)). These  $T_m$  values are lower when compared with other protein scaffolds that have  $T_m$  higher than  $55^\circ\text{C}$ , for example, ABD ( $48\text{--}58^\circ\text{C}$ ) (Jonsson et al., 2008), Affibody ( $55\text{--}57^\circ\text{C}$ ) (Nygren, 2008), or Affilin ( $56\text{--}80^\circ\text{C}$ ) (Fiedler et al., 2006). Nevertheless, to stabilize the scaffold, we could incorporate a Cys in the end terminals to promote a disulfide bond and introduce more rigidity in the sequence, as it was described for the Pin1 scaffold by Patel et al. (2013).

We also assessed the binding affinity between CW3S\_Cys and HSA by microscale thermophoresis (MST) (Figure S21). Using the equipment software, data was automatically fitted with a Hill model, and EC50 was estimated to be  $0.96\text{ }\mu\text{M}$  and Hill coefficient at 0.75, indicating non-cooperative binding. Furthermore, the estimated values for  $K_D$  and  $K_a$  were  $284 \pm 86\text{ nM}$  and  $3.5 \times 10^6\text{ M}^{-1}$ , respectively. We concluded that the WWp5\_4 naïve phage display library generated a lead molecule that efficiently recognized HSA. The binding constant calculated for the complex is within the range of other albumin binding domains reported in the literature from evolved protein scaffolds (G148-ABD<sub>Y22A</sub> (5 KDa), 330 nM) to bind HSA (Table S3). Nevertheless, to improve the affinity of CW3S, it would be necessary to perform affinity maturation experiments or start off-rate selections with the third round outputs from phage display selection.

## 2.5 | Potential application for HSA capture

WW domains can be used for the capture of phosphorylated proteins and proline-rich peptides (Dias et al., 2015, 2016). Thus, we decided to test if the novel domain could selectively capture HSA after immobilization on a solid support. This property would be interesting for future applications, since there are several commercially available kits to capture HSA, either for depletion (e.g., ProteoPrep® Blue Albumin Depletion Kit, Sigma-Aldrich) or purification (e.g., Affibody anti-HSA immobilized in SulfoLink® Coupling Gel, Affibody). For that, we used the CW3S\_Cys, since this sequence has a Cys residue in the N-terminal, which allows for oriented protein immobilization into the chromatographic support. We applied the same immobilization protocol already developed by our team for hPin1\_WW and hYAP65\_WW, through the Sulfo-SMCC chemistry (sulfosuccinimidyl 4-[N-maleimidomethyl]cyclohexane-1-carboxylate) (Dias et al., 2015, 2016). We were able to immobilize  $1.35 \times 10^{-3}\text{ }\mu\text{mol peptide/mg support}$ , in the same range as observed in the previous works. This novel support (CW3S\_Cys\_Ag) was then tested for binding HSA and human IgG (here considered as the



**FIGURE 5** Selective capture of human serum albumin (HSA) and human Immunoglobulin G (IgG) at 4 and 23°C using a matrix with immobilized CW3S\_Cys. (a) HSA and (b) IgG. The plotted data corresponds to the average of triplicate assays and error bars represent the standard deviation. The binding buffer used was PBS1× pH 7.4.

contaminant) at two distinct temperatures typically used in affinity capture (Figure 5). The best results were obtained at 4°C, where the CW3S\_Cys\_Ag bound  $88.50\% \pm 0.67\%$  of loaded HSA (Figure 6a), thus capturing 291 ng protein/mg support (versus agarose with  $24.50\% \pm 16.20\%$ , 80.30 ng protein/mg agarose support), and bound  $32.1\% \pm 5.40\%$  of loaded IgG (107 ng protein/mg support). This is almost negligible since unmodified agarose binds IgG with  $23 \pm 3.4\%$  (76.40 ng protein/mg support). At 23°C, for both targets the binding is lower than 5%. These results suggest that the immobilized CW3S selectively captures HSA at low temperature. It is important to note that the chemical immobilization of a protein ligand on a chromatographic matrix creates local microenvironments and partitioning effects, can potentially alter the conformation, and affect the accessibility to interact with the target. Thus, it is not possible to directly correlate the temperature effect on the binding between HSA and CW3S immobilized on chromatographic beads with the stability of the structure in solution as observed by CD spectroscopy.

## 2.6 | Understanding structural features of the interaction between CW3S and HSA

We started by assessing the stability of CW3S folding and compared it with the other mutants (CW3 and

CW3S\_Cys). The results show that all three molecules have a similar stability in the last 10 ns of the simulation ( $\text{RMSD} < 0.5$ ) and the DSSP analysis shows that these structures maintain the folding with  $\beta$ -sheets (Figures S22–S24).

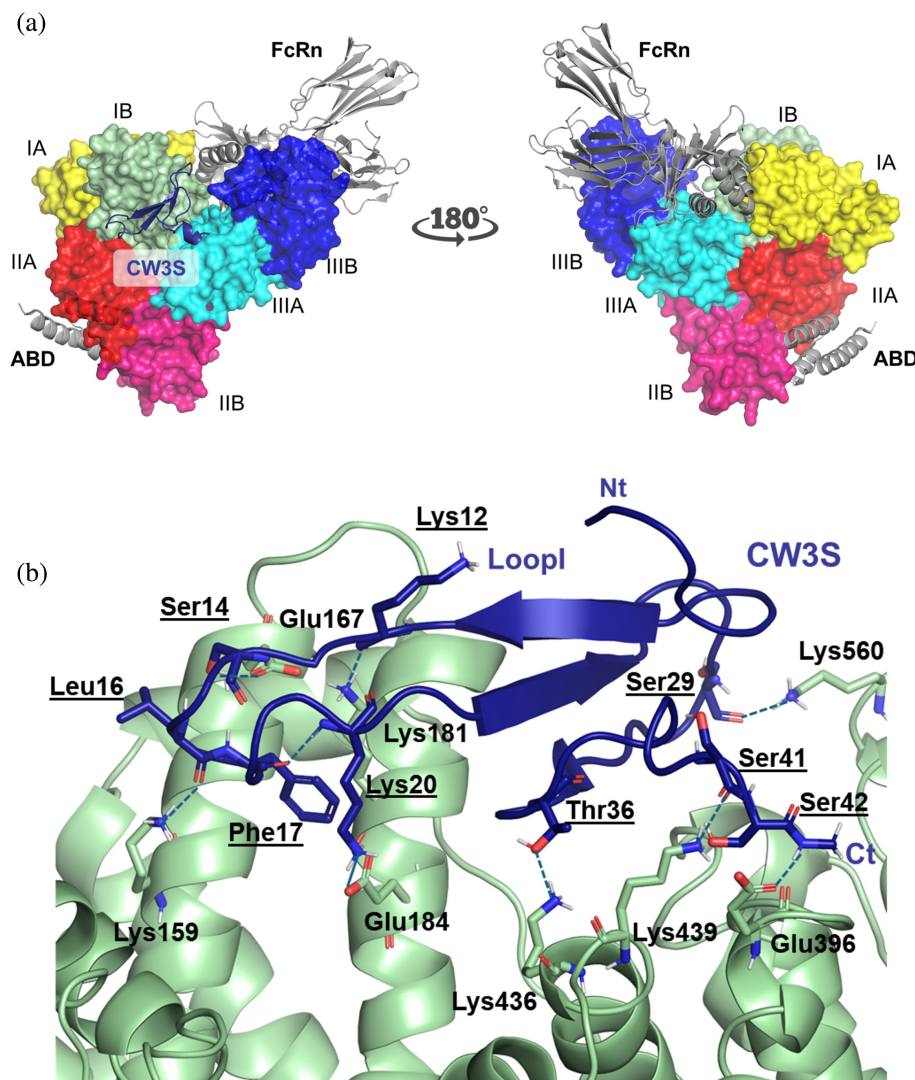
Second, to have a prevision of the possible intermolecular interactions, a model of CW3S in complex with HSA was assembled using AlphaFold3 (Abramson et al., 2024). According to this model, CW3S interacts with domains IB, IIIB, and IIIA from HSA. Thus, CW3S is expected to interact with HSA in different sites than other ligands, as the albumin binding domain (ABD) (Nilvebrant & Hober, 2013) (domains IIA and IIB) and on similar domains but on the opposite side from the neonatal Fc receptor (FcRn) (Oganessian et al., 2014) (domains IIIB, IIIA, IB, IA) (Figure 6a).

To test the stability of the predicted binding mode of CW3S to HSA, molecular dynamics simulations were performed. After 100 ns of MD, the small protein remains in the predicted region and slightly adapts to fit the crevice created by the interface of the three domains, IB, IIIA, IIIB. The molecular dynamics simulations trajectories were clustered, and the clusters with >10% frames were further analyzed. During 100 ns of MD, the CW3S structure remained stable (Figures S25, S26) and the N and C terminals were not facing the recognition surface (Figure 6b).

Table 3 summarizes the amino acids identified in the contacts between the CW3S and HSA for the three most populated clusters from the MD simulation. In bold are identified the amino acids that were present in all three clusters. Regarding the contacts between the WW domain and HSA, it can be noted that loop I Ser14, Leu16, Phe17 residues interact with the HSA domain IB region supporting the importance of the modified loop I for the binding, with its residues contributing energetically to HSA binding (Table 3 and for more details see Tables S4–S6). Important hydrogen bond networks around Phe17 backbone and Ser14 side chain, and ionic interactions with Lys20 are among the contacts that most contribute to the binding. In fact, if we look into the sequences discovered in this work (Figure 3b, c), the best binders always have in Loop I a hydrogen donor that is typically a Ser (Clone 3, 4, and 5), thus indicating that this is critical for binding against HSA. Loop II interfaces HSA at domain IIIB, mainly through a network of hydrogen bonds and hydrophobic interactions around Ser29. The  $\beta$ -sheet regions of CW3S interact with domains IB and IIA of HSA, residues Tyr22, Thr32, Trp35, and Thr36, mainly through hydrogen bonds and hydrophobic interactions.

## 3 | CONCLUSIONS

In this work we designed a protein scaffold inspired by the WWp sequence, but possessing extended loops to increase diversity and geometric flexibility to bind non-



**FIGURE 6** Binding characterization between CW3S and HSA. (a) Predicted binding region for CW3S (deep blue) to HSA (surface) by AlphaFold3 and comparison to superimposed co-crystallized structures for ligands ABD (PDBID: 1TF0) and FcRn (PDBID: 4K71) (gray). HSA domain color code (as per Sugio et al., 1999b): IA—yellow, IB—green, IIA—red, IIB—magenta, IIIA—cyan, IIIB—blue. (b) Most relevant interactions (Energy < −5.0 kcal/mol) from molecular dynamics simulations, of cluster 1 central structure, between CW3S (dark blue, underlined residues) and HSA (green); salt bridge—continuous line, hydrogen bond—dashed line.

cognate targets. After extensive study of the loop size we identified that the ideal size to maintain the scaffold stability is in—Loop I 5 to 7 residues and Loop II—4 to 6 residues. Inspired by the maximum loop size found in Nature, we designed the new scaffold named WWp5\_4 with 5 and 4 residues in loops I and II, respectively. We generated a naïve library based on the WWp5\_4 sequence with full randomization of the loops, which evolved through phage display against HSA. After the third round of selection, we identified one lead ligand (CW3S) that binds HSA. This protein was recombinantly expressed alone (CW3S), as a fusion protein to GFP (CW3S\_GFP), as well as chemically synthesized (CW3S\_Cys). The binding constant between CW3S and HSA was established in the nM range. The CW3S\_Cys maintained the characteristic folding of other native WW domains. After immobilization in a chromatographic support, it selectively captured HSA at 4°C (291 ng protein/mg support). The binding mode of CW3S with the target HSA was predicted in silico

showing that Loop I, followed by Loop II, are the regions which mostly contribute to these interactions.

These encouraging results show the potential of a totally synthetic design, non-existing in nature WW domain library for the development of affinity reagents. Due to their small size, we envision these reagents in various constructs, as in tandem, in fusion, or conjugated with probes. Thus, they can generate innovative reagents with applications in different biotechnological fields, such as affinity chromatography, in vitro diagnostics, or therapeutics.

## 4 | MATERIALS AND METHODS

### 4.1 | Molecular modeling

*WW Loop elongation studies:* The structure of the WWp (PDB code: 1E0M (Macias et al., 2000)) was initially engineered with Ala residues to generate several

**TABLE 3** Amino acids in HSA surface that interact with CW3S (energy lower than  $-0.5$  kcal/mol) from central structure of MD clusters with more than 10% of frames.

| Cluster number (%frames) | Amino acids in HSA surface  | Amino acids in CW3S  |
|--------------------------|---|--|
| Cluster 1 (50%)          | <b>Asp173</b> , Lys181, Glu167<br>Glu167, Lys159, Lys181, Arg160, Glu184, Ala164, Leu185<br>Glu184, Ala176, Ala175<br>Lys560<br><b>Glu518</b> , Ser517, Pro180, Lys181, Glu184, Lys436, Phe395, Lys439, Glu396  | N-Ter/ $\beta$ -sheet I: Tyr11, Lys12, <b>Ser13</b><br>Loop I: <b>Ser14</b> , <b>Leu16</b> , <b>Phe17</b> , Leu18<br>$\beta$ -sheet II: Lys20, <b>Tyr22</b> , <b>Tyr24</b><br>Loop II: <b>Ser29</b><br>$\beta$ -sheet III/C-Ter: <b>Thr32</b> , Trp35, Thr36, Met40, Ser41, Ser42                  |
| Cluster 2 (34%)          | Lys560, Glu556, Val555, Arg521, <b>Glu518</b> , <b>Asp173</b><br>Asp173, Gln170, Ala171, Cys177<br>Asp173, Leu179, <b>Glu518</b> , Lys519<br>Thr515, Glu119, Lys519<br>Glu119, Arg117, Leu179, Asp183, <b>Glu518</b> , Ala176, Asp549, Lys402, Gln397, Leu398 | N-Ter/ $\beta$ -sheet I: Gly3, Leu4, Pro5, Pro6, Trp8, <b>Ser13</b><br>Loop I: <b>Ser14</b> , <b>Leu16</b> , <b>Phe17</b><br>$\beta$ -sheet II: <b>Tyr22</b> , <b>Tyr24</b> , Asn25<br>Loop II: Asn28, <b>Ser29</b> , Pro30<br>$\beta$ -sheet III/C-Ter: <b>Thr32</b> , Ser33, Thr36, Asp37, Met40 |
| Cluster 3 (12%)          | <b>Asp173</b> , Glu167<br>Glu167, Gln170, Thr166<br>Cys177, Leu179, Lys519<br>Lys519, <b>Glu518</b> , Ser517, Lys432, Glu400<br>Asp183, Glu520, Pro180, Glu184, Lys181  | N-Ter/ $\beta$ -sheet I: Lys12, <b>Ser13</b><br>Loop I: <b>Ser14</b> , <b>Leu16</b> , <b>Phe17</b><br>$\beta$ -sheet II: <b>Tyr22</b> , <b>Tyr24</b> , His26<br>Loop II: Asp27, Asn28, <b>Ser29</b><br>$\beta$ -sheet III/C-Ter: Lys31, <b>Thr32</b> , Ser33, Trp35                                |

Note: In bold are the amino acids that are present in three clusters as contact points.

structures with distinct lengths in loops I and II. The homology models were created with SwissModel (Arnold et al., 2006) using the 1E0M structure as a template and automatic sequence alignment from SwissModel. The obtained models of different loop lengths were prepared in PyMol software (DeLano, 2002) by incorporating in the N and C terminals capping groups, an acetyl or amine group respectively. The proteins were solvated in a box with 1.2 nm minimum distance between box faces and protein surface. The molecular dynamics Simulations were carried out on the GRO-MACS (Lindahl et al., 2020) 2023.5 simulation package. The force field used was Charmm36 all-atom (Huang & Mackerell, 2013), TIP3P water model, and the simulations were performed in triplicates during 100 ns. The systems were previously minimized with steepest descent and conjugate gradient algorithms, to a convergence criterion of 1000 and 400 kJ/mol·nm<sup>-1</sup>, respectively. The minimized systems were submitted to a heating and equilibration step. The systems were heated using annealing to 298 K during 100 ps and equilibrated for another 100 ps at 298 K, with Berendsen temperature coupling on an NVT ensemble. The system was further equilibrated in NPT for 200 ps using V-rescale thermostat at 298 K and C-rescale barostat at 1.0 bar. LINCS constraints were used for hydrogen bonding atoms, Particle Mesh Ewald for electrostatics with a cut-off of 1.0 nm and a Van der Waals cut-off of 1.0 nm. Triplicates of molecular dynamics runs were done from the heating step, where velocities were

randomly assigned. For visualization of the results, we used PyMol 1.3 and 3 (DeLano 2002), VMD 1.9 (Humphrey et al., 1996), and Moe (Anon, 2024).

**WWp5\_4 library design:** The selected length WW prototype-based scaffold with 5 and 4 amino acids in loops I and II, respectively, was proposed and designated as the WWp5\_4 scaffold. Then, two controls and 10 mutants were generated using SwissModel, as previously described, to assess resilience to chemical diversity in the loops: (i) WW prototype; (ii) WWp5\_4 with loops substituted with Ala residues; (iii) six WWp5\_4 mutants with loops substituted by combinations of four amino acids (YADS—Tyr, Ala, Asp and Ser); (iv) four WWp5\_4 mutants with loops substituted by a random selection of amino acids (Table S3). The mutant structures were generated by homology modeling, as previously described. The obtained mutant models of each structure were prepared and followed by molecular dynamics simulations, as described above. The selected binder from phage display CW3 and its mutant sequences (CW3S and CW3S\_Cys) were also submitted to the same model assembly and molecular dynamics simulations as the other mutants.

**CW3S and HSA complex.** The clone CW3S interaction with HSA was predicted using AlphaFold3 (Abramson et al., 2024). The sequences of CW3S and HSA (Uniprot: P02768) were submitted to the webserver. The obtained complex (pTM of 0.82) was also simulated using the previously described pipeline for 100 ns in triplicates. The analysis of all simulations was

performed using rms, rmsf, energy, and dssp tools from the GROMACS package (Lindahl et al., 2020) and Moe (Anon, 2024).

## 4.2 | Phage display selection of WWp5\_4 library

Library construction is described in detail in Supplementary Information—Materials and Methods (Table S1). For the phage display selection of the WWp5\_4 library against HSA, the target was dissolved in PBS1× and immobilized overnight into two wells of the microplate (Corning 3690, high-binding microplates) using 1 µg (first panning), 0.5 µg (second panning) and 0.25 µg (third panning) at 4°C. The target was removed, and the wells washed with PBS1× sterile, the plate was blocked with 150 µL 5% Soya Milk (dilute from 6.9% total protein) in PBS1× filtered with 0.22 µm. This solution was incubated for 60 min at room temperature. The blocking solution was removed and 100 µL of phages library solution was added to the microplate wells and incubated for 2 h at room temperature. The non-bounded phages were removed, and the wells were washed with: PBS1×-0.05% Tween-20. This procedure was repeated 5× (first panning), 10× (second panning) and 15× (third panning) and between each wash the plate was incubated 1 min with washing solution. Finally, the wells were washed with PBS1×. The elution of the bounded phages was achieved with a trypsin solution in PBS1× (10 mg/mL), the microplate wells were incubated for 30 min at 37°C. The eluted fractions were resuspended and collected, the wells were washed with PBS1× and the solution recovered. The eluted phages were used to infect 2 mL of exponentially growing *E. coli* ER2738 ( $OD_{600nm} = 0.6–0.7$ ) and were incubated for 15 min (first panning and 30 min in third panning), 37°C. From the infected culture 2 µL were collected to perform dilutions to titer the number of output clones, these cultures were plated in LB ampicillin plates. The titer calculations were performed as described in Barbas III et al., (2001). The remaining infected culture was used to continue for the next panning step. For that it was prepared as described in Barbas III et al., (2001). Afterwards, the second round of panning started as described above, after the third panning the eluted phages were used to infect *E. coli* TOP10F' to proceed for protein expression. The details for the clones protein expression and ELISA screening please see the Supplementary Information—Materials and Methods.

## 4.3 | CW3S-GFP expression, purification and characterization

The gene of CW3 was mutated in loop II to change a Cys to a Ser to avoid dimerization and was named

CW3S. This gene was cloned into an expression vector that contained the gene of GFP (pET21c\_GFP) (Pina et al., 2014). CW3S was cloned into pET21c\_GFP, and one positive clone was transformed into *E. coli* BL21 (DE3) to proceed for protein expression. Protein expression using pET21c\_CW3S\_GFP was performed using 1 L of LB culture medium as described before for WW domain (human Pin1 WW domain) in Fernandes et al., (2014). After expression, the bacterial extracts were treated, as in Fernandes et al., (2014). The protein content in the SF and in the insoluble fractions (ISF) was characterized by SDS-PAGE and Western blot as described in Fernandes et al., (2014). The SF was further purified using AKTA Pure System using agarose resin modified with Ugi Ligand A4C7, a 5.4 mL of a 50% slurry, as described in Pina et al. (2015). The purification protocol was followed as described in Fernandes et al. (2014). After purification, CW3S\_GFP was quantified by total protein, and fluorescence intensity was determined. More details regarding protein expression and purification are described in Supplementary Information—Materials and Methods.

The purified protein CW3S\_GFP was used to study the binding toward the target HSA. To assess selectivity, we used the ELISA scheme shown in Figure S1. The negative controls used were: (i) one well with blocking solution PBS1×-5%Soya Milk; well with HSA (1 µg/well) immobilized with GFP protein (1 µg/well), instead of CW3S\_GFP; and (ii) one well without a target where the CW3S\_GFP or GFP was added, after which the wells were blocked, and all other components were added. To determine the equilibrium constant of association, we employed an ELISA-based method. Assays were made in triplicate. The full details of the method can be found in Supplementary Information.

## 4.4 | Solid-phase peptide synthesis, purification, and characterization of CW3S

The sequence of CW3S was modified at the N-terminal to incorporate a Cys residue that will be useful for immobilization in a solid-support (CW3S\_Cys), as already described for hPin1 and hYAP65 WW domains (Dias et al., 2015, 2016). The chemical synthesis protocol, as well as characterization by CD and Mass Spectrometry, is described in detail in the Supplementary Information—Materials and Methods. MST was used to determine the affinity between CW3S\_Cys peptide labeled with a fluorescent agent (Fluorescein Diacetate 5-Maleimide) and HSA, the target molecule. Briefly, the peptide was labeled with Fluorescein Diacetate 5-Maleimide (Sigma-Aldrich) and purified by PD10 desalting column (GE Healthcare Life Sciences) in accordance with manufacturer instructions. MST assays were performed with Monolith NT.115 Standard Treated Capillary K002 in NT.115 Nanotemper, with

filter BLUE/RED (for fluorescein, blue filter was used). A more detailed description can be found in Supplementary Information—Materials and Methods.

## 4.5 | Solid-support immobilization of CW3S for HSA capture

The CW3S\_Cys produced chemically was immobilized through the sulfhydryl group of a Cys in an aminated matrix (Sephacrose CL6B) using Sulfo-SMCC chemistry. The protocol for immobilization was the same as described for hPin1\_WW domain in Dias et al. (2015). Thus, a novel support CW3S\_Cys\_Ag was generated. The CW3S\_Cys\_Ag (50 mg) was used to determine the binding capacity against HSA (20 µg in 300 µL in PBS1× pH 7.4). We further studied the support specificity by using a solution of human Immunoglobulin G (IgG) in the same conditions. Full details of the protocol can be found in Supplementary Information.

## AUTHOR CONTRIBUTIONS

**Ana Margarida Gonçalves Carvalho Dias:** Conceptualization; formal analysis; writing – original draft; investigation; writing – review and editing. **Gonçalo Duarte Gomes Teixeira:** Investigation. **Arménio Jorge Moura Barbosa:** Investigation; formal analysis. **Joao Goncalves:** Supervision. **Olga Iranzo:** Supervision; funding acquisition. **Ana Cecilia Afonso Roque:** Funding acquisition; validation; supervision; project administration; writing – review and editing; writing – original draft; conceptualization; methodology; resources.

## ACKNOWLEDGEMENTS

The authors would like to acknowledge the financial support of FCT—Fundação para a Ciência e a Tecnologia, I.P., in the scope of the project UIDP/04378/2020 and UIDB/04378/2020 of the Research Unit on Applied Molecular Biosciences—UCIBIO and the project LA/P/0140/2020 of the Associate Laboratory Institute for Health and Bioeconomy—i4HB, projects UIDB/04138/2020 e UIDP/04138/2020 of the Research Institute for Medicines (iMed) and Gonçalo Teixeira PhD grant (PD/BD/139800/2018; COVID/BD/152648/2022). This work has received funding from the European Union's Horizon 2020 programme under grant agreement No. 899732 (PURE Project) and from Fundação para a Ciência e Tecnologia (Portugal) and ERDF under the PT2020 Partnership Agreement (LISBOA-01-0145-FEDER-028878) for funding the Sea2See project (PTDC/BII-BIO/28878/2017) and Proteios project (PTDC/CTM-CTM/3389/2021).


## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## ORCID

Ana Margarida Gonçalves Carvalho Dias  <https://orcid.org/0000-0001-7416-7215>

Gonçalo Duarte Gomes Teixeira  <https://orcid.org/0000-0001-7675-5926>

Arménio Jorge Moura Barbosa  <https://orcid.org/0000-0002-9756-4958>

Joao Goncalves  <https://orcid.org/0000-0002-1245-3715>

Olga Iranzo  <https://orcid.org/0000-0001-8542-2429>

Ana Cecilia Afonso Roque  <https://orcid.org/0000-0002-4586-3024>

## REFERENCES

- Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*. 2024;630:493–500.
- Aires F, Corte-real S, Gonçalves J. Recombinant antibodies as therapeutic agents pathways for modeling new biodrugs. *BioDrugs*. 2008;22:301–14.
- Anon. Molecular Operating Environment (MOE). 2024.
- Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*. 2006;22:195–201.
- Barbas CF III, Burton DR, Scott JK. Silverman GJ phage display: a laboratory manual. 1st ed. New York: Cold Spring Harbor Laboratory Press; 2001.
- Cooper BM, Iegre J, O'Donovan DH, Öwegård Halvarsson M, Spring DR. Peptides as a platform for targeted therapeutics for cancer: peptide-drug conjugates (PDCs). *Chem Soc Rev*. 2021; 50:1480–94.
- Dalby PA, Hoess RH, DeGrado WF. Evolution of binding affinity in a WW domain probed by phage display. *Protein Sci*. 2000;9:2366–76.
- DeLano W. The PyMOL Molecular Graphics System. 2002.
- Dias A. CIS display: DNA-based technology as a platform for discovery of therapeutic biologics. In: Iranzo O, Roque ACA, editors. Peptide and protein engineering: from concepts to biotechnological applications. Berlin, Germany: Springer Science+Business Media, LLC, part of Springer Nature; 2020. p. 1–18.
- Dias A, dos Santos R, Iranzo O, Roque A. Affinity adsorbents for proline-rich peptide sequences: a new role for WW domains. *RSC Adv*. 2016;6:68979–88.
- Dias A, Iranzo O, Roque A. An in silico and chemical approach towards small protein production and application in phosphoproteomics. *RSC Adv*. 2015;5:19743–51.
- Dias A, Roque A. The future of protein scaffolds as affinity reagents for purification. *Biotechnol Bioeng*. 2017;114:481–91. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27568828>
- Elsadek B, Kratz F. Impact of albumin on drug delivery—new applications on the horizon. *J Control Release*. 2012;157:4–28.
- Espanel X, Navin N, Kato Y, Tanokura M, Sudol M. Probing WW domains to uncover and refine determinants of specificity in ligand recognition. *Cytotechnology*. 2003;43:105–11.
- Fernandes CSM, Pina AS, Dias AMGC, Branco RJF, Roque ACA. A theoretical and experimental approach toward the development of affinity adsorbents for GFP and GFP-fusion proteins purification. *J Biotechnol*. 2014;186:13–20.
- Fiedler E, Fiedler M, Proetzel G, Scheuermann T, Fiedler U, Rudolph R. Affilin™ molecules novel ligands for bioseparation. *Food Bioprod Process*. 2006;84:3–8.
- Gebauer M, Skerra A. Engineering of binding functions into proteins. *Curr Opin Biotechnol*. 2019;60:230–41.
- Gray A, Bradbury A, Knappik A, Plückthun A, Borrebaeck C, Dübel S. Animal-free alternatives and the antibody iceberg. *Nat Biotechnol*. 2020;38:1234–8.

- Gunasekaran K, Ramakrishnan C, Balaram P. Beta-hairpins in proteins revisited: lessons for de novo design. *Protein Eng.* 1997; 10:1131–41.
- Huang J, Mackerell AD. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J Comput Chem.* 2013;34:2135–45.
- Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph.* 1996;14:33–8.
- Jäger M, Zhang Y, Bieschke J, Nguyen H, Dendle M, Bowman ME, et al. Structure-function-folding relationship in a WW domain. *Proc Natl Acad Sci USA.* 2006;103:10648–53.
- Jiang H, Jude KM, Wu K, Fallas J, Ueda G, Brunette TJ, et al. De novo design of buttressed loops for sculpting protein functions. *Nat Chem Biol.* 2024;20:974–80.
- Jonsson A, Dogan J, Herne N, Abrahamsen L, Nygren P-A. Engineering of a femtomolar affinity binding protein to human serum albumin. *Protein Eng Des Sel.* 2008;21:515–27.
- Kaul R, Angeles AR, Jager M, Powers ET, Kelly JW. Incorporating b-turns and a turn mimetic out of context in loop 1 of the WW domain affords cooperatively folded b-sheets. *J Am Chem Soc.* 2001;123:5206–12.
- Koepl EK, Petrassi HM, Sudol M, Kelly JW. WW: an isolated three-stranded antiparallel b-sheet domain that unfolds and refolds reversibly; evidence for a structured hydrophobic cluster in urea and GdnHCl and a disordered thermal unfolded state. *Protein Sci.* 1999;8:841–53.
- Lindahl E, Abraham MJ, Hess B, van der Spoel D. GROMACS 2020 Source Code. 2020.
- Linsky TW, Vergara R, Codina N, Nelson JW, Walker MJ, Su W, et al. De novo design of potent and resilient hACE2 decoys to neutralize SARS-CoV-2. *Science.* 2020;370:1208–14.
- Macias MJ, Gervais V, Civera C, Oshkinat H. Structural analysis of WW domains and design of a WW prototype. *Nat Struct Biol.* 2000;7:375–9.
- Macias MJ, Wiesner S, Sudol M. WW and SH3 domains, two different scaffolds to recognize proline-rich ligands. *FEBS Lett.* 2002;513:30–7.
- Martínez-Lumbreras S, Träger LK, Mulorz MM, Payr M, Dikaya V, Hipp C, et al. Intramolecular autoinhibition regulates the selectivity of PRPF40A tandem WW domains for proline-rich motifs. *Nat Commun.* 2024;15:3888.
- Millioni R, Tolín S, Puricelli L, Sbrignadello S, Fadini GP, Tessari P, et al. High abundance proteins depletion vs low abundance proteins enrichment: comparison of methods to reduce the plasma proteome complexity. *PLoS One.* 2011;6:e19603.
- Morgan AA, Rubenstein E. Proline: the distribution, frequency, positioning, and common functional roles of proline and polyproline sequences in the human proteome. *PLoS One.* 2013;8:e53785.
- Mouratou B, Schaeffer F, Guilvout I, Tello-manigne D, Pugsley AP, Alzari PM. Remodeling a DNA-binding protein as a specific in vivo inhibitor of bacterial secretin PulD. *Proc Natl Acad Sci.* 2007;104:17983–8.
- Muyldermans S. A guide to: generation and design of nanobodies. *FEBS J.* 2021;288:2084–102.
- Nilvebrant J, Hober S. The albumin-binding domain as a scaffold for protein engineering. *Comput Struct Biotechnol J.* 2013;6:e201303009.
- Nygren P-A. Alternative binding proteins: affibody binding proteins developed from a small three-helix bundle scaffold. *FEBS J.* 2008;275:2668–76.
- Oganesyan V, Damschroder MM, Cook KE, Li Q, Gao C, Wu H, et al. Structural insights into neonatal fc receptor-based recycling mechanisms. *J Biol Chem.* 2014;289:7812–24.
- Patel S, Mathonet P, Jaulent AM, Ullman CG. Selection of a high-affinity WW domain against the extracellular region of VEGF receptor isoform-2 from a combinatorial library using CIS display. *Protein Eng Des Sel.* 2013;26:307–15. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23378640>
- Piana S, Sarkar K, Lindorff-larsen K, Guo M, Gruebele M, Shaw DE. Computational design and experimental testing of the fastest-folding  $\beta$ -sheet protein. *J Mol Biol.* 2011;405:43–8.
- Pina AS, Dias AMGC, Ustok FI, El Khoury G, Fernandes CSM, Branco RJF, et al. Mild and cost-effective green fluorescent protein purification employing small synthetic ligands. *J Chromatogr A.* 2015;1418:83–93. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0021967315013424>
- Pina AS, Pereira AS, Branco RJF, El KG, Lowe CR. A tailor-made “tag–receptor” affinity pair for the purification of fusion proteins. *Chembiochem.* 2014;15:1423–35.
- Price JL, Powers DL, Powers ET, Kelly JW. Glycosylation of the enhanced aromatic sequon is similarly stabilizing in three distinct reverse turn contexts. *Proc Natl Acad Sci USA.* 2011;108:14127–32.
- Salah Z, Alian A, Aqeilan RI. WW domain-containing proteins: retrospectives and the future. *Front Biosci.* 2012;17:331 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22201747>
- Sudol M, Hunter T. NeW wrinkles for an old domain Minireview. *Cell.* 2000;103:1001–4.
- Sugio S, Kashima A, Mochizuki S, Noda M, Kobayashi K. Crystal structure of human serum albumin at 2.5 Å resolution. *Protein Eng.* 1999a;12:439–46.
- Sugio S, Kashima A, Mochizuki S, Noda M, Kobayashi K. Crystal structure of human serum albumin at 2.5 Å resolution. *Protein Eng.* 1999b;12:439–46.
- Villegas-Méndez A, Fender P, Garin MI, Rothe R, Liguori L, Marques B, et al. Functional characterisation of the WW minimal domain for delivering therapeutic proteins by adenovirus dodecahedron. *PLoS One.* 2012;7:e45416.
- Wang J, Lisanza S, Juergens D, Tischer D, Watson JL, Anishchenko I, et al. (2022) scaffolding protein functional sites using deep learning. *Science.* 1979;377:387–94.
- Yanagida H, Matsuura T, Yomo T. Compensatory evolution of a WW domain variant lacking the strictly conserved Trp residue. *J Mol Evol.* 2008;66:61–71.
- Zimmermann I, Egloff P, Hutter CAJ, Kuhn BT, Bräuer P, Newstead S, et al. Generation of synthetic nanobodies against delicate proteins. *Nat Protoc.* 2020;15:1707–41.
- Zorzi A, Linciano S, Angelini A. Non-covalent albumin-binding ligands for extending the circulating half-life of small biotherapeutics. *Medchemcomm.* 2019;10:1068–81.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Dias AMGC, Teixeira GDG, Barbosa AJM, Goncalves J, Iranzo O, Roque ACA. Design and evolution of a synthetic small protein scaffold based on the WW domain. *Protein Science.* 2025;34(6):e70164. <https://doi.org/10.1002/pro.70164>