



**HAL**  
open science

# Hydrolytic Zinc Metallopeptides Using a Computational Multi-State Design Approach

Henrique Carvalho, Ricardo J.F. Branco, Fábio Leite, Manolis Matzapetakis,  
Ana Cecilia A Roque, Olga Iranzo

► **To cite this version:**

Henrique Carvalho, Ricardo J.F. Branco, Fábio Leite, Manolis Matzapetakis, Ana Cecilia A Roque, et al.. Hydrolytic Zinc Metallopeptides Using a Computational Multi-State Design Approach. *Catalysis Science & Technology*, 2019, 10.1039/C9CY01364D . hal-02335583

**HAL Id: hal-02335583**

**<https://hal.science/hal-02335583>**

Submitted on 14 Dec 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.

# Hydrolytic Zinc Metallopeptides Using a Computational Multi-State Design Approach

Henrique F. Carvalho,<sup>a,b</sup> Ricardo J. F. Branco,<sup>a</sup> Fabio Leite,<sup>a</sup> Manolis Matzapetakis,<sup>b</sup> A. Cecília A. Roque,<sup>a</sup> Olga Iranzo<sup>c\*</sup>

Hydrolytic zinc enzymes are common targets for protein design. The versatility of the zinc chemistry can be combined with the usage of small protein scaffolds for biocatalytic applications. Despite this, the computational design of metal-containing proteins remains challenging due to the need to properly model protein-metal interactions. We addressed these issues by developing a computational multi-state design approach of artificial zinc hydrolases based on small protein scaffolds. The zinc-finger peptide Sp1f2 was redesigned to accommodate a catalytic zinc centre and the villin headpiece C-terminal subdomain HP35 was *de novo* designed for metal-binding and catalytic activity. Both metallopeptides exhibited metal-induced folding ( $K_{ZnP,app} \approx 2 \times 10^5 \text{ M}^{-1}$ ) and hydrolytic activity ( $k_2 \approx 0.1 \text{ M}^{-1}\text{s}^{-1}$ ) towards an ester substrate. By focusing on the inherent flexibility of small proteins and their interactions with the metal ion by molecular dynamics simulations and spectroscopic studies, we identified current limitations on computational design of metalloenzymes and propose how these can be overcome by integrating information of protein-metal interactions in long time scale simulations.

## Introduction

Metalloenzymes are highly specific and active protein catalysts found in biological systems. The versatile chemical properties of these catalysts are widely explored in research and industry for the production of biologically-active and valuable chemicals.<sup>1</sup> Not surprisingly, metalloenzymes are common targets of protein design endeavours<sup>2</sup>, given the possibility to explore different metal ion chemistries along with protein sequence modifications. Among them, those containing zinc have been the subject of several designs since this metal ion plays a catalytic role in all classes of enzymes.<sup>3</sup> Computational Enzyme Design (CED) approaches have contributed significantly to advancements in the field of biocatalysis by aiding to redesign metal centres in native proteins or by *de novo* design of non-native scaffolds.<sup>4-7</sup>

Peptides and small proteins are interesting scaffolds for CED approaches given their reduced size and simple fold. Moreover, full sequence modifications and introduction of other non-proteinogenic chemical functionalities can be readily achieved through chemical synthesis. This makes the use of such small scaffolds promising for technological applications, since one can design the catalytic machinery while exploring a broad substrate scope. Artificial zinc hydrolases based on peptides and small proteins are one example of this, such as the *de novo* designed peptide with a zinc-finger fold<sup>8</sup> and amyloid-forming heptapeptides<sup>9</sup> by CED approaches, and also without explicit computational modelling.<sup>10-15</sup> Nonetheless, the catalytic efficiencies of such artificial hydrolases are still far from the ones of native zinc hydrolases by one or more orders of magnitude, thus putting into evidence the current limitations of current rational design methods.<sup>16</sup> This gap in terms of catalytic proficiency is not entirely understood. Explicit modelling of both first and second sphere interactions in catalytic zinc sites may be a requirement to achieve productive catalytic interactions between the catalyst and substrates.<sup>17</sup> Although in proteins the extensive network of residue contacts in active sites may compensate for improper modelling of metal-protein interactions, in small scaffolds the chemical environment around the metal ions needs to be properly defined.<sup>18</sup> On this

regard, CED approaches are well-suited to address these issues since interactions of the metal ion with both the protein and substrates can be modelled explicitly.

In this work, we sought to explore CED of small scaffolds (up to 64 amino acid chain length) to develop new artificial zinc hydrolases. We employed the Rosetta software<sup>19-21</sup> to screen and design a set of 43 scaffolds using a Multi-State Design (MSD) approach. MSD takes into account the inherent dynamic properties of protein scaffolds in solution and it has been recently shown to improve CED approaches where backbone and side chain reorientations overlap with the timescales of protein-substrate interactions.<sup>22-24</sup> Despite its potential, MSD has not been widely employed for small scaffolds, which typically exhibit high conformational dynamics in solution. As a model for the development of a zinc hydrolase by MSD, we focused on the conserved chemical and geometric features that make up the catalytic machinery of a type of hydrolases with broad substrate scope, the zinc metalloproteases (MPs). MPs are widely found in nature<sup>25,26</sup> and have a wide range of industrial<sup>1,27,28</sup> and biotechnological<sup>29,30</sup> applications. These enzymes hydrolyse peptide bonds through a nucleophilic attack of the scissile carbonyl moiety via a water molecule activated by the catalytic zinc metal ion and glutamate residue.<sup>31-35</sup> We explored two different approaches. In the first, we redesigned the zinc site of a native zinc-finger metallopeptide, the finger 2 of the Sp1 transcription factor (Sp1f2).<sup>36</sup> In the second, a native scaffold – villin headpiece C-terminal subdomain (HP35)<sup>37</sup> – was *de novo* designed to accommodate the catalytic zinc centre.

Along with the MSD approach, we further explored the dynamical properties of the designed metallopeptides with molecular dynamics (MD) simulations and solution structure properties using spectroscopic methods, as well as their metal-binding properties and hydrolytic activity. The details will be described in the following sections.

## Experimental

### Computational Studies

**Computational enzyme design.** Full details of the employed CED are given in Electronic supplementary information (ESI). In overview: to develop a general active site model of MPs - MA(M):diAla - structures of MP-inhibitor complexes from the MA(M) subclass were selected from the MEROPS database.<sup>25</sup> The dipeptide Acetyl-Ala-Ala-NH<sub>2</sub> (diAla) bound to zinc was modelled as an ensemble of 256 conformers of the *gem*-diol intermediate. Details of modelled properties are given in Figure 1. The zinc metal ion was modelled as part of the substrate molecule, as done similarly in the redesign of adenosine deaminase.<sup>38</sup> The PDB files of the selected scaffolds (35 NMR structures, containing between 10-38 states each, and 8 X-ray structures) were used as input for MSD with the *matcher* executable with no energy minimization.<sup>19,39</sup> The “secondary algorithm” was used in the first stage for the combinatorial screen of all residues in each scaffold (total of  $2.2 \times 10^4$  C $^{\alpha}$  positions for 552 structure files). This yielded active site models (hits) that could be crafted with proper Glu<sub>cat</sub> geometries and relaxed His<sub>1-3</sub> geometries. During this stage, hits with at least one residue in the termini were discarded due to the high flexibility of these regions. Hits where two His were modelled consecutively in sequence since were also discarded since this pattern produced highly distorted His geometries and is not found in MA(M) active sites. The “classical algorithm” was used in the second stage to filter hits with proper geometry for all residues. Only combinations of residues where Glu<sub>cat</sub> could be modelled in the first step were screened for each structure. The *enzyme design* executable was used for the resulting hits, with 4 cycles of repacking and design of residues close to the modelled active site in order to optimize catalytic interactions. The outputted designs were evaluated by principal component analysis (PCA), using a set of 16 Rosetta scoring function parameters (Rosetta energy units, REU).<sup>40,41</sup> For the screening of peptide/small-protein scaffolds, only 1 round of design was used in a total of 542 hits. For design, 10 rounds were made and sequence logos generated to select the most frequent residue substitutions. A final visual inspection was made and the candidate sequences were selected for synthesis.

**MD simulations.** peptides and astacin were simulated in explicit solvent under periodic boundary conditions with the GROMACS 5.1.2 simulation package.<sup>42,43</sup> The employed force field was AMBER99SB\*-ILDN<sup>44</sup> modified to include the Cationic Dummy Atom (CaDA) approach<sup>45</sup>, with the zinc ion coordinated to four dummy atoms in a tetrahedron-shaped geometry and coordinating histidines in double-deprotonated state (charge -1). Input structure files corresponded to native Sp1f2 (PDB ID: 1VA2, NMR state 27), HP35 (PDB ID: 1UNC, NMR state 6) and Astacin (PDB ID: 1AST) or the outputted structures with best Score<sub>total</sub> for RD01 and RD02 scaffolds. For systems containing the zinc ion, the metal coordinates from the original file were replaced by a zinc-CaDa model. Two independent simulation trajectories of 1  $\mu$ s each for peptides and two of 350 ns for astacin in the NPT ensemble were done, with a total simulated time 2  $\mu$ s for each peptide and 700 ns for astacin. Details of system setup and cluster analysis given in ESI.

## Materials and methods

**Peptide Synthesis.** Peptides were produced by automated solid-phase peptide synthesis using standard Fmoc-protocols. Crudes were purified by preparative reversed-phase high-pressure liquid chromatography (HPLC) with a C18 column. Collected peaks were analysed by analytical HPLC using C12 or C18 columns to determine purity. The identity of peptides (Sp1f2: 3841.9 Da, RD01: 3743.9 Da, HP35: 3998.1 Da and RD02: 3730.9 Da) was confirmed by mass spectrometry (MS) analysis using time-of-flight with electrospray ionisation or matrix-assisted laser desorption/ionization. Full details of reagents, equipment and experimental conditions are given in ESI.

**Stock solutions.** Zinc stock solutions (ZnCl<sub>2</sub>) of 10.33 or 1.033 mM were prepared from a 103.3 mM stock solution (standardized by inductively coupled plasma MS). Peptide stock solutions were prepared by dissolving lyophilized peptide directly in MilliQ H<sub>2</sub>O and their concentrations were determined by readings at 280 nm in 6 M guanidium chloride, considering absorbance contributions of tryptophan ( $\epsilon=5690 \text{ M}^{-1}\cdot\text{cm}^{-1}$  for RD01, RD02 and HP35), and also tyrosine ( $\epsilon=1280 \text{ M}^{-1}\cdot\text{cm}^{-1}$  for RD01) residues.<sup>46</sup> For Sp1f2, the concentration was determined by the Ellman’s test (cysteine content determination).<sup>47</sup>

## Spectroscopic studies

**UV-Vis spectroscopy.** UV-Vis absorbance assays were done in either quartz cuvettes ( $V_T=700$  or  $900 \mu\text{L}$ , *path-length* = 1 cm) or in 96-well plates, ( $V_T=300 \mu\text{L}$ , *path-length*  $\approx 0.7$  cm). Absorbance values were recorded at 25 °C. 4-Nitrophenyl acetate (4-nPA) was prepared by diluting the solid in acetonitrile up to 100 mmol/L concentration. For peptide-zinc assays, the peptide was first added and equilibrated in buffer solution prior to addition of the metal. Afterwards, at least 1h of incubation was allowed before the beginning of assays to ensure that the peptide-zinc complex formation reached equilibrium. Control assays of the catalysed reaction in buffer were done with the addition of only 4-Nitrophenyl acetate (4-nPA) or together with 10  $\mu\text{M}$  ZnCl<sub>2</sub>. The final volume percentage of acetonitrile, used as co-solvent, was kept constant (5%  $V_T$ ) by adding variable amounts of 100 mmol/L 4-nPA and acetonitrile according to initial assay conditions. Formation of the product 4-nitrophenol or phenolate was followed by single-wavelength readings of absorbance at 400 nm (405 nm in plate-well assays) at fixed time intervals. Initial rates were measured until *ca.* 2-3% total product formation under pseudo-first order reaction conditions. After addition of 4-nPA and signal stabilization, the  $A_{400}$  value was used as  $A_{400}(0)$ . Second-order rate constant  $k_2$  (or turnover number) were calculated by fitting to a linear model. The data points in the plot figures correspond to average values from at least 2 replicate assays and error bars correspond to the standard error of the mean. Amidase and protease screening assays were tested under similar conditions.

**CD spectroscopy.** Far-UV circular dichroism (CD) assays were done in 1 mm path-length cells,  $V_T=300 \mu\text{L}$ . Spectra were baseline corrected to subtract buffer contributions in the “Far-

UV" region (203-280 nm). For variable temperature assays, spectra were obtained after 5 min equilibration, from 5 to 95 °C, in intervals of 10 °C and 2°C/min ramp. Fitting of the data was made to a 1:1 binding model described elsewhere<sup>48,49</sup> with the Levenberg-Marquardt algorithm.

**NMR spectroscopy.** <sup>1</sup>H nuclear magnetic resonance (NMR) assays were made in D<sub>2</sub>O 50mM NaCl at pH ≈7.5, adjusted by additions of concentrated NaOH and HCl solutions. Peptide stock solutions were added up to 150 μM (RD01) or 1 mM (RD02), followed by stepwise additions of 10.33 mM ZnCl<sub>2</sub> (V<sub>T</sub>= 550 μL). Spectra were recorded in the 10 to 60 °C temperature range using 3-(Trimethylsilyl)propanoic acid (48 μM) as reference (0 ppm).

## Results and discussion

### Computational Enzyme Design

The MSD approach developed in this work included two challenges: the first was to adequately model both first and second sphere interactions of metalloenzymes at the crafted active site, using common CED methods that rely on knowledge-based potentials to treat metal-protein interactions. The second was to account for the inherent flexibility of small protein scaffolds. In order to build a model active site, we focused on the stabilization of the most energetic state along the hydrolytic peptide-bond cleavage mechanism<sup>50,51</sup>, which in the case of MPs typically corresponds to the formation of a *gem-diol* intermediate upon nucleophilic attack of an activated water molecule to the zinc-bound peptide substrate.<sup>31-35</sup> For this we obtained a consensus description of first and second sphere interactions in 10 active sites from MP-inhibitor complexes of the MA(M) subclan, since in most cases these molecules are modelled after the hypothesized *gem-diol* intermediate (Figure 1 and Table S1).<sup>52</sup> The first sphere interactions consisted of the tetrahedral/trigonal bipyramidal coordination geometry of the metal bound to histidines (His<sub>1-3</sub>), while the second sphere included both the orientation of the catalytic glutamate (Glu<sub>cat</sub>) in respect to the metal ion, as well as the position of the hydroxide ion and carbonyl oxygen upon *gem-diol* intermediate formation (Figure 1a). A oxyanion hole was not included in the model since it is not conserved throughout the MA(M) subclan, varying in terms of exposure to solvent and identity of constituting residues.<sup>33,34,53,54</sup> The capped diAla, with N-terminal acetylation and C-terminal amidation (diAla), was chosen since it has only one target peptide bond and allows to sample side chain and backbone conformations (Figure 1b). The resulting model containing the abovementioned elements was termed MA(M):diAla in this work (Figure 1c).

The Rosetta software package was used to design peptide scaffolds using the *matcher* and *enzyme design* programs. As a MSD approach, NMR structures were used as inputs, whenever possible, to sample multiple backbone conformations. As a separate control, the model reproduced the active site geometry of the MA(M) subclan archetype, astacin (PDB ID 1QJ1), using native catalytic residues positions (not shown). The

generation of hits for the Sp1f2 peptide (PDB ID 1VA2) was limited, with only one MA(M):diAla placement containing small atomic fluctuations for a single NMR state (Figure 2). This highly constrained design may be attributed to the C10H sequence modification, corresponding to the bulkier His<sub>3</sub> in the hydrophobic core. The designed active site presented a mixed coordination to zinc: Ne<sup>2</sup> (H10, H23) and Nδ<sup>1</sup> (H27) atoms vs. Ne<sup>2</sup> (H23, H27) in native Sp1f2. The resulting hits were modified with an additional C5G modification to allow for a bulk water molecule to coordinate with the zinc ion. In this way, the native coordinating histidines were kept while the cysteine residues were redesigned, rendering the zinc ion more surface-exposed than in native scaffold. The Glu<sub>cat</sub> was designed by a T26E modification, thus recapitulating the HEXXH<sub>n</sub>H coordination motif found in native MPs. Additional M4T (β1) and R13T (β2) modifications were introduced, as threonine mutations in β-sheets are known to increase the stability of designed zinc-fingers.<sup>55</sup> Two additional sequence modifications appeared in the design stage, K12V or K12H, the latter being discarded in order to exclude additional zinc coordinating residues. The designed valine side chain interacted with the capped N-terminal of the diAla substrate. The final peptide sequence was termed RD01.

In addition to the redesign of a native metalloprotein, a set of 42 peptides and small-protein scaffolds available in the Structural Classification Of Proteins – extended (SCOPe) database<sup>56</sup> of protein structures were screened and designed. These scaffolds varied in terms of chain length (20-64 residues), architecture (all-α, α+β and all-β) and contained no metal sites or disulphide bridges, as summarized in Table 1 and described in detail on Table S4. Two stages of screening were done with the *matcher* program: the first – “Secondary Algorithm” - allowed for distorted His<sub>1-3</sub> geometries; the second – “Classical Algorithm” - only allowed for properly defined geometries. The number of tested conformations was typically ≈ 10<sup>9</sup> per modelled residue/substrate/scaffold position. A high number of hits were produced for the majority of screened scaffolds, indicating that MA(M):diAla can be geometrically crafted in smaller systems with reduced number of secondary structure elements. All hits obtained were designed with the *enzyme design* program, including the native astacin as a positive control (MP activity) and hits from RD01. This yielded 5494 designs, which correspond to all NMR states, sequence variants and alternative residue/ligand conformers of all hits (Supplementary List 1). Given the high number of designs, a detailed evaluation of each one was not pursued. The selection of the candidate scaffold consisted first in identifying which of the 16 selected Rosetta parameters could be used to best evaluate 5494 designs by principal component analysis (PCA, full details in ESI). The PC1-PC2 subspace encoded 42% of total variance in the dataset and contained two distinct clusters corresponding to astacin and remaining peptide/small-protein designs (Figure S1). The pairs of parameters that presented the highest loadings along the two clusters were (Figure 3a): *i*) Score<sub>total</sub> and sequence length L (*correlation*, *corr*= -0.52), interpreted as larger scaffolds being more favourable; *ii*) constraints *k* and Score<sub>E<sub>diAla</sub></sub> (*corr*= 0.27), interpreted as designs with higher constraints tending to present less favourable diAla

scores. The first decile of designs ( $DE_{P10}$ ) with lower  $Score_{total}$  (more negative) and lower  $\log k/Score_{diAla}$  was projected along the original subspace of pair  $i$ ) and a linear combination of pair  $ii$ ).<sup>‡</sup> This set consisted in designs from 11 scaffolds, with varying sequence length and predominantly an all- $\alpha$  architecture. Inclusion of RD01 results was made for comparison purposes, revealing a high positive  $Score_{total}$  and high  $\log k/Score_{diAla}$  combination. The scaffold which presented best  $Score_{total}/L$  values, even better than for native astacin, was HP35 (PDB ID 1UNC) - although with relatively higher  $\log k/Score_{diAla}$ .

Table 1 - Summary from the screening and design of peptide and small-protein scaffolds.

	Structures	Number of Scaffolds
SCOPE class $g,j,k$	34 NMR, 8 X-ray	42 (33 all- $\alpha$ , 6 $\alpha+\beta$ , 3 all- $\beta$ )
	Secondary algorithm 279	35
Matcher	Classical Algorithm 122 (542 hits)	27
	DE 5494	27 (20 all- $\alpha$ , 4 $\alpha+\beta$ , 3 all- $\beta$ )
Enzyme Design	$DE_{P10}$ 555	11 (8 all- $\alpha$ , 2 $\alpha+\beta$ , 1 all- $\beta$ )

We decided to further explore the HP35 scaffold since it is a well-known model of protein folding by both computational<sup>57,58</sup> and experimental approaches.<sup>59-63</sup> It turned out to be quite “designable” since five additional sequence variants were obtained, although not present in the  $DE_{P10}$  set. Those included in the  $DE_{P10}$  set correspond to two sequence variants for the designed His<sub>1</sub> residue, F6H and F17H. The remaining His<sub>2</sub>, His<sub>3</sub> and Glu<sub>cat</sub> residues were designed at identical positions in the scaffold. Both variants presented at least one sequence modification for the highly conserved F6, F10 and F17 residues, whose hydrophobic collapse is attributed to be major driving force of HP35 folding.<sup>60</sup> Substitution of these phenylalanine residues for leucine is known to lead to significant destabilization of the scaffold, with F17L leading to the most destabilizing variants (substitution for other amino acids leads to unfolded or misfolded variants).<sup>60</sup>

Hits obtained for the HP35-F6H displayed coordination to zinc by the  $N\epsilon^2$  atoms of His<sub>1-3</sub> residues, as in native MA(M) enzymes. Indeed, a higher number of hits was produced in comparison with RD01, corresponding to scaffolds with slight coordinate variations of designed residues and four different diAla conformers (> 4000 placements) over four NMR states. Regarding interactions with the substrate, the F6H variant presented diAla in an extended conformation and placed along the scaffold surface, while in the F17H variant it was positioned closer to the first  $\alpha$ -helix in a bent conformation. Given the least destabilizing sequence modification and diAla placement resembling those found in protease-substrate complexes<sup>64</sup>, the F6H variant was selected. Designs from the F6H variant yielded 10 sequence modifications with more than one proposed amino acid identity (Figure 3b). This highlights the importance of employing a MSD approach, since the different amino acids proposed resulted from the design of more than on NMR state.

Native residue identity was kept when it was outputted by the *enzyme design* program, such as F10, P14, R22 and L34. The proposed S18A was excluded since the native serine residue interacts through a hydrogen bond with native R22 residue and to avoid formation of poly-alanine sequence motifs. The F17A sequence modification was accepted over the F17L in order to decrease steric clashing with the metal centre. L1G and F35S were accepted to optimize interactions with diAla by reducing side chain size in the termini of  $\alpha_1$  and  $\alpha_3$ . Additional modifications were done to address the stability of the scaffold: D5A ( $\alpha_1$ ) to eliminate competing zinc binding residues in the vicinity of designed His<sub>3</sub>; K24M and N27A ( $\alpha_3$ ) to increase scaffold stability, since they are known to contribute for the formation of hyper stable HP35 variants.<sup>57</sup> The scores of designed sequence, termed RD02, remained practically unchanged since the initial screening through the design stage and after the additional D5A, K24M and N27A modifications ( $Score_{total}$  variation from -33.9 to -33.2 REU), suggesting that no destabilizing modifications were introduced during this step.

### Zinc binding and folding

The RD01 and RD02 peptides were experimentally characterized in terms of their affinity for zinc metal ion and stability of the respective peptide-metal complexes. Far-UV circular dichroism (CD) spectroscopy was used to directly monitor changes of secondary structure content of the peptides upon zinc addition and thus determine the affinity constants of the peptides for the metal ( $K_{ZnP,app}$ ) at pH 7.5 (Table 2).<sup>65,66</sup>

Table 2 – Zinc binding and structural properties of RD peptides and native HP35 at pH 7.5.

	SP1F2	RD01	HP35	RD02
$K_{ZnP,app} (M^{-1})^a$	n.a.	$2.39 \pm 0.31 \times 10^5$	n.a.	$2.54 \pm 0.20 \times 10^5$
$T_m (^\circ C)$	n.a.	$46.9 \pm 2.4^b$	$41.5 \pm 1.7^c$	$49.6 \pm 4.5^c$
$\Delta H_{Tm} (kcal/mol)$	n.a.	$-13.3 \pm 1.3$	$-15.6 \pm 1.6$	$-10.8 \pm 1.4$
$\Delta G_{25^\circ C} (kcal/mol)$	n.a.	$-0.90 \pm 0.10$	$-0.81 \pm 0.08$	$-0.82 \pm 0.10$

a. 10 mM HEPES 50 mM NaCl, pH 7.5 at 25 °C. b. 10 mM TRIS 50 mM NaCl, pH 8.0. c. 10 mM HEPES 50 mM NaCl, pH 7.5.

The RD01 peptide presented spectral signatures characteristic of random coil conformation, with a large negative ellipticity band at 204 nm (Figure 4a). Upon addition of zinc there was an increase of negative ellipticity at 222 nm, concomitant with a decrease of the 204 nm band and an isodichroic point at *ca.* 206 nm. The RD01-Zn complex showed larger ellipticity at 204 and 222 nm compared to the Sp1f2-Zn complex (Figure S2), suggesting increased helical content. Increased helical content has also been observed for other Sp1f2 variants lacking one coordinating cysteine residue, presumably resulting from increased flexibility of the backbone, which allows for extension of the  $\alpha$ -helix.<sup>67</sup> The RD02 peptide also adopted a random coil conformation, with a large negative band at 204 nm (Figure 4b). However, upon additions of zinc this band decreased, together with an increase in negative ellipticity at 222nm and an isodichroic point at 210 nm. This

pointed to a folded  $\alpha$ -helical structure for RD02-Zn complex given the two negative bands at 222 and 208 nm with a  $[\theta]_{222}/[\theta]_{208}$  of 0.91. The native HP35 peptide also presented the same ratio between band intensities although with increased band intensities under the tested conditions<sup>59</sup>. Upon metal additions, there was only a small overall decrease in ellipticity with no observed isodichroic point (Figure S3).

Data of both RD peptides was fitted to a 1:1 peptide-zinc complex formation model, yielding  $K_{ZnP,app}$  values in the  $10^5$  M<sup>-1</sup> range at pH 7.5. Binding of zinc was also monitored at different pH values (range 7.0 - 9.0) and the formation of peptide-zinc complexes was observed in all cases (Figure S4). While for RD02-Zn this corresponds to the introduction of a zinc-binding site, for RD01 the binding constant is more than 4 orders of magnitude lower than for other zinc-finger peptides.<sup>68</sup> A decrease in the affinity for zinc in the Sp1f2 scaffold was expected given the removal of cysteines, which tend to form stronger interactions with the metal ion. The (His)<sub>3</sub>-zinc coordination motif used in RD peptides is found in native zinc metalloenzymes and has been the focus of several design efforts. These include a small redesigned zinc-finger<sup>69</sup> and conotoxins<sup>70</sup>, designed coiled coils<sup>49,71</sup> and helix bundles<sup>17,72,73</sup>, engineered iron-containing proteins<sup>74</sup> and antibodies.<sup>75</sup> Affinity constants can span up to 4 orders of magnitude in the micro- to nano-molar range and do not correlate necessarily with scaffold size. Indeed, the metal binding constants of RD peptides are in the same range as that of a designed TIM barrel fold using a similar CED approach.<sup>76</sup> This is in contrast with the sub-nanomolar affinities found in native metalloenzymes such as carbonic anhydrase II<sup>77</sup>, putting into evidence the important role of fine-tuned first and second sphere interactions as well as the surrounding chemical environment.<sup>78</sup>

The thermal stability of RD peptides was evaluated by variable-temperature CD assays at pH 7.5. Both peptide-zinc complexes showed reversible folding and the data could be fitted to a two-state model, thus allowing to determine the temperature of melting ( $T_m$ ), and the free energy ( $\Delta G$ ) and enthalpy of folding ( $\Delta H_{Tm}$ ). Under the tested conditions, the Sp1f2-Zn complex presented no unfolding up to 95 °C (data not shown). However, in the case of the RD01-Zn complex (Figure 5a) there were clear spectral changes indicating unfolding of the peptide. Nonetheless, at the end-point temperature the CD spectrum did not resemble that of the free peptide, suggesting no release of zinc upon RD01-Zn unfolding despite the lower thermal stability in comparison to Sp1f2-Zn. In contrast to Sp1f2-Zn, the native HP35 peptide showed unfolding under the tested conditions (Figure S5). Most reports on villin headpiece subdomain have been done at mild acidic conditions where the peptide tends to be more stable and no unfolding is observed.<sup>59</sup> The RD02-Zn complex presented unfolding as well (Figure 5b), with spectra at 75 °C maintaining features of that obtained for HP35 albeit distinct from the free peptide at 25 °C. The derived  $T_m$  of RD02-Zn complex is higher than the one obtained for HP35 at pH 7.5 although both peptides present similar enthalpies and free energies of unfolding. Stability of the zinc complexes was also addressed in the presence of acetonitrile, since it was used as a co-solvent in catalytic assay due to substrates solubility

issues. No significant spectral changes were observed after addition of 5% acetonitrile (data not shown).

The similarity of physicochemical properties between RD01-Zn and RD02-Zn complexes shows the robustness of the employed MSD approach. While the RD01 peptide folds upon binding to the metal ion (metal-coupled folding) as in native Sp1f2 peptide, in the case of RD02 an additional design challenge was introduced, a metal-dependent folding is achieved in the scaffold of human HP35 whose native driving force of folding is the hydrophobic collapse.<sup>79</sup> Moreover, both peptide-zinc complexes converged to similar thermal stabilities although having distinct sequence length and architecture, *i.e.*  $\Delta G_{25\text{ °C}} \approx 0.9$  kcal/mol, which suggests a prevalent contribution of the designed zinc-centres to fold stability.

### Hydrolytic activity

The hydrolytic activity of RD peptides and the corresponding zinc complexes, as well as native HP35 peptide was screened and characterized using the chromogenic substrate 4-Nitrophenyl acetate (4-nPA) as model substrate (Figure 6 and Table 3). The measured  $V_{cat}$  values varied linearly within the range of tested 4-nPA and catalyst concentrations and therefore the corresponding  $k_2$  values were calculated. Native Sp1f2-Zn peptide was not tested since it has been reported no hydrolytic activity towards 4-nPA under similar experimental conditions.<sup>10</sup> In microplate assays at pH 7.5 and with low excess of peptide in relation to zinc (1:2 peptide-to-metal ratio), RD01 and RD02 peptides presented higher  $k_2$  (turnover number) than HP35 and control zinc ion in both free and metal-bound assays (Figure 6a). To minimize contributions from the free peptide, and to ensure the role of zinc-complex forms as active catalysts, the assays for RD01-Zn and RD02-Zn were repeated in cuvette format with a higher peptide-to-metal ratio (1:4, Figure 6b). Under these conditions, the metal-bound fraction was higher than 90% by considering the previously determined  $K_{ZnP,app}$  values of both peptide-zinc complexes. The free peptides presented higher  $k_2$  values than the zinc-complex forms, with RD01 displaying slightly higher values than RD02 in both cases (Table 3). The approximately 2-fold difference between the two forms indicates that the zinc-complex is an active catalyst, given that at the assay conditions only 10% of the free peptide is present. The  $k_2$  values increased with higher pH for both RD-Zn complexes. However, for the free peptides no clear trend in  $k_2$  values was observed (Figure 6c).

Table 3 – Hydrolytic activity of free RD peptides and corresponding Zn-complexes at pH 7.5.

	SP1F2	RD01	HP35	RD02
$k_2$ (s <sup>-1</sup> ) peptide	0.41±0.01 <sup>a</sup>	0.29±0.01 <sup>b</sup>	0.02±0.02 <sup>c</sup>	0.23±0.00 <sup>b</sup>
$k_2$ (s <sup>-1</sup> ) Zn-complex	0 <sup>a</sup>	0.15±0.00	0.01±0.02	0.12±0.01

a. values obtained in ref 10. b. 10 mM HEPES 50 mM NaCl, pH 7.5 at 25 °C. b. Values obtained in microplate assay, 5  $\mu$ M peptide and 1:2 peptide-zinc ratio.

The obtained  $k_2 \approx 0.1$  s<sup>-1</sup> for 4-nPA hydrolysis at pH 7.5 are comparable with other redesigned zinc fingers, including Sp1f2

variants with three histidine residues and the CP1-Zn consensus peptide not specifically designed towards 4-nPA hydrolysis.<sup>10,11</sup> In the case of RD01, the employment of MSD with the MA(M):diAla model did not lead to improvements when compared to other Sp1f2 variants. Indeed, RD-Zn complexes present similar values as the organic complex cyclen-Zn, suggesting only modest contributions of the introduced Glu<sub>cat</sub> residue.<sup>80</sup> These values are one order of magnitude lower than those of the BBA-B3 zinc-finger peptide designed specifically towards the hydrolysis of 4-nPA<sup>8</sup> and up to two orders of magnitude in the case of designed protein-zinc complexes with higher structural complexity, such as the coiled coils<sup>14</sup>, calmodulin variants<sup>81</sup>, assemblies of helical dimers<sup>13,82</sup>, tetramers<sup>15,83</sup> and amyloid-forming heptapeptides.<sup>9</sup> With more complex structures the network of interacting residues increases, leading to the possibility of establishing favourable protein-substrate interactions or proper activation of the catalytic species.<sup>49</sup> Designed hydrolases are nonetheless still far from the efficiencies of native systems, such as in the case of Carbonic Anhydrase II<sup>16</sup>, whose rates of hydrolysis towards 4-nPA can be up to 4 orders of magnitude higher than the ones obtained for RD peptides. Indeed, both RD01-Zn and RD02-Zn presented relatively constant hydrolytic activity levels across the range of pH values tested, showing only a clear increase at pH 9.0 which suggests that zinc-bound hydroxide ions act as the active catalytic species rather than the designed glutamate residue.<sup>84</sup>

### NMR spectroscopy

In order to obtain more insights to rationalize the catalytic activities obtained for the zinc complexes of the RD peptides, their structural properties were addressed by NMR spectroscopy. A complete structural elucidation of RD01-Zn and RD02-Zn complexes was not possible by <sup>1</sup>H-NMR spectroscopy due to extensive signal overlap and broadening. This was attributed to both nuclear relaxation phenomena and zinc binding/release processes occurring in the millisecond regime, leading to the overlap of signals originated from the free and metal-bound peptide and possibly additional intermediate states. We therefore focused our analysis on specific features of the peptide-zinc complex formation.

In the case of RD01-Zn (Figure 7) a new peak in the aliphatic region appeared at 1.05 ppm upon zinc addition, which corresponds to the methyl side chain groups of L20 residue and thus indicates the formation of  $\alpha$ 1 (Figure 7a). The data could be fitted to a 1:1 peptide-zinc complex with a  $K_{ZnP,app}$  of  $1.41 \times 10^5 \text{ M}^{-1}$  (Figure S6). This value is in accordance with the values determined by CD spectroscopy. Signal broadening was quite evident in the NH region, suggesting major backbone readjustments upon metal binding. The W7 signal at 10.0 ppm presented broadening and splitting upon complex formation, suggesting multiple states of the  $\beta$ 1/ $\beta$ -turn interface where this residue is located. Temperature effects indicate the presence of a single state only at low temperatures where scaffold flexibility is restrained (Figure 7b). New signals associated with residues in  $\beta$ -sheet conformation were observed for higher

temperatures in the H $\alpha$  spectral region concomitant with new signals in the aliphatic region, pointing to major fold readjustments close to the determined  $T_m$ .

For RD02-Zn complex (Figure 8), W23 showed a characteristic sharp signal at 10.25 ppm corresponding to the free peptide in solution. At sub-equimolar zinc additions this signal is shifted and broadened, corresponding to either a transient species or a dimer. Concomitant with this was the appearance of a new broad signal that did not reach full intensity up to 1.75 molar excess of zinc, and may correspond to the zinc-complex. Determination of  $K_{ZnP,app}$  could not be accessed in the case of RD02 and therefore the existence of multiple of oligomeric states cannot be ruled out.

Overall, the NMR data could not provide complete information on the solution structures of RD zinc complexes, and therefore we turned our focus on their assessment by simulation methods, described in the next section.

### Molecular dynamics simulations

Additional structural insights of the RD-Zn complexes were obtained by MD simulations in the microsecond time-scale. With this method, conservation of the catalytic interactions introduced during the design stage could be further probed in light of the intrinsic dynamical features of the designed scaffolds.<sup>85,86</sup> To probe metal-protein interactions we employed the Amber force field parameters for the Cationic Dummy Atom (CaDA) approach, where a non-bonded description of the zinc metal ion is made by the inclusion of charged virtual particles that mimic the orientations of the unoccupied 4s4p<sup>3</sup> orbitals of the closed zinc 3d<sup>10</sup> system.<sup>45,87</sup> This method has been used in simulations of metalloproteins, including native zinc metalloenzymes, and has shown to reasonably capture the structural and electrostatic effects involved in metal-protein interactions, including ligand-exchange events.<sup>88–90</sup> Simulations of native astacin (350 ns) were used as a control to probe active site geometries.

Cluster analysis of the simulations revealed that RD01-Zn and RD02-Zn complexes diverged from the original modelled structures and adopted multiple conformations with major backbone rearrangements (Figure 9 and Table S6). In the case of the RD01-Zn complex, the  $\beta$ -turn region adopted multiple conformations (Figure 9a), consistent with the NMR results. The designed sequence modifications M4T and R13T led to disruption of  $\beta$ 1 and  $\beta$ 2 secondary structure elements, which adopted mixed turn/coil configurations. This is in contrast with the native Sp1f2-Zn complex, where only minor readjustments of the C-terminal region occurred over the entire  $\beta\beta\alpha$  fold. For HP35 there was only small changes in the  $\alpha\alpha\alpha$  fold (Figure 9b), with readjustments occurring mostly at  $\alpha$ 1 in the top populated clusters. On the other hand, RD02-Zn complex presented major disruptions of  $\alpha$ 1, helical reconfigurations in  $\alpha$ 2 and partial disruption of  $\alpha$ 3 towards the C-terminal. The penalizing effect of removing the highly conserved phenylalanine residues was clear: the F6H replacement led to disruption of  $\alpha$ 1 and the F17A led to reconfiguration of  $\alpha$ 2. In contrast, the region near to the K24M and N27A sequence modifications remained relatively

stable throughout the trajectories. The simulation results indicate slightly less  $\alpha$ -helix content in the RD02-Zn complex than in HP35 as also observed in CD spectroscopy. This is in line with findings reported for the native HP24 stab structure, a HP35-derived peptide which lacked  $\alpha 1$  but still formed supersecondary structures resembling the native topology.<sup>91</sup> It is therefore acknowledged the small contribution of  $\alpha 1$  to the native fold stability. For native astacin, only local loop reconfigurations and small adjustments at the active site were observed in the 350 ns control simulations (Figure S7).

Active site conformations were probed by a low dimensional description of the subspace of residue-zinc interactions relevant for catalysis (Figure 10). Comparison with MA(M) consensus catalytic machinery, and in particular with astacin, proved to be useful in understanding how first and second sphere interactions were disrupted in RD-Zn complexes. First sphere interactions were partially conserved in both RD01-Zn and RD02-Zn, with His<sub>1</sub> and His<sub>2</sub> presenting only small deviations from the MA(M):diAla geometry given their location in conserved  $\alpha$ -helices. For RD01-Zn the His<sub>3</sub> geometry was entirely disrupted, which is in line with the multiple states observed by NMR for the  $\beta$ -turn where this residue is closely positioned. In RD02-Zn complex, the geometrical interactions were nonetheless close to the ones obtained for astacin, which maintained His<sub>1-3</sub> distances close to the MA(M):diAla geometry. In both zinc complexes, second sphere interactions were particularly affected, since the Glu<sub>cat</sub> residue drifted away from the modelled MA(M):diAla distances. While in astacin the Glu<sub>cat</sub> is constrained by nearby secondary structure elements which limit major distance fluctuations, in RD01-Zn and RD02-Zn complexes this residue is solvent-exposed and therefore free to sample a higher number of conformations. Given its high flexibility, it is unlikely that this residue would form proper hydrogen-bond networks with the approaching substrate and the zinc-bound water molecule, presumed to be required upon transition-state formation.

The MD simulations highlight some important aspects found during the design stage. Proper MA(M):diAla geometry was reproduced in only 1 out of 31 NMR states for Sp1f2/RD01, while in HP35/RD02 the number raised to 4 out of 25. Upon simulation of the systems under conditions close to the experimental ones, the active site geometry was lost in RD01-Zn to a higher extent than for RD02-Zn. This supports the argument that RD02 is a less restrained design, which correlates with its more favourable Rosetta score.

Employment of the CaDA non-bonded model proved to be quite adequate since it reproduced realistic aspects of first sphere interactions, such as internal rotations of the metal-dummy particles and switch of coordinating atoms throughout the trajectory. In the case of astacin, the His<sub>3</sub> switched coordination to zinc between N $\epsilon^2$  and N $\delta^1$  atoms by rotation of the imidazole ring. The free zinc-coordinating position was occupied by bulk water molecule and exchange phenomena were observed in the ns timescale. Moreover, for systems with sub-micromolar affinities for zinc (Sp1f2 and astacin) the first coordination sphere remained stable. Ligand interchange between first and second coordination sphere also occurred, as in the case of Glu<sub>cat</sub>-zinc distance shortening observed in

astacin. Nonetheless, this method still has limitations, since the tetrahedral coordination geometry was kept throughout the trajectory and therefore other degenerate geometries were not sampled.<sup>92</sup> While the implementation of these features in current CED methods is not straightforward, it could be useful in the sequence optimization of RD peptides or other metalloprotein designs. Nonetheless, the incorporation of these MD simulations during the design stage could aid the identification of major conformational changes occurring upon incorporation of the metal binding site.

## Conclusions

In this work we show that multi-state design of hydrolytic metallopeptides can be achieved with the commonly used Rosetta enzyme design approach as in the case of villin headpiece subdomain – HP35. Our studies also reveal limitations on the current treatment of protein-metal interactions employed in CED approaches. We show how this can be partially overcome by incorporating information on long-timescale conformational dynamics by MD simulations using non-bonded zinc models during the design stage. The good agreement between the computational data and the experimental results is an indicator of the robust design strategy employed to develop catalytic metallopeptides, despite the catalytic efficiencies obtained for RD-zinc complexes still lagging behind those of native metal-dependent hydrolases.

The combination of multi-state design and molecular dynamics simulations can thus be extended to other metal-dependent systems and thus contribute to the development of successful strategies for the rational design of artificial metalloenzymes.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

OI and ACAR acknowledge the support from Centre national de la recherche scientifique (CNRS) and Fundação para a Ciência e a Tecnologia (FCT) through the Programme International de Coopération Scientifique – Project PICS-147340. This work was supported by Unidade de Ciências Biomoleculares Aplicadas, UCIBIO financed by national funds from FCT/MEC (UID/Multi/04378/2019) and co-financed by the ERDF under the PT2020 Partnership Agreement (POCI-01-0145-FEDER-007728). The authors are grateful to FCT for funding through the project ERA-IB-2/0001/2013, fellowships SFRH/BD/90644/2012 to HFC and SFRH/BPD/69163/2010 to RJFB. The authors also thank Spectropole (Aix-Marseille Université) and UniMS Mass spectrometry Unit (ITQB/IBET), Laboratório de Análises (REQUIMTE) and BioLab (UCIBIO).

## Notes and references



‡ A linear combination of parameters from pair ii) was chosen since these are directly correlated: an increase in one parameter reflects an increase in the other. The parameter  $k$  was converted to  $\log k$  for best dispersion of values.

- 1 J. Polaina and A. P. MacCabe, *Industrial Enzymes*, Springer Netherlands, Dordrecht, 2007.
- 2 Y. W. Lin, *Coord. Chem. Rev.*, 2017, **336**, 1–27.
- 3 C. Andreini and I. Bertini, *J. Inorg. Biochem.*, 2012, **111**, 150–156.
- 4 G. Kiss, N. Celebi-Ölçüm, R. Moretti, D. Baker and K. N. Houk, *Angew. Chem. Int. Ed. Engl.*, 2013, **52**, 5700–25.
- 5 A. Zanghellini, *Curr. Opin. Biotechnol.*, 2014, **29**, 132–138.
- 6 V. Muñoz Robles, E. Ortega-Carrasco, L. Alonso-Cotchico, J. Rodríguez-Guerra, A. Lledós and J. D. Maréchal, *ACS Catal.*, 2015, **5**, 2469–2480.
- 7 T. Heinisch, M. Pellizzoni, M. Dürrenberger, C. E. Tinberg, V. Köhler, J. Klehr, D. Häussinger, D. Baker and T. R. Ward, *J. Am. Chem. Soc.*, 2015, **137**, 10414–10419.
- 8 K. R. Srivastava and S. Durani, *PLoS One*, 2014, **9**, e96234.
- 9 C. M. Rufo, Y. S. Moroz, O. V. Moroz, J. Stöhr, T. A. Smith, X. Hu, W. F. DeGrado and I. V. Korendovych, *Nat. Chem.*, 2014, **6**, 303–309.
- 10 A. Nomura and Y. Sugiura, *Inorg. Chem.*, 2004, **43**, 1708–13.
- 11 A. N. Besold, L. R. Widger, F. Namuswe, J. L. Michalek, S. L. J. Michel and D. P. Goldberg, *Mol. BioSyst.*, 2016, **12**, 1183–1193.
- 12 D. Árus, N. V. Nagy, Á. Dancs, A. Jancsó, R. Berkecz and T. Gajda, *J. Inorg. Biochem.*, 2013, **126**, 61–9.
- 13 B. S. Der, D. R. Edwards and B. Kuhlman, *Biochemistry*, 2012, **51**, 3933–40.
- 14 M. L. Zastrow, A. F. a Peacock, J. a Stuckey and V. L. Pecoraro, *Nat. Chem.*, 2012, **4**, 118–23.
- 15 W. J. Song and F. A. Tezcan, *Science (80-. )*, 2014, **346**, 1525–1528.
- 16 A. Innocenti, A. Scozzafava, S. Parkkila, L. Puccetti, G. De Simone and C. T. Supuran, *Bioorganic Med. Chem. Lett.*, 2008, **18**, 2267–2271.
- 17 S. Studer, D. A. Hansen, Z. L. Pianowski, P. R. E. Mittl, A. Debon, S. L. Guffy, B. S. Der, B. Kuhlman and D. Hilvert, *Science (80-. )*, 2018, **362**, 1285–1288.
- 18 A. J. Leguto, M. A. Smith, M. N. Morgada, U. A. Zitire, D. H. Murgida, K. M. Lancaster and A. J. Vila, *J. Am. Chem. Soc.*, 2019, **0**, jacs.8b12335.
- 19 A. Leaver-Fay, M. Tyka, S. M. Lewis, O. F. Lange, J. Thompson, R. Jacak, K. W. Kaufman, P. D. Renfrew, C. A. Smith, W. Sheffler, I. W. Davis, S. Cooper, A. Treuille, D. J. Mandell, F. Richter, Y.-E. A. Ban, S. J. Fleishman, J. E. Corn, D. E. Kim, S. Lyskov, M. Berrondo, S. Mentzer, Z. Popović, J. J. Havranek, J. Karanicolas, R. Das, J. Meiler, T. Kortemme, J. J. Gray, B. Kuhlman, D. Baker and P. Bradley, *Methods Enzymol.*, 2011, **487**, 545–74.
- 20 A. Leaver-Fay, M. J. O’Meara, M. Tyka, R. Jacak, Y. Song, E. H. Kellogg, J. Thompson, I. W. Davis, R. A. Pache, S. Lyskov, J. J. Gray, T. Kortemme, J. S. Richardson, J. J. Havranek, J. Snoeyink, D. Baker and B. Kuhlman, in *Methods in Protein Design*, ed. A. E. Keating, Academic Press, 2013, vol. 523, pp. 109–143.
- 21 F. Richter, A. Leaver-Fay, S. D. Khare, S. Bjelic and D. Baker, *PLoS One*, 2011, **6**, e19230.
- 22 P. Löffler, S. Schmitz, E. Hupfeld, R. Sterner and R. Merkl, *PLOS Comput. Biol.*, 2017, **13**, e1005600.
- 23 M. Karimi and Y. Shen, *Bioinformatics*, 2018, **34**, i811–i820.
- 24 J. Ludwiczak, A. Jarmula and S. Dunin-Horkawicz, *J. Struct. Biol.*, 2018, **203**, 54–61.
- 25 N. D. Rawlings, A. J. Barrett and R. Finn, *Nucleic Acids Res.*, 2016, **44**, D343–D350.
- 26 A. Messerschmidt, R. Huber, T. Poulos and K. Wieghardt, *Handbook of Metalloproteins*, John Wiley & Sons, Ltd, Chichester, 2006.
- 27 O. A. Adekoya and I. Sylte, *Chem. Biol. Drug Des.*, 2009, **73**, 7–16.
- 28 J.-W. Wu and X.-L. Chen, *Appl. Microbiol. Biotechnol.*, 2011, **92**, 253–262.
- 29 X. Gao, K. Bain, J. B. Bonanno, M. Buchanan, D. Henderson, D. Lorimer, C. Marsh, J. A. Reynes, J. M. Sauder, K. Schwinn, C. Thai and S. K. Burley, *J. Struct. Funct. Genomics*, 2005, **6**, 129–134.
- 30 D. R. Durham, D. Z. Fortney and L. B. Nanney, *J. Burn Care Rehabil.*, 1993, **14**, 544–51.
- 31 M. R. A. Blomberg, T. Borowski, F. Himo, R. Liao and P. E. M. Siegbahn, *Chem. Rev.*, 2014, **114**, 3601–3658.
- 32 J. Blumberger, G. Lamoureux and M. L. Klein, *J. Chem. Theory Comput.*, 2007, **3**, 1837–1850.
- 33 S.-L. Chen, Z.-S. Li and W.-H. Fang, *J. Inorg. Biochem.*, 2012, **111**, 70–79.
- 34 T. Vasilevskaya, M. G. Khrenova, A. V. Nemukhin and W. Thiel, *J. Comput. Chem.*, 2016, **37**, 1801–1809.
- 35 V. Navrátil, V. Klusák and L. Rulišek, *Chem. - A Eur. J.*, 2013, **19**, 16634–16645.
- 36 S. Oka, Y. Shiraishi, T. Yoshida, T. Ohkubo, Y. Sugiura and Y. Kobayashi, *Biochemistry*, 2004, **43**, 16027–35.
- 37 W. Vermeulen, P. Vanhaesebrouck, M. Van Troys, M. Verschuere, F. Fant, M. Goethals, C. Ampe, J. C. Martins and F. A. M. Borremans, *Protein Sci.*, 2004, **13**, 1276–1287.
- 38 S. D. Khare, Y. Kipnis, P. Greisen, R. Takeuchi, Y. Ashani, M. Goldsmith, Y. Song, J. L. Gallaher, I. Silman, H. Leader, J. L. Sussman, B. L. Stoddard, D. S. Tawfik and D. Baker, *Nat. Chem. Biol.*, 2012, **8**, 294–300.
- 39 A. Zanghellini, L. Jiang, A. M. Wollacott, G. Cheng, J. Meiler, E. A. Althoff, D. Röthlisberger and D. Baker, *Protein Sci.*, 2006, **15**, 2785–2794.
- 40 R. F. Alford, A. Leaver-Fay, J. R. Jeliazkov, M. J. O’Meara, F. P. DiMaio, H. Park, M. V. Shapovalov, P. D. Renfrew, V. K. Mulligan, K. Kappel, J. W. Labonte, M. S. Pacella, R. Bonneau, P. Bradley, R. L. Dunbrack, R. Das, D. Baker, B. Kuhlman, T. Kortemme and J. J. Gray, *J. Chem. Theory Comput.*, 2017, **13**, 3031–3048.
- 41 M. J. O’Meara, A. Leaver-Fay, M. D. Tyka, A. Stein, K. Houlihan, F. Dimairo, P. Bradley, T. Kortemme, D. Baker, J. Snoeyink and B. Kuhlman, *J. Chem. Theory Comput.*, 2015, **11**, 609–622.
- 42 H. J. C. Berendsen, D. van der Spoel and R. van Drunen, *Comput. Phys. Commun.*, 1995, **91**, 43–56.
- 43 M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, *SoftwareX*, 2015, **1–2**, 19–25.
- 44 A. E. Aliev, M. Kulke, H. S. Khaneja, V. Chudasama, T. D. Sheppard and R. M. Lanigan, *Proteins Struct. Funct. Bioinforma.*, 2014, **82**, 195–215.
- 45 Y.-P. Pang, *J. Mol. Model.*, 1999, **5**, 196–202.
- 46 S. C. Gill and P. H. von Hippel, *Anal. Biochem.*, 1989, **182**, 319–26.
- 47 G. L. Ellman, *Arch. Biochem. Biophys.*, 1959, **82**, 70–77.
- 48 P. Thordarson, *Chem. Soc. Rev.*, 2011, **40**, 1305–1323.

- 49 M. L. Zastrow and V. L. Pecoraro, *J. Am. Chem. Soc.*, 2013, **135**, 5895–5903.
- 50 X. Zhang and K. N. Houk, *Acc. Chem. Res.*, 2005, **38**, 379–85.
- 51 M. Garcia-Viloca, J. Gao, M. Karplus and D. G. Truhlar, *Science (80-. )*, 2004, **303**, 186–95.
- 52 L. Englert, K. Silber, H. Steuber, S. Brass, B. Over, H.-D. Gerber, A. Heine, W. E. Diederich and G. Klebe, *ChemMedChem*, 2010, **5**, 930–40.
- 53 V. Pelmentschikov and P. E. M. Siegbahn, *Inorg. Chem.*, 2002, **41**, 5659–5666.
- 54 F. X. Gomis-Rüth, *Mol. Biotechnol.*, 2003, **24**, 157–202.
- 55 M. D. Struthers, R. P. Cheng and B. Imperiali, *J. Am. Chem. Soc.*, 1996, **118**, 3073–3081.
- 56 N. K. Fox, S. E. Brenner and J.-M. Chandonia, *Nucleic Acids Res.*, 2014, **42**, D304-9.
- 57 Y. Bi, J.-H. Cho, E.-Y. Kim, B. Shan, H. Schindelin and D. P. Raleigh, *Biochemistry*, 2007, **46**, 7497–7505.
- 58 S. Piana, K. Lindorff-Larsen and D. E. Shaw, *Proc. Natl. Acad. Sci.*, 2012, **109**, 17845–17850.
- 59 J. C. McKnight, D. S. Doering, P. T. Matsudaira and P. S. Kim, *J. Mol. Biol.*, 1996, **260**, 126–134.
- 60 B. S. Frank, D. Vardar, D. Buckley and C. J. McKnight, *Protein Sci.*, 2002, **11**, 680–687.
- 61 R. Godoy-Ruiz, E. R. Henry, J. Kubelka, J. Hofrichter, V. Muñoz, J. M. Sanchez-Ruiz and W. A. Eaton, *J. Phys. Chem. B*, 2008, **112**, 5938–5949.
- 62 S. Xiao and D. P. Raleigh, *J. Mol. Biol.*, 2010, **401**, 274–285.
- 63 C. J. McKnight, P. T. Matsudaira and P. S. Kim, *Nat. Struct. Biol.*, 1997, **4**, 180–184.
- 64 J. D. Tyndall, T. Nall, D. P. Fairlie and P. K. Madala, *Chem. Rev.*, 2010, **110**, PR1-31.
- 65 K. Suzuki, H. Hiroaki, D. Kohda, H. Nakamura and T. Tanaka, *J. Am. Chem. Soc.*, 1998, **120**, 13008–13015.
- 66 O. Sénèque, E. Bonnet, F. L. Joumas and J.-M. Latour, *Chem. - A Eur. J.*, 2009, **15**, 4798–4810.
- 67 A. Nomura and Y. Sugiura, *Inorg. Chem.*, 2002, **41**, 3693–3698.
- 68 O. Sénèque and J.-M. Latour, *J. Am. Chem. Soc.*, 2010, **132**, 17760–17774.
- 69 M. D. Struthers, R. P. Cheng and B. Imperiali, *J. Am. Chem. Soc.*, 1996, **118**, 3073–3081.
- 70 C. Vita, C. Roumestand, F. Toma, A. Ménez, A. Menez and A. Ménez, *Proc. Natl. Acad. Sci.*, 1995, **92**, 6404–6408.
- 71 T. Kiyokawa, K. Kanaori, K. Tajima, M. Koike, T. Mizuno, J.-I. Oku and T. Tanaka, *J. Pept. Res.*, 2004, **63**, 347–353.
- 72 V. M. Cangelosi, A. Deb, J. E. Penner-Hahn and V. L. Pecoraro, *Angew. Chemie Int. Ed.*, 2014, **53**, 7900–7903.
- 73 L. Regan and N. D. Clarke, *Biochemistry*, 1990, **29**, 10878–10883.
- 74 H. N. Müller and a Skerra, *Biochemistry*, 1994, **33**, 14126–14135.
- 75 J. T. Adams and J. A. Deweese, *J. Am. Chem. Soc.*, 1994, **53**, 1745–1747.
- 76 S. L. Guffy, B. S. Der and B. Kuhlman, *Protein Eng. Des. Sel.*, 2016, **29**, 327–338.
- 77 J. A. Hunt, M. Ahmed and C. A. Fierke, *Biochemistry*, 1999, **38**, 9054–9062.
- 78 T. Kochańczyk, A. Drozd and A. Krężel, *Metallomics*, 2015, **7**, 244–257.
- 79 F. Polticelli, *Biochem. Mol. Biol. Educ.*, 2001, **29**, 16–20.
- 80 M. Subat, K. Woinaroschy, S. Anthofer, B. Malterer and B. König, *Inorg. Chem.*, 2007, **46**, 4336–4356.
- 81 Y. S. Moroz, T. T. Dunston, O. V. Makhlynets, O. V. Moroz, Y. Wu, J. H. Yoon, A. B. Olsen, J. M. McLaughlin, K. L. Mack, P. M. Gosavi, N. A. J. van Nuland and I. V. Korendovych, *J. Am. Chem. Soc.*, 2015, **137**, 14905–14911.
- 82 K. S. Broo, L. Brive, P. Ahlberg and L. Baltzer, *J. Am. Chem. Soc.*, 1997, **119**, 11362–11372.
- 83 Y. Bai, Y. Ling, W. Shi, L. Cai, Q. Jia, S. Jiang and K. Liu, *Chembiochem*, 2011, **12**, 2647–58.
- 84 X. Wang, R. Li, W. Cui, Q. Li and J. Yao, *Sci. Rep.*, 2018, **8**, 7042.
- 85 H. K. Privett, G. Kiss, T. M. Lee, R. Blomberg, R. a Chica, L. M. Thomas, D. Hilvert, K. N. Houk and S. L. Mayo, *Proc. Natl. Acad. Sci.*, 2012, **109**, 3790–3795.
- 86 S. Osuna, G. Jiménez-Osés, E. L. Noey and K. N. Houk, *Acc. Chem. Res.*, 2015, **48**, 1080–1089.
- 87 Y. P. Pang, K. Xu, J. E. Yazal and F. G. Prendergas, *Protein Sci.*, 2000, **9**, 1857–65.
- 88 P. Oelschlaeger, R. D. Schmid and J. Pleiss, *Protein Eng.*, 2003, **16**, 341–350.
- 89 P. Oelschlaeger, R. D. Schmid and J. Pleiss, *Biochemistry*, 2003, **42**, 8945–8956.
- 90 J. G. Park, P. C. Sill, E. F. Makiyi, A. T. Garcia-Sosa, C. B. Millard, J. J. Schmidt and Y. P. Pang, *Bioorganic Med. Chem.*, 2006, **14**, 395–408.
- 91 H. G. Hocking, F. Häse, T. Madl, M. Zacharias, M. Rief and G. Žoldák, *Biophys. J.*, 2015, **108**, 678–686.
- 92 Q. Liao, S. C. L. Kamerlin and B. Strodel, *J. Phys. Chem. Lett.*, 2015, **6**, 2657–2662.

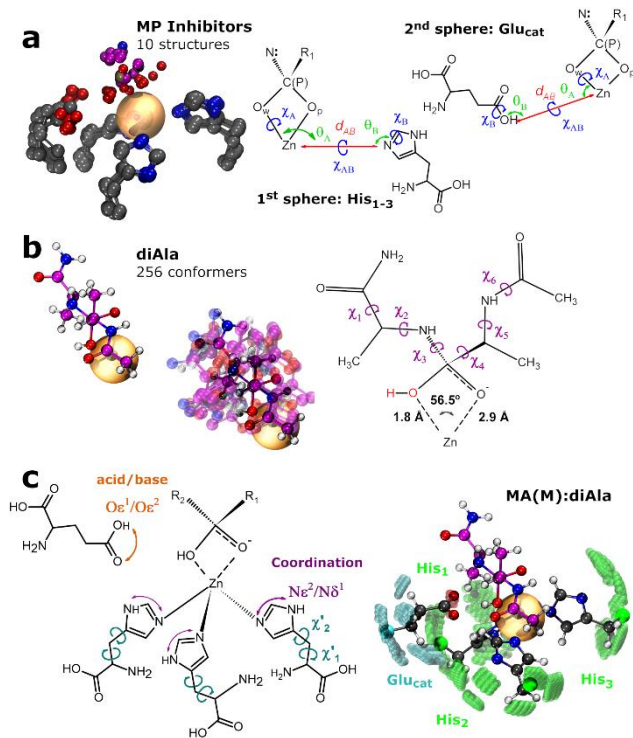


Figure 1 – (a) Conserved catalytic machinery of metalloproteases from the MA(M) subclan. Structural alignment of 10 active sites from MP-inhibitor complexes. Geometrical parameters defined for first sphere and second sphere interactions: one distance ( $d_{AB}$ ), two angles ( $\theta_A$  and  $\theta_B$ ) and three dihedrals ( $\chi_A$ ,  $\chi_{AB}$ ,  $\chi_B$ ) between protein, zinc and inhibitor atoms. (b) Modelling of diAla substrate model. Representative subset of 256 conformers in transparent representation used for clarity. (c) Distance and angles measured in the astacin-PFK complex (PDB 1QJ1). MA(M):diAla model. Interactions with more than one possible atom identified as coloured double arrows. Representative subset of alternative C $\alpha$  positions in coloured spheres.

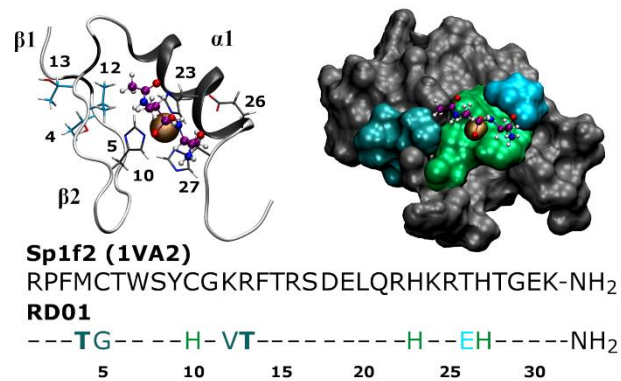


Figure 2 - RD01 model based on Sp1f2 scaffold. Sequence modifications from *matcher* step in green (1<sup>st</sup>-sphere) and cyan (2<sup>nd</sup>-sphere), from *enzyme design* step in turquoise and for stability increase in bold.

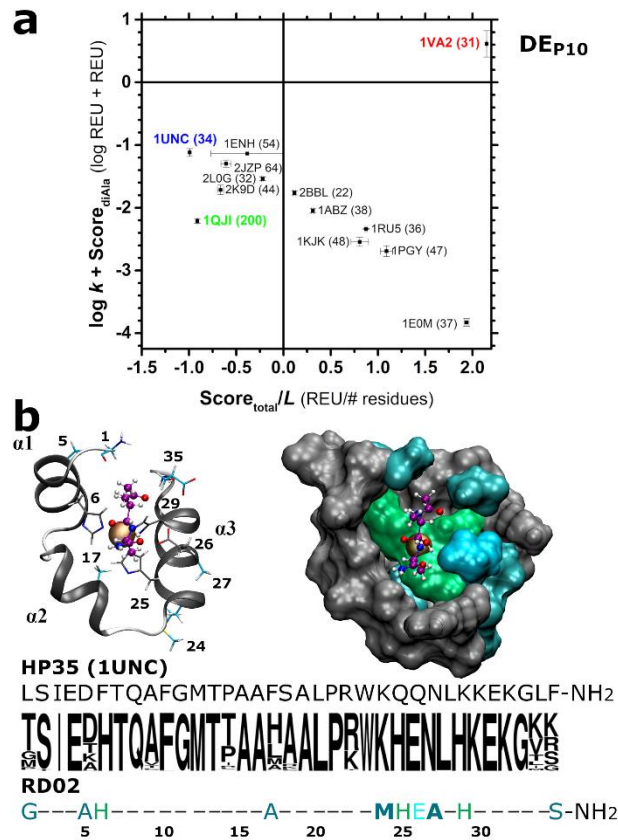


Figure 3 – (a) Projection of DEs along the subspace of parameters  $Score_{total}$ , chain length ( $L$ ), and linear combination of constraints ( $\log k$ ) and  $Score_{diAla}$ . Designs from RD01 represented in red and from astacin as green. Error bars corresponds to values from all DE designs. Labels corresponds to PDB identifier of input scaffold and chain length in parenthesis. (b) Best candidate scaffold 1UNC identified in blue. RD02 model (F6H variant) based on HP35 scaffold. Sequence logo of designs obtained during screening step. Consensus RD02 sequence after the design, coloured as described in Figure 2.

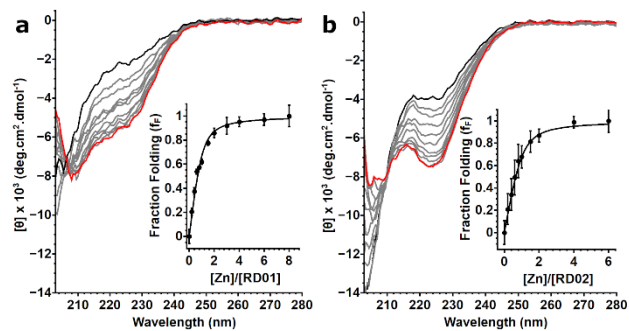


Figure 4 - Far-UV CD spectra obtained for the titration of  $ZnCl_2$  to (a)  $25 \mu M$  RD01 and (b)  $25 \mu M$  RD02 in 10 mM HEPES 50 mM NaCl at  $25^\circ C$ , pH 7.5. Black line corresponds to free peptide form, red line to end-point of titration and grey lines to intermediate additions of  $ZnCl_2$ . Insets: Corresponding fraction of folding upon addition of zinc. Solid lines correspond to fitted 1:1 complex formation model.

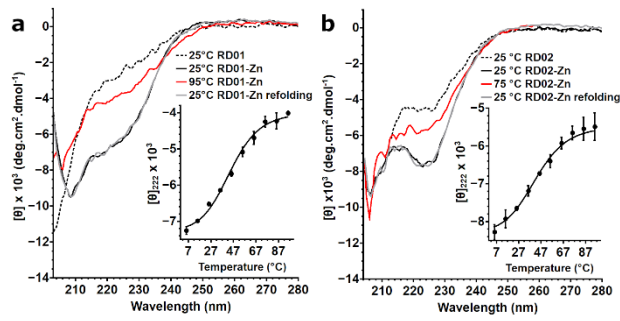


Figure 5 - Variable temperature spectra of (a) 25  $\mu\text{M}$  RD01-Zn in 10 mM TRIS 50 mM NaCl at pH 8 and (b) 25  $\mu\text{M}$  RD02-Zn in 10 mM HEPES 50 mM NaCl at pH 7.5. Insets: Corresponding ellipticity values at 222 nm as a function of temperature, solid lines corresponding to a two-state transition model.

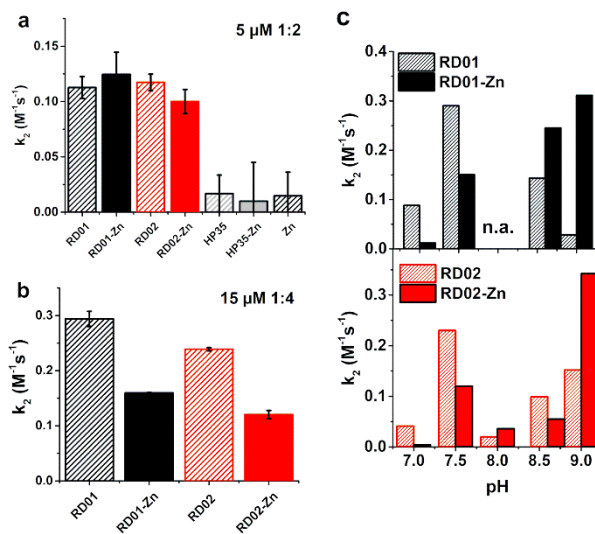


Figure 6 - (a) Second-order rate constants  $k_2$  for 5  $\mu\text{M}$  peptide and corresponding 1:2 zinc-complexes at room temperature. Control values obtained for 5  $\mu\text{M}$   $\text{ZnCl}_2$ . (b) Second-order rate constants  $k_2$  for 15  $\mu\text{M}$  peptide and corresponding 1:4 zinc-complexes at 25  $^\circ\text{C}$ . Peptide concentrations and peptide-zinc ratios used indicated in top right. Assays performed in a 0.25 to 2 mM 4-nPA concentration range. (c) Second-order rate constants  $k_2$  values obtained at different pH values at 25  $^\circ\text{C}$  (see methods and ESI for experimental details).

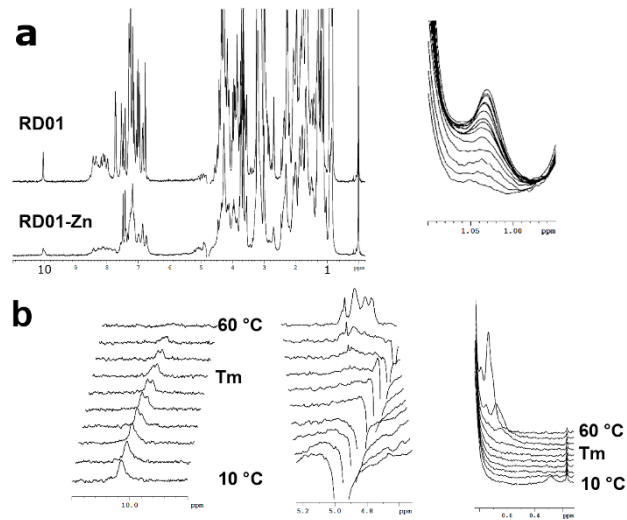


Figure 7 - (a) Spectra of 150  $\mu\text{M}$  RD01 before and after addition of 0-534  $\mu\text{M}$  of  $\text{ZnCl}_2$  at 25  $^\circ\text{C}$  (in 50 mM NaCl, pH 7.5). Right: Details of aliphatic region signal changes upon metal additions. (b) Temperature effects in 10 to 60  $^\circ\text{C}$  range for W7 on the left, H $\alpha$  region in the centre and aliphatic region on the right.

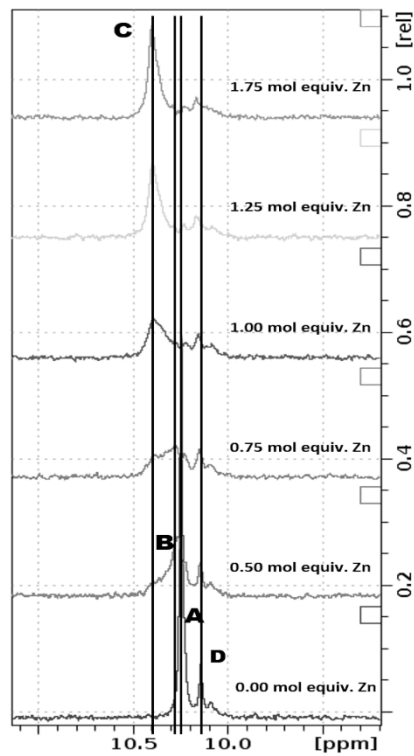


Figure 8 - Spectra showing the W23 signal of 1 mM RD02 peptide in 50 mM NaCl at 25  $^\circ\text{C}$ , pH 7.5 after 0-1.75 mM additions of  $\text{ZnCl}_2$ . Labels: (A) corresponds to the free peptide, (B) to a transient species at sub-equimolar Zn concentrations, (C) to the metal-complex form and (D) to a small impurity or negligible free peptide state.

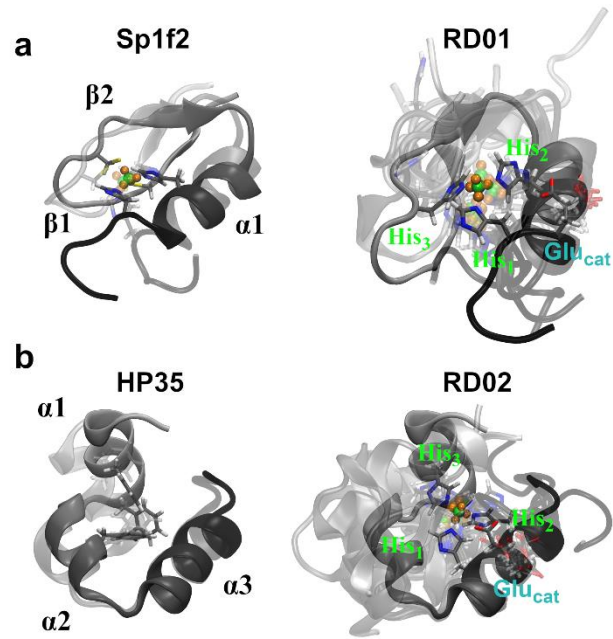


Figure 9 – Cluster analysis of MD simulations. Top populated clusters (>80% populated time) of (a) native Sp1f2-Zn and RD01-Zn and (b) native HP35 and RD02-Zn. Backbone in cartoon representation and colored based on residue index. Residues involved in metal interactions and folding shown in licorice, atoms from the CaDA model shown in spheres, green for zinc ion and orange for dummy atoms. Centroids of top populated clusters (transparent) were aligned in the  $\alpha 1$  (Sp1f2) and  $\alpha 3$  (HP35) regions to the initial conformation (solid).

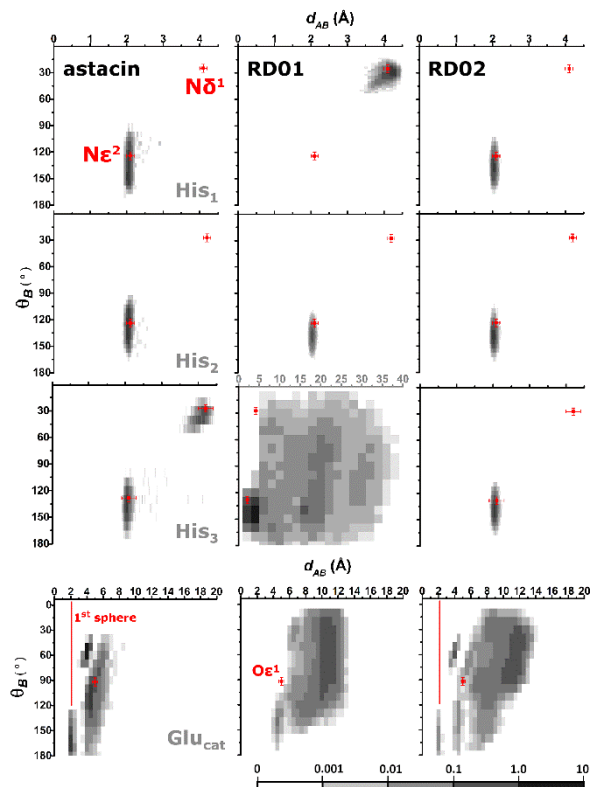


Figure 10 - Geometrical parameters of first sphere and second sphere interactions, corresponding to one distance ( $d_{AB}$ ) and one angle ( $\theta_B$ ) between residues and the metal ion. Values correspond to time percentage values for a total of 2  $\mu$ s (peptides) or 750 ns (astacin) from two replicate simulations. Data from MA(M):diAla model included for comparison.