PCCP

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal

Cite this: DOI: 10.1039/d1cp02536h

Insights on peptide topology in the computational design of protein ligands: the example of lysozyme binding peptides†

Cristina Cantarutti, *[a] M. Cristina Vargas, [b] Cedrix J. Dongmo Foumthuim, [ac] Mireille Dumoulin,[d] Sara La Manna,[e] Daniela Marasco,[e] Carlo Santambrogio,[f] Rita Grandori,[f] Giacinto Scoles,[a] Miguel A. Soler, [ag] Alessandra Corazza[a] and Sara Fortuna *[agh]

Herein, we compared the ability of linear and cyclic peptides generated *in silico* to target different protein sites: internal pockets and solvent-exposed sites. We selected human lysozyme (HuL) as a model target protein combined with the computational evolution of linear and cyclic peptides. The sequence evolution of these peptides was based on the PARCE algorithm. The generated peptides were screened based on their aqueous solubility and HuL binding affinity. The latter was evaluated by means of scoring functions and atomistic molecular dynamics (MD) trajectories in water, which allowed prediction of the structural features of the protein–peptide complexes. The computational results demonstrated that cyclic peptides constitute the optimal choice for solvent exposed sites, while both linear and cyclic peptides are capable of targeting the HuL pocket effectively. The most promising binders found *in silico* were investigated experimentally by surface plasmon resonance (SPR), nuclear magnetic resonance (NMR), and electrospray ionization mass spectrometry (ESI-MS) techniques. All tested peptides displayed dissociation constants in the micromolar range, as assessed by SPR; however, both NMR and ESI-MS suggested multiple binding modes, at least for the pocket binding peptides. A detailed NMR analysis confirmed that both linear and cyclic pocket peptides correctly target the binding site they were designed for.

Received 6th June 2021,
Accepted 16th September 2021

DOI: 10.1039/d1cp02536h

rsc.li/pccp

1. Introduction

The growing interest in peptides as potential drugs,[1–6] along with emerging technologies such as cell penetrating peptides and peptide–drug conjugates[3] and their use as modulators of protein/protein interactions,[7] calls for novel design strategies. Highly effective binders capable of capturing large biomolecules are typically optimized either experimentally[8,9] or computationally[10–14] by generating, screening, and selecting the best candidate out of a large number of possible solutions.

Computational tools are particularly appealing as *in silico* design allows promising candidates to be pre-selected for wet lab validation, thus resulting in a more time- and cost-effective selection process of potential hits.[15,16] Recent developments in this field include a number of evolutionary algorithms for peptide design based either on docking[17] or molecular dynamics (MD)[18,19] for the identification of short peptide sequences capable of binding small molecules and proteins.[10–12,20] These methods are capable of tailoring the binding affinity of a peptide towards optimal values by iteratively mutating its sequence and adapting its conformation to

[a] Department of Medicine, University of Udine, Piazzale M. Kolbe 4, 33100 – Udine, Italy. E-mail: cristina.cantarutti@uniud.it, s.fortuna@units.it
[b] Departamento de Física Aplicada, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional (Cinvestav), Unidad Mérida, Apartado Postal 73 "Cordemex", 97310, Mérida, Mexico
[c] Department of Molecular Sciences and Nanosystems, Ca' Foscari University of Venice, Campus Scientifico – Via Torino 155, 30172 Mestre, Italy
[d] Centre for Protein Engineering, InBios, Department of Life Sciences, University of Liege, Liege, Belgium
[e] Department of Pharmacy – University of Naples "Federico II", 80134, Naples, Italy
[f] Department of Biotechnology and Biosciences, University of Milano-Bicocca, Piazza della Scienza, Milan, Italy
[g] Italian Institute of Technology (IIT), Via Melen – 83, B Block, 16152 – Genova, Italy
[h] Department of Chemical and Pharmaceutical Sciences, University of Trieste, Via L. Giorgieri 1, 34127 Trieste, Italy

† Electronic supplementary information (ESI) available: Fig. S1: Identification and ranking of binding sites of HuL for conformation sampled along the 50 ns molecular dynamics trajectory; Fig. S2: Overlay of aromatic and aliphatic ^{1}H NMR spectra of peptide 140 at 0.2 mM and 0.6 mM; Fig. S3: CSP bar plots recorded at a protein : peptide ratio of 1 : 3 with peptide 410 and peptide 140. Fig. S4: Representative CSP of some outlier residues as a function of concentration of peptides 410, 140, and 368; Fig. S5. Blind docking results. See DOI: 10.1039/d1cp02536h

the chosen target. The binding affinity is generally evaluated through one or more scoring functions, which guarantee a fast, albeit rough, estimation of the expected binding affinity of the peptide/target complex.[21] While the evaluation of a single scoring function sufficed for the design of peptides,[10–13,17–20] for the design of larger systems – such as antibody fragments – a collection of scoring functions is best suited.[22,23]

The emergence of novel molecular modelling tools[10,11,24] allows addressing a burdening problem: the control of the target binding site. This aspect is particularly important for the design of binders targeting surface binding sites, which are often involved in the protein–protein interactions that regulate biological functions.[25] These sites are often considered ''undruggable'' by small molecule approaches, but can be explored by rational peptide design.[7] The computational design of novel peptides requires (i) a model for the target protein, ideally derived from an experimentally determined structure (NMR spectroscopy, X-ray crystallography) or, in its absence, a model built by homology, and (ii) an effective exploration of the peptide conformational space,[26] together with sequence optimization.[27] One open question is the correct peptide topology to employ for targeting sites of different nature. It could be argued that cyclic peptides would guarantee an entropic gain over linear peptides, due to the restricted conformational space assured by cyclization.[7] However, we will show that this is not always true and that the optimum peptide topology (linear or cyclic) depends on the location and the nature of the binding site chosen on the target. We will prove this point by designing peptides for different binding sites of a well-characterized system: human lysozyme (HuL).

Lysozyme has been an excellent model system as a target for (i) the development of novel computational methods,[22,28] including new methods for predicting binding affinities[29] or for free energy calculations,[30] and (ii) the study of drug/protein interactions.[31–34] Indeed, lysozyme possesses a druggable pocket, similar to human and egg proteins, and a large solvent-exposed area comprising the whole protein surface which can be assumed to contain a number of possible binding sites. In this study, we will explore both types of binding sites on HuL. For each site, we designed both cyclic and linear peptides to explore the effect of the binding site nature (pocket or surface-exposed) on the choice of ligand topology. The design was based on PARCE,[35] a software developed to optimize sequences and conformations of randomly generated peptides,[10,11] as well as larger binders,[23] towards a preselected site on a target protein. The idea was to explore the results obtained by choosing a random peptide with an arbitrarily chosen starting conformation close to the selected binding site, and drive the peptide sequence optimization without human intervention. Peptides were then screened using MD simulations to identify good candidates. The binding affinities of the selected peptides were evaluated by surface plasmon resonance (SPR) and electrospray ionization mass spectrometry (ESI-MS). The mechanism of recognition of two pocket binding peptides was structurally characterized using nuclear magnetic resonance (NMR).

2. Materials and methods

2.1 Materials

Unlabelled human lysozyme (HuL) was purchased from SIGMA. Peptides with 95% purity were purchased from ProteoGenix SAS (Schiltigheim, France), with the exception of the cyclic peptides 368 and 278 which were purchased from NovoPro Bioscience (Shanghai, China).

2.2 Protein preparation and binding site selection

The HuL structure with PDB code 1JSF[36] was first minimized using the steepest descent minimization method, then placed into a cubic box with a water layer of 0.7 nm and Na^+ Cl^- ions to neutralize the system, and a second minimization was performed. *NVT* and *NPT* equilibrations for 100 ps, followed by 50 ns *NPT* production run at 300 K, were performed as described below. 10 conformations were sampled at constant time intervals from the trajectory. The protein surface of each MD conformation was explored using the Peptimap[37] tool to identify possible peptide binding sites. Peptimap is a fast Fourier transform based grid-sampling method that allows exploring the protein target surface to identify druggable sites specific for peptides. The binding sites were grouped according to their location on the protein surface (see Section 3.1), and scored according to the rank provided by Peptimap in each exploration. The PPI mode, which is a minor modification in Peptimap to the scoring function that reduces the weight of a cavity term and thus optimizes the results on a protein–protein interaction test set, was activated.

2.3 Peptides design

A starting peptide CAAAAAAAAAAC, with an arbitrary selected conformation, was put close to the chosen binding site and PARCE[10,11,23] was used to evolve its conformation and sequence. PARCE is a Monte Carlo (MC) optimization method looking for sequences that best bind a target protein. Each MC step starts with a molecular complex associated with a binding score for the receptor–peptide pair. At each step, an amino acid of the peptide is replaced by another residue, randomly. The new system is minimized. Then, the binding conformation of the protein target and the mutated peptide is sampled. Here, with respect to the standard PARCE implementation, a short replica exchange MD (REMD) scheme was employed; the REMD configurations were clustered and the binding score of each cluster-representative conformation was assessed. It is worth noticing that the REMD exchange ratio was kept to a minimum, as the goal of this step was not to build a statistical ensemble at a certain temperature, but to generate conformations. If the computational resources were limited, this step could as well be replaced by a set of parallel MD – without exchanges. The score is then compared with the score of the structure before the point mutation on the peptide. The MC step ends with a decision on whether to accept or reject the mutation. The acceptance probability P_{acc} depends both on the energy change associated with the proposed mutation and on a parameter

T which modulates the strictness of acceptance probability itself as in a standard MC scheme:[38]

$$P_{acc} = \min[1, \exp[-(E_{new} - E_{old})/T]]$$

where E_{old} is the estimated binding affinity of the previous configuration, and E_{new} is that of the attempted mutation.

Each optimization consisted of three MC paths, run at three different T values and, thus, at three different acceptance probabilities. A parallel tempering scheme was implemented to allow for exchanges between the paths after each attempted mutation. Exchanges between two randomly selected paths were attempted after each step.

At each step of the design, a random mutation of one random peptide residue (excluding the cysteines at each extremity) was performed by using the AmberTools program.[39] After each mutation, the structure was fully relaxed by three successive minimizations: (i) a partial minimization only for the side chain of the mutated residue, (ii) a partial minimization for the mutated amino acid and nearest neighboring residues, and (iii) a global minimization. This procedure was followed by a REMD simulation with 8 *NVT* replicas at temperatures of 375, 391, 407, 423, 440, 458, 477 and 495 K (predicted exchange rate of 0.08, 40 exchanges in the simulated time, according to the webtool http://folding.bmc.uu.se/remd [40]), with all atomic bonds constrained by using LINCS algorithm,[41] and with the HuL backbone restrained to initial conformation by a harmonic potential with a force constant of 1000 kJ mol^{-1} nm^{-2}. Each replica was run for 1 ns, with a time step of 2 fs and an attempt of exchange every 2 ps. We clustered the peptide–protein samples obtained from all replicas by using the Daura method[42] as implemented in the g_cluster program (part of the GROMACS package) with a cutoff of 0.105 nm. We discarded clusters containing less than 10 structures for the scoring evaluation. The score of the peptide/protein complexes was estimated using AutoDock Vina.[43] The new configuration was accepted or rejected following the standard Metropolis rule, then an exchange between two randomly selected runs was attempted, and accepted or rejected following the standard parallel tempering scheme. After 500 PARCE steps, the lowest energy peptide conformations were selected. The solubility of the peptides was assessed with the online tool www.pepcalc.com.

2.4 Computational screening

Each peptide/HuL complex was minimized using the steepest descent method, then placed into a cubic box with a water layer of 0.7 nm and Na^+ Cl^- ions to neutralize the system, and a second minimization was performed. Four *NVT* equilibrations were performed followed by one *NPT* using the leap-frog Verlet integrator with a time step of 1 fs. A first equilibration of 25 ps was done by freezing both HuL and peptide. The temperature of the previously minimized system was raised from 0 to 100 K using velocity rescaling. A second equilibration of 50 ps was performed by keeping the temperature of solvent plus ions constant at 100 K and increasing the temperature of the protein and peptide from 0 to 200 K. In a third equilibration of 50 ps, we raised the temperature from 100 to 200 K and from 200 to 300 K for water plus ions and protein/peptide, respectively. In a fourth equilibration of 50 ps, the full system was equilibrated to 300 K. The fifth and last equilibration of 100 ps was done in *NPT* keeping the pressure constant to a reference value of 1 bar using the Parrinello–Rahman pressure coupling, while the temperature remained constant at 300 K. Production runs consisted of 50 ns long *NPT* simulations with a 2 fs time step. Selected systems were run for further 250 or 500 ns, as required. In all cases, configurations and energies were sampled every 10 ps. All simulations were performed with amber99sb-ildn force fields.[44] The systems were solvated using the tip4p water model.[45,46] The Particle Mesh Ewald summation accounted for long range electrostatic interactions. All the calculations as well as their analysis were performed using Gromacs-4.6.2.[47] The score of the peptide/protein complexes was estimated using AutoDock Vina.[43]

2.5 ESI-MS

HuL and peptides were suspended in 10 mM ammonium acetate pH 7 and mixed at different final molar ratios (17 : 0, 17 : 22 and 17 : 111 for peptide 410; 17 : 0, 17 : 25 and 17 : 128 for peptide 140). The resulting samples were incubated at room temperature for 15 min and then injected into a hybrid quadrupole – time-of-flight mass spectrometer (QSTAR Elite, AB Sciex, Framingham, MA, USA) equipped with a nano-electrospray ion source. Metal-coated borosilicate capillaries with medium length emitter tips of 1 μm internal diameter (Thermo Fisher Scientific, Waltham, MA, USA) were used to directly infuse the sample into the spectrometer. The following instrumental setting was applied: source heater OFF, declustering potential 80 V, ion spray voltage 1.1–1.2 kV and curtain gas-pressure 20 psi. Spectra were averaged over 1 min acquisition.

2.6 Surface plasmon resonance

Real time binding assays were performed at 25 °C using a Biacore 3000 Surface Plasmon Resonance (SPR) instrument (GE Healthcare). Human lysozyme was immobilized at 1095 RU, on a CM5 Biacore sensor chip in 10 mM sodium acetate at pH 5.5, by using the EDC/NHS chemistry, at a flow rate of 5 μL min^{-1} and an injection time of 7 min. Binding assays were carried out by injecting 90 μL of the analyte, at a flow rate of 20 μL min^{-1} using HBS buffer at pH = 7.4 with an association phase of 270 s, and a dissociation phase of 300 s. A regeneration of the sensor surface was performed with 20 μL of 10 mM NaOH. Each peptide was injected at the following concentrations: (i) 47.6, 68, 204, 340, 700 and 900 μM for peptide 140; (ii) 40, 100, 200, 300 and 500 μM for peptides 410 and 368; and (iii) 80, 200, 300, 400 and 500 μM for peptide 278. The experiments were carried out in duplicate. Kinetic parameters (k_{on} and k_{off}) were estimated assuming a 1 : 1 binding model and using the version 4.1 Evaluation Software (GE Healthcare).

2.7 Protein expression

Uniformly ^{15}N labelled HuL was expressed in *Pichia pastoris* and purified as previously described.[48] Protein purity was

>95% based on SDS-PAGE analysis and the molecular mass was confirmed by MS.

2.8 NMR experiments

NMR experiments were recorded at 310 K using a Bruker Avance spectrometer operating at 500 MHz (^{1}H). 1D ^{1}H spectra were acquired with 4096 data points, a spectral width of 16 ppm and 4096 scans. Water suppression was achieved by excitation sculpting scheme.[49] 2D [^{1}H, ^{15}N] HSQC spectra were acquired with 2048 and 400 points in the direct and indirect dimensions, respectively, and 32 scans, over spectral widths of 16 and 37 ppm in the ^{1}H and ^{15}N dimensions, respectively. NMR samples were prepared in 50 mM sodium phosphate buffer at pH 6.5 and contained 8% D_2O for locking purposes. The protein concentrations were 299 μM and 239 μM for the titration with peptide 410 and peptide 140, respectively. The data were processed with Topspin 2.1 and analyzed with Sparky.[50] Combined chemical shift perturbation (CSP) was calculated as $\Delta\delta$ (ppm) = $[(\Delta\delta_H)^2 + (\Delta\delta_N/6.5)^2]^{1/2}$ where $\Delta\delta_H$ and $\Delta\delta_N$ are the chemical shift variations for ^{1}H and ^{15}N, respectively.[51] The HuL assignment was based on the one deposited in the Biological Magnetic Resonance Data Bank (code 5130).

3. Results

3.1 Target preparation and binding site selection

To pinpoint suitable binding sites for peptide design, we first run 50 ns MD atomistic simulations of HuL (PDB ID 1JSF[36]) in full water solvent. We sampled 10 conformations along the generated MD trajectory and explored the protein surface using Peptimap[37] (Fig. S1, ESI†). We identified three recurrent sites (Fig. 1a): a pocket site (indicated as P site), followed by two surface sites labelled as R (named after ''right-side region'', Fig. 1b) and B (named after ''bottom-side'' region, Fig. 1c). Within the P site is located the known active site of the protein for polysaccharide hydrolysis,[52] while the surface sites R and B play no known physiological role.

We found that the binding site P (pocket) has the largest binding surface area that can reach up to 1042.116 Å^2 by involving about 20 residues of the protein pocket. This size allows hosting peptide chains as long as 12 residues, corresponding to a surface area range of 1000–2000 Å^2. The most recurrent surface site along the trajectory was the one labelled as ''B''. The binding site B has a surface area of 949.5 Å^2 involving up to 14 protein residues. This size is compatible with a 10-residue cyclic peptide, the surface of which ranges from 1000–1700 Å^2. Peptides were designed for the pocket site P and the surface exposed site B. For ease of comparison, we employed 10mer peptides capped with cysteines on both ends for all sites.

3.2 Generation of the peptide sequences

We started from two dodecapeptides composed by ten alanines sandwiched by two terminal cysteines. The cyclic variants contain a disulphide bridge between these two terminal Cys, while linear sequences have free thiol cysteines. The starting dodecapeptides, in either linear or cyclic form, were indicated as 0 (''zero'') in Table 1. Four different peptide/HuL starting configurations were assembled by placing a linear/cyclic starting peptide in the proximity of a P or B binding site on HuL.

To generate high affinity ligands, we run multiple PARCE-based optimizations of both cyclic and linear peptides by targeting either the pocket P site (Fig. 2a–c) or the solvent-exposed site B (Fig. 2d–g). PARCE was run with a single scoring function (AutoDock Vina[43]) and by using the parallel tempering scheme with three parallel Monte Carlo (MC) runs at T = 0.3, 0.6, and 0.9. Each optimization consisted of 500 steps, corresponding to 500 subsequent attempted mutations.

To assess the peptide-optimization procedure, we kept track of the binding score evolution along each MC path. In each PARCE run, three MC paths, each associated with a different

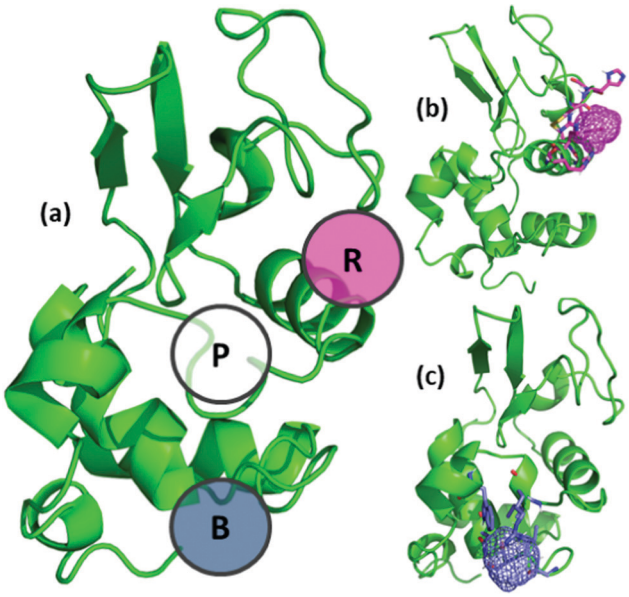

Fig. 1 (a) Suitable peptide binding sites on the HuL surface identified using Peptimap:[37] pocket (P), right site (R), and a bottom site (B), and two examples of the Peptimap output corresponding to the sites identified: R (pink) and B (blue). Molecular dynamics snapshots taken at (b) 20 ns and (c) 15 ns.

Table 1 Main features of the peptides analyzed in this study. Id indicates the step number in the mutation cycle. The binding sites are labelled as in Fig. 1: pocket (P), and surface exposed (B) sites; (me), *S*-methylated

Id	Topology	Binding site (label)	Sequence
0	Cyclic	P/B	CAAAAAAAAAAC (S-S between C1/C12)
0	Linear	P/B	CAAAAAAAAAAC
410	Cyclic	P	CPEYFEYWEQQC (S–S between C1/C12)
140	Linear	P	C(me)QNGKDFWSRWC(me)
368	Cyclic	B	CHEFDSLDFEYC (S-S between C1/C12)
278	Cyclic	B	CHEFDSLEFHYC (S-S between C1/C12)

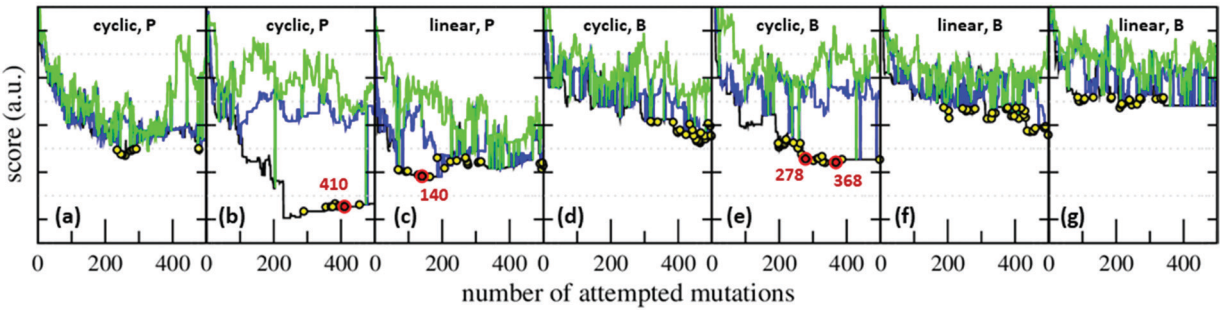

Fig. 2 PARCE runs for peptide optimization: evolution of the score along 500 MC optimization steps, paths at: T = 0.3 (black), T = 0.6 (blue), and T = 0.9 (green) for (a and b) cyclic peptides targeting the pocket site P, (c) linear peptides targeting the pocket site P, (d and e) cyclic peptides targeting the surface exposed site B, and (f and g) linear peptides targeting the surface exposed site B. Sequences corresponding to peptides selected for MD screening are marked (circles) and those tested experimentally are further highlighted (red). Scores are expressed in arbitrary units (a.u.).

acceptance probability, are clearly distinguishable (Fig. 2a–g). The highest path corresponds to the highest acceptance probability (*i.e.* T = 0.9). In the highest path (in green in Fig. 2), peptides are likely to mutate also towards unfavorable scores allowing for a thorough exploration of the mutational space. The lowest path is instead associated with the lowest acceptance probability (*i.e.* T = 0.3). As unfavorable mutations are quite unlikely in this path, the peptide is more likely to mutate towards more favorable scores. This kind of path can easily settle and get trapped, at a sub-optimum sequence. This is why exchanges with the other optimization paths, which are less strict, are necessary to overcome the trap. The exchange events are visible as vertical lines connecting the paths of the PARCE runs (for instance the green line at step 200 in Fig. 2b). An optimization was considered concluded when the path score settled to a constant value.

3.3 Computational screening

From each optimization run, we selected a set of sequences associated with the lowest scores in the MC optimization (highlighted by circles in Fig. 2a–g) and predicted to be water soluble. The associated peptide/HuL complexes underwent 50 ns of MD simulations in full water solvent to pre-screen all candidates. A number of parameters were monitored: (i) the peptide distance from the binding site, (ii) the number of hydrogen bonds between the peptide and HuL, (iii) the backbone root mean squared deviation (RMSD) of both the peptide and HuL, and (iv) the binding affinity score. Overall, a consistently low and stable binding affinity score turned out to be the only relevant indicator for a successful binder.

To compare several peptides, the binding affinity scores averaged along the MD trajectories can be effectively represented by histogram channels with error bars indicating their respective standard deviations, as reported in Fig. 3a and b for the pocket peptides. Surprisingly, this average value also correlated well with the optimization output (Fig. 2a–c), suggesting this short screening to be almost unnecessary for selecting binders for the pocket sites, such as site P. However, a longer MD trajectory of few well-behaved candidates turned out to be an important step to fully characterize the binders, especially those designed for the more labile surface exposed sites.

Peptides targeting the pocket site P. Among cyclic peptides, peptide 390 and peptide 408 had the lowest scores (Fig. 3a), but the large variation in their scores along the simulated time suggested they are too labile. Peptide 410, despite having a less favorable score (−19 a.u. to be compared with −24 a.u. for peptide 390 and −21 a.u. for peptide 408) better behaved as indicated by its score error bar of ±2 a.u. (to be compared with ±4 a.u. for both peptide 390 and peptide 408). Linear peptides, even if endowed with less favorable scores, presented smaller error-bars (Fig. 3b) and the one with the lowest score was chosen as optimum binder. This corresponds to peptide 140 (Table 1), associated with a score of −21 ± 2 a.u. Longer MD simulations, accounting for 250 ns, were run on peptides 410 and 140. The distance profile between the peptide centers of mass and their binding site along the trajectory (Fig. 3c and d) indicated that the cyclic peptide 410 changed its conformation along the simulated time, while peptide 140 maintained its position inside the pocket site. Simulation snapshots further revealed that peptide 410 binds to the pocket with one of its tyrosines (Y7) (Fig. 3e) but can also expand outside the pocket by anchoring itself to the HuL surface with the other tyrosine (Y4) (Fig. 3f). Instead, the linear peptide 140 kept its position close to the target binding site along the entire simulated time (Fig. 3g and h). While the linear peptide 140 kept its position on HuL binding site, it also showed a less favorable score with respect to the cyclic peptide, making it difficult to predict *a priori* which would be the optimum binder. Both peptides 140 and 410 underwent further experimental analyses.

Peptides targeting the solvent-exposed site B. The MD screening showed a distinct advantage of the cyclic peptides with respect to the linear ones to target the solvent exposed site B. The dotted line in Fig. 4a and b allows the scores reached by cyclic peptides (Fig. 4a) to be compared with the optimum score reached by linear peptides (−14 a.u. for peptides 106 and 338, Fig. 4b). The cyclic peptides are consistently predicted to bind HuL with a stronger affinity with respect to the linear peptides. However, longer simulations were required for an effective screening of the peptides targeting the solvent exposed site

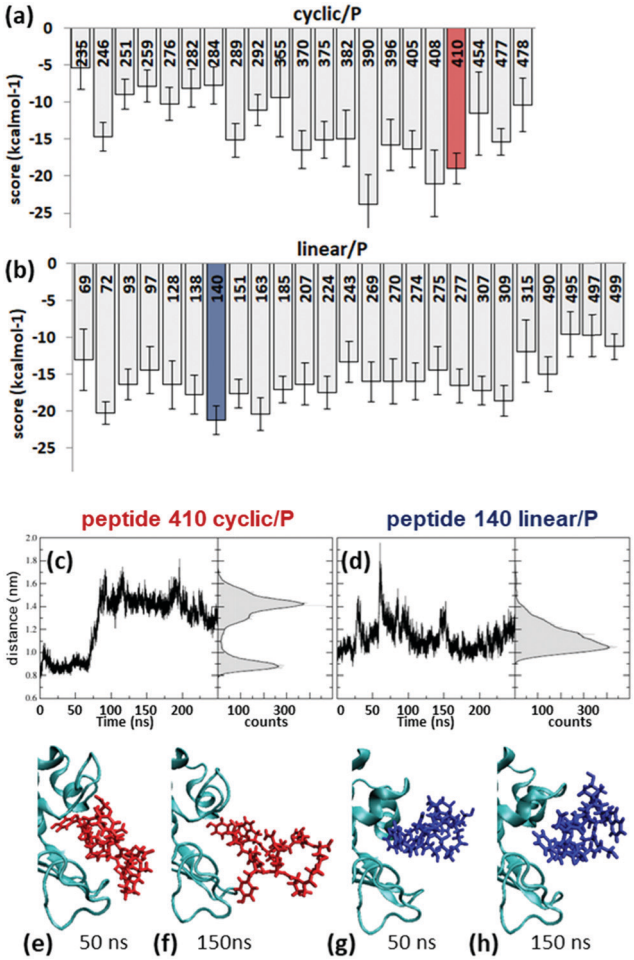

Fig. 3 Computational screening of the pocket peptides. (a and b) First screening: average scores along 50 ns MD simulations for (a) cyclic peptides and (b) linear peptides for the HuL pocket site; scores are expressed in arbitrary units; error bars are standard deviation. The two peptides selected for the second screening and subsequent experimental validation are highlighted. (c and d) Second screening: distance of the peptide from the binding site for the optimal peptides of each set and selected simulation snapshots for (e and f) the cyclic peptide 410 and (g and h) the linear peptide 140. Color code: HuL (cyan), cyclic peptide 410 (red), and linear peptide 140 (blue).

since they tended to leave their site at longer timescales. In fact, the optimal peptide designed for the solvent exposed site (*i.e.*, the cyclic peptide 313) initially selected after the first 50 ns long simulation, shows a high lability as confirmed by NMR (data not shown) and by longer MD simulations (Fig. 4c).

From a set of 50 ns long MD simulations for the solvent-exposed B site, we selected 15 peptides (marked by asterisks in Fig. 4a) that were further screened by longer MD simulations. In the additional 500 ns timeframe, several peptides left their binding sites (Fig. 4c), while seven sequences (259, 262, 278, 317, 329, 368, and 386) remained at their initial positions (Fig. 4d). The RMSD, which measures how much the backbone atoms of a protein change their position at each timestep with respect to their initial coordinates, showed that these seven peptides did not undergo major structural rearrangements (Fig. 4e). Their RMSD in the protein framework (thus calculated after HuL backbone alignment) also confirmed that they did not move away from their binding site throughout the whole simulation time (Fig. 4f). Among them, the peptide 278 was associated with the lowest score, as calculated along the whole trajectory, followed by the peptide 368 (Fig. 4g). Both peptides 278 and 368 were chosen for initial experimental characterization.

Overall, on the basis of computational studies, peptides 410, 140, 278, and 368 (Table 1) were chosen to be experimentally investigated by means of different techniques. While for optimization and screening we employed free cysteins, in all experiments, *S*-methylated residues were introduced to avoid disulphide bond formation in the linear peptides upon air oxidation.

3.4 Affinity measurements

SPR assays were carried out to evaluate the ability of the selected peptides to bind to HuL. All peptides exhibited a dose–response variation of RU intensity *versus* the peptide concentration (Fig. 5). The kinetic parameters derived from a global fitting of the curves, allowed the estimation of the dissociation constants (K_D) for peptide 410, peptide 368 and peptide 278. For peptide 140, too fast association and dissociation phases negatively affect the fitting of experimental curves, thus not allowing for the estimation of K_D. The 1 : 1 Langmuir equation was employed for the estimation of K_Ds as reported in Table 2. In all three cases, comparable K_D values in high micromolar range were estimated, in line with previous studies.[10–13]

3.5 Peptides binding mode by NMR titration and ESI-MS

All peptides, with the exception of peptide 278, were soluble under the conditions employed for NMR investigations and did not oligomerise, as thoroughly verified for peptide 410 (Fig. S2, ESI†).

^{1}H–^{15}N HSQC experiments were recorded using ^{15}N labelled HuL at increasing concentrations of the peptides with the peptide : protein ratios up to 3 for the peptides 410 and 368 and up to 5.6 for peptide 140. While with the peptides 410 and 140, a significant variation of the chemical shift ($\Delta\delta$) of some protein backbone amide peaks was observed (Fig. 6), with peptide 368, under the employed experimental conditions, a trend in the chemical shift deviations was observed (Fig. S3c, ESI†), although the maximum values reached were extremely small and comparable with the threshold of significance. A possible explanation for the disagreement between NMR and SPR might be due to a more favorable entropic factor when the protein is constrained on the chip surface of SPR in a position exposing the 368 binding site.

For peptide 140 and peptide 410, the presence (Fig. 6) of single shifted peaks, whose positions are the weighted average of the free and bound species, indicates that the exchange between the two states is fast on the NMR timescale.[53] In other terms, in both cases, the exchange rate, k_{ex}, of the peptide : protein complex formation is greater (at least 10 times) than

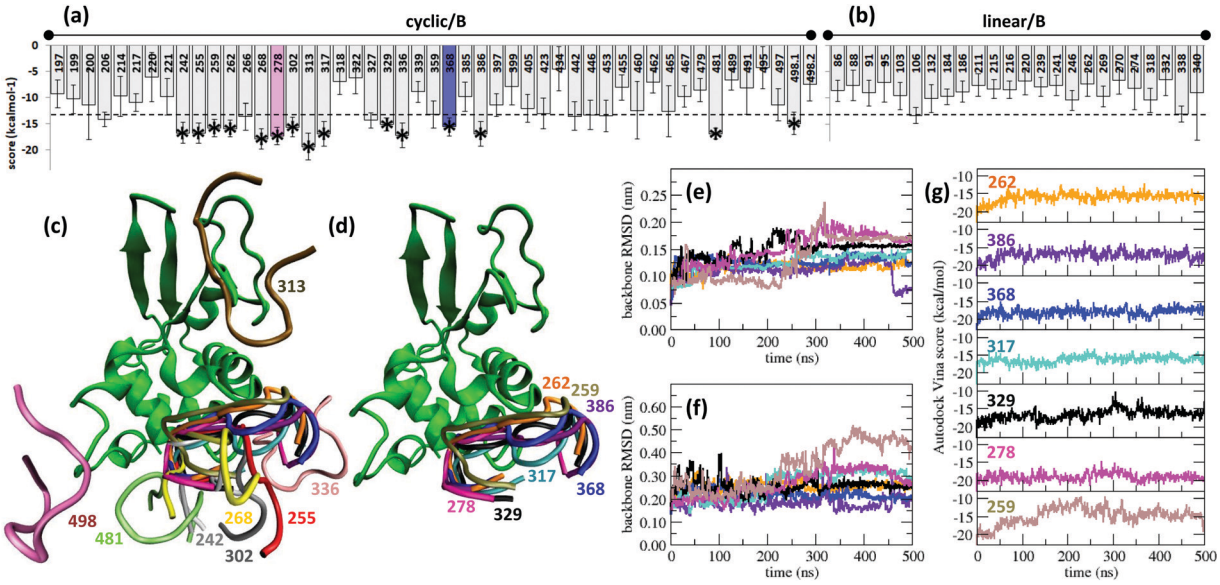

Fig. 4 Computational screening for the solvent exposed binding site. (a and b) First screening: average score along 50 ns MD simulations for (a) cyclic peptides and (b) linear peptides. Scores are expressed in arbitrary units, error bars are standard deviations, a dashed horizontal line marks the lowest average score achieved by the lowest scoring linear peptide. Peptides selected for the second computational screening are marked with an asterisk and peptides selected for experimental validation are further highlighted (the peptide 278 in pink, and the peptide 368 in blue). (c and d) Second screening: (c) simulation snapshot at 500 ns MD simulation of the 15 cyclic peptides selected from the first screening and (d) the same configurations for the subset of 7 peptides that did not leave their binding site. Color code: HuL (green) and peptides 313 (ochre), 242 (white), 255 (red), 259 (tan), 262 (orange), 268 (yellow), 278 (magenta), 302 (silver), 317 (cyan), 329 (black), 336 (pink), 368 (blue), 386 (purple), 481 (lime), and 498 (mauve). (e) RMSD of the peptide backbone with respect to the peptide initial configuration, and (f) in the frame of the HuL backbone, (g) scores of the 7 peptides that did not leave their binding site along the 500 ns long trajectories.

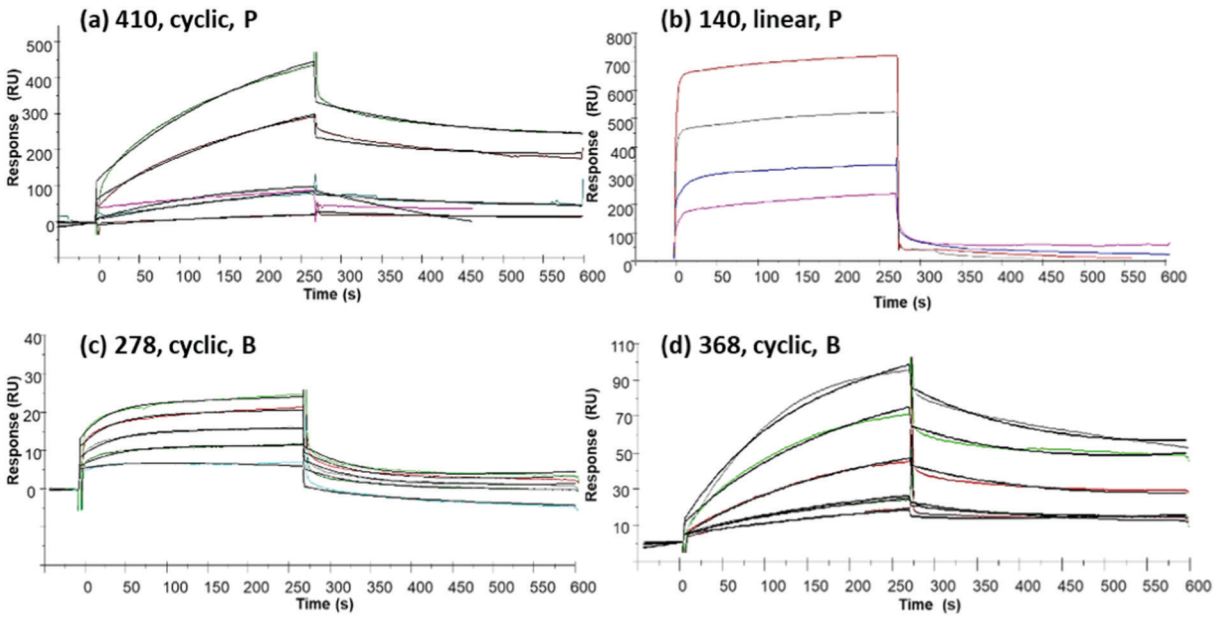

Fig. 5 Overlay of SPR sensograms showing the binding and dissociation of different peptides (a) 410, (b) 140, (c) 278, and (d) 368, to immobilized HuL. The colored lines show experimental measurements and black lines correspond to the fitting of the results (with the exception of panel B).

the difference of the chemical shift of the bound and unbound proteins ($\Delta\omega$ in rad s^{-1}).[53] In the hypothesis of a diffusion-limited ligand : protein interaction, the on-rate constant, k_{on}, is typically around 10^9 M^{-1} s^{-1};[54] therefore, the dissociation constant, K_D, can be approximated to $k_{off}/10^9$, where k_{off} is the off-rate constant in s^{-1}. Since in the fast exchange regime,

Table 2 SPR based kinetic parameters, equilibrium dissociation constants (K_D) for the interaction of HuL with peptides using the BIA evaluation v.4.1 software and 1 : 1 binding model. The dataset fit to the model, estimated by χ^2, is also indicated

Peptide	k_{on} (M^{-1} s^{-1})	k_{off} ($\times 10^{-3}$ s^{-1})	K_D (μM)	χ^2
410	10.6	1.7	160.4	4.5
278	25.7	4.6	179.5	7.0
368	15.4	2.6	165.9	4.7

$k_{ex} \gg \Delta\omega$ and the maximum $\Delta\omega^{1H}$ found with peptide 140 was 785 rad s^{-1}, we infer a lower limit for k_{ex} in the order of 8×10^3 s^{-1}. It can be shown that at high ligand concentrations, k_{ex} and k_{off} are of the same order of magnitude, leading to a K_D lower limit of about 8×10^{-6} M.[53] The same calculation performed with the proton maximum chemical shift perturbation (CSP) value obtained with peptide 410 gave a lower limit for K_D in the same order of magnitude.

The CSP obtained in the presence of peptide 140 is higher than that obtained in the presence of peptide 410; this observation suggests a larger perturbation of the protein residues when bound to the linear peptide (Fig. S3, ESI†). Instead, a plateau was reached earlier with peptide 410 (peptide : protein = 3) than with peptide 140 (peptide : protein = 5.6) indicating a lower affinity of the latter. Moreover, the chemical shift plots of the most perturbed residues in the presence of peptides 410 and 140 highlight a complex binding scheme, not consistent with a simple quadratic curve representing a 1 : 1 binding model (Fig. S4, ESI†). This result suggests the involvement of multiple binding modes or of cooperative binding; the latter is however less likely due to the relatively weak binding.

To further investigate the binding mode of peptides 410 and 140, the protein : peptide mixtures were analyzed by native MS (Fig. 7). This technique can capture non-covalent complexes by preserving weak interactions during an ionization/desolvation process conducted under mild conditions.[55–57] The spectra of the free HuL are similar to those previously reported and are typical of a well-folded globular protein.[58,59] New peaks, specific for the protein : peptide complexes, appear upon the addition of either peptide 410 or peptide 140, allowing a direct

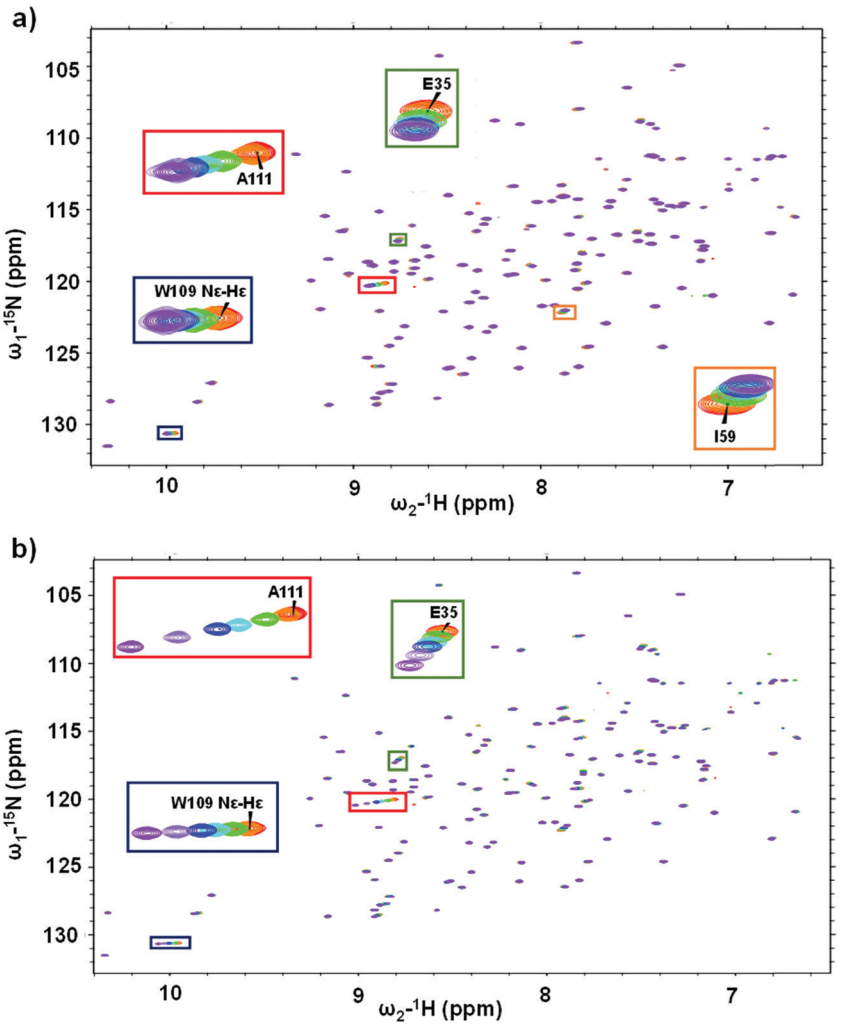

Fig. 6 Overlay of $^1H-^{15}N$ HSQC spectra of ^{15}N labelled HuL with (a) peptide 410 and (b) peptide 140 at peptide : protein ratios of 0 (red), 0.6 (dark orange), 1 (green), 1.6 (cyan), 2 (blue), 2.6 (medium purple) and 3 (purple). Some representative peak shifts are reported in the colored boxed insets.

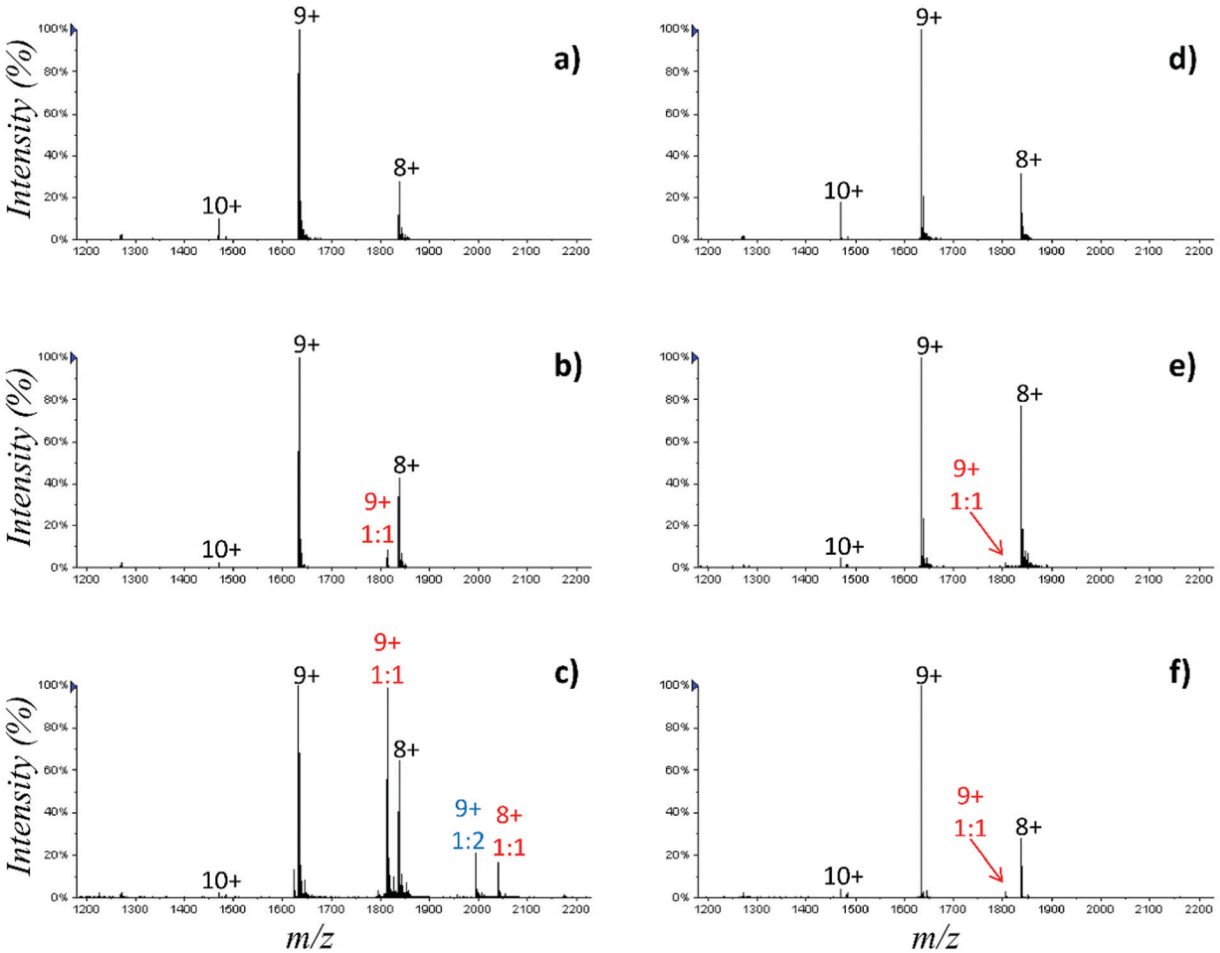

Fig. 7 Native ESI-MS results for 17 μM HuL with increasing concentrations of (a–c) the cyclic 410 and (d–f) linear 140 pocket peptides. The peptide : protein ratios employed are 0 (a and d), 1.3 (b), 6.5 (c), 1.5 (e) and 7.5 (f). Each protein peak is labelled by the corresponding charge state. Black labels correspond to the free protein; red labels to the 1 : 1 protein : peptide complex and blue labels to the 1 : 2 protein : peptide complex.

visualization of the supramolecular assemblies. Major differences are, however, observed between the two peptides. In first place, peptide 410 exhibits a higher apparent affinity, with more predominant complex-specific peaks than peptide 140 under similar conditions. This result is in line with the saturation curves obtained by NMR (Fig. S4, ESI†) as peptide 410 reached saturation at a lower peptide : protein ratio than peptide 140. Secondly, clear evidence of both 1 : 1 and 1 : 2 binding modes is offered by titration with peptide 410. Again, native MS is in agreement with NMR, indicating more complex schemes than the simple 1 : 1 stoichiometry and suggesting the existence of a secondary binding site. This conclusion, however, is supported only for peptide 410, since the low intensity of the HuL : 140 peaks does not allow to draw conclusions on possible signals of the 1 : 2 complex.

The ESI-MS result, which is in agreement with the NMR observations, is in disagreement with the SPR measures. Indeed, while the χ^2 value obtained by fitting the SPR data to the 1 : 1 binding model suggests that the majority of the peptide 410 bind in the ratio of 1 : 1 to the protein when the protein is immobilised on a surface; however, this does not hold true in solution where all accessible binding sites are available. Docking results[60,61] confirmed the presence of competing binding sites on the HuL surface (Fig. S5, ESI†).

3.6 Locating the binding site by analysing the NMR chemical shift perturbation (CSP)

We were able, based on the literature,[62,63] to fully assign the HSQC spectrum of apo-HuL and to identify the backbone amides perturbed the most by the addition of the peptides in the holo-HuL. The CSP distribution identifies several residues which are mostly involved in the binding of peptides (Fig. S3, ESI†). These residues can be grouped in three classes (Table 3): (i) $\Delta\delta > \Delta\delta_{av} + \sigma$, (ii) $\Delta\delta > \Delta\delta_{av} + 2\sigma$ and (iii) $\Delta\delta > \Delta\delta_{av} + 3\sigma$,

Table 3 Residues showing a significant chemical shift deviation $\Delta\delta$ with the addition of 3 equivalents of each peptide

	Peptide 410	Peptide 140
$\Delta\delta > \Delta\delta_{av} + 3\sigma$	A111, W109Nε-Hε	A111, W109Nε-Hε
$\Delta\delta > \Delta\delta_{av} + 2\sigma$	I59, R21	I59, E35
$\Delta\delta > \Delta\delta_{av} + \sigma$	E35, V100, W109	V100, W109, R113, R50, H78

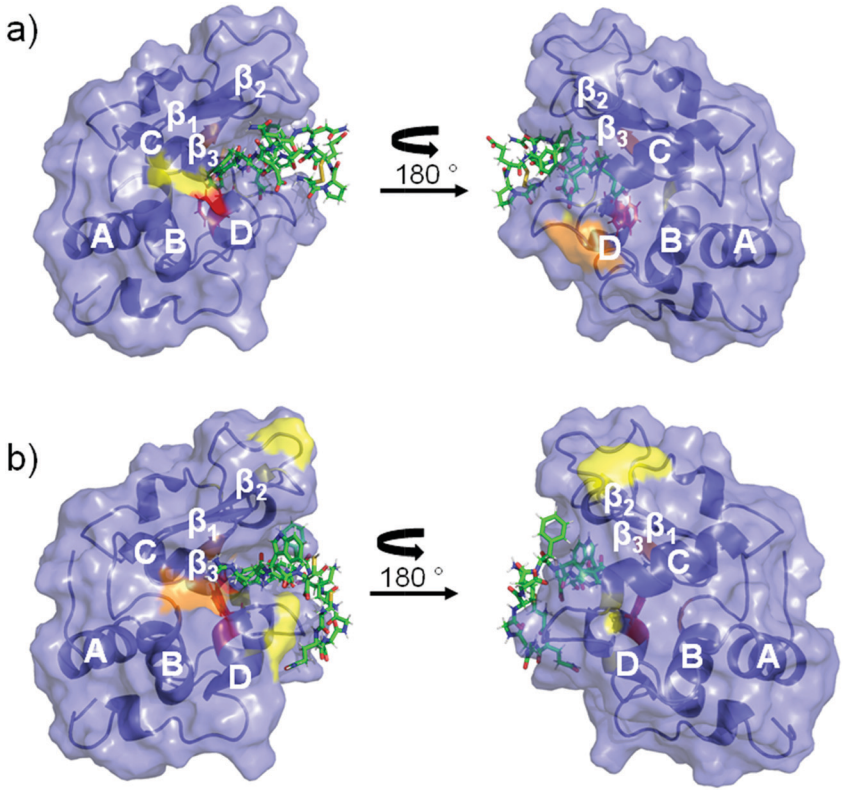

Fig. 8 Interaction of (a) peptide 410 and (b) peptide 140 with HuL according to CSP analysis. Residues that exhibit $\Delta\delta > \Delta\delta_{av} + \sigma$ (yellow), $\Delta\delta > \Delta\delta_{av} + 2\sigma$ (orange), or $\Delta\delta > \Delta\delta_{av} + 3\sigma$ (red) are highlighted. The secondary structure elements are labeled in white according to H. Kumeta *et al.*[64] Peptides are positioned according to PARCE outcome (peptide green licorice and protein blue cartoon).

where $\Delta\delta_{av}$ is the average CSP and σ is the standard deviation. The identified residues allowed locating the possible peptide binding site on HuL. To this aim, the residues were highlighted on the protein structure and compared with conformations predicted by the design algorithm PARCE (Fig. 8).

Fig. 8 shows that the perturbed region is located between the α-helix-rich and the β-strand-rich domains, where the active site cleft has previously been identified.[64] With both peptides (410 in Fig. 8a and 140 in Fig. 8b) the C-terminus of helix B, the N-terminus of helix D, and the strand β_3 are involved in the peptide binding and they are all reported to belong to the enzyme pocket.[64] In the presence of peptide 140 (Fig. 8b) another two protein portions are perturbed: the loop between strand β_3 and helix C and the loop β_1–β_2, identifying a possible second interaction site or conformational variations following the binding site occupancy.

3.7 Structural considerations

As shown above, most of the perturbed residues unveiled *via* NMR analysis clearly identifies the HuL binding pocket as the interaction site for peptides 410 and 140. This is supported by the CSP of HuL-E35 which is one of the two enzyme catalytic residues, the other being D53. Another residue of the pocket which is mostly affected by the binding is W109, along with its indole ring. In the HuL : peptide complexes obtained by simulations, the W109 side chain can interact through sandwich π-stacking with peptide 410-Y7 (distance 4.5 Å, Fig. 9a) and through an edge-to-face interaction with peptide 140-W8 (distance 3.5 Å, Fig. 9b). Moreover, HuL-I59 appears to be perturbed by both peptides. I59 is very close to the pocket and its sec-butyl side chain is positioned 3 Å apart from the aromatic ring of 410-Y7 (Fig. 9a) and 4 Å from the aromatic ring of 140-W8, allowing the formation of a CH–π interaction (Fig. 9b). Interestingly, although not belonging to the binding site, A111 is the residue that exhibits the highest CSP, with both peptides. It is located just opposite to E35 and its backbone NH forms a hydrogen bond (H-bond) with the E35 carboxyl group. These observations suggest the possibility of a conformational change of the target protein upon binding with the peptide. Notably, the docking of the peptide 140 is consistent with the formation of a salt bridge between peptide 140-R10 and HuL-E35, as distances between carboxyl oxygens and amino groups hydrogens are shorter than 3 Å. The higher perturbation of HuL-A111 recorded in the presence of the peptide 140 could be due to the fact that HuL-E35 interacts with both amino groups of 140-R10 forming a salt bridge (a stronger bond than a H-bond) and as a consequence HuL-A111 loses its H-bond with HuL-E35. On the other hand, in the presence of peptide 410, HuL-E35 forms an H-bond with an indolic NH (distance 2.3 Å), which still allows the formation of a H-bond between the other carboxyl oxygen of E35 and the amino group of A111. This structural consideration can explain the higher CSP of A111 with peptide 140, despite an

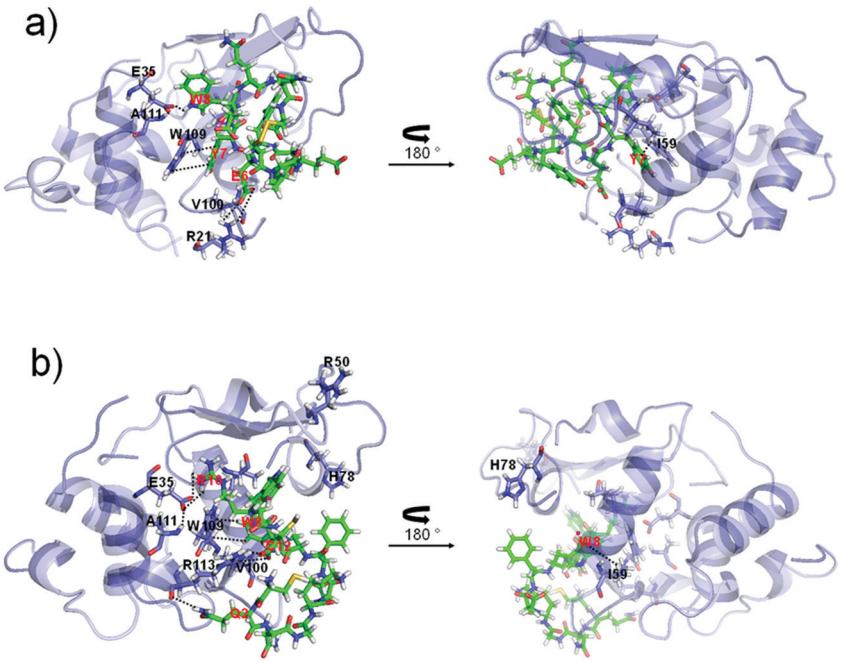

Fig. 9 HuL : peptide structures according to the design algorithm output for (a) peptide 410 and (b) peptide 140. All peptide residues (green), as well as the most perturbed residues of HuL identified by NMR analysis (blue) are highlighted (licorice) and labelled (black for protein and red for peptides). The distances between interacting atoms are indicated (dotted black lines).

observed lower binding affinity indicated by a slower saturation under NMR conditions. Furthermore, the simulation of peptide140–HuL complex is congruent with other two interactions involving the perturbed residue R113: a salt bridge between HuL-R113 side chain and 140-C12 C-terminal carboxylate (distances of 3.4 and 3.8 Å) and a H-bond between HuL-R113 carbonyl and 140-Q2 side chain amino group (distance 2.6 Å). As stated above, significant variations of HuL-H78 and HuL-R50 were observed with peptide 140. To rule out possible effects due to small pH variations during titration, we calculated HuL-H78 pK_a with and without the peptide obtaining similar values (6.29 and 6.30, respectively).[65] In the case of peptide 410, a salt bridge between peptide 410-E6 and HuL-R21 (distances of 3.2 and 4.0 Å) is also in agreement with the observed CSP. HuL-R21 is located in the loop between helix A and B and does not belong to the binding pocket, but its long side chain is positioned to possibly interact with the peptide located inside the pocket. The perturbation of HuL-V100, which has its carbonyl group close to the amino group of the HuL-R21 side chain (3.6 Å) and its amino group close to the peptide 410-E6 carboxylic group (3.8 Å), can be attributed to the general rearrangement of the H-bond/salt bridge network in that region of the protein.

4. Discussion and conclusions

The goal of this study was to define a set of guidelines for the computational design of peptides for biomolecular recognition with focus on the peptide structure to be employed for concave binding sites inside protein pockets or for surface-exposed, flat binding sites. To this aim, we chose human lysozyme (HuL) as a test case. We have computationally designed and screened both linear and cyclic peptides for two distinct binding sites of HuL. The best *in silico* performing binders were tested experimentally and an improved protocol for peptide selection was defined. We have shown that the optimum peptide topology (linear or cyclic) critically depends on the binding site to be targeted. Cyclic peptides, due to the restricted conformational space they can explore, generally guarantee a reduced entropy loss upon binding.[66,67] Overall, we have confirmed that cyclic peptides should be employed when targeting surface-exposed sites, while linear and cyclic peptides are both well-suited for pocket sites. In a pocket, greater backbone peptide flexibility seems important to achieve optimal interactions between the peptide and the protein. This factor might counter-balance the greater loss of entropy upon binding by linear peptides.

Two pocket peptides were thoroughly characterized. In particular, NMR confirmed the pocket targeting of both linear (140) and cyclic (410) peptides; however, this study indicated a more complex binding mode than the 1 : 1 stoichiometry. The multimodal behaviour was confirmed both *in silico* and by ESI-MS. SPR and NMR gave contrasting results in terms of binding affinities and kinetic parameters, making it difficult to determine which of the two candidates was the optimum binder. In NMR experiments, for which both peptide and HuL were free in solution, a peptide : protein ratio of 3 was sufficient to reach the plateau for peptide 410, while with peptide 140 it was necessary to increase the ratio up to 5.6.

For surface exposed binding sites, the design algorithm evolved cyclic peptides with much favourable predicted affinity

towards HuL than the generated linear peptides. Only the cyclic peptides were then characterized. During computational screening we found that peptides targeting surface-exposed sites required much longer simulated times for a reliable ranking to be achieved before being translated into the wet lab with respect to those targeting pockets. Indeed, peptides targeting the HuL pocket required only 50 ns long molecular dynamics (MD) trajectories to be chosen for experimental validation, while peptides designed for the nearby surface site required as much as 500 ns.

While cyclic peptides were computationally confirmed to be better candidates than linear ones for surface exposed binding sites, targeting such sites with *ex novo* design methods remains elusive when both recognition partners are intended to be free in solution. While binding affinity measurements were feasible when the target was surface-bound (namely, in the SPR setup where experiments confirmed that the selected peptides could bind the target with micromolar binding affinity), no reliable estimate was possible when both peptide and HuL were free in solution, as in the NMR setup. In particular the CSPs recorded in the presence of peptide 368, while consistent with the computational results, was negligible under the experimental conditions used and peptide 278 showed solubility problems at the typical concentrations used for NMR. This behaviour was consistent with previous observations: when targeting proteins, measurements on computationally designed peptides were feasible when at least one of the binding partners was surface bound, thus restricting the conformational space explored by the peptide.[10,11]

Another aspect explored in this paper was to test whether generating a binder from an arbitrarily chosen starting conformation close to the selected binding site was feasible without human intervention. To this aim we run multiple optimisations of an arbitrarily chosen initial peptide configuration which led to different sets of peptides, as expected from the stochastic nature of the optimisation. A different choice of starting configurations, for instance by randomising all initial conformations, would have been another suitable alternative allowing to further explore the peptides conformational space. Intuitively, as the optimisation process at its core is Monte Carlo based, and a Monte Carlo optimisation is known to lead to a minimum (not necessarily corresponding to the global minima) randomising the initial conformation of each stochastic optimisation path could open the possibility to identify further peptides candidates.

Another interesting aspect of PARCE, which is now based on a molecular dynamics (MD) engine, is that it allows for induced fit effects, an effect that can be controlled by choosing appropriate restrains on target conformation. In this work all optimisations were here run by restraining the target backbone to reduce protein conformational changes, but depending on the particular application the peptides will be employed in (drug design, protein immobilisation, and protein reactivation) different restrains can be employed. For instance, if the binding affinity should be maximized regardless of the protein conformation, induced fit can be introduced by releasing all restrains or by restraining only the structured segments of the protein.

Overall, the *in silico* process of evolving high-affinity binders remains highly promising by giving correct prediction on the geometry of the system. The NMR analysis underlined the need for an improvement in the design of peptides targeting surface-exposed sites. Further control on the binding site selectivity can be achieved by further algorithmic development or by changing the ligand to achieve higher affinities not always achievable by small peptides. While the computational method does not necessarily aim to outperform simple intuitive design, the low measured binding affinities point to a problem in the correlation between computational predictions and measures. The choice of larger and less degradable binders, such as antibody-derived ligands,[23] would also allow outcompeting physiologically relevant binding events. We are thus confident that the introduction of a consensus acceptance probability, based on multiple scoring functions,[35,68] coupled with the optimisation of larger, less flexible ligands holds the potential to further improve the binding affinity of the designed binders and break the micromolar limit.

Author contributions

C. V. and M. A. S. optimized the peptides; C. J. D. F., S. F. and M. A. S. screened and selected the peptides and analyzed the computational data; S. L. M. and D. M. performed the SPR assays; M. D. produced the ^{15}N labelled HuL; C. S. and R. G. designed the ESI-MS experiment and analysed the data; C. C. and A. C. designed and performed the NMR experiments and analyzed the NMR data; G. S. and S. F. designed the research. The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Abbreviations

HuL	Human lysozyme
SPR	Surface plasmon resonance
NMR	Nuclear magnetic resonance
ESI-MS	Electrospray ionization mass spectrometry
MC	Monte Carlo
MD	Molecular dynamics
PDB	Protein data bank
REMD	Replica exchange molecular dynamics
RMSD	Root mean squared deviation
HSQC	Heteronuclear single quantum correlation experiment
CSP	Chemical shift perturbation
H-bond	Hydrogen bond

Conflicts of interest

The authors declare no competing financial interest.

Acknowledgements

This work has been funded by the Italian Association for Cancer Research (AIRC) through the grant ''My First AIRC grant'' Rif.18510 and the grant ''Associazione Italiana per la Ricerca sul Cancro 5 per mille'', by the Royal Society of Chemistry (RSC) Research Fund grant, and by POR CAMPANIA FESR 2014/2020 ''Combattere la resistenza tumorale: piattaforma integrata multidisciplinare per un approccio tecnologico innovativo alle oncoterapie-Campania Oncoterapie'' (Project N. B61G18000470007). Computing resources have been provided by the Cinvestav del IPN (Mexico). We further acknowledge the CINECA Award No. HP10CIWPJ4 and HP10CRVL7F for the availability of high performance computing resources and support. CJDF acknowledges the TRIL fellowship for having provided support during the first part of this work under the ICTP TRIL programme, Trieste, Italy. Sara La Manna was supported by an AIRC fellowship for Italy. MD is a Research Associate supported by the FNRS (Fonds National de la Recherche Scientifique, Belgium).

References

1 M. Ciemny, M. Kurcinski, K. Kamel, A. Kolinski, N. Alam, O. Schueler-Furman and S. Kmiecik, Protein–peptide docking: opportunities and challenges, *Drug Discovery Today*, 2018, **23**(8), 1530–1537.

2 D. J. Diller, J. Swanson, A. S. Bayden, M. Jarosinski and J. Audie, Rational, computer-enabled peptide drug design: principles, methods, applications and future directions, *Future Med. Chem.*, 2015, **7**(16), 2173–2193.

3 K. Fosgerau and T. Hoffmann, Peptide therapeutics: current status and future directions, *Drug Discovery Today*, 2015, **20**(1), 122–128.

4 S. La Manna, E. Lee, M. Ouzounova, C. Di Natale, E. Novellino, A. Merlino, H. Korkaya and D. Marasco, Mimetics of suppressor of cytokine signaling 3: Novel potential therapeutics in triple breast cancer, *Int. J. Cancer*, 2018, **143**(9), 2177–2186.

5 S. La Manna, C. Di Natale, D. Florio and D. Marasco, Peptides as therapeutic agents for inflammatory-related diseases, *Int. J. Mol. Sci.*, 2018, **19**(9), 2714.

6 A. C.-L. Lee, J. L. Harris, K. K. Khanna and J.-H. Hong, A comprehensive review on current advances in peptide drug development and design, *Int. J. Mol. Sci.*, 2019, **20**(10), 2383.

7 A. Russo, C. Aiello, P. Grieco and D. Marasco, Targeting ''Undruggable'' Proteins: Design of Synthetic Cyclopeptides, *Curr. Med. Chem.*, 2016, **23**(8), 748–762.

8 G. P. Smith and V. A. Petrenko, Phage display, *Chem. Rev.*, 1997, **97**(2), 391–410.

9 C. Tuerk and L. Gold, Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase, *Science*, 1990, **249**(4968), 505–510.

10 M. A. Soler, S. Fortuna and G. Scoles, Computational design of peptides as probes for the recognition of protein biomarkers, *Eur. Biophys. J.*, 2015, **44**, S149.

11 M. A. Soler, A. Rodriguez, A. Russo, A. F. Adedeji, C. J. D. Foumthuim, C. Cantarutti, E. Ambrosetti, L. Casalis, A. Corazza, G. Scoles, D. Marasco, A. Laio and S. Fortuna, Computational design of cyclic peptides for the customized oriented immobilization of globular proteins, *Phys. Chem. Chem. Phys.*, 2017, **19**(4), 2740–2748.

12 A. Russo, P. L. Scognamiglio, R. P. H. Enriquez, C. Santambrogio, R. Grandori, D. Marasco, A. Giordano, G. Scoles and S. Fortuna, *In Silico* Generation of Peptides by Replica Exchange Monte Carlo: Docking-Based Optimization of Maltose-Binding-Protein Ligands, *PLoS One*, 2015, **10**(8), e0133571.

13 A. F. A. Olulana, M. A. Soler, M. Lotteri, H. Vondracek, L. Casalis, D. Marasco, M. Castronovo and S. Fortuna, Computational evolution of beta2-microglubulin binding peptides for nanopatterned surface sensors, *Int. J. Mol. Sci.*, 2021, **22**(2), 812.

14 A. Lathbridge and J. M. Mason, Computational Competitive and Negative Design to Derive a Specific cJun Antagonist, *Biochemistry*, 2018, **57**(42), 6108–6118.

15 I. D'Annessa, F. S. Di Leva, A. La Teana, E. Novellino, V. Limongelli and D. Di Marino, Bioinformatics and biosimulations as toolbox for peptides and peptidomimetics design: where are we?, *Front. Mol. Biosci.*, 2020, **7**, 66.

16 A. Obarska-Kosinska, A. Iacoangeli, R. Lepore and A. Tramontano, PepComposer: computational design of peptides binding to a given protein surface, *Nucleic Acids Res.*, 2016, **44**(W1), W522–W528.

17 R. P. Hong Enriquez, S. Pavan, F. Benedetti, A. Tossi, A. Savoini, F. Berti and A. Laio, Designing short peptides with high affinity for organic molecules: a combined docking, molecular dynamics, and Monte Carlo approach, *J. Chem. Theory Comput.*, 2012, **8**(3), 1121–1128.

18 M. Del Carlo, D. Capoferri, I. Gladich, F. Guida, C. Forzato, L. Navarini, D. Compagnone, A. Laio and F. Berti, *In Silico* Design of Short Peptides as Sensing Elements for Phenolic Compounds, *ACS Sensors*, 2016, **1**(3), 279–286.

19 I. Gladich, A. Rodriguez, R. P. Hong Enriquez, F. Guida, F. Berti and A. Laio, Designing High-Affinity Peptides for Organic Molecules by Explicit Solvent Molecular Dynamics, *J. Phys. Chem. B*, 2015, **119**(41), 12963–12969.

20 L. A. Chi and M. C. Vargas, *In silico* design of peptides as potential ligands to resistin, *J. Mol. Model.*, 2020, **26**, 1–14.

21 C. Geng, L. C. Xue, J. Roel-Touris and A. M. Bonvin, Finding the $\Delta\Delta G$ spot: Are predictors of binding affinity changes upon mutations in protein–protein interactions ready for it?, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2019, **9**(5), e1410.

22 M. A. Soler, S. Fortuna, A. De Marco and A. Laio, Binding affinity prediction of nanobody–protein complexes by scoring of molecular dynamics trajectories, *Phys. Chem. Chem. Phys.*, 2018, **20**(5), 3438–3444.

23 M. A. Soler, B. Medagli, M. Semrau, P. Storici, G. Bajc, A. de Marco, A. Laio and S. Fortuna, A consensus protocol for in-silico optimisation of antibody fragments, *Chem. Commun.*, 2019, **55**, 14043–14046.

24 V. Salmaso, M. Sturlese, A. Cuzzolin and S. Moro, Exploring protein-peptide recognition pathways using a supervised molecular dynamics approach, *Structure*, 2017, **25**(4), 655.e2–662.e2.
25 G. Zinzalla and D. E. Thurston, Targeting protein–protein interactions for therapeutic intervention: a challenge for the future, *Future Med. Chem.*, 2009, **1**(1), 65–93.
26 M. A. Soler, J. Zúñiga, A. Requena and A. Bastida, Understanding the connection between conformational changes of peptides and equilibrium thermal fluctuations, *Phys. Chem. Chem. Phys.*, 2017, **19**(5), 3459–3463.
27 R. Ochoa, M. A. Soler, A. Laio and P. Cossio, Assessing the capability of in silico mutation protocols for predicting the finite temperature conformation of amino acids, *Phys. Chem. Chem. Phys.*, 2018, **20**(40), 25901–25909.
28 M. A. Soler, A. De Marco and S. Fortuna, Molecular dynamics simulations and docking enable to explore the biophysical factors controlling the yields of engineered nanobodies, *Sci. Rep.*, 2016, **6**, 34869.
29 W. Jiang, C. Chipot and B. Roux, Computing relative binding affinity of ligands to receptor: An effective hybrid single-dual-topology free-energy perturbation approach in NAMD, *J. Chem. Inf. Model.*, 2019, **59**(9), 3794–3802.
30 T. B. Steinbrecher, M. Dahlgren, D. Cappel, T. Lin, L. Wang, G. Krilov, R. Abel, R. Friesner and W. Sherman, Accurate binding free energy predictions in fragment optimization, *J. Chem. Inf. Model.*, 2015, **55**(11), 2411–2420.
31 G. Ferraro, I. De Benedictis, A. Malfitano, G. Morelli, E. Novellino and D. Marasco, Interactions of cisplatin analogues with lysozyme: A comparative analysis, *Biometals*, 2017, **30**(5), 733–746.
32 D. Marasco, L. Messori, T. Marzo and A. Merlino, Oxaliplatin *vs.* cisplatin: competition experiments on their binding to lysozyme, *Dalton Trans.*, 2015, **44**(22), 10392–10398.
33 I. R. Krauss, L. Messori, M. A. Cinellu, D. Marasco, R. Sirignano and A. Merlino, Interactions of gold-based drugs with proteins: the structure and stability of the adduct formed in the reaction between lysozyme and the cytotoxic gold(III) compound Auoxo3, *Dalton Trans.*, 2014, **43**(46), 17483–17488.
34 A. Niitsu, S. Re, H. Oshima, M. Kamiya and Y. Sugita, De Novo Prediction of Binders and Nonbinders for T4 Lysozyme by gREST Simulations, *J. Chem. Inf. Model.*, 2019, **59**(9), 3879–3888.
35 R. Ochoa, M. A. Soler, A. Laio and P. Cossio, PARCE: Protocol for Amino acid Refinement through Computational Evolution, *Comput. Phys. Commun.*, 2021, **260**, 107716.
36 K. Harata, Y. Abe and M. Muraki, Full-matrix least-squares refinement of lysozymes and analysis of anisotropic thermal motion, *Proteins*, 1998, **30**(3), 232–243.
37 A. Lavi, C. H. Ngan, D. Movshovitz-Attias, T. Bohnuud, C. Yueh, D. Beglov, O. Schueler-Furman and D. Kozakov, Detection of peptide-binding sites on protein surfaces: The first step toward the modeling and targeting of peptide-mediated interactions, *Proteins*, 2013, **81**(12), 2096–2105.
38 K. Binder, D. Heermann, L. Roelofs, A. J. Mallinckrodt and S. McKay, Monte Carlo simulation in statistical physics, *Comput. Phys.*, 1993, **7**(2), 156–157.
39 C. Schafmeister, W. Ross and V. Romanovski, *LEAP*, University of California San Francisco, San Francisco, USA, 1995.
40 A. Patriksson and D. van der Spoel, A temperature predictor for parallel tempering simulations, *Phys. Chem. Chem. Phys.*, 2008, **10**(15), 2073–2077.
41 B. Hess, H. Bekker, H. J. Berendsen and J. G. Fraaije, LINCS: a linear constraint solver for molecular simulations, *J. Comput. Chem.*, 1997, **18**(12), 1463–1472.
42 X. Daura, K. Gademann, B. Jaun, D. Seebach, W. F. van Gunsteren and A. E. Mark, Peptide folding: when simulation meets experiment, *Angew. Chem., Int. Ed.*, 1999, **38**(1–2), 236–240.
43 O. Trott and A. J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, *J. Comput. Chem.*, 2010, **31**(2), 455–461.
44 K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror and D. E. Shaw, Improved side-chain torsion potentials for the Amber ff99SB protein force field, *Proteins*, 2010, **78**(8), 1950–1958.
45 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys.*, 1983, **79**(2), 926–935.
46 W. L. Jorgensen and J. D. Madura, Temperature and size dependence for Monte Carlo simulations of TIP4P water, *Mol. Phys.*, 1985, **56**(6), 1381–1392.
47 B. Hess, C. Kutzner, D. Van Der Spoel and E. Lindahl, GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation, *J. Chem. Theory Comput.*, 2008, **4**(3), 435–447.
48 C. L. Hagan, R. J. Johnson, A. Dhulesia, M. Dumoulin, J. Dumont, E. De Genst, J. Christodoulou, C. V. Robinson, C. M. Dobson and J. R. Kumita, A non-natural variant of human lysozyme (I59T) mimics the in vitro behaviour of the I56T variant that is responsible for a form of familial amyloidosis, *Protein Eng., Des. Sel.*, 2010, **23**(7), 499–506.
49 T.-L. Hwang and A. Shaka, Water suppression that works. Excitation sculpting using arbitrary waveforms and pulsed field gradients, *J. Magn. Reson., Ser. A*, 1995, **112**(2), 275–279.
50 T. Goddard and D. Kneller, *Sparky* **3**, University of California, San Francisco, USA, 2000.
51 F. A. Mulder, D. Schipper, R. Bott and R. Boelens, Altered flexibility in the substrate-binding site of related native and engineered high-alkaline Bacillus subtilisins, *J. Mol. Biol.*, 1999, **292**(1), 111–123.
52 H. Song, K. Inaka, K. Maenaka and M. Matsushima, Structural changes of active site cleft and different saccharide binding modes in human lysozyme co-crystallized with hexa-N-acetyl-chitohexaose at pH 4.0, *J. Mol. Biol.*, 1994, **244**(5), 522–540.
53 M. P. Williamson, Using chemical shift perturbation to characterise ligand binding, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2013, **73**, 1–16.

54 A. Fersht, *Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding*, Freeman, New York, 1999.
55 A. Natalello, C. Santambrogio and R. Grandori, Are charge-state distributions a reliable tool describing molecular ensembles of intrinsically disordered proteins by native MS?, *J. Am. Soc. Mass Spectrom.*, 2016, **28**(1), 21–28.
56 R. Marchese, R. Grandori, P. Carloni and S. Raugei, A computational model for protein ionization by electrospray based on gas-phase basicity, *J. Am. Soc. Mass Spectrom.*, 2012, **23**(11), 1903–1910.
57 M. Šamalikova and R. Grandori, Role of opposite charges in protein electrospray ionization mass spectrometry, *J. Mass Spectrom.*, 2003, **38**(9), 941–947.
58 M. Šamalikova, I. Matečko, N. Müller and R. Grandori, Interpreting conformational effects in protein nano-ESI-MS spectra, *Anal. Bioanal. Chem.*, 2004, **378**(4), 1112–1123.
59 C. Santambrogio, A. Natalello, S. Brocca, E. Ponzini and R. Grandori, Conformational Characterization and Classification of Intrinsically Disordered Proteins by Native Mass Spectrometry and Charge-State Distribution Analysis, *Proteomics*, 2019, **19**(6), 1800060.
60 Y. Zhang and M. F. Sanner, Docking flexible cyclic peptides with AutoDock CrankPep, *J. Chem. Theory Comput.*, 2019, **15**(10), 5161–5168.
61 Y. Zhang and M. F. Sanner, AutoDock CrankPep: combining folding and docking to predict protein–peptide complexes, *Bioinformatics*, 2019, **35**(24), 5121–5127.
62 T. Ohkubo, Y. Taniyama and M. Kikuchi, ^{1}H and ^{15}N NMR study of human lysozyme, *J. Biochem.*, 1991, **110**(6), 1022–1029.
63 M. Dumoulin, A. M. Last, A. Desmyter, K. Decanniere, D. Canet, G. Larsson, A. Spencer, D. B. Archer, J. Sasse and S. Muyldermans, A camelid antibody fragment inhibits the formation of amyloid fibrils by human lysozyme, *Nature*, 2003, **424**(6950), 783–788.
64 H. Kumeta, A. Miura, Y. Kobashigawa, K. Miura, C. Oka, N. Nemoto, K. Nitta and S. Tsuda, Low-temperature-induced structural changes in human lysozyme elucidated by three-dimensional NMR spectroscopy, *Biochemistry*, 2003, **42**(5), 1209–1216.
65 L. Wang, M. Zhang and E. Alexov, DelPhiPKa web server: predicting p K a of proteins, RNAs and DNAs, *Bioinformatics*, 2016, **32**(4), 614–615.
66 F. Fogolari, C. J. Dongmo Foumthuim, S. Fortuna, M. A. Soler, A. Corazza and G. Esposito, Accurate Estimation of the Entropy of Rotation–Translation Probability Distributions, *J. Chem. Theory Comput.*, 2016, **12**(1), 1–8.
67 F. Fogolari, A. Corazza, S. Fortuna, M. A. Soler, B. VanSchouwen, G. Brancolini, S. Corni, G. Melacini and G. Esposito, Distance-based configurational entropy of proteins from molecular dynamics simulations, *PLoS One*, 2015, **10**(7), e0132356.
68 R. Ochoa, A. Laio and P. Cossio, Predicting the Affinity of Peptides to Major Histocompatibility Complex Class II by Scoring Molecular Dynamics Simulations, *J. Chem. Inf. Model.*, 2019, **59**(8), 3464–3473.