Computational design of cyclic peptides for the customize oriented immobilization of globular proteins†

Miguel A. Soler,[ab] Alex Rodriguez,[b] Anna Russo,[a] Abimbola Feyisara Adedeji,[ac] Cedrix J. Dongmo Foumthuim,[a] Cristina Cantarutti,[a] Elena Ambrosetti,[cd] Loredana Casalis,[d] Alessandra Corazza,[a] Giacinto Scoles,[a] Daniela Marasco,[e] Alessandro Laio[b] and Sara Fortuna‡*[a]

[a] *Department of Medical and Biological Sciences, University of Udine, Piazzale Kolbe, 4 – 33100 Udine, Italy. E-mail: sara.fortuna@ung.si*

[b] *SISSA, Via Bonomea 265, I-34136 Trieste, Italy*

[c] *PhD School of Nanotechnology, Department of Physics, University of Trieste, Via Valerio 2, 34127 Trieste, Italy*

[d] *Nanoinnovation Lab – Elettra-Sincrotone S.C.p.A., ss 14 km 163, 5 in AREA Science Park I-34149, Basovizza-Trieste, Italy*

[e] *Department of Pharmacy, CIRPEB: Centro Interuniversitario di Ricerca sui Peptidi Bioattivi-University of Naples ''Federico II'', DFM-Scarl, 80134, Naples, Italy*

† Electronic supplementary information (ESI) available: Table S1. Scoring energies (*E*, in kcal mol^{-1}) and sequences of the soluble peptides with the lowest scoring *E*. Fig. S1. Primary screening: MD simulation results. Fig. S2. SPR results. Table S2. Peptide sequences elected through SPR secondary screening, pI (isoelectric point) and dissociation constants toward β2m are also reported (K_d), determination of the dissociation constant by atomic force microscopy (AFM). Fig. S4. AFM-based quantitative approach for analyzing surface-based cyclic peptide–DNA assemblages before and after β2m recognition at different target (β2m) concentration. Fig. S5. An example of a subsequent height measurement procedure employed in determining the average height difference across the topographic AFM micrograph of cyclic peptide–DNA assemblage before and after β2m recognition (100 mM concen- tration).

‡ Now at: Center for Biomedical Sciences and Engineering, University of Nova Gorica, Dvorec Lanthieri, Glavni Trg 8, 5271 Vipava, Slovenia.

DOI: 10.1039/c6cp07807a

Accepted by PCCP journal: 21st December 2016

0. Abstract

The oriented immobilization of proteins, key for the development of novel responsive biomaterials, relies on the availability of effective probes. These are generally provided by standard approaches based on *in vivo* maturation and *in vitro* selection of antibodies and/or aptamers. These techniques can suffer technical problems when a non-immunogenic epitope needs to be targeted. Here we propose a strategy to circumvent this issue by *in silico* design. In our method molecular binders, in the form of cyclic peptides, are computationally evolved by stochastically exploring their sequence and structure space to identify high-affinity peptides for a chosen epitope of a target globular protein: here a solvent-exposed site of b2-microglobulin (β2m). Designed sequences were screened by explicit solvent molecular dynamics simulations (MD) followed by experimental validation. Five candidates gave dose–response surface plasmon resonance signals with dissociation constants in the micromolar range. One of them was further analyzed by means of isothermal titration calorimetry, nuclear magnetic resonance, and 250 ns of MD. Atomic-force microscopy imaging showed that this peptide is able to immobilize

β2m on a gold surface. In short, we have shown by a variety of experimental techniques that it is possible to capture a protein through an epitope of choice by computational design.

1. Introduction

Developing novel biomaterials and platforms for the detection of biomarkers relies on strategies for capturing and immobilizing target proteins with well-defined orientations on solid supports.[1] This goal is currently achieved by chemical modifications either at functional groups or at the termini of the protein to be immobilized.[1c,2] Scanning probe or dippen lithography allows the fabrication of nanopatterned surfaces consisting of antibodies[3] for protein detection, of streptavidin for capturing biotinylated protein,[4] or of peptide nanoarrays.[5]

Maturation of antibodies and their *in vitro* selection[6] are the most common tools providing highly effective binders. These techniques requires a careful planning of the selection procedure[7] to pinpoint the specific epitope to which the anti-body will bind to (or binding site). This procedure can lead to antibodies reaching up to femtomolar affinity.[8] Typically dissociation constants (K_d) between antibodies and the target protein are in the nano- pico-molar range.[6,7] The recognition of antibodies toward their antigens mainly involves their six hyper-variable loops. Assuming that each of them equally contributes to the binding, this averages to K_d values of 10^{-2} –10^{-3} M per loop.

Other examples of binders are antibodies fragments, such as single domain antibodies (also known as nanobodies), that present three hypervariable loops providing affinities in the nanomolar range.[9] Similar affinities can be reached when only two loops actively participates in the binding with an average binding affinity of up to 10^{-5} M per loop.[10]

Antibody loops can be thought of as short cyclic peptides and it has been demonstrated that a cyclic peptide mimicking a single loop can reach up to micromolar affinity ($K_d \sim 10^{-6}$ M) as for the α-amylase binding peptide selected among mutants of the α-amylase inhibitor.[11] Often more efficient binders are obtained from mutated fragments of natural binders. For instance, several cyclic peptides for the detection of cysteine-rich intestinal protein 1 were selected by phage display technique and displayed K_d B10^{-4}–10^{-5}, a value that dropped to 10^{-6} M after their *in silico* optimization.[12] While the optimization of natural binders targeting existing protein pockets, known as ''active'' sites, is considered a ''solved issue'', the design of molecules capable of capturing a protein through a chosen binding site and achieving affinities similar to those exhibited by natural binders is still an open challenge. In this context, the computational design of short peptides as probes offers an exciting opportunity.[13]

Here we show that it is possible to target a surface exposed binding site to design *ex novo* a binder able to immobilize its target protein on a surface. To this aim, we applied a stochastic evolutionary algorithm.[14] This algorithm is capable of generating peptides suited for capturing a protein in solution reaching the same micromolar affinity as that obtained by selecting peptides either by phage display library or by natural evolution.

To push forward the binding affinity of the optimized probes towards their target and to address some of the drawbacks accounted in previous design strategies,[14a,c] we have developed, optimized, and tested a variant of the algorithm (in Fig. 1) based on sampling the bound state by molecular dynamics (MD).[18]

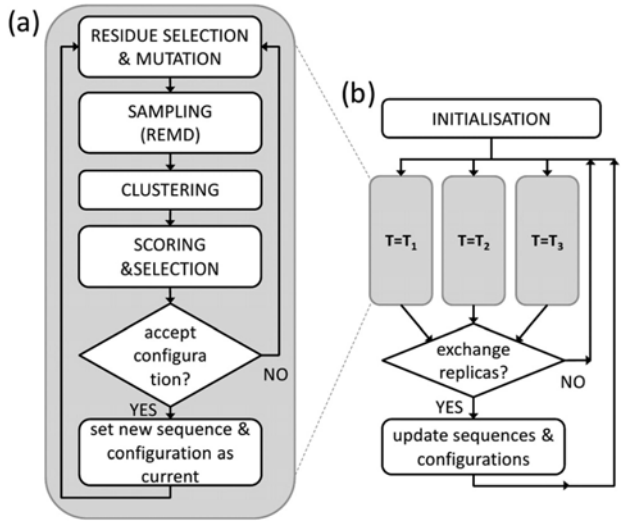

Fig. 1 (a) Optimisation kernel. (b) Overall algorithm.

As a test case we have chosen β2-microglobulin (β2m), a well characterized[15] 12 kDa protein included in a test panel for the prediction of ovarian cancer aggressiveness.[16] β2m also accumulates in patients undergoing hemodialysis, eventually leading to amyloidosis.[17] As a binding site we selected the region known to bind the human histocompatibility antigen.[15b] The selected binding site is exposed to the solvent and easily accessible to the designed peptides (Fig. 2a). For that binding site we designed cyclic peptides closed by a disulfide bridge.[18] The cyclic structure is expected to reduce the conformational freedom of the unbound peptide reducing the entropic contribution unfavorable to the complex formation.[19]

2. Results

2.1 Peptides design

The design approach is based on a combination of replica exchange molecular dynamics (REMD),[20] cluster analysis,[21] and replica exchange Monte Carlo.[22] The algorithm, summarized in Fig. 1, has been implemented in a bash script embedding functionalities from GROMACS,[23] AMBERTOOLS12[24] and AutoDock Vina.[25] (See the Materials and Methods section for a detailed description on the computational procedures.) At each step, as can be seen in Fig. 1a: (i) one peptide residue is randomly selected, mutated to other random residue, and subsequently energy minimization is carried out in the new structure (RESIDUE SELECTION & MUTATION), (ii) a replica exchange molecular dynamics (REMD) simulation is performed to sample the binder/target conformational space (SAMPLING), (iii) representative conformations are selected by clustering the obtained configurations with the Daura's algorithm,[21] (CLUSTERING), and (iv) the binding affinities of the cluster representatives are scored with AutoDock Vina and the conformation with the lowest binding score value (from now on scoring energy) is chosen (SCORING & SELECTION). Finally, the attempted mutation is accepted or rejected following the Metropolis criteria:

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

where E_{old} is the scoring energy of the old configuration, E_{new} is that of the attempted mutation, and T_{MC} is here a parameter for tuning the Metropolis acceptance probability P_{acc}.

As represented in Fig. 1b, the optimization was carried out in parallel at three values of T_{MC} (0.3, 0.6 and 0.9 kcal mol^{-1}). Random exchanges among replicas are accepted by using the standard parallel tempering scheme.[22]

We run the optimization algorithm twice, in both instances by starting from the same polyalanine/β2m configuration (CAAAAAAAAAAC peptide in the inset of Fig. 2b). One of the two optimization runs is shown in Fig. 2b where the evolution of the scoring energy between the β2m and the peptide is plotted along the mutation number. The score appears to reach a plateau after 200 mutations where the scoring energy settles to values between −24 and −25 kcal mol^{-1}. We note here that the scoring energy is subjected to (possibly large) errors[19] due to sampling fluctuations and that we will use it only to score the relative quality of different peptides. In the subsequent phase, that lasts about 300 mutations attempts, the scoring energy does not change significantly indicating the system reached equilibrium allowing to sample configurations. In all configurations the peptides stay well in contact with the protein surface.

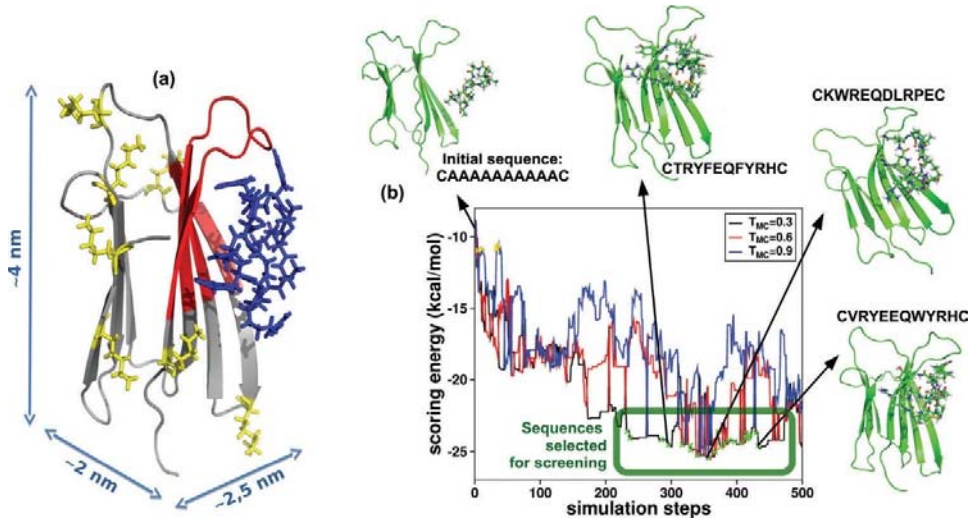

Fig. 2 (a) Structure of β2-microglobulin (gray) bound to the optimized peptide 461 (blue). The β2m binding site is composed by A, B, E, D b-sheet-forming strands (red). The β2m lysine residues are also highlighted (yellow). (b) Evolution of the scoring energies of the three Monte Carlo replicas (T_{MC} = 0.3, 0.6, 0.9 kcal mol^{-1}) during a typical run. Low score and soluble peptides (green crosses) are selected for successive screening. Selected configurations along the optimization path are shown in the insets.

2.2 Peptides screening

The solubility of the peptide sequences found in the plateau (see Fig. 2b) was assessed with the Innovagen's peptide calculator webtool.[7] Then, among the peptides predicted to have ''Good water solubility'', we chose the 20 peptides with the lowest scoring energy towards β2m (Fig. 2b). Their sequences are shown in Table S1 (ESI†).

We then carried out 50 ns MD simulations of each peptide/ β2m complex in water to identify the optimum candidate binders for these trajectories. We analyzed a number of observables involving the protein, its binding site, and each of the candidate peptide binders: the number of hydrogen bonds between the peptide and β2m, the solvent accessible surface area (SASA) of the hydrophobic groups of the binding site of β2m and of the peptide, and the average root mean square fluctuations (RMSF) of peptide backbone showed negligible differences among candidates, while the scoring energy, the distance and root mean square deviation (RMSD) between each peptide and the β2m binding site appeared valuable guides in selecting good candidates for experimental screening (Fig. S1, ESI†). In particular, we selected, for the experimental validation and secondary screening, peptides with both low values and a stable evolution of these descriptors along the simulated time. In more particular we selected peptides exhibiting scoring energies lower than —8 kcal mol^{-1}, and with energy fluctuations smaller than 4 kcal mol^{-1} (Fig. 3a). We excluded peptides with a separation distance from their center of mass to that of their binding site greater than 1.2 nm, selecting those with distances between 0.6 nm and 1.2 nm (Fig. 3b). Finally, as the RMSD indicates whether either the peptide or the binding site change their conformation (Fig. 3c), peptides with large conformational changes (more than 0.5 nm) were discarded, as well as those whose RMSD or separation distance was increasingly diverging along the simulation time. Moreover, we discarded peptides differing by only one residue of similar chemical properties, such as Thr/Ser. This primary screening led to eleven peptide sequences (listed in Table S1, ESI†).

A secondary screening was carried out employing an experimental binding assay based on SPR. β2m was immobilized on a CM5 chip as described in the Method section, with an immobilization level of 3200 RU. The β2m binding site did not contain lysine residues (Fig. 2a) ensuring its right exposure after the amine-coupling employed for the surface immobilization of the protein.

The eleven selected sequences were preliminarily assayed at six different concentrations of peptides (from 20 to 500 mM). Histograms of Response Unit (RU_{max}) for each concentration (Fig. 3d) show that six sequences gave dose response signals and were further analyzed. The experimental curves corresponding to different concentrations of peptides (Fig. 3e for peptide 461 and in Fig. S2 (ESI†) for the other sequences) were fitted according to a single exponential binding model with 1 : 1 stoichiometry. This kinetic analysis led to the estimation of dissociation constant values reported in Table S2 (ESI†), all in the micromolar range.

Sequence 461 was chosen as, among the peptides with a linear SPR response, it had the highest signal/noise response and does not contain His residue, thus reducing possible problems of pH control during long MD simulations. To evaluate the affinity of peptide 461 with a completely solution binding assay, we performed ITC experiments. By titrating aliquots of 461 into β2m solution, the downward ITC titration peaks demonstrated the exothermic association between peptide and β2m (Fig. 4a). The ITC data were best fitted by a nonlinear least-squares approach to the 'one set of sites' binding model (see Materials and methods), providing a dissociation constant (K_d) value of 62 mM (Fig. 4b), in agreement with the affinities showed for the kinetic based evolution of SPR assays. The 'one set of sites' model assumes a number of equally contributing binding sites having the same K_d and D*H*. The fitted number of sites was 1.34 T 0.14.

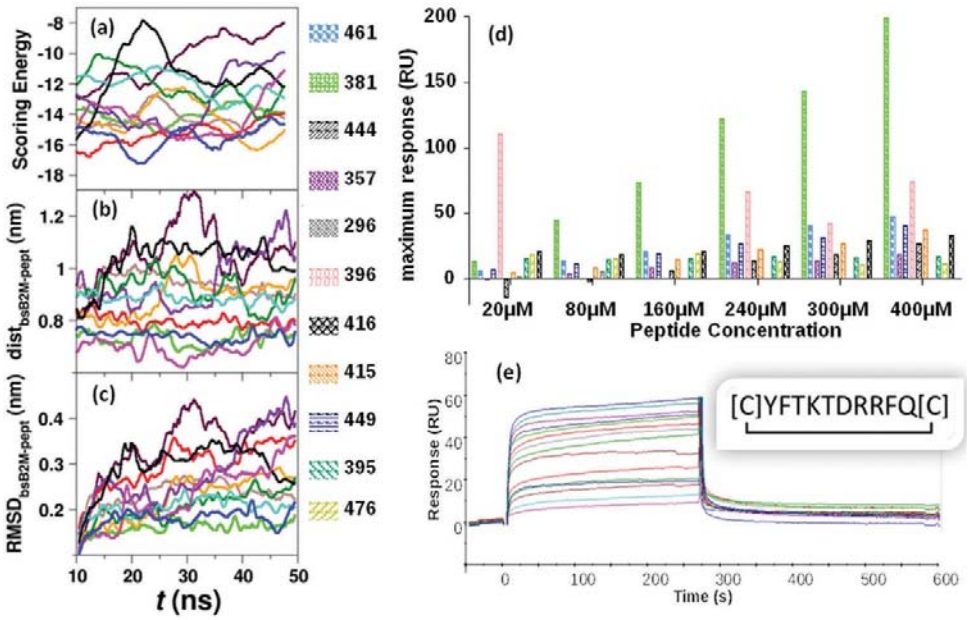

Fig. 3 (a–c) Primary screening: MD simulation results. (a) Average scoring energy, (b) separation distance between β2m binding site (BS-B2M) and peptide, and (c) RMSD of BS-B2M + peptide along the simulation time. (d) Secondary screening: SPR signals of eleven selected peptides at six different concentrations. (e) Overlay of sensograms for the sequence 461, reported in the inset.

2.3 Protein/peptide binding conformations

We performed NMR analysis to see whether the designed peptide was pinpointing the binding site it was designed for. ^{15}N NMR was showing very minor chemical shifts in the 2D [^{15}N,^{1}H] HSQC spectra during the titration of ^{15}N labeled β2m with the unlabeled 461 peptide (Fig. 5a) and no saturation was reached in the chemical shift deviations plots for the shifted residues (Fig. S3, ESI†), an effect not observed in SPR (Fig. S2, ESI†). Nevertheless, the NMR results showed that the affected amino acids were located on, or in close proximity to the A, B, E, D β-sheets forming strands to which the putative binding site belongs (Fig. 5b). When compared to the RMSF variation of the backbone amide nitrogens, the ^{15}NH cross peaks of D53, D59 and E74 that grow in intensity and H51, S52, R97 that show a chemical shift deviation also display a N-RMSF variation upon binding (Fig. 5c). Long MD simulations (250 ns, Fig. 5d) revealed that the peptide 461 was highly mobile on the β2m binding site and could adopt several conformations. The distance between the centers of mass of the peptide and the β2m binding site was distributed between 0.75 and 1.00 nm, with a distribution (which is the convolution of at least two gaussians, one centered at about 0.8 nm and one at about 0.9 nm) indicating multiple binding configurations located in the proximity of the chosen binding site (Fig. 5d). Overall the 461 peptide, while being highly mobile on the surface exposed site, tends to explore configurations close to the binding site it was designed for.

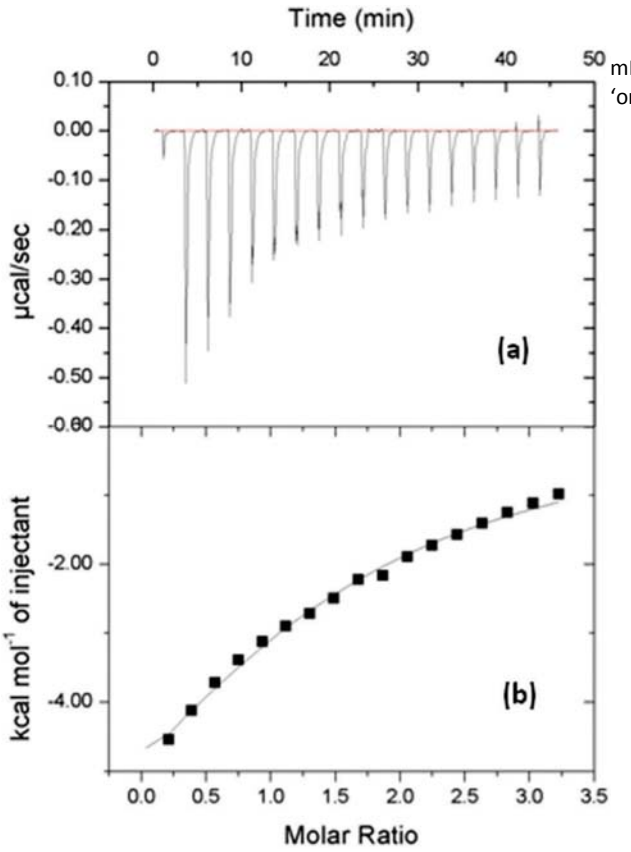

Fig. 4 ITC calorimetric (a) raw and (b) integrated data for β2m (42 mM) titration with peptide 461 (1 mM). Data fitting was achieved with 'one set of sites' model.

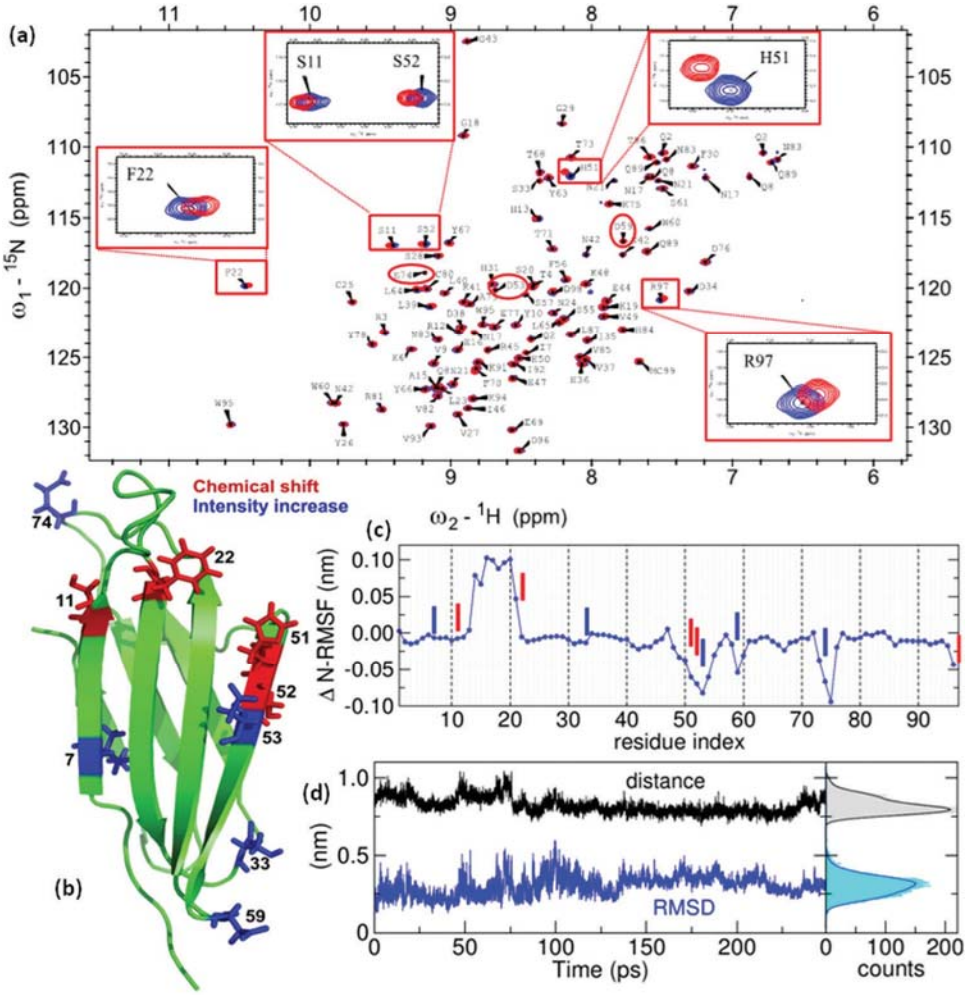

Fig. 5 (a) Overlay of 2D [1H, 15N] Heteronuclear Single-Quantum Correlation (HSQC) spectra of 15N-labeled β2m 45 mM in the free form in blue and in the presence of peptide 461 in red (peptide/protein = 36.5/1). The peaks presenting the largest chemical shift deviations are boxed in red and enlarged in the excerpts. The signals affected by the most consistent intensity variations are highlighted with red circles. (b) NMR results depicted on β2m structure: residues showing larger chemical shift deviations and intensity variations in the 2D [^{1}H, ^{15}N] HSQC experiments are colored in red and in blue, respectively. (c) Nitrogen root mean square fluctuation difference (DN-RMSF) between free and peptide bound β2m with vertical lines indicating amino acids presenting either chemical shift (red) or intensity increase (blue) upon binding, averages over 250 ns MD trajectories. (d) Peptide root mean square deviation (RMSD) at fixed β2m binding site (blue) and distance between the peptide and its binding site (black) along the MD trajectory (left panel) and their distributions (right panel).

2.4 Protein immobilization

To investigate the ability of the computationally designed peptide to immobilize β2m with a selective orientation we used an Atomic Force Microscopy (AFM) based quantitative method. In this approach, both NanoGrafting and DNA-directed immobilization (DDI)[26] were synergistically employed to generate surface-tethered cyclic peptide–DNA assemblages on an ultra-flat gold surface as described in the Materials and method section and the step-by-step procedure schematically represented in Fig. 6a–e. AFM topographic images of the assemblage before (Fig. 6f) and after (Fig. 6g) exposure to β2m, evidences an increased contrast of the patch with respect to the surrounding matrix, indicating β2m recognition by the peptide. The change in height, constant along the patch (Fig. 6h) and equal to 2 nm, corresponded to the size of the β2m orthogonal to the direction of the capturing peptide (Fig. 2a).

Further, several 461 peptide–DNA assemblages (see ESI,† Fig. S4) were generated and tested for β2m

recognition at different concentration of β2m (10–200 μM). Using an AFM height measurements (see Fig. S5, ESI†), the differences in height profiles (Δ*H*), as calculated by eqn (S1) and (S2) (ESI†), were plotted as a function of the concentration of β2m. This resulted into the binding curve reported in Fig. 6i. The experimental data were further fitted with Hill's equation from which we obtained a dissociation constant in the micro molar range, with K_d = (84.1 ± 0.4) μM, consistent with the SPR and ITC results.

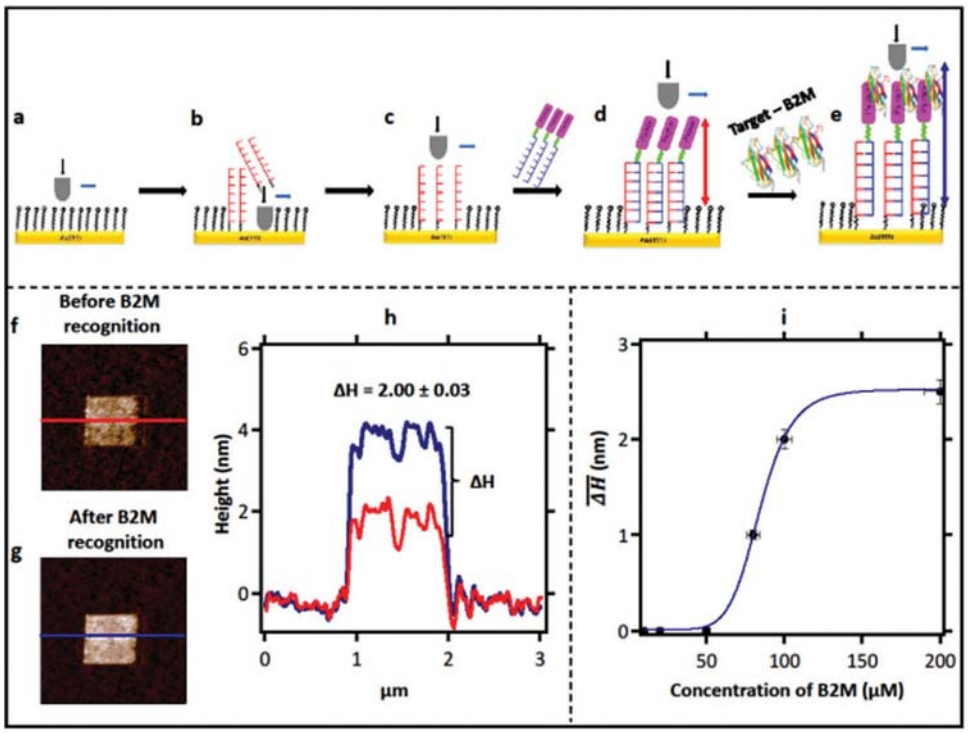

Fig. 6 AFM-based assay for immobilization of cyclic peptide and detection of β2m–cyclic peptide interactions on surfaces. The schematics represent the step-by-step procedures of the AFM nanolithography approach termed NanoGrafting; (a) scan of the AFM tip at a low nominal load (≈0.2 nN) over a given section of a monolayer of protein-repellent ethylene glycol-terminated alkylthiol on an ultra-flat gold surface, followed by the removal of the alkylthiols from the pre-selected section of the ultra-flat gold substrate at high nominal force (≈100 nN) and grafting of C6-thiol modified single stranded (ss)DNA in place of the previous alkylthiol on the surface as shown in (b). This yields laterally confined ssDNA molecules with respect to neighboring monolayer of protein-repellant alkylthiol as represented in (c). Subsequently, the peptide covalently-conjugated with complementary DNA was immobilized onto (c) by means of DNA-directed immobilization (DDI) through sequence-specific base pairing, and resulted into laterally confined cyclic peptide–DNA as shown in (d). After the successful generation of (d), the cyclic peptide–DNA assemblage is ready to be used for protein recognition (e), in our case β2-microglobulin (β2m). The AFM micrographs corresponding to (d–e) are shown in (f and g). AFM topographic height measurements were carried out on (f and g) and the change in height is shown in (h). The change in height became the binding signal of cyclic peptide towards β2m in standard solution condition. (i) Graph of overall differential height (see ESI,† Fig. S5) associated with β2m–cyclic peptide interactions as a function of β2m concentration. The experimental data (closed black spheres) were fitted using Hill's equation.

3. Conclusions

Overall we have shown it is possible to computationally evolve binders capable of immobilizing a protein through an *a priori* defined binding site. To test our method we have chosen a well characterized globular protein, the β2-microglobulin. By optimizing a simple polyalanine we reached the μM binding affinity as verified by a SPR/ITC/AFM platform. The result is in itself remarkable since it proves that a single loop can be optimized to reach the average binding affinity of those evolved by antibodies, overcoming the limit of binding only to a subset of binding sites accessible by *in vivo* selection. In this first proof-of-concept the optimization did not include any bias to guarantee the peptides selectivity, which is reflected on their tendency to form aggregates at high concentration. Nevertheless, the designed peptides have shown to be capable of binding their target

whether surface adsorbed or free in solution. By circumventing the problem of generating probes for a pre-selected epitope with an *in silico* approach, our method represents a step further in the development of novel biomaterials for biomarker detection.

Materials and methods

Peptides design

The β2m structure with PDB code 1LDS[27] was first minimized with the steepest descent minimization method. A peptide was then put close to the chosen binding site, and at each step of the design a random mutation of one random peptide residue (excluding the extreme cysteine amino acids) was performed by using the AmberTools program.[24] The mutated 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, with all bond atoms constrained by using LINCS algorithm,[28] and with the backbone of the β2m restrained 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[21] 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.

Molecular dynamics simulations

Each β2m/peptide complex was first minimized using the steepest descent method. 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 β2m 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 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 and 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 to 300 K. Production runs consisted of 50 ns long *NPT* simulations with 2 fs time step (plus additional 250 ns for the peptide 461). Configurations and energies were sampled every 10 ps. All simulations were performed with amber99sb-ildn force fields.[29] The system was solvated using the tip4p water model.[30] The Particle Mesh Ewald summation accounted for long range electrostatic interactions. All the calculations as well as their analysis were performed as implemented by Gromacs-4.6.2.[23] The binding energy of the peptide/protein complexes was estimated using Autodock Vina.[25]

Materials

β2m was purchased by SIGMA, while cyclic peptides were purchased as 95% purity from ProteoGenix SAS (Schiltigheim, France). Thiol modified (with C6 linker) single stranded (ss) DNA (SH-(CH2)6-5^0-CTTCACGATTGCCACTTTCCAC-3^0) and its corresponding complementary DNA with amino-C6-link (NH2- (CH2)6 -5'-GTGGAAAGTGGCAATCGTGAAG-3^0) were purchased as HPLC purified grade from Biomers GmbH (Ulm, Germany). Top-oligo-ethylene glycol (TOEG 6 (1-mercaptoundec-11-yl) hexa(ethyleneglycol), (HS-((CH2)11)-(O-CH2-CH2)6-OH)) was purchased from Prochimia Surfaces (Poland), sodium chloride (NaCl), TRIS-EDTA, sodium phosphate dibasic (NaH_2PO_4), sodium phosphate monobasic ($NaHPO_4$), and absolute ethanol (99.8%) were all purchased from Sigma Aldrich. All buffer solution were prepared in ultra-pure water (miliQ-H_2O), of resistivity 18.2 MΩ cm at 25 1C and filtered before use with sterile syringe filter (of 0.22 μm pore size).

Surface plasmon resonance

The interactions between the protein and computationally optimized peptides were measured using the BIAcore 3000 (GE Healthcare Milano, Italy). β2m was immobilized at a concentration of 100 μg mL^{-1} in 10 μM acetate buffer pH 5 (flow rate 5 μL min^{-1}, time injection 7 min) on a CM5 Biacore sensor chip,[31] using EDC/NHS chemistry following the manufacturer's instructions. Residual reactive groups were deactivated by treatment with 1 M ethanolamine hydrochloride, pH 8.5. The reference channel was activated with EDC/NHS and deactivated with ethanolamine. The binding assays were carried out at 20 mL min^{-1} at 25 °C, with 4.5 min contact-time. Peptides were diluted in the HBS running buffer (10 mM Hepes, 150 mM NaCl, 3 mM EDTA, pH 7.4). Analyte injections were performed at the indicated concentrations. The sensor surface was regenerated by using 10 mM NaOH for 1 minute. The association phase (k_{on}) was followed for 250 s, whereas the dissociation phase (k_{off}) was followed for 300 s. The instrument BIA evaluation analysis package (version 4.1, GE Healthcare, Milano, Italy) was used to subtract the signal of the reference channel.

ITC isothermal titration calorimetry

ITC experiments were carried out with an iTC200 calorimeter (Microcal/GE Healthcare). 461 peptide (1 mM) was titrated into a solution of β2m protein (42 μM). Data were fitted with 'one set of sites' model with ORIGIN software (GE Healthcare). Protein was dialysed against 10 mM phosphate buffer (pH 7.2) overnight, and all further dilutions of protein and peptide for ITC were made using the leftover external dialysate.

NMR

NMR experiments were recorded at 298 K on a Bruker Avance spectrometer operating at 500 MHz (^{1}H). 1D ^{1}H spectra were acquired with 4096 data point, a spectral width of 16 ppm and 4096 scans. The water suppression was achieved by excitation sculpting scheme. 2D [^{1}H, ^{15}N] HSQC spectra were acquired with 1024 and 128 points in the direct and indirect dimensions, respectively, and 400–1600 scans depending on the sample concentration, over spectral widths of 16 and 37 ppm in the ^{1}H and ^{15}N dimension, respectively. NMR samples were prepared in 25 mM sodium phosphate buffer and 50 mM NaCl at pH 6.95 and contained 8% D_2O for lock purposes. The data were processed with Topspin 2.1 and analyzed with Sparky.[32] Chemical shift deviations were 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.[33]

The β2m assignment was based on the file deposited on the Biological Magnetic Resonance Data Bank (Accession Code: 17165).

DNA–peptide conjugation

SoluLink's superior bioconjugation method was exploited to prepare the peptide–oligonucleotide conjugates in three steps. (i) Peptide modification: a volume that represents 10 mole equivalents of HyNic/mole protein was added to the peptide (concentrated 2.0 mg mL^{-1}) and mixed. The reaction was carried out at room temperature for 1.5 hours. (ii) Oligonucleotide modification: the oligonucleotide F9-amine (Biomers GmbH Ulm, Germany) was desalted into nuclease free water using a 5 K MWCO VivaSpin diafiltration apparatus and OD μL^{-1} concentration at 260 nm was determined (0.4 OD μL^{-1}). A volume containing 20 equivalents S-4FB was added to the oligonucleotide solution and incubated at room temperature for 2 hours. The 4FB-modified oligonucleotide was desalted into conjugation buffer (100 mM phosphate, 150 mM NaCl, pH 6.0). (iii) Peptide–oligo conjugation: taking into account the concentration and the mass of the HyNic-modified peptide to be functionalized and concentration and MSR of 4FB-modified oligonucleotide, volumes of the two components to be mixed are determined; 1/10 volume 10× TurboLink Catalyst Buffer (100 mM aniline, 100mM phosphate, 150 mM NaCl, pH 6.0) was added to the conjugation solution and the reaction was carried out overnight at 4 °C. The conjugation reaction can be visualized spectrophotometrically by determining the absorbance at A354 due to the formation of the chromophoric conjugate bond.

Immobilization of cyclic peptide on ultra-flat gold (Au(111)) surface

ssDNA solution (22 bps with C6-linker, 1 μM ssDNA in IM NaCl, TE1X) was dispensed onto the top-oligo-ethylene glycol-terminated alkanethiol, TOEG6, passivated ultra-flat gold surface prior to the nanografting process as depicted in Fig. 6(a–c). After the immobilization of thiolated ssDNA, the remaining ssDNA solution was removed and the sample was rinsed three times with DNA free buffer (1 M NaCl, TE1X, pH 7.2). This was followed by a fast scan at low force (~0.2 nN) of the grafting area (where immobilization took place) in DNA-free buffer (1 M NaCl, TE1X, pH 7.2). Afterwards, the hybridization process with cyclic peptide–DNA conjugate (100 nM cyclic peptide–DNA in 1 M NaCl, TE1X, pH 7.2), which is complementary to the ssDNA on surface, took place at 37 °C for 1 hour. At the end of the hybridization process, the excess cyclic-peptide conjugate solution was removed and the sample was rinsed three times with peptide–DNA-free buffer (1 M NaCl, TE1X, pH 7.2).

AFM

All AFM experiments (starting from scanning, nanolithography, and height detection of β2m–cyclic peptide binding interactions) were carried out with MFP-3D-BIO AFM, (Asylum research, Santa Barbara, CA). For nano-grafting, commercially available pyramidal silicon etched probes NSC 18/no Al of spring constant of 2.8 N m^{-1} (Mikro-Masch, Germany) was used. For contact mode imaging, soft probe (CSC 38/no Al, spring constant 0.03 N m^{-1}, Mikro-Masch, Germany) was used.

β2-Microglobulin (β2m)–cyclic peptide AFM binding assay

The binding buffer, which is phosphate buffer (25 mM sodium phosphate, 50 mM NaCl, pH 7) was introduced onto the sample after the removal of the buffer used for imaging and nano-grafting process (1 M NaCl, TE1X, pH 7.2). To equilibrate the microenvironment on the sample for binding assay, the sample was rinsed (twice) and incubated with phosphate buffer for 10 minutes. Afterwards, the binding assay was carried in contact with β2m solution (10–200 μM β2m in phosphate buffer (25 mM sodium phosphate, 50 mM NaCl, pH 7)) at 25 °C for period of 1 hour. The grafting section (where cyclic peptide was immobilized) was imaged at low force (~0.2 nN) in contact mode. AFM topographic height measurement (using Igor Pro, 6.3A) was employed to measure the height of cyclic peptide–DNA assemblage before and after binding assay with β2m.

Author contributions

GS and AL proposed the setup, SF devised and coordinated the project, MAS and ARO implemented the code; MAS computationally generated the peptides; CJDF computationally screened the peptides; SF computationally characterized the lead binder; CJDF MAS ARO SF AL analyzed the computational data; ARU and DM run the SPR experiments and analyzed the data; DM run the ITC experiments and analyzed the data; CC and AC run the NMR experiments and analyzed the data; EA prepared the conjugates, AFA performed the AFM experiments, AFA and GS analyzed the AFM data. MAS, DM, AFA, LC, AC, GS, AL, SF wrote the manuscript. All authors reviewed the manuscript and have given approval to the final version of the manuscript.

Funding sources

This work has been funded by the ERC Advanced Grant MoNaLiSA: QUIDPROQUO (proposal no. 269025, 2011-2016, PI: Giacinto Scoles). We further acknowledge financial support from the grant ‘‘Associazione Italiana per la Ricerca sul Cancro 5 per mille’’, Rif. 12214. ARu further acknowledges the financial support of the European Fund for Regional Development – Cross-Border Cooperation Programme Italy-Slovenia 2007–2013, (Project PROTEO, Code No. CB166). CJDF acknowledges the TRIL fellowship for having provided support during the first part of this work under the ICTP TRIL programme, Trieste, Italy. SF further acknowledges partial support from the TALENTS3 program (European Social Fund 2014/2020 Regional Operative Programme, Axis 3 – Education and Training Call n. 782 of 13/04/2015, PI Sara Fortuna).

Abbreviations

β2m	β2-Microglobulin
MD	Molecular dynamics
SPR	Surface plasmon resonance
NMR	Nuclear magnetic resonance
AFM	Atomic force microscopy

Acknowledgements

We gratefully acknowledge Federico Fogolari (University of Udine, Italy) and Cristina Vargas (CINESTAV, Mexico) for fruitful discussions on this work and the availability of computational resources. We acknowledge the CINECA Awards No. HP10CCSKB9, No. HP10CLACPE, and No. HP10CQZV2H, 2014, for the availability of high performance computing resources and support. We acknowledge PRACE for awarding SF access to resource MareNostrum III based in Spain at Barcelona Supercomputing Center (BSC-CNS).

References

1 (*a*) S. Chen, L. Liu, J. Zhou and S. Jiang, *Langmuir*, 2003, 19, 2859–2864; (*b*) P. Peluso, D. S. Wilson, D. Do, H. Tran, M. Venkatasubbaiah, D. Quincy, B. Heidecker, K. Poindexter, N. Tolani and M. Phelan, *Anal. Biochem.*, 2003, 312, 113–124; (*c*) F. Rusmini, Z. Zhong and J. Feijen, *Biomacromolecules*, 2007, 8, 1775–1789.

2 W. H. Scouten, J. H. Luong and R. S. Brown, *Trends Biotechnol.*, 1995, 13, 178–185.

3 J. R. Kenseth, J. A. Harnisch, V. W. Jones and M. D. Porter, *Langmuir*, 2001, 17, 4105–4112.

4 J. Hyun, S. J. Ahn, W. K. Lee, A. Chilkoti and S. Zauscher, *Nano Lett.*, 2002, 2, 1203–1207.

5 (*a*) B. A. Williams, K. Lund, Y. Liu, H. Yan and J. C. Chaput, *Angew. Chem.*, 2007, 119, 3111–3114; (*b*) K. L. Christman, V. D. Enriquez-Rios and H. D. Maynard, *Soft Matter*, 2006, 2, 928–939.

6 H. R. Hoogenboom, *Nat. Biotechnol.*, 2005, 23, 1105–1116.

7 H. R. Hoogenboom and P. Chames, *Immunol. Today*, 2000, 21, 371–378.

8 E. T. Boder, K. S. Midelfort and K. D. Wittrup, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, 97, 10701–10705.

9 S. Muyldermans, C. Cambillau and L. Wyns, *Trends Biochem. Sci.*, 2001, 26, 230–235.

10 K. Decanniere, A. Desmyter, M. Lauwereys, M. A. Ghahroudi, S. Muyldermans and L. Wyns, *Structure*, 1999, 7, 361–370.

11 M. Takahashi, K. Nokihara and H. Mihara, *Chem. Biol.*, 2003, 10, 53–60.

12 J. Hao, A. Serohijos, G. Newton, G. Tassone, Z. Wang, D. C. Sgroi, N. V. Dokholyan and J. P. Basilion, *PLoS Comput. Biol.*, 2008, 4, e1000138.

13 P. Vanhee, A. M. van der Sloot, E. Verschueren, L. Serrano, F. Rousseau and J. Schymkowitz, *Trends Biotechnol.*, 2011, 29, 231–239.

14 (*a*) R. P. H. Enriquez, S. Pavan, F. Benedetti, A. Tossi, Savoini, F. Berti and A. Laio, *J. Chem. Theory Comput.*, 2012, 8, 1121–1128; (*b*) I. Gladich, A. Rodriguez, R. P. Hong Enriquez, F. Guida, F. Berti and A. Laio, *J. Phys. Chem. B*, 2015, 119, 12963–12969; (*c*) A. Russo, P. L. Scognamiglio, R. P. H. Enriquez, C. Santambrogio, R. Grandori, D. Marasco, A. Giordano, G. Scoles and S. Fortuna, *PLoS One*, 2015, 10, e0133571.

15 (*a*) G. Verdone, A. Corazza, P. Viglino, F. Pettirossi, S. Giorgetti, P. Mangione, A. Andreola, M. Stoppini, V. Bellotti and G. Esposito, *Protein Sci.*, 2002, 11, 487–499; (*b*) M. Saper, P. Bjorkman and D. Wiley, *J. Mol. Biol.*, 1991, 219, 277–319.

16 E. T. Fung, *Clin. Chem.*, 2010, 56, 327–329.

17 F. Locatelli, F. Mastrangelo, B. Redaelli, C. Ronco, D. Marcelli, G. La Greca and G. Orlandini, *Kidney Int.*, 1996, 50, 1293–1302.

18 A. V. Collis, A. P. Brouwer and A. C. Martin, *J. Mol. Biol.*, 2003, 325, 337–354.
19 A. Russo, C. Aiello, P. Grieco and D. Marasco, *Curr. Med. Chem.*, 2016, 23, 748–762.
20 Y. Sugita and Y. Okamoto, *Chem. Phys. Lett.*, 1999, 314, 141–151.
21 X. Daura, K. Gademann, B. Jaun, D. Seebach, W. F. van Gunsteren and A. E. Mark, *Angew. Chem., Int. Ed.*, 1999, 38, 236–240.
22 Y. Iba, *Int. J. Mod. Phys. C*, 2001, 12, 623–656.
23 B. Hess, C. Kutzner, D. Van Der Spoel and E. Lindahl, *J. Chem. Theory Comput.*, 2008, 4, 435–447.
24 C. E. A. F. Schafmeister, W. S. Ross and V. Romanovski, *LEaP*, University of California, San Francisco, 1995.
25 O. Trott and A. J. Olson, *J. Comput. Chem.*, 2010, 31, 455–461.
26 (*a*) F. Bano, L. Fruk, B. Sanavio, M. Glettenberg, L. Casalis, C. M. Niemeyer and G. Scoles, *Nano Lett.*, 2009, 9, 2614–2618; (*b*) M. Ganau, A. Bosco, A. Palma, S. Corvaglia, P. Parisse, L. Fruk, A. P. Beltrami, D. Cesselli, L. Casalis and G. Scoles, *Nanomedicine*, 2015, 11, 293–300.
27 C. H. Trinh, D. P. Smith, A. P. Kalverda, S. E. Phillips and S. E. Radford, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, 99, 9771–9776.
28 B. Hess, H. Bekker, H. J. Berendsen and J. G. Fraaije, *J. Comput. Chem.*, 1997, 18, 1463–1472.
29 K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror and D. E. Shaw, *Proteins: Struct., Funct., Bioinf.*, 2010, 78, 1950–1958.
30 (*a*) W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, 79, 926–935; (*b*) W. L. Jorgensen and J. D. Madura, *Mol. Phys.*, 1985, 56, 1381–1392.
31 G. Sreejit, A. Ahmed, N. Parveen, V. Jha, V. L. Valluri, S. Ghosh and S. Mukhopadhyay, *PLoS Pathog.*, 2014, 10, e1004446.
32 T. D. Goddard and D. G. Kneller, *Sparky–NMR assignment and integration software*, University of California, 2006.
33 F. A. Mulder, D. Schipper, R. Bott and R. Boelens, *J. Mol. Biol.*, 1999, 292, 111–123.