

Thermodynamically–Weighted Conformational Ensemble of Cyclic RGD Peptidomimetics from NOE Data

F. Vasile,^{*,†} M. Civera,[†] L. Belvisi,[†] D. Potenza,[†] and G. Tiana^{*,‡}

[†]*Department of Chemistry, Università degli Studi di Milano, via Golgi 19, 20133 Milano, Italy*

[‡]*Center of Complexity and Biosystems and Department of Physics, Università degli Studi di Milano and INFN, via Celoria 16, 20133 Milano, Italy*

E-mail: francesca.vasile@unimi.it; guido.tiana@unimi.it

Abstract

In the case of flexible molecules, the standard approach of transforming NOE intensities into spatial restraints and of building conformational models minimizing these restraints greatly neglects the richness of molecular conformations. Making use of NOE intensities measured in triplicate and of an iterative molecular–dynamics scheme we optimized a force field to generate a set of conformations whose ensemble is compatible with the experimental data and which are weighted according to the Boltzmann distribution. This scheme is applied to two cyclic peptidomimetic ligands of integrins. Their difference in binding affinity is recapitulated in terms of their difference in conformational fluctuations.

Introduction

In recent years, the focus of NMR spectroscopy shifted considerably from the determination of static structures to the elucidation of the structural and dynamical properties of molecules displaying multiple equilibrium conformations.¹ A preeminent example is the study of the conformational ensemble populated by short peptides or intrinsically-disordered proteins on the basis of chemical shifts and of residual dipolar couplings. Several theoretical approaches have been successfully developed to turn these kinds of NMR data into ensembles of representative conformations.²⁻⁴

Chemical shifts and residual dipolar couplings have been used extensively to study the conformational properties of flexible biomolecules.⁵ On the other hand, the use of NOE intensities has so far been limited, mainly because detectable NOEs are usually restricted to short-range couplings, and consequently do not provide enough information to specify the molecular conformations. Although it has been long known that the interpretation of NOEs in terms of static restraints is not adequate in the case of fluctuating proteins,^{6,7} only rarely NOEs have been used in the case of non-structured proteins, often in combination with other types of NMR data.⁸ Moreover, in the case of protein, spin diffusion plays an important role and makes the interpretation of NOEs in terms of atomic distances even more problematic^{9,10} (see ref. 11 for a review).

Different is the case of small cyclic molecules, which can often fluctuate among a limited number of different conformations, and still their conformational properties can be exhaustively monitored by NOEs.¹² Also in this case, algorithms were designed to obtain ensembles of conformations which are qualitatively compatible with NOEs.¹³⁻¹⁵ However, the qualitative compatibility of a conformational ensemble with the experimental data is a first necessary step, but does not guarantee that its thermodynamic properties are quantitatively correct, in the sense that the statistical weights of the different conformations of the ensemble are those given by the Boltzmann distribution. A simple example is when two nuclei in a flexible molecule populate preferentially two possible conformations with respect to each other, and

negligibly other ones. If NOEs are obtained in conditions of rapid exchange, a single signal is obtained from the experiment. An ensemble of homogeneous conformations in which the nuclei populate with probability one a single conformation, intermediate between the two, will give the same NOE intensity. But this approximation of the real situation neglects the thermal fluctuations which can be important to understand the behaviour of that molecule.

Moreover, at variance with proteins, in the case of small molecules with limited conformational freedom, spin diffusion is usually negligible.¹⁶ We show below that this is the case for our molecules of interest. Consequently, each conformation of the ensemble is expected to contribute to each observed NOE intensity with a quantity which is the inverse sixth power of the associated distance, correctly weighted according to the principles of thermodynamics.

The goal of the present work is to develop a scheme to generate an ensemble of conformations which is not only compatible with NOE intensities, but also displays a statistical weight equal to their Boltzmann probability according to a realistic energy function, and consequently displays consistent thermodynamical properties.

The molecules of interest in this work, DKP5-RGD and DKP2-RGD^{17,18} (see Fig. 1), are ligands of the $\alpha_v\beta_3$ integrin, showing low nanomolar IC50 values. They are 17-membered (in terms of number of backbone atoms) cyclic peptide mimics, containing the peptidic Arginine-Glycine-Aspartic acid (RGD) sequence and a (trans 3R, 6S) bifunctional diketopiperazine (DKP) scaffold. They were chosen because they are structurally similar and still they display a different binding behavior. In fact, they only differ in the functionalization at the DKP nitrogen atoms with two or one benzylic groups, respectively, but nonetheless they display different IC50 values. Since the conformational preferences of the RGD-tripeptide recognition motif is heavily involved in the interaction with the integrin receptor, we investigated the conformational features of these two molecules in order to rationalize their difference in binding affinity for their target, and thus to the associated IC50 (IC50=12.2 \pm 5.0 nM for DKP5-RGD and IC50=3.2 \pm 2.7 nM for DKP2-RGD). One should anyway remember that, although the IC50 is usually correlated to the binding affinity, they express different

properties of the system.

An interesting feature of these two molecules is that they are not rigid. The analysis of the associated NMR parameters (chemical shift and temperature variation of the amide protons) indicates the presence of multiple conformations in equilibrium.¹⁸ Consequently, the construction of a structural model to describe them is not straightforward.

The standard procedure of transforming NOE intensities into spatial restraints¹⁹ and to look for the structures which optimize them is thermodynamically grounded only for rigid structures. It corresponds to a mean-field approximation which regards the thermodynamic averages $\langle d^{-6} \rangle$ associated with the interatomic distances d as $\langle d \rangle^{-6}$, neglecting their fluctuations. The result is to predict a single average conformation (or few homogeneous conformations) even when the molecule displays multiple equilibrium states.

The use of molecular-dynamics (MD) simulations would make it possible, in principle, to reproduce the correct thermodynamic weight of multiple molecular conformations. However, imperfections in the force fields employed by MD often make its ability to reproduce experimental data only qualitative. Specifically, in the case of cyclic RGD peptidomimetics it was possible to match the NOE distance constrains, that is upper bounds in the associated distances, with only few violations,²⁰ but with poor correlation between the experimental and the back-calculated intensities (cf. also Figs. S1 and S2 in the Supplementary Material). In fact, it was shown²¹ that matching the distance constrains associated with NOEs is not a stringent requirement for the validity of MD results.

In the present work we develop a computational scheme to correct the force field used in MD simulations to match the experimental NOEs at quantitative grade. Thus, the MD simulation results in an ensemble of conformations which display the correct statistical weights according to the modified force field, and that reproduce the experimental NOEs. Under specific conditions, the potential converges to a unique and robust optimal set of parameters, although this is system-dependent and must be re-calculated for each different molecule.

The scheme consists in an iterative modification of some of the parameters of the force

field following a conformational sampling of the system and takes advantage of an efficient reweighting of the sampled conformations.²² This kind of approach gave good results in connection with other kind of systems.^{23,24} In the present case it is applied to the dihedral terms of the force fields associated with the atoms involved in the NOEs.

We show that DKP2-RGD displays an overall rigid structure, while DKP5-RGD fluctuates between two main different conformations and a less populated state, a fact which affects its affinity for $\alpha_v\beta_3$. We also show that it is possible to take into account this conformational heterogeneity in docking calculations, thus interpreting affinity data in a more realistic way.

Materials and Methods

NMR experiments. DKP2-RGD and DKP5-RGD were synthesized as described in ref. 18.

NMR spectra of DKP2-RGD and DKP5-RGD were obtained using a 10 mM solution in H₂O/D₂O 9:1. NMR experiment were acquired using a Bruker Avance 600 MHz instrument at 283 K. The TOCSY spectra, with a mixing time of 80 ms, made it possible a complete assignment of all the protons. The chemical shifts assignments are reported in Supplementary Information (Table S1 and S2). For the conformational analysis three independent NOESY spectra (with 64 scans and 256 increments) were collected using a mixing times of 700 ms. Water suppression was achieved by excitation–sculpting pulse sequence. The 1D and NOESY spectra of both compounds are reported in Supplementary Information (Figs. S3 and S4 for DKP5-RGD and Figs. S5 and S6 for DKP2-RGD). From each NOESY spectrum, the intensity of NOE signals was calculated in terms of height of the peak.²⁵ The average and the standard deviation of the NOE intensity are calculated over the three replicated spectra (Tables S3 and S4 in the Supporting Information).

To be sure that, as expected for this kind of molecules,¹⁶ spin diffusion can be neglected,

we studied the dependence of intensity of selected NOEs as a function of mixing time (see Fig. S7 in the Supplementary Materials). At the mixing time of 700 ms used for the following analysis, the system is still in the linear regime.

Initial model. The interactions of each molecule are first described by the GAFF force field.²⁶ Charges are obtained with the RESP procedure of the antechamber program (AmberTools 1.5).²⁷ Solvation effects are described by the water GBSA implicit solvent model.²⁸ Gromacs input files (i.e. topology and coordinates files) were generated using the Acepype script.²⁹ Notice that the simulations, performed in implicit water, match the experimental conditions of pure water, corresponding to a low ionic strength.

MD simulations. Conformational sampling is carried out with (second-order) Langevin dynamics at 283K with a time step of 2 fs making use of Gromacs 4.5.5. In each simulation, which lasts for 40 ns, the coordinates $\{r^n\}$ of set of $N_c = 5000$ conformations are recorded to be analyzed.

Forward model. From the recorded conformations, the NOE intensity associated with atoms i and j is obtained as thermodynamic average

$$I(i, j) = \frac{I_0}{N_c} \sum_{n=1}^{N_c} d(r_i^n, r_j^n)^{-6}, \quad (1)$$

where $d(r_i, r_j)$ indicates the distance between the two atoms. To obtain I_0 we compare the NOE intensity measured for the HG-HE atom pair in the side chain of arginine with the thermodynamic average $\langle d^{-6} \rangle$ of the associated distance d calculated in a ACE-GLY-ARG-GLY-NME peptide in explicit solvent with the Amber99SB force field. The reason is that this force field applied to peptidic molecules has been widely tested³⁰ in terms of side-chain dynamics, and consequently is more reliable than GAFF as a reference. Moreover, the absence of spin diffusion suggests that the NOE intensity of the HG-HE pair can be interpreted in terms of inverse sixth power of their distance. We found that $\langle d^{-6} \rangle = 2322 \text{ nm}^{-6}$, resulting in $I_0 = 156.0 \text{ nm}^6$ for DKP5-RGD and $I_0 = 221.9 \text{ nm}^6$ for DKP2-RGD.

Iterative update of the force field. An iterative set of simulations is carried out to sample the conformational space of the molecule according to the Boltzmann probability associated with the potential which is operative at that iteration; the NOE are back-calculated from the simulation and the potential is then modified to improve the match between the back-calculated and the experimental NOE.

Each step of the iterative scheme starts from the last conformation of the previous step. The new step involves a 10^5 -step steepest-descent minimization of the initial conformation and a MD sampling, as described above. From the simulation $N_c = 5000$ conformations are recorded. To avoid distortion of the peptidic bonds the recorded conformations undergo a 100-step constrained minimization, in which the distances between the pairs of atoms which define the NOE are constrained with a harmonic potential with restraint distance equal to that found in that specific conformation and harmonic constant 10^5 kJ/mol/nm², and to peptide bonds is applied a cosine dihedral potential with energy constant 10^4 kJ/mol.

The NOEs intensities $I^{calc}(i, j)$ directly associated with the sampled conformations are back-calculated by Eq. (1) and compared with the experimental one through

$$\chi^2 = \frac{1}{N} \sum_{ij} \frac{[I^{calc}(i, j) - I^{exp}(i, j)]^2}{\sigma(i, j)^2}, \quad (2)$$

where i, j are the identifiers of the atoms for which a NOE is available, N is the total number of such pairs, $I^{exp}(i, j)$ is the experimental value of the NOE, $\sigma(i, j)$ is its standard error calculated from the triplicate experiment. The dihedral potential is then modified to minimize the χ^2 , following the procedure described below.

In the Gromacs porting of GAFF, the dihedral potential is defined by the Ryckaert-Bellemans function

$$V(\phi_{ijkl}) = \sum_{n=0}^5 C_n^{ijkl} [\cos(\phi_{ijkl} - 180^\circ)]^n, \quad (3)$$

where ϕ_{ijkl} is the dihedral defined by atoms i, j, k and l , and C_n^{ijkl} are the energy parameters which define the potential. We shall modify only those $C_n^{ijkl} \neq 0$ associated with dihedrals

which involve at least one atom present in the list of experimental NOEs.

A random minimization of the χ^2 is performed for 500 steps, modifying the potential (3) according to a reweighting procedure.^{22,31} Specifically, a random change of a C_n^{ijkl} is attempted within 2.5 kJ/mol from the current value; the putative NOE intensities are calculated as

$$I^{calc'}(i, j) = \frac{I_0}{Z \cdot N_c} \sum_{n=1}^{N_c} d(r_i^n, r_j^n)^{-6} \cdot \exp \left[\frac{-V'(r^n) + V(r^n)}{k_B T} \right], \quad (4)$$

where

$$Z = \sum_{n=1}^{N_c} \exp \left[\frac{-V'(r^n) + V(r^n)}{k_B T} \right], \quad (5)$$

$V(r^n)$ is the original dihedral energy of the n th recorded conformations, calculated with the potential used in the MD, $V'(r^n)$ is that calculated with the modified potential, and k_B is Boltzmann constant. This expression allows the algorithm to back-calculate the NOE intensity of an ensemble of conformation interacting with a potential V' even if the conformations r^n were obtained from a sampling carried out a different potential V , thus avoiding the need of a new sampling each time the potential is modified. The χ^2 is calculated between the new intensities $I^{calc'}$ and the experimental ones, and the change in C_n^{ijkl} is accepted if the χ^2 decreases.

The above scheme is then iterated, performing a new MD simulation and a new optimization of the χ^2 until this reaches values comparable to unity. The new MD simulation is necessary because, after the potential has been changed consistently during the minimization procedure from that used for the sampling, the sampled conformations are no longer representative of the equilibrium distribution associated with the modified potential.

Restraint minimization. The results of the present calculations are compared with those resulting from a standard restraint minimization algorithm. A molecular dynamics analysis

starting from the integrated NOE peaks was performed using the program DYANA,¹⁹ that uses simulated annealing combined with molecular dynamics in torsion angle space (torsion angle dynamics), i. e., the numerical solution of the classical mechanical equations of motion (Lagrange equations) with torsion angles as generalized coordinates. The target function takes the role of the potential energy, and the system is coupled to a temperature bath which is cooled down slowly from its initial high temperature, thereby allowing the system to cross barriers between local minima of the target function. The topologies of diketopiperazine scaffolds of RGD mimics were added to the standard amino acid library. DYANA calculations were started from 100 randomized conformations. The 10 conformers with the lowest residual target function values were analyzed.

Docking The Glide program (version 4.5)³² employed in docking calculations holds the protein rigid, while permitting torsional flexibility of the ligand. However, during the docking process, the program considers the cyclopeptide rigid and prevents switching between backbone conformations, whereas the side chains remain free to rotate.

The crystal structure of the extracellular segment of integrin $\alpha_v\beta_3$ in complex with the cyclic peptide Cilengitide (cyclo(-RGDf-NMe-V), 1L5G.pdb)³³ was taken as a reference model for the interpretation of the docking results in terms of ligand–protein interactions. This high affinity ligand (IC₅₀=0.6 nM for the inhibition of biotinylated vitronectin binding to $\alpha_v\beta_3$ receptor¹⁸) in the X-ray crystal structure, bound to the head group of the integrin, features an almost extended conformation of the RGD sequence, with a C β (Arg)–C β (Asp) distance of 0.96 nm.

Glide results for each input ligand conformations were sorted by GlideScore, an empirical scoring function that approximates the ligand binding free energy.

Results

NOEs induce a unique potential which reproduces the experimental intensities.

The iterative modification of the force field is applied to minimize the difference between experimental NOEs and those back-calculated from the model conformations, quantified by a χ^2 . Different runs can reach values of $\chi^2 \approx 1$ (see Fig. S8 in the Supporting Information). Since the χ^2 is weighted by the experimental error estimated in the triplicate NOESY experiment, a $\chi^2 \approx 1$ indicates that the agreement between experimental and model data is within the experimental precision.

A comparison between the experimental and the model NOE intensities, back-calculated at the end of the iterative scheme, is displayed in Fig. 2. The corresponding χ^2 are 1.13 for DKP5-RGD and 0.81 for DKP2-RGD.

An important issue to investigate is to which extent the modified force field is unique. If we had modified in the force field the Lennard–Jones terms, which display the same r^{-6} spatial dependence as the NOEs used as proxy for the quality of the conformational ensemble, a maximum–entropy argument would guarantee the uniqueness of the optimal solution.^{34,35} However, the modification of the Lennard–Jones terms allowed us to decrease the χ^2 only to ≈ 10 (see Fig. S9 in the Supporting Information).

Thus, we have checked explicitly the similarity between the parameters which define the dihedral potentials in different minimizations (see Fig. S10 in the Supporting Information). For DKP5-RGD the correlation coefficient between the set of parameters found in the optimal minimization (which reached $\chi^2 = 1.13$) and the first suboptimal ($\chi^2 = 1.30$) is 0.989, while for DKP2-RGD is 0.971, supporting the unicity of the optimized parameters set.

Moreover, we checked that the energy parameters are robust with respect to the NOEs used for the optimization. Two independent χ^2 minimizations were carried out for DKP5-RGD using, instead of all the available NOEs, a subset in which, respectively, 4 and 8 NOEs were deleted. The resulting energy parameters (cf. Fig. S11) display a correlation of 0.972 and 0.945, respectively, with those obtained using the full set.

DKP5-RGD displays two main conformational states. The ensemble of conformations generated by the simulation can be analyzed from the perspective that each

conformation appears with a probability which is a good approximation of the actual thermodynamic probability. The heatmap reported in Fig. 3 displays the root mean square difference (RMSD) between each pair of conformations, calculated on the atoms of the cycle only. Two clusters of conformations similar within an RMSD of 0.05 nm (cyan points) are apparent; the similarity of the cycles of conformations belonging to different clusters is larger than 0.10 nm. The statistical weight of the two clusters, labelled as A and B, is 64% and 36%, respectively. The two central conformations of each of them is also reported in Fig. 3.

The extension of the side chain of arginine displays a two-state behaviour (cf. lower panel in Fig. 3) and results correlated to the state of the cycle; in state A the side chain is preferentially elongated, while in state B it bends back towards the aromatic ring (cf. also Fig. S12 in the Supporting Information). Vice versa, the side chain of the aspartic acid is rather rigid.

The only long range NOE of DKP5-RGD is between the amide protons NHArg and NHGly (H69-H23). This contact seems to be a good proxy for the formation of the hydrogen bond between H69(NHGly) and O17 (DKPC=O) and corresponding to the β -turn at DKP-Arg of type III H-bonding pattern. In the upper panel of Fig. 4 it is displayed the free energy as a function of the distances NHGly-DKPC=O and NHGly-NHArg. The plot displays two minima, corresponding to clusters A and B identified above. The state B, in which NHGly and NHArg are at a distance ≤ 0.2 nm, also displays NHGly and DKPC=O closer than 0.3 nm. Vice versa, conformations in state A display H69 and H23 further than 0.4 nm, and H69 and O17 further than 0.5 nm. Only a weakly-populated region displays contact NHGly-NHArg formed but not contact NHGly-DKPC=O.

The lower panel of Fig. 4 displays the distribution of the long range NOE distance between NHGly and NHArg. The distribution has a bimodal shape, corresponding to the two states A and B. In the plot we also show with a red line the single distance that one would obtain in the mean-field approximation $\langle d^{-6} \rangle = \langle d \rangle^{-6}$. In other words, when transforming NOE intensities into spatial restrains, one would classify the coupling between NHGly and

NHArg and concluding that the two atoms are permanently in contact, thus neglecting the state in which they are apart. The reason for this effect is the $1/d^6$ modulation of the distance distribution typical of NOE magnetic dipolar coupling (cf Figs. S13-S15 in the Supporting Information).

Among the type-B conformations, some structures display an hydrogen bond between ArgNH and GlyC=O (H23 and O64). We displayed in Fig. 5 the free energy profile of DKP5-RGD with respect to the β -turn distance (NHGly–DKPC=O) and the distance ArgNH and GlyC=O (H23–O64). The two main free-energy minima correspond to the state B, forming the β -turn at DKP-Arg (NHGly–DKPC=O distance < 0.3 nm), and the state A, with neither hydrogen bonds formed (NHGly–DKPC=O and ArgNH–GlyC=O distances > 0.4 nm). The plot displays a third minimum, B2, less deep than the others in which the β -turn is not formed, and its place is taken by the hydrogen bond between ArgNH and GlyC=O (H23–O64 distance < 0.3 nm). The population of this state is 7%. Consequently, the state we labelled as B in Fig. 4 can be further divided into two sub-states, the most populated B1 in which β -turn at DKP-Arg is formed, and B2 in which the hydrogen bond switches to ArgNH and GlyC=O.

DKP2-RGD displays essentially a single state.

The same analysis was applied to DKP2-RGD. The RMSD calculated over the atoms of the cycle between each pair of sampled conformation, displayed in the upper part of Fig. 6, is rather homogeneous and is centered around 0.05 nm. The macrocycle conformation of DKP2-RGD is characterized by an hydrogen bond between NHGly and DKPC=O, closely resembling the type III H-bonding pattern. Also the relative position of the side chains of arginine and aspartic acid is rather fixed (cf. lower panel in Fig. 6); in particular the side chain of arginine is in an elongated position similar to state A of DKP5-RGD (cf. Fig. 3), while the side chain of the aspartic acid is bent back towards the arginine.

As a matter of fact, in the case of DKP2-RGD the distribution of distances between the protons associated with the measured NOEs is mostly unimodal, and the identification

of $\langle d^{-6} \rangle$ with $\langle d \rangle^{-6}$ is a reasonable approximation for this molecule (cf. Fig. S14 in the Supporting Information).

None of the sampled conformations is identical to those obtained with restraint-based techniques.

The ensembles of conformations obtained for the two molecules were also compared with the conformations resulting from restraint minimization, in which restraints are calculated from the NOE intensities following a standard algorithm.¹⁹ The analysis of the 10 best conformers gave a single family of structures for both molecules, displaying a RMSD (calculated on the atoms of macrocycle) of 0.06 nm for DKP5-RGD and 0.09 nm for DKP2-RGD. It is important to notice that there is not any physical reason to regard this set of conformations minimizing the restraints as the thermodynamic fluctuations of the molecule, something that is true only for the ensemble generated with the algorithm described above.

The RMSD between the ensemble of conformations obtained with the present model and the minimal-restraints ones (i.e. that with the best target function according to Dyana) is shown in Fig. 7. In the case of DKP2-RGD the distribution is rather peaked, centered around 0.4 nm. The sampled conformations are, as discussed above, structurally homogeneous, but markedly different from that obtained from restraint minimization. Nonetheless, the restraint-minimized structures of DKP2-RGD showed the formation of the pseudo β -turn motif with the formation of H-bond H56-O17 (NHGly and DKP C=O) for DKP2-RGD (type III H-bonding pattern), also present in our conformational ensemble.

In the case of DKP5-RGD the peak is centered around 0.15 nm. The presence of a single peak in spite of the two conformational clusters observed for this molecule suggests that the minimized-restraint conformation is equidistant from the two clusters described in Fig. 3.

In ref. 18 a slightly different strategy was followed. MC/SD (Monte Carlo/Stochastic Dynamics) simulations were run in implicit water with OPLS/A force field, imposing as single restraint the long range NOE contact between NHGly and NHArg. In these simulations, for both DKP5-RGD and DKP2-RGD more than the 90% of the sampled structures formed

the hydrogen bond between NHGly and DKP C=O (type III H-bonding pattern) also found minimizing all constrains (see above). In particular, the superposition of type III representative conformations of those simulations to those obtained with the present protocol shows for DKP2-RGD an RMSD of 0.08 nm, calculated on C and N atoms of macrocycle, due to a different arrangement of the macrocycle around the Asp residue, while in the case of DKP5-RGD the better overlap was found with state B1 (RMSD=0.05 nm).

Docking calculation into $\alpha_v\beta_3$ binding site.

A study of the role of conformational heterogeneity in the binding of DKP5-RGD and DKP2-RGD to their biological target has been carried out making use of docking calculations. The average structures of DKP2-RGD and of the three states of DKP5-RGD were docked to $\alpha_v\beta_3$ as described in the Materials and Methods (see also caption of Fig. S16 in the Supplementary Information).

The results of the docking were evaluated comparatively to the crystal structure of the potent cyclic pentapeptide ligand of $\alpha_v\beta_3$ (Cilengitide) whose structure is similar to that of DKP5-RGD and DKP2-RGD. In the case of Cilengitide, the positively charged Arg guanidinium group interacts with the negatively charged side chains of Asp218 and Asp150 in the α unit. One carboxylate oxygen atom of the Asp side chain is coordinated to the metal cation in the metal-ion-dependent adhesion site (MIDAS) region of the β unit, while the second carboxylate oxygen atom forms hydrogen bonds with the backbone amides of Asn215 and Tyr122 in the β unit. Further stabilizing interactions can occur that involve the formation of hydrogen bonds between the ligand backbone NH group of the Asp residue and the backbone carbonyl group of Arg216 in the β unit. Assuming that the X-ray crystal structure describes the best interaction mode of the RGD sequence with the $\alpha_v\beta_3$ receptor, the top-ranked poses resulting from docking calculations of DKP2-RGD and DKP5-RGD macrocycle states are evaluated for their ability to reproduce the Cilengitide binding mode.

In the case of DKP5-RGD, the state A produces top-ranked poses forming all the relevant interactions of Cilengitide in the X-ray complex. Besides the classical ligand-receptor elec-

trostatic clamp, further stabilizing interactions occur between one ligand aromatic moiety and the charged side chain of Lys253 β and Arg248 α and between the second aromatic group and the adjacent to MIDAS cation (ADMIDAS) and the residues coordinating the metal (see Fig. 8). This conformation displays an extended arrangement of the RGD sequence, with a C β (Arg)/C β (Asp) distance of 0.91 nm, similar to the RGD bound conformation of Cilengtide (0.89 nm).

Also the state B1 of DKP5-RGD perfectly fits the RGD motif into the receptor site and forms in the top-ranked poses the key polar interactions observed in the X-ray complex (Fig. S16 in the Supporting Information). In general, for this type of macrocycle conformation we obtain docking results similar to those previously achieved using the MC/SD structures. The β -turn at DKP-Arg favors an extended RGD conformations (C β (Arg)/C β (Asp) distance of 0.89 nm).

On the other hand, docking calculation starting from the less populated state B2 of DKP5-RGD shows top-ranked poses conserving only one of the two important electrostatic interactions with the $\alpha_v\beta_3$ (see Fig. S17 of the Supporting Information). The short C β (Arg)-C β (Asp) of 0.75 nm distance probably prevents the guanidine and carbonyl groups from achieving the required separation for binding to the $\alpha_v\beta_3$ integrin.

DKP2-RGD adopts a single conformation forming the same β -turn at DKP-Arg obtained by MC/SD restrained simulations. As expected, this extended structure (C β -C β distance of 0.87 nm) forms in the top-ranked poses all the key crystallographic interactions (see Fig. S18 of the Supporting Information). However, compared to MC/SD geometry, docking results of DKP2-RGD starting from the structure obtained by our protocol show a different orientation of the molecule within the binding site (see Fig. S19 of the Supporting Information). Probably due to the different arrangement of the macrocycle around the Asp residue, to form the classical electrostatic clamp with the receptor the molecule shifts the Gly residue away from the receptor and places the aromatic ring toward the ADMIDAS cation.

Overall, the presence in DKP5-RGD of a conformational state with suboptimal binding

mode to its target could explain its slightly higher (12.2 ± 5.0 nM) IC₅₀ with respect to DKP2-RGD (3.2 ± 2.7 nM).

Discussion and Conclusion

The biological activity of compounds is related to their physicochemical, structural and conformational properties in solution. NMR is the principal biophysical technique used to analyze such properties, by delineating the molecular structure in solution by through-space connectivities. Small molecules can show a high internal mobility, with transitions between several conformational states. If the conformations are being averaged over the mixing time of the experiment, the NOEs will reflect the average of the inverse sixth power of interatomic distances and the conformational properties of the molecules will be difficult to elucidate (cf Fig. S13 in the Supporting Information). In particular, the standard interpretation of NOEs will result in the determination of a single conformation, in which inter-nuclear distances will appear smaller than their true average value. In this work we describe an integrated computational/experimental approach to optimize the exploration of the conformational space of a small molecule.

Computational methods are in principle the perfect complement to NMR data to obtain structural information about flexible systems, particularly in the case of small molecules, for which the computational sampling is particularly handy. However, the approximations contained in general-purpose force fields make the use of computational methods limited to provide a qualitative picture of the conformational properties of the molecules. This is particularly true for non-peptidic molecules, whose force fields did not experience the intense development that peptidic force fields did. In fact, the simulation of the dynamics of DKP5-RGD and DKP2-RGD controlled only by the force field, produced NOEs which are not compatible at a quantitative grade with those measured experimentally.

An iterative molecular-dynamics scheme was applied to correct the GAFF force field to

match experimental NOEs to quantitative degree in implicit solvent. At variance with the standard approach of optimizing NOE-derived restrained, the present scheme produces an ensemble of conformations weighted by Boltzmann distribution, and thus thermodynamically self-consistent. Here it is not the NOE intensity calculated on a single conformation which matches the experimental value but, more realistically, the NOE intensities averaged over a large number of conformations.

This scheme is very versatile, and can be applied not only to NOEs, but also to an heterogeneous set of experimental inputs, including also J-couplings, residual dipolar couplings, etc. For larger molecules, whose NOEs can be affected by spin diffusion, it can be used straightforwardly in connection with more sophisticated NOESY experiments, like eNOEs.²⁵

It is complementary to the approach based on optimizing the match between observables averaged simultaneously on several replicas of the system and the experimental data.⁴ With respect to this, our scheme is computationally heavier but displays the advantage of ending up not only in a set of conformations compatible with the data but also into an effective potential. This potential is optimized once, and eventually can be used for multiple tasks, like calculating free-energy differences, binding properties to other molecules, etc. Moreover, it is a starting point to create a portable potential which is compatible with the experimental data in a quantitative way. The reason why the energy parameters of the dihedral potential are different for the same quadruplets of atoms in two different molecules is likely to be a dependence on further degrees of freedom, beyond the quadruplet. A systematic application of our strategy to a large number of molecules could help to highlight such a dependence.

A modest aggravation of the present method with respect to the standard treatment of NOEs is the need to repeat independent acquisitions of the NOESY spectrum to calculate the standard error on the NOE intensities. However, this is an important step in a quantitative approach like the present, because the standard errors allow one to weight the corresponding data in a more correct way. Other techniques based on Bayesian inference have been recently

developed to treat experimental uncertainty,³⁶⁻³⁹ which could be eventually integrated with the present one.

Two ligands of $\alpha_v\beta_3$ integrin were analyzed and the ensemble of obtained conformations was used to rationalize their binding affinity. The two molecules differ only for a benzylic substitution of DKP scaffold. Common methods for conformational analysis, based on the interpretation of NOEs in terms of distance restraints, gave a single conformation for both molecules. Overall, the nanomolar affinity of these RGD peptidomimetics for $\alpha_v\beta_3$ could be attributed to their high structural pre-organization. The docking calculations performed with these preorganized conformation show a good positioning of both ligands into $\alpha_v\beta_3$ binding site (both molecules reproduce the relevant interactions observed for the reference ligand in the X-ray complex), which can be correlated with their nanomolar affinity in the competitive binding assays with biotinylated vitronectin and isolated receptor. Moreover, compared to DKP2-RGD, DKP5-RGD is about four times less effective as ligand of $\alpha_v\beta_3$. The slight lower affinity of DKP5-RGD was here explained in terms of a high conformational equilibrium and in particular due to the presence of a less populated state (B2) with a low pre-organization for the binding to $\alpha_v\beta_3$ integrin.

The ability of the present strategy to interpret NOE intensities in terms of different conformational states, each associated with its equilibrium probability, makes it a useful complement to docking calculations. Docking algorithms used to interpret the affinity data usually lack a rigorous sampling of molecular conformers, specially for conformationally constrained macrocycles. Moreover, the starting conformation strongly affects docking results when the transition between conformers is prevented by high energy barriers or if it requires the formation of transient distorted structures to occur. As recently discussed in ref.,⁴⁰ in the case of acyclic ligands all possible conformations can theoretically be obtained through systematic scan of all rotatable bonds. However, the transition from one macrocycle conformer to another typically requires the concerted rotation of multiple dihedrals. Docking analyses that systematically scan predetermined rotamer libraries for side chain and acyclic ligand

conformations are unable to resolve the concerted motions required for macrocycle conformation sampling. As a result, most docking procedures do not rigorously sample macrocycle conformations, leading to incomplete description of relevant binding modes. The use of multiple conformers in docking calculations has become the best solution to this limitation and a robust treatment of macrocycle conformation is increasingly regarded as an essential step prior to docking efforts. The algorithm we suggested to achieve thermodynamically-weighted conformational ensembles from NOE data of cyclic RGD peptidomimetics offers a valuable and thorough solution to tackle this key step in macrocycle docking.

Supporting Information. Figures with details of the experiments and of the calculations are provided as Supporting Information. This material is available free of charge via the Internet at <http://pubs.acs.org/>.

References

- (1) Venditti, V.; Egner, T. K.; Clore, G. M. Hybrid approaches to structural characterization of conformational ensembles of complex macromolecular systems combining NMR residual dipolar couplings and solution X-ray scattering. *Chem. Rev.* **2016**, *116*, 6305–6322.
- (2) Ringkjøbing Jensen, M.; Salmon, L.; Nodet, G.; Blackledge, M. NMR characterization of long-range order in intrinsically disordered proteins. *J. Am. Chem. Soc.* **2010**, *132*, 1270–1272.
- (3) Ozenne, V.; Bauer, F.; Salmon, L.; Huang, J.; Ringkjøbing Jensen, M.; Segard, S.; Beranrdó, P.; Charavay, C.; Blackledge, M. Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. *Bioinformatics* **2012**, *28*, 1463-1470.
- (4) Camilloni, C.; De Simone, A.; Vranken, W. F.; Vendruscolo, M. Determination of

- secondary structure populations in disordered states of proteins using NMR chemical shifts. *Biochemistry* **2012**, *51*, 2224–2231.
- (5) Eliezer, D. Biophysical characterization of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **2009**, *19*, 23–30.
- (6) Fennen, J.; Torda, A. E.; van Gunsteren, W. F. Structure refinement with molecular-dynamics and a Boltmann-weighted ensemble. *J. Biomol. NMR*, **1995**, *6*, 163–170.
- (7) Torda, A. E.; Scheek, R. M.; van Gunsteren, W. F. Time-averaged nuclear Overhauser effect distance restraints applied to tendamistat. *Chem. Phys. Lett.*, **1989**, *157*, 289–294.
- (8) Mantsyzov, A. B.; Maltsev, A. S.; Ying, J.; Shen, Y.; Hummer, G.; Bax, A. A maximum entropy approach to the study of residue-specific backbone angle distributions in α -synuclein, an intrinsically disordered protein. *Protein Sci.* **2014**, *23*, 1275–1290.
- (9) Vogeli, B.; Segawa, T. F.; Leitz, D.; Sobol, A.; Choutko, A.; Trzesniak, D.; van Gunsteren, W.; Riek, R. Exact distances and internal dynamics of perdeuterated ubiquitin from NOE buildups. *J. Am. Chem. Soc.* **2009**, *131*, 17215–17225.
- (10) Vögeli, B.; Olsson, S.; Güntert, P.; Riek, R. The exact NOE as an alternative in ensemble structure determination. *Biophys. J.*, **2016**, *110*, 113–126.
- (11) Vögeli, B. The Nuclear Overhauser Effect from a quantitative perspective *Prog. Nuc. Magn. Res. Spectr.* **2014**, *78*, 1–46.
- (12) Kessler, H. Conformation and biological activity of cyclic peptides *Angew. Chem.* **1982**, *21*, 512–523.
- (13) Bruschweller, R.; Blackledge, M.; Ernst, R. R. Multi-conformational peptide dynamics derived from NMR data: a new search algorithm and its application to antamanide *J. Biomol. NMR* **1991**, *1*, 3–11.

- (14) Cicero, D. O.; Barbato, G.; Bazzo, R. NMR Analysis of molecular flexibility in Solution: A new method for the study of complex distributions of rapidly exchanging conformations. Application to a 13-residue peptide with an 8-residue loop *J. Am. Chem. Soc.* **1995**, *117*, 1027–1033.
- (15) Kolmer, A.; Edwards, L. J.; Kuprov, I.; Thiele, C. M. Conformational analysis of small organic molecules using NOE and RDC data: A discussion of strychnine and alpha-methylene-gamma-butyrolactone. *J. Magn. Res.* **2015**, *261*, 101–190.
- (16) Boelens, R.; Vuister, G. W.; Konig, T. M. G.; Kaptein, R. Observation of spin diffusion in biomolecules by three-dimensional NOE–NOE spectroscopy. *J. Am. Chem. Soc.* **1989**, *111*, 8525–8526.
- (17) Ressurreicao, A. S. M.; Vidu, A.; Civera, M.; Belvisi, L.; Potenza, D.; Manzoni, L.; Ongeri, S.; Gennari, C.; Piarulli, U. Cyclic RGD-Peptidomimetics Containing Bifunctional Diketopiperazine Scaffolds as New Potent Integrin Ligands. *Chemistry* **2009**, *45*, 12184-12188.
- (18) Marchini, M.; Mingozi, M.; Colombo, R.; Guzzetti, I.; Belvisi, L.; Vasile, F.; Potenza, D.; Piarulli, U.; Arosio, D.; Gennari, C. Cyclic RGD Peptidomimetics containing bifunctional diketopiperazine scaffolds as new potent integrin ligands. *Chemistry* **2012**, *19*, 3563-3567.
- (19) Güntert, P.; Mumenthaler, C.; Wüthrich, K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* **1997**, *273*, 283–298.
- (20) Wakefield, A. E.; Wuest, W. M.; Voelz, V. A. Molecular simulation of conformational pre-organization in cyclic RGD peptides. *J. Chem. Inf. Model.* **2015**, *55*, 806–813.
- (21) Zagrovic, B.; van Gunsteren, W. F. Comparing atomistic simulation data with the NMR experiment: how much can NOEs actually tell us? *Proteins Struct. Funct. Bioinf.* **2006**, *63*, 210–218.

- (22) Norgaard, A. B.; Ferkinghoff-Borg, J.; Lindorff-Larsen, K. Experimental parameterization of an energy function for the simulation of unfolded proteins *Biophys. J.* **2008**, *94*, 182–192.
- (23) Giorgetti, L.; Galupa, R.; Nora, E. P.; Pilot, T.; Lam, F.; Dekker, J.; Tiana, G.; Heard, H. Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. *Cell* **2014**, *157*, 950-963.
- (24) Capelli, R.; Paissoni, C.; Sormanni, P.; Tiana, G. Iterative derivation of effective potentials to sample the conformational space of proteins at atomistic scale *J. Chem. Phys.* **2014**, *140*, 195101-1–195101-9.
- (25) Vögeli, B.; Segawa, T. F.; Leitz, D.; Sobol, A.; Choutko, A.; Trzesniak, D.; van Gunsteren, W.; Riek, R. Exact distances and internal dynamics of perdeuterated ubiquitin from NOE buildups. *J. Am. Chem. Soc.* **131**, 17215–17225.
- (26) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general Amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- (27) Case, D. A.; Cheatham, T. E. ; Darden, T.; Gohlke, H.; Luo, H.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26*, 1668–1688.
- (28) Onufriev, A.; Bashford, D.; Case, D. Exploring protein native states and large-scale conformational changes with a modified generalized Born model. *Proteins: Struct. Funct. Gen.* **2004**, *55*, 383–394.
- (29) da Silva A. W. S.; Vranken W. F. ACPYPE-Antechamber Python parser interface. *BMC Research Notes* **2012**, *5*, 367-1–367-12.
- (30) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.;

- Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010**, *78*, 1950–1958.
- (31) Tiana, G.; Villa, F.; Zhan, Y.; Capelli, R.; Paissoni, C.; Sormanni, P.; Heard, E.; Giorgetti, L.; Meloni, R. MonteGrappa: An iterative Monte Carlo program to optimize biomolecular potentials in simplified models. *Comp. Phys. Comm.* **2015**, *186*, 93-104
- (32) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem.* **2004**, *47*, 1750–1759.
- (33) Xiong, J. P.; Stehle, T.; Zhang, R.; Joachimiak, A.; French, M.; Goodman, S. L.; Arnaout, M. A. Crystal structure of the extracellular segment of integrin alpha Vbeta3 in complex with an Arg-Gly-Asp ligand. *Science* **2002**, *296*, 151–155.
- (34) Pitner, J. W.; Chodera, J. D. *J. Chem. Theo. Comp.* **2012**, *8*, 3445–3451.
- (35) Tiana, G.; Amitai, A.; Pollex, T.; Piolot, T.; Holcman, D.; Heard, E.; Giorgetti, L. Structural fluctuations of the chromatin fiber within topologically associating domains *Biophys. J.* **2016**, *110*, 1234–1245.
- (36) Voelz, V. A.; Zhou, G. Bayesian inference of conformational state populations from computational models and sparse experimental observables *J. Comp. Chem.* **2014**, *35*, 2215–2224.
- (37) Beauchamp, K. A.; Pande, V. S.; Das, R. Bayesian energy landscape tilting: Towards concordant models of molecular ensembles *Biophys. J.* **2014**, *106*, 1381–1390.
- (38) Bonomi, M.; Camilloni, C.; Cavalli, A.; Vendruscolo, M. MetaInference: A Bayesian inference method for heterogeneous systems. *Science Adv.* **2016**, *2*, e1501177.
- (39) MacCallum, J. L.; Perez, A.; Dill, K. A. Determining protein structures by combining

semireliable data with atomistic physical models by Bayesian inference. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 6985–6990.

- (40) Allen, S. E.; Dokholyan, N. V.; Bowers, A. A. Dynamic docking of conformationally constrained macrocycles: Methods and applications. *ACS Chem. Biol.* **2016**, *11*, 10–24.

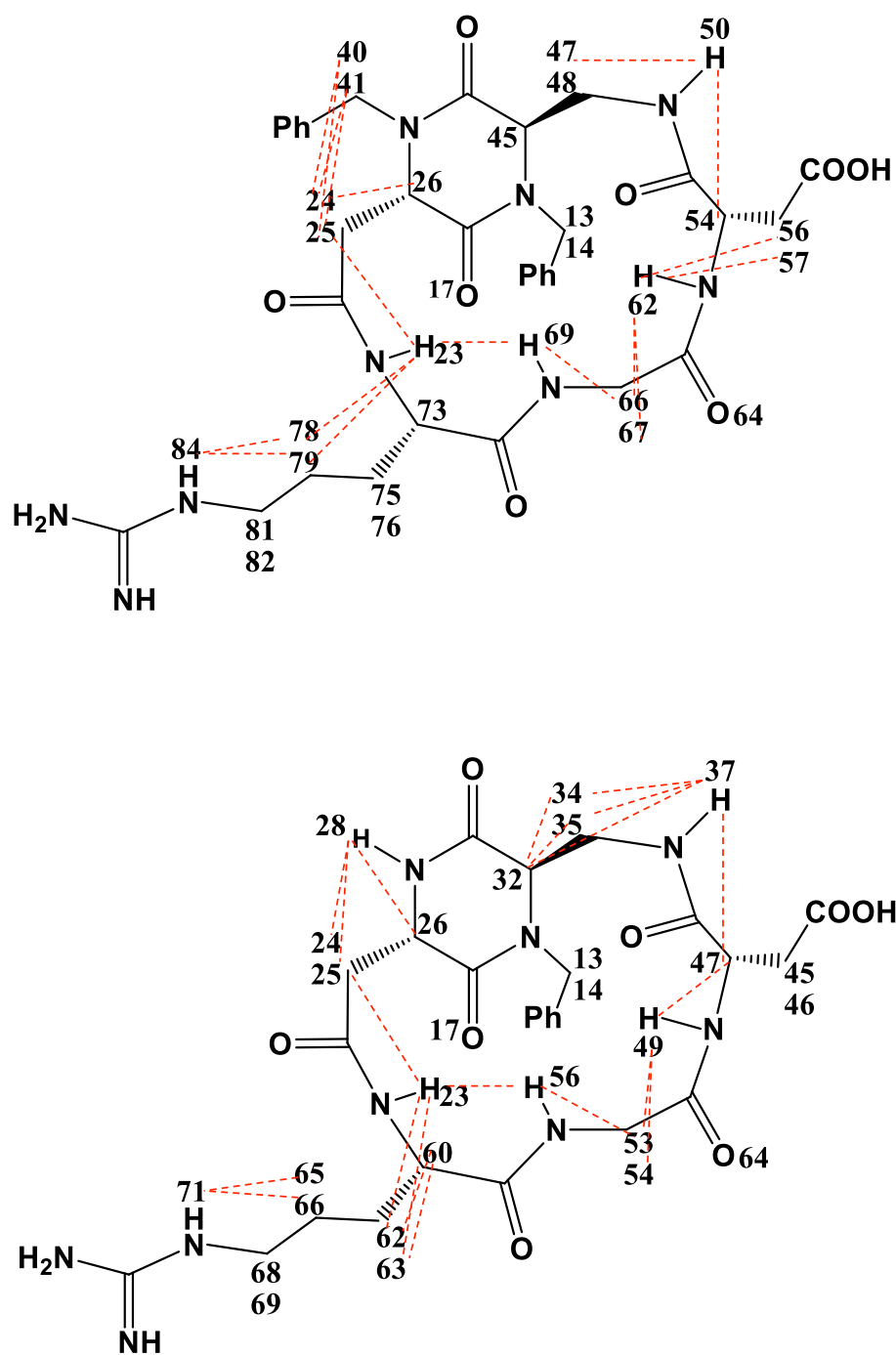


Figure 1: The structure of DKP5-RGD (upper panel) and DKP2-RGD (lower panel). Dashed red lines indicate the available NOEs .

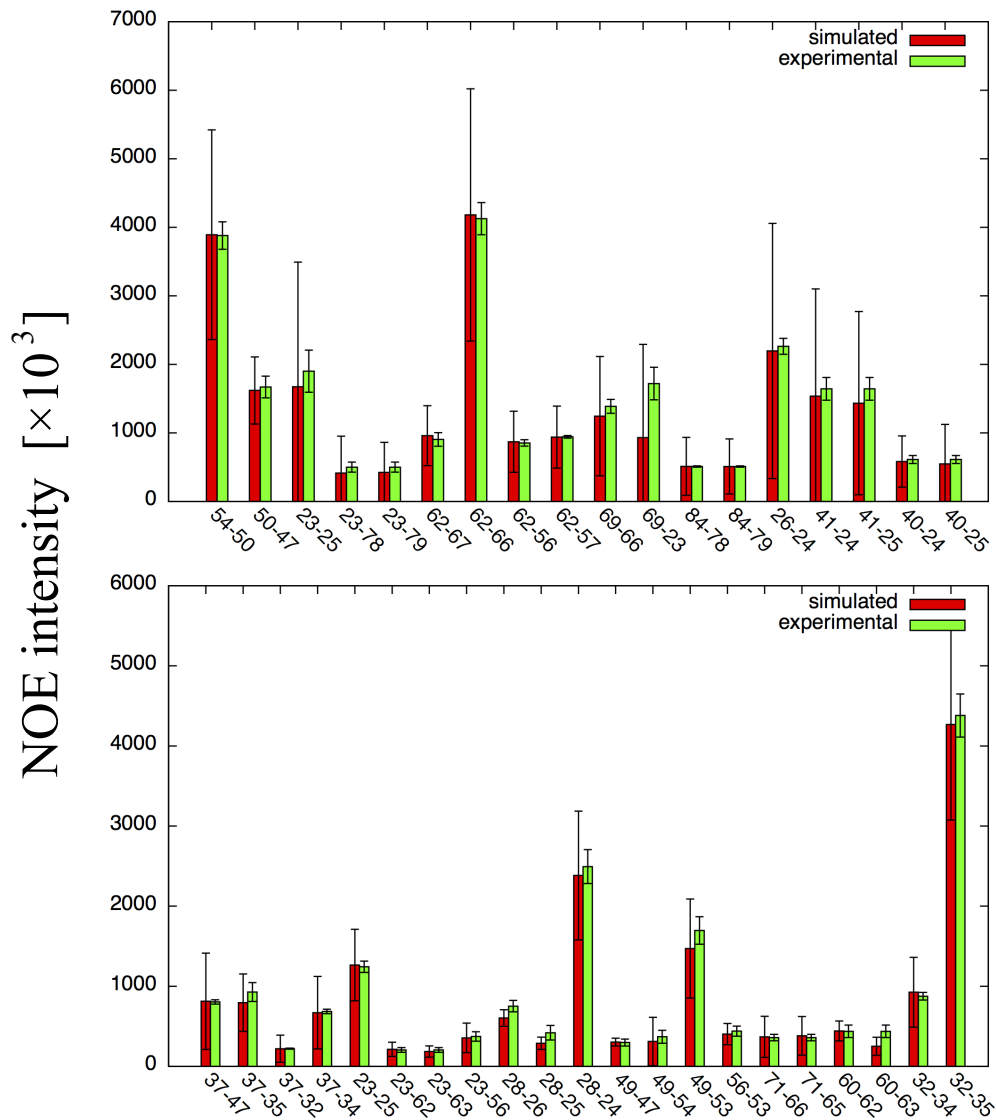


Figure 2: A comparison between the experimental (green bars) and back-calculated (red bars) NOE intensities for DKP5-RGD (above) and DKP2-RGD (below). The error bars in the experimental data indicate the standard error obtained from the triplicate experiment; those in the calculated intensities indicate the standard deviation associated with conformational fluctuations.

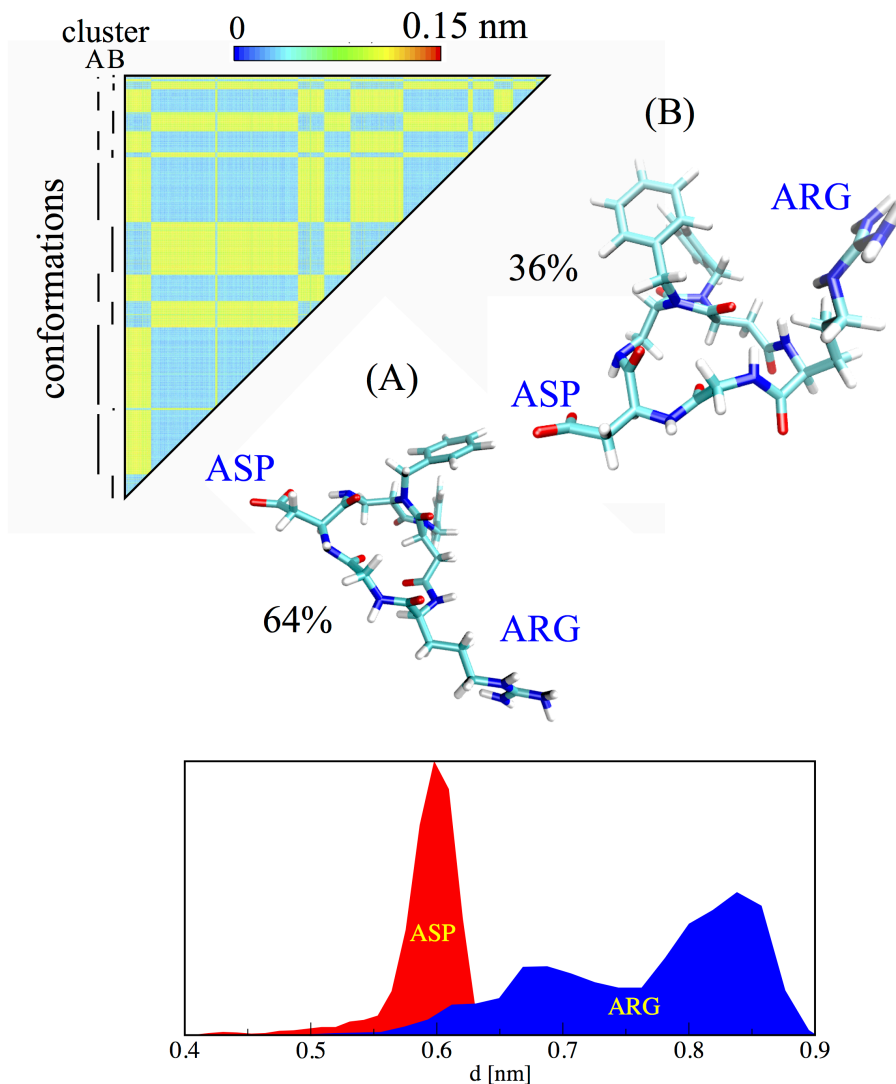


Figure 3: The DKP5–RGD heatmap reports the RMSD calculated on the atoms belonging to the cycle between each pair of conformations generated by the simulation. The reported molecular models are the central conformations of each of the two clusters arising from the heatmap. The lower histogram displays the distribution of distances between the centre of mass of the side chain of arginine (in blue) and of the aspartic acid (in red), and that of the cycle.

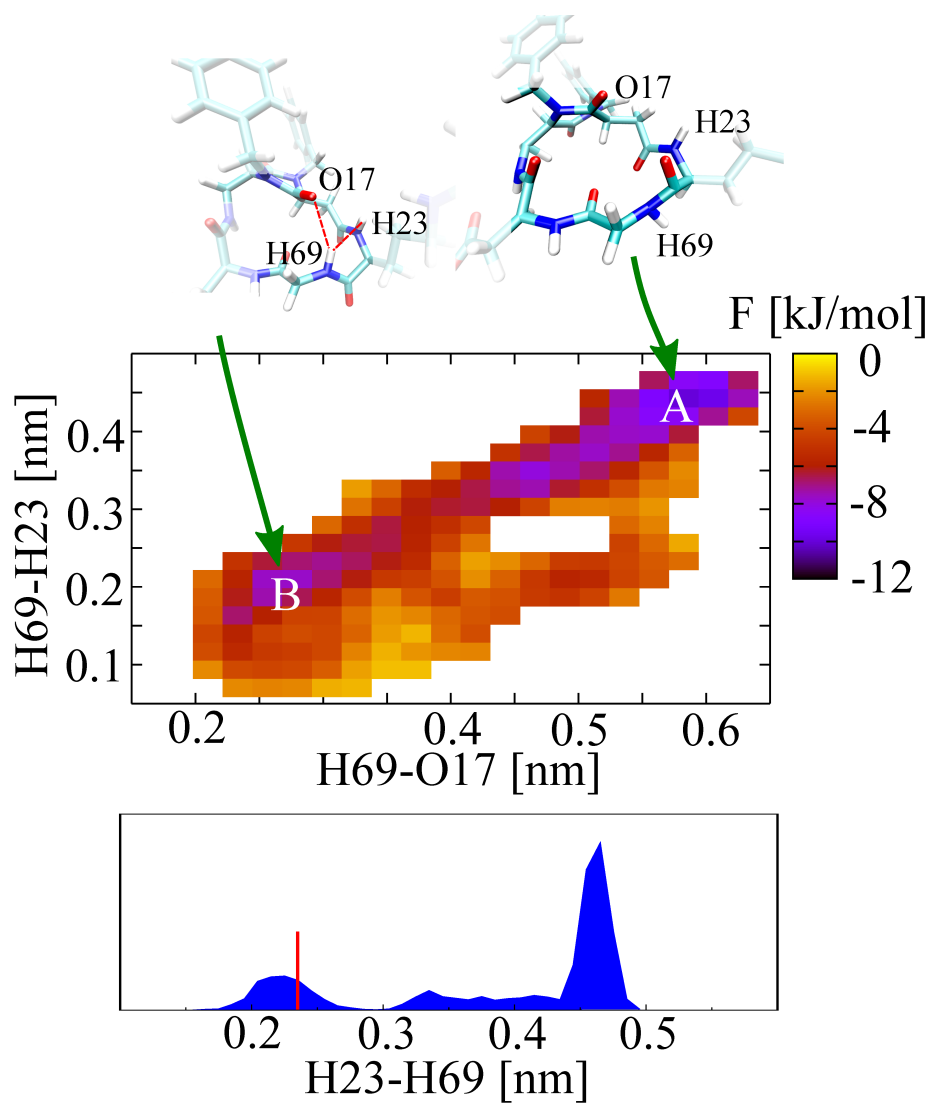


Figure 4: The free energy profile of DKP5–RGD as a function of the distance between H69 and O17 and of the distance between H69 and H23. The plot displays two main minima, corresponding to the two states of the hydrogen bond between O17 and H69.

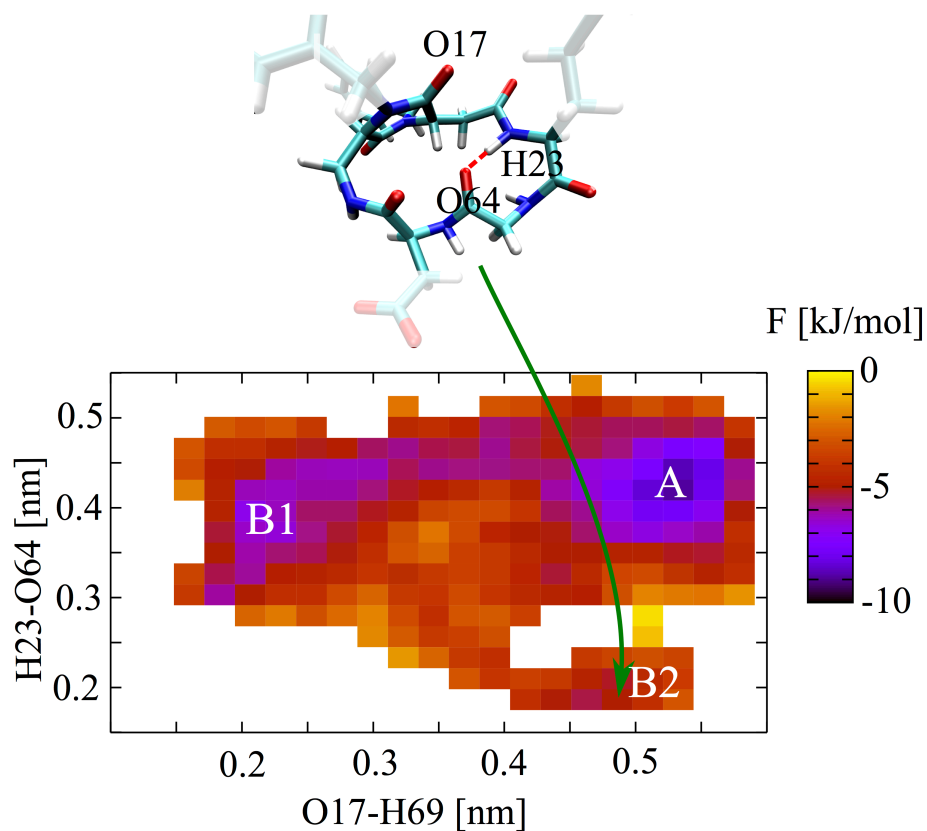


Figure 5: The free energy profile of DKP5-RGD as a function of the distance between β -turn at DKP-Arg (H69-O17) and of the distance between ArgNH and GlyC=O (H23-O64). The plot displays two main minima: the most populated A with no β -turn (H69-O17 > 0.4 nm) and B1 with the β -turn (H69-O17 < 0.3 nm). A third minority minimum B2 is shown with no β -turn and with the hydrogen bond ArgNH/GlyC=O .

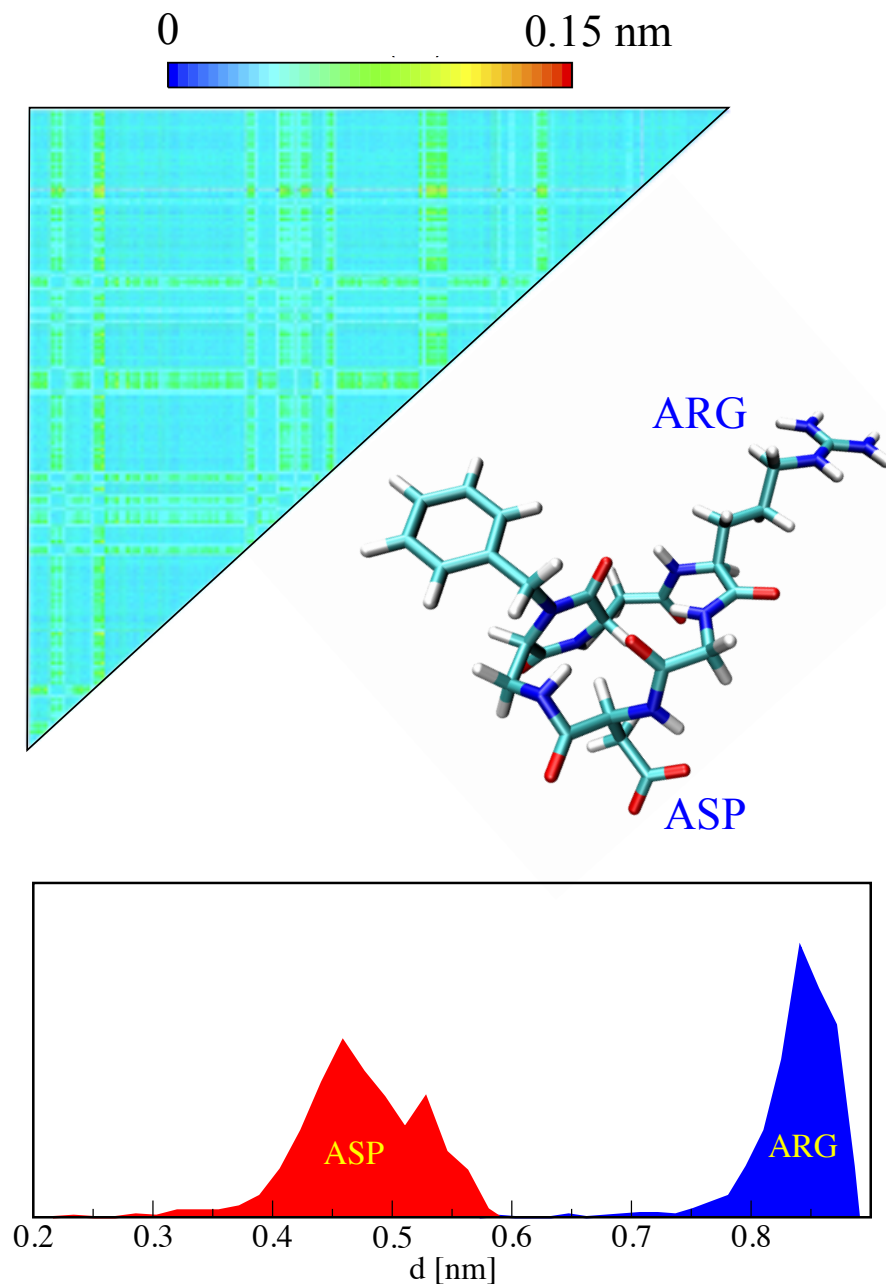


Figure 6: The DKP2-RGD heatmap shows the mutual RMSD of the cycle atoms of each pair of sampled conformation. The displayed structure is the central structure obtained from the above heatmap. The histogram below displays the distribution of distances between the centre of mass of the side chain of arginine (in blue) and aspartic acid (in red) with respect to the centre of mass of the cycle.

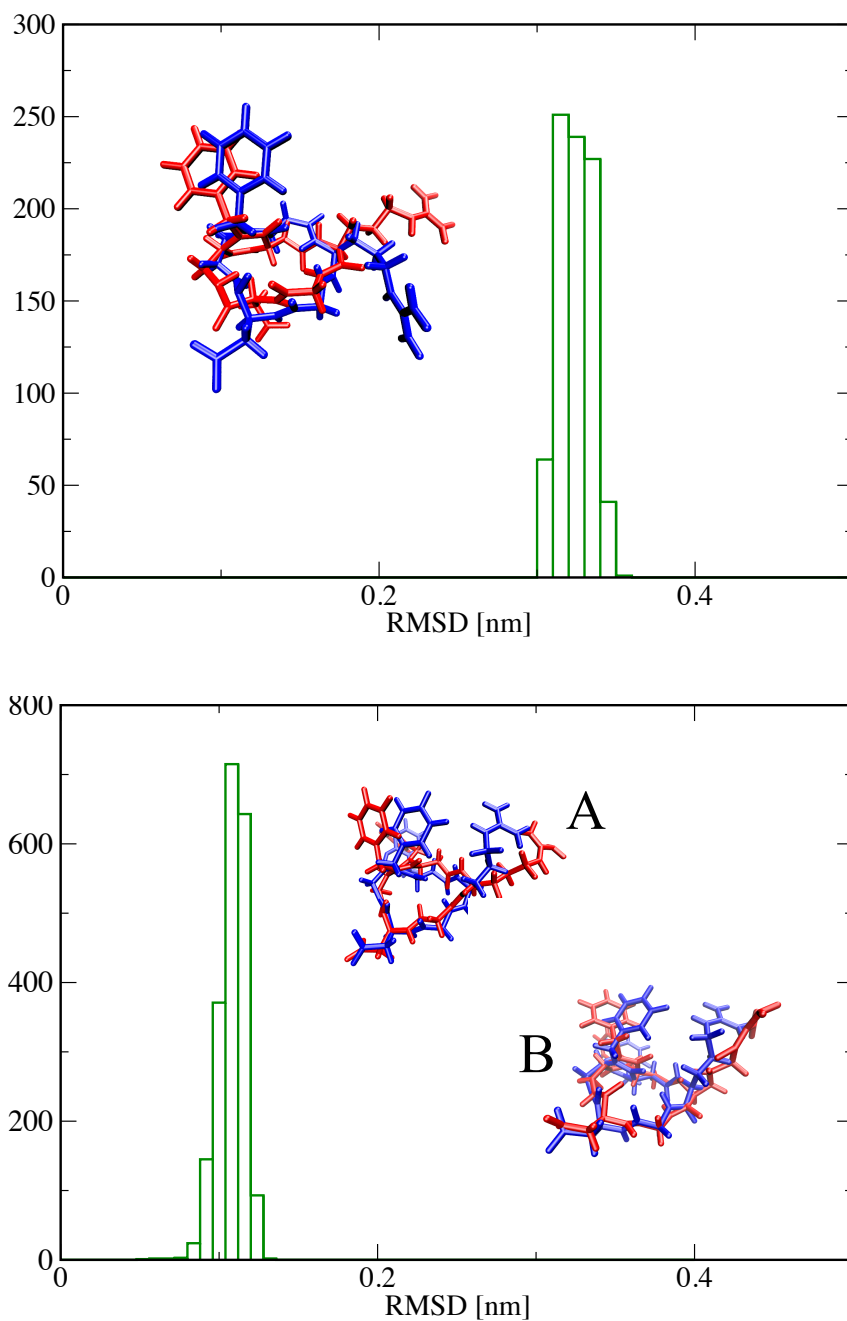


Figure 7: The distribution of RMSD over backbone atoms between the conformation obtained by restraint minimization and the set of conformations obtained with the present model for DKP2-RGD (above) and DKP5-RGD (below). The blue structures are those obtained with restraint minimization, while the red structures are the central conformations obtained from the sampling.

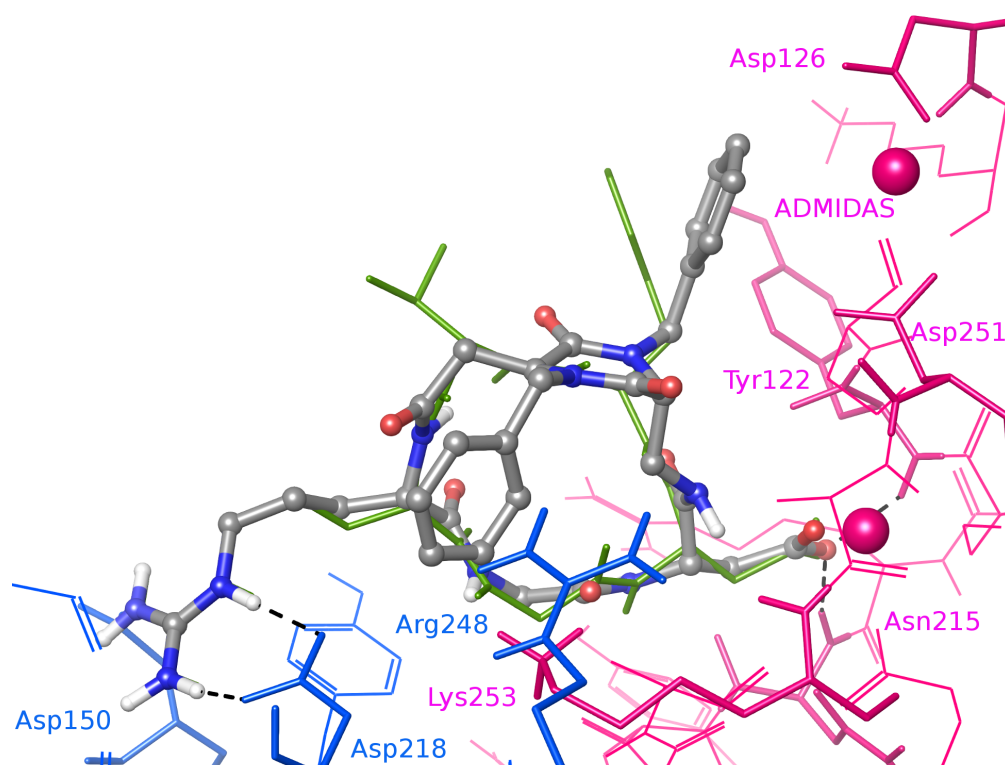


Figure 8: Best pose of DKP5-RGD state A (grey, Gscore -9.3 kcal/mol) into $\alpha_v\beta_3$ integrin overlaid to Cilengitide (green). Residues involved in the interactions with ligand are represented as tube and metal ions as CPK sphere (α subunit in light blue, β subunit in purple).

Table of Contents Image

