Electronic Supplementary Information

Mechanism of RGD-conjugated nanodevice binding to its target protein integrin $\alpha_V\beta_3$ by atomistic molecular dynamics and machine learning

Giulia Frigerio,[a] Edoardo Donadoni,[a] Paulo Siani,[a] Jacopo Vertemara,[b] Stefano Motta,[c] Laura Bonati,[c] Luca De Gioia,[b] Cristiana Di Valentin[a,d*]

[a] *Dipartimento di Scienza dei Materiali, Università di Milano-Bicocca, via R. Cozzi 55, 20125 Milano, Italy.*
[b] *Dipartimento di Biotecnologie e Bioscienze, Università di Milano-Bicocca, Piazza della Scienza 1, 20126 Milan, Italy.*
[c] *Dipartimento di Scienze dell'Ambiente e del Territorio, Università di Milano-Bicocca, Piazza della Scienza 1, 20126 Milan, Italy.*
[d] *BioNanoMedicine Center NANOMIB, Università di Milano-Bicocca, Italy.*
[*] *Corresponding author: cristiana.divalentin@unimib.it*

S1. Setting up the MD simulations

S1.1 Rigid-receptor docking calculation details

Docking calculations were performed through the Schrödinger Suite programs included in Maestro (release 2021-1). The integrin structure was extracted from the crystal structure PDBID:1L5G and prepared using the *Protein Preparation Wizard*.[1] In this step, protein protonation states were assigned using Epik for pH=7.0. The ligand molecule was prepared with the *LigPrep* tool. The receptor grid box for docking was obtained with *Receptor Grid Generation* tool and centered on the centroid of cilengitide, the ligand which was co-crystallized with integrin $\alpha_V\beta_3$ in PDBID:1L5G. Rigid-receptor docking was performed at extra-precision (XP) with the *Glide* software[2–4] and OPLS_2005 force field (FF).[5] One metal–protein bond constraint was used for the central metal atom of the pocket at MIDAS.

S1.2 Rigid-receptor docking results

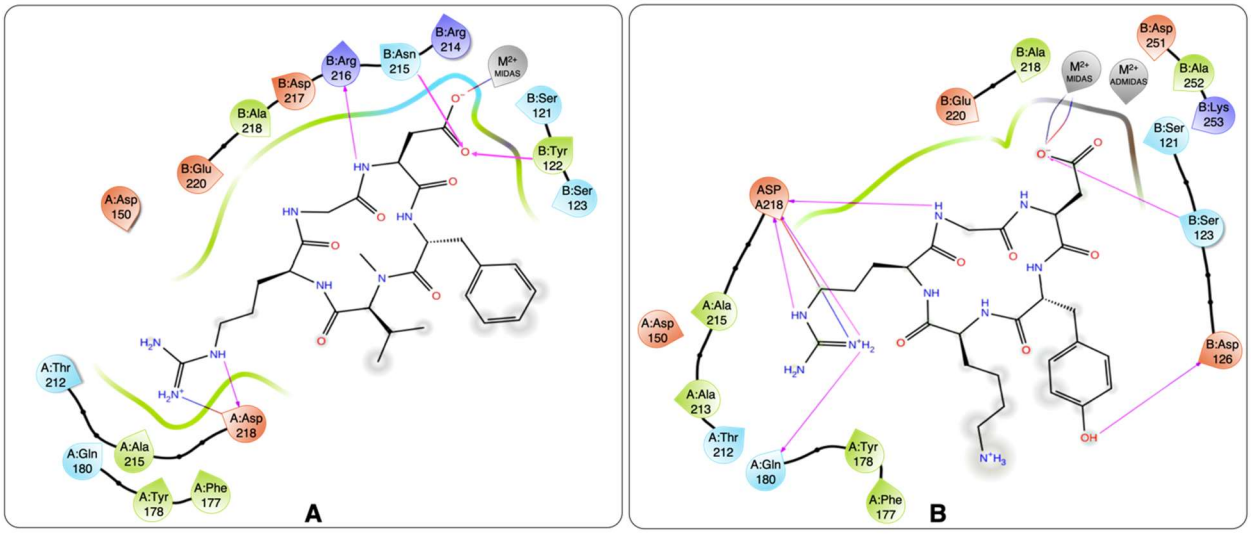

Figure S1. Interaction diagrams of (A) cilengitide in complex with $\alpha_V\beta_3$ as found in the crystal structure corresponding to PDBID: 1L5G and (B) of cRGD in complex with $\alpha_V\beta_3$ as found in the best docking pose from our calculations (Glidescore: –9.1 kcal mol^{-1}). Both diagrams are adapted from Maestro interaction diagrams. Colors code: pink arrows are drawn from donors to acceptors of H-bonds, blue-to-red shadowing lines represent salt bridges between positively and negatively charged groups, respectively, grey lines represent interactions with ions, and grey shadows highlight the atoms exposed to the solvent. Orange is used for negatively charged polar residues, blue for positively charged polar ones, light blue for neutral polar ones and green for non-polar ones. The number of the residues is preceded by the letters A and B when the residue belongs to α and β subunit, respectively. M^{2+} is the divalent cation at MIDAS which interacts with Asp side chain.

As the common routine prescribes for protein–ligand interactions, we first tried to dock cRGD into the binding pocket of integrin $\alpha_V\beta_3$ structure, which was taken from the crystal structure of the extracellular segment of integrin $\alpha_V\beta_3$ in complex with cilengitide (CGT).[6] We performed the docking calculation, as described in Section 2.1.1 of the main text, in order to get a reasonable starting-point geometry for the *cRGD MD simulation*. Figure S1 shows the interactions between CGT and integrin

$\alpha_V\beta_3$ binding site found in the crystal structure corresponding to PDBID:1L5G (Figure S1A) and the best docking pose of cRGD into the integrin $\alpha_V\beta_3$ binding site of the latter crystal structure (Figure S1B). Our best docking pose reproduces the main interactions found in the reference crystal structure both for the arginine with A:Asp218, on the left side of the molecule, and for the aspartate with the divalent ion at MIDAS (M^{2+}), on the right side of the molecule. However, it misses the H-bonds of one of the two O atoms of the aspartate carboxylate group with B:Asn215 and with B:Tyr122, which are crucial for the binding. Therefore, we decided to start all our MD simulations from the crystal structure geometry for the RGD portion of cRGD, which is a more reliable one, on the basis of the results in Section 3.1 of the main text.

S1.3 REMD computational details

We ran two REMD simulations, one for cRGD and one for PEG_{150}-cRGD, which is composed of c(RGDyK) and a three-monomeric-unit long PEG chain (MW ~ 150 Da). Each molecule was explicitly solvated in a TIP3P water dodecahedron box and a Cl^- ion was added to counterbalance the charge in the case of cRGD. In the case of PEG_{150}-cRGD, the PEG tail was kept fixed in order to mimic the effect of being constrained to a heavy NP. The two systems were minimized. To explore the T range 300-440 K with an exchange probability of ~0.3, we set up 32 and 48 replicas for cRGD and PEG_{150}-cRGD, respectively. We ran a 2 ns equilibration for each replica and then ran the REMD simulations trying the exchanges and saving the trajectory every 1ps for a simulation time of 100 ns per replica, yielding a total simulation time of 3.2 µs and 4.8 µs for cRGD and PEG_{150}-cRGD, respectively. We confirm that we obtained an average exchange probability of ~0.3 and ~0.4 for cRGD and PEG_{150}-cRGD, as a consequence of the good overlap between the energy histograms at different temperature values in Figure S2. All simulations were performed with GROMACS 2022 code.

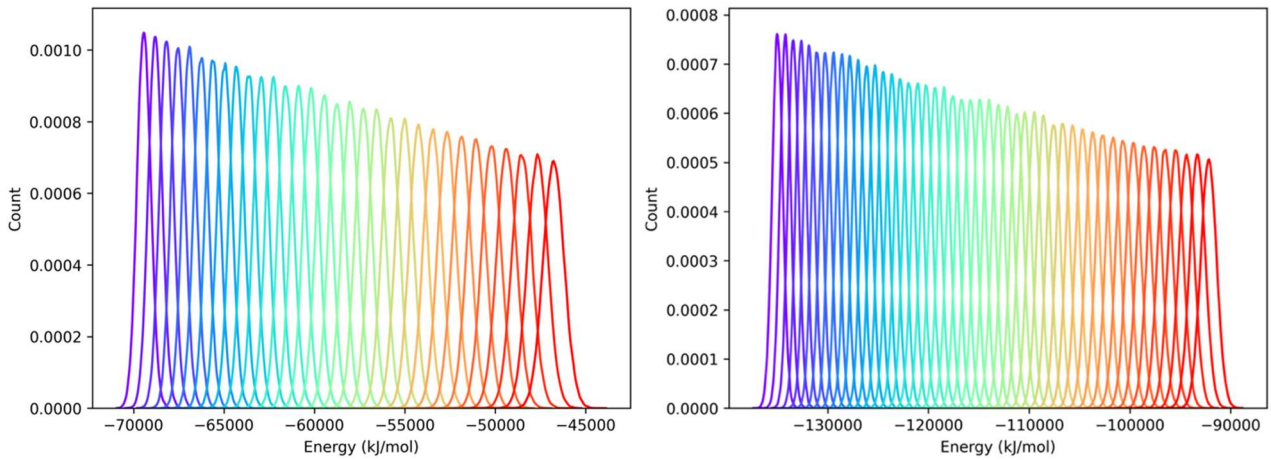

Figure S2. Energy histograms for the 32 and 48 replicas of cRGD (left) and PEG_{150}-cRGD (right), respectively. The profiles are colored according to the increasing temperature value going from violet (T ~ 300 K) to red (T ~ 440 K).

In both the REMD simulations, we found that the replicas are exploring all the temperatures range, in the sense that the structure starting at a certain T value assumes all the other T values (Figure S3)

and, on the other side, a certain T value is assumed by all the replicas at different simulation times (Figure S4).

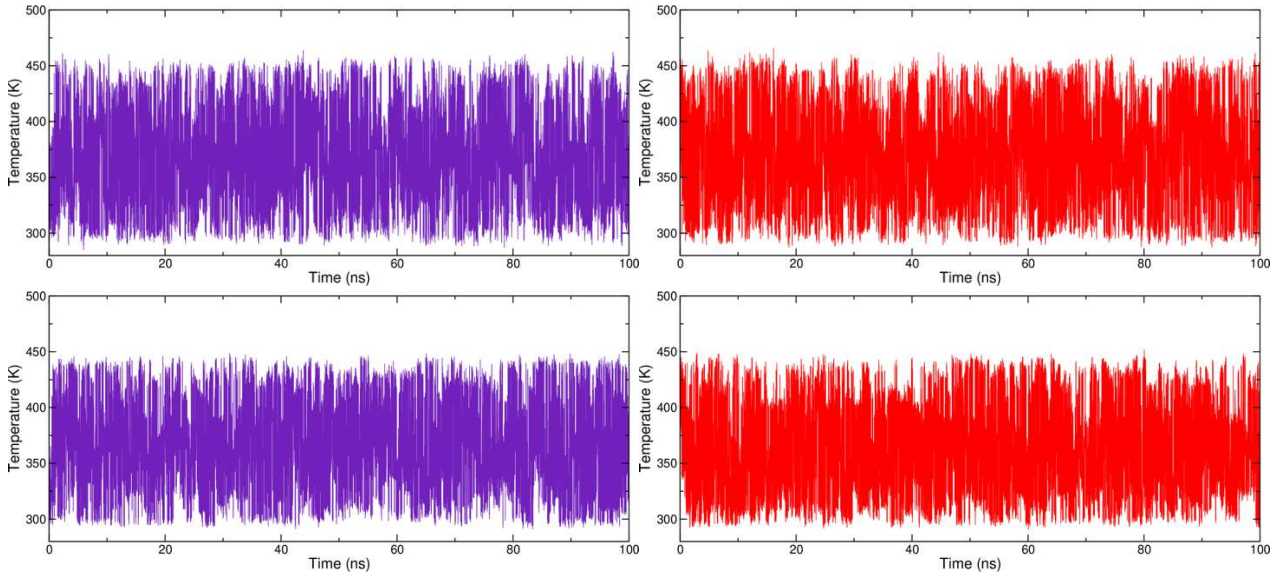

Figure S3. Temperature assumed by the structure starting at the lowest T (left) and at the highest T (right) for cRGD (top) and PEG_{150}-cRGD (bottom).

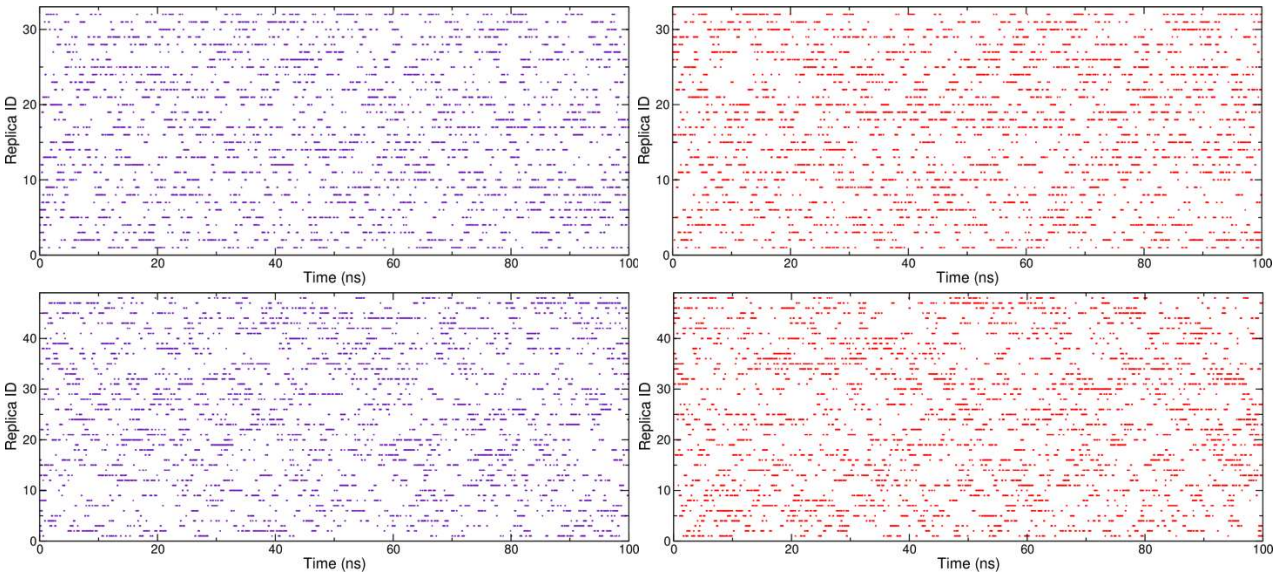

Figure S4. Replica ID at the lowest T value (left) and at the highest T value (right) along the simulation for cRGD (top) and PEG_{150}-cRGD (bottom).

This means that a good annealing of the system is realized and, therefore, we decided to use only the lowest T replica for the analysis, i.e., the trajectory composed of all the frames at T = 300 K. We measured the backbone dihedral angles values along the above-mentioned trajectory with *CPPTRAJ* and we performed a conformational cluster analysis on the basis of the coordinates of backbone heavy atoms through *gmx cluster* with the GROMOS algorithm as implemented in the GROMACS 2022 code with a cutoff of 0.07 nm.

Upon finding that Arg ψ and Gly ϕ angles can assume two principal values for cRGD in solution (Figure 3 of main text), we checked the convergence of our simulations by proving that Arg ψ and

Gly ϕ dihedral angles are switching from one value to the other one both in the case of cRGD and of PEG_{150}-cRGD (Figure S5).

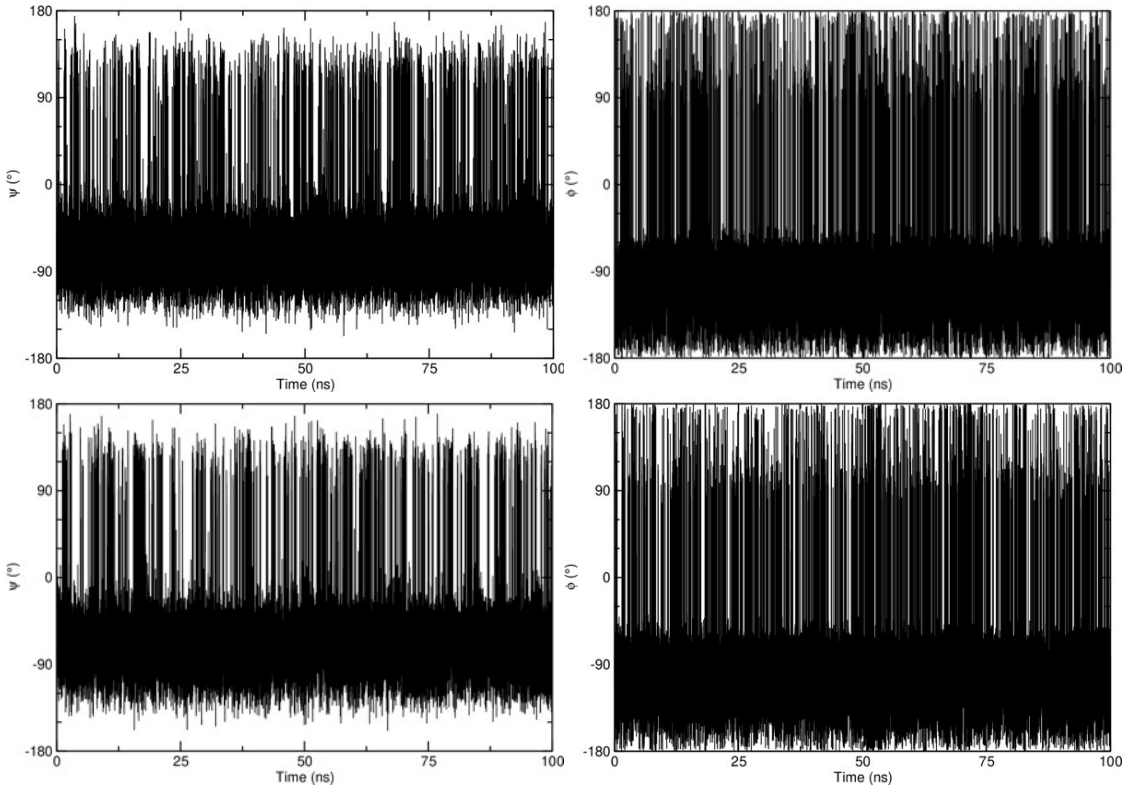

Figure S5. Arg ψ (right) and Gly ϕ (left) dihedral angles along the lowest T replica simulation for cRGD (top) and PEG_{150}-cRGD (bottom).

S1.4 Ions parameters

Being the Asp–Ca^{2+}_{MIDAS} interaction a critical binding feature, we tested two different sets of parameters for Ca^{2+} ions running a 200 ns long MD-simulation for each set and using the crystal structure as the starting-point geometry, as it has been done for all the simulations performed in this work. Table X reports the two sets of parameters tested: Ca^{2+} original CHARMM36m parameters[7–9] and Ca^{2+} CHARMM36m parameters with the NBFIX [10] correction for LJ cross-interactions between Ca^{2+} and O atoms of carboxylate groups (atom name: OD). This NBFIX correction, which is usually included in CHARMM36m FF, consists in increasing by 5% the value of σ with respect to the one calculated with Lorentz-Berthelot combining rules. Figure S6 shows the distance from the aspartate carboxylate O atoms (OD1 and OD2) to Ca^{2+}_{MIDAS} ion and to the backbone N atoms of A:Asn215 the MD simulations with the two sets of parameters.
It is evident that the original CHARMM36 parameters (Figure S6A) perform better than the corrected ones in keeping stable H-bonds and metal–O interactions.

Table S1. FF parameters tested in this work for the cations found in integrin $\alpha_V\beta_3$ structure.

Force Field	CHARMM36m	CHARMM36m + NBFIX
Ion M^{2+}	Ca^{2+}	Ca^{2+}
q (e)	+2.0	+2.0
σ (Å)	2.43572	2.43572
ε (M^{2+}-M^{2+}) (kJ/mol)	0.50208	0.50208
σ (M^{2+}-$O_{carboxylate}$) (Å)	2.73239	2.87938 (+5%)
ε (M^{2+}-$O_{carboxylate}$) (kJ/mol)	0.50208	0.50208

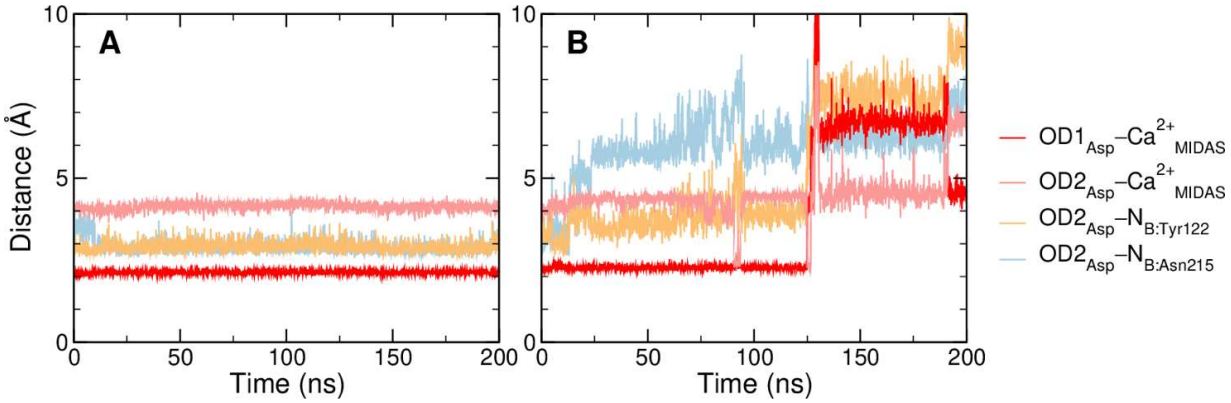

Figure S6. Distances from the aspartate carboxylate O atoms ($OD1_{Asp}$ and $OD2_{Asp}$) to $Ca^{2+}{}_{MIDAS}$ ion and to the backbone N atoms of A:Asn215 the MD simulations with the two sets of parameters: (A) CHARMM36m, (B) CHARMM36m with NBFIX.

To validate the FF parameters of the selected ions, we calculated the following descriptors for the three divalent cations adjacent to the binding site: RDF, coordination number and water survival probability.

Figure S7 shows the radial distribution functions g(r) and the cumulative between $Ca^{2+}{}_{MIDAS}$ and surrounding O atoms. The g(r) profiles feature a sharp peak between 2.0 and 2.5 Å from the ion corresponding to the first solvation shell and a second broad peak between 4.0 and 5.0 Å for the second solvation shell. The g(r) behavior is the one expected for an ion partially coordinated by the carboxylate oxygen atoms and partially solvated by water,[11] however it does not converge to the bulk value of 1 for water because the ion is immersed in a quite different environment from the one typical of a water solution, it is rather buried between the protein and the ligand.

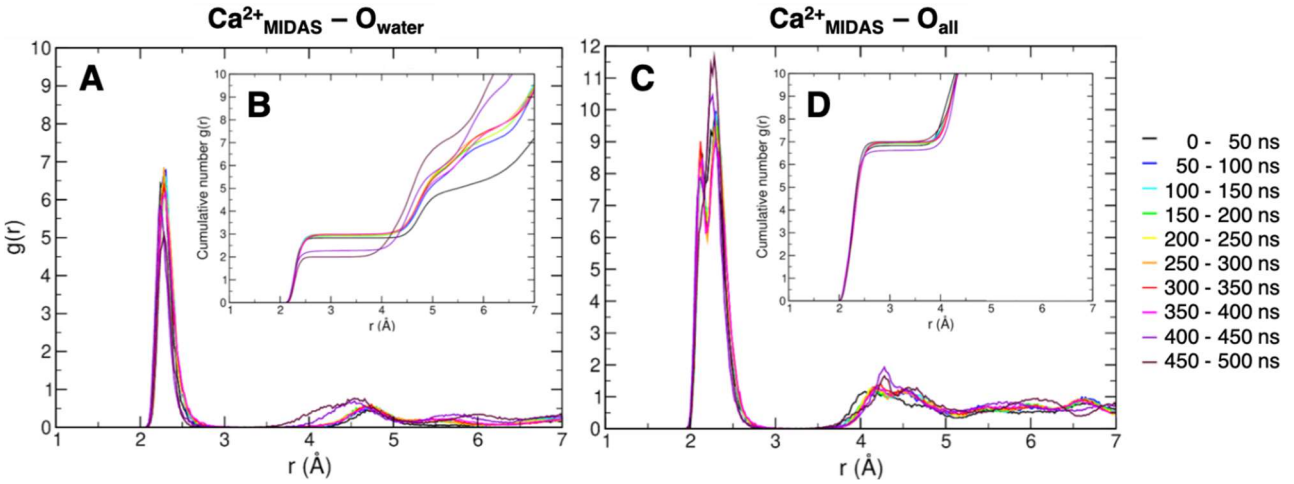

Figure S7. Radial distribution functions g(r) and cumulative g(r) functions, in the insets, between Ca^{2+}_{MIDAS} and O atoms of water (A,B) or all O atoms (C,D) for subsequent time intervals.

From the cumulative g(r) functions in the insets of Figure S7, it is possible to calculate Ca^{2+}_{MIDAS} coordination number, which is the value of the function in the plateau region at r values between the first and the second peak of g(r). The number of water molecules around the ion is equal to 3 for the *bridging* binding mode (black to pink lines in Figure S7B) and decreases to 2 for the *chelating* binding mode (violent and maroon lines in Figure S7B), simultaneously with the formation of an additional cRGD–ion interaction by Asp – see Section 3.2.1 of the main text for the definition of *bridging* and *chelating* binding modes. The overall coordination number is constant and equal to 7 for all Ca^{2+} ions (Figure S7D, Table S2) as previously found for Ca^{2+} ions in water, either free or coordinated by carboxylate groups.[11,12] The number of water molecules belonging to the first solvation shell of the cations at both Ca^{2+}_{LIMBS} and $Ca^{2+}_{ADMIDAS}$ sites is equal to 2, with the other coordinating groups belonging to the protein. Interestingly, the water molecule which is lost by the Ca^{2+}_{MIDAS} ion when shifting from the *bridging* to the *chelating* mode goes to the Ca^{2+}_{LIMBS} ion.
Craig et al. suggested that one water molecule tightly coordinated to the Ca^{2+}_{MIDAS}, which is found also in our simulations, may have the functional role of shielding the latter interaction from water molecules attacks, which are frequent being integrin binding pocket rather shallow.[13]

Table S2. Coordination number of cations at LIMBS, MIDAS and ADMIDAS.

Coordination site	cRGD O	Protein O	Water O	Total
Ca^{2+}_{LIMBS}	0	*Bridging*: 6 *Chelating*: 5	*Bridging*: 1 *Chelating*: 2	7
Ca^{2+}_{MIDAS}	*Bridging*: 1 *Chelating*: 2	*Bridging*: 3 *Chelating*: 3	*Bridging*: 3 *Chelating*: 2	7
$Ca^{2+}_{ADMIDAS}$	0	5	2	7

Another important descriptor is the water survival probability (SP), or residence probability, around a cation for a prolonged time. The survival probability of water around cations is defined as the probability for a water molecule to remain in a sphere of a radius of 3.0 Å centered on the cation. The SP is given by:

$$SP(\tau) = \frac{1}{T}\sum_{t=1}^{T}\frac{N(t, t+\tau)}{N(t)}$$

where T is the maximum time of simulation, τ is the timestep, $N(t)$ the number of particles at time t, and $N(t,t+\tau)$ is the number of particles at every frame from t to τ. The SP was calculated with the MDAnalysis *waterdynamics* module dedicated in its latest version.[14] The higher is the probability at higher times, the lower is the exchange rate of water molecules around the ion. In other words, if the probability decays fast it means that water molecules are exchanging fast in the ion coordination shell, and vice versa. In fact, the residence probability profile for Ca^{2+}_{LIMBS} ion decays slower than for Ca^{2+}_{MIDAS} ion, and the latter one in turn goes to zero slower that the one of $Ca^{2+}_{ADMIDAS}$ ion (Figure S8). This means that the water mobility around the three ions increases in this order: Ca^{2+}_{LIMBS} < Ca^{2+}_{MIDAS} < $Ca^{2+}_{ADMIDAS}$, as expected from their exposure to the solvent.

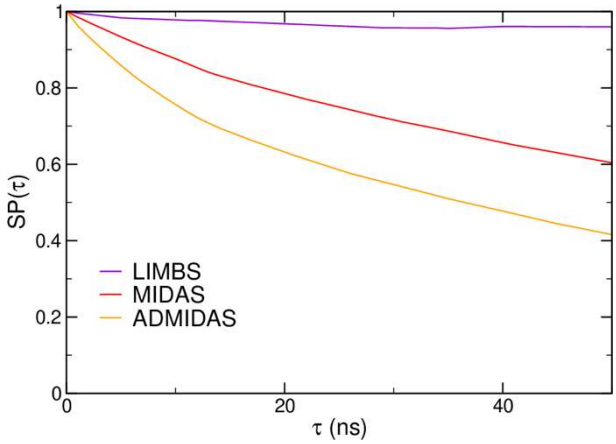

Figure S8. Survival probability of water molecules within a spherical region with a radius of 3.0 Å around Ca^{2+} ions at different sites.

S2. cRGD backbone conformational analysis

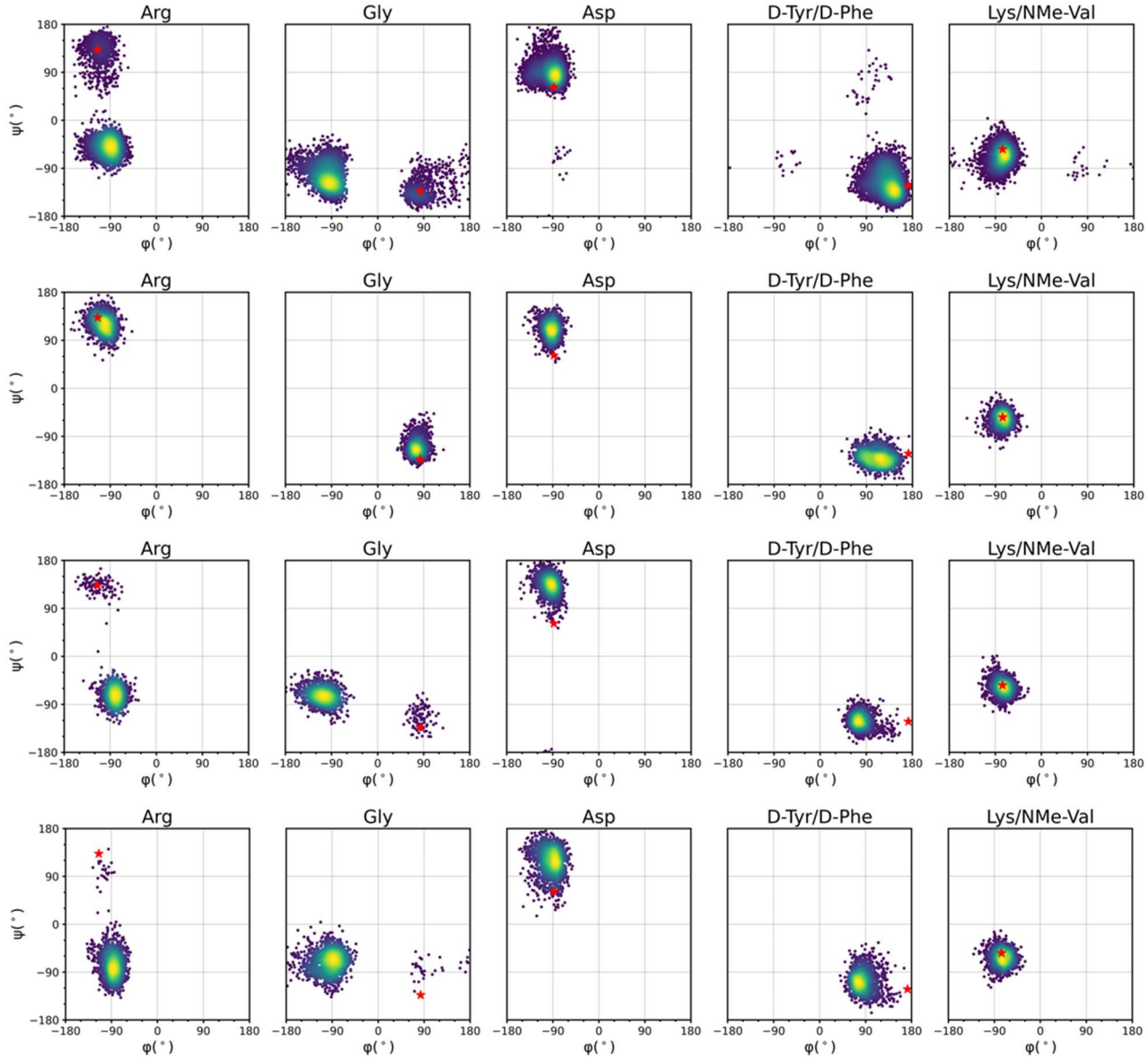

Figure S9. Ramachandran dihedral angles density plots of the ligand in *cRGD MD simulation* and *NP-cRGD MD simulations* at increasing cRGD density, from top to bottom, respectively.

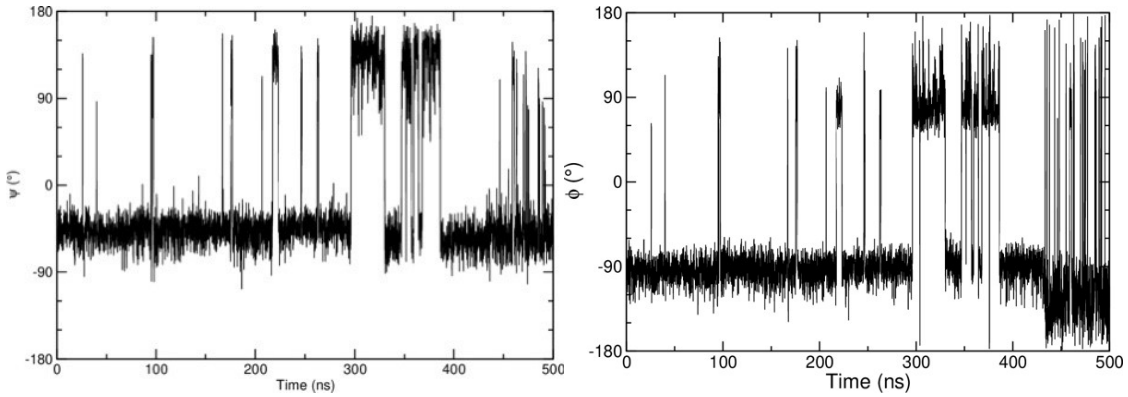

Figure S10. Time evolution of Arg ψ (left) and Gly ϕ (right) dihedral angles in *cRGD MD simulation*.

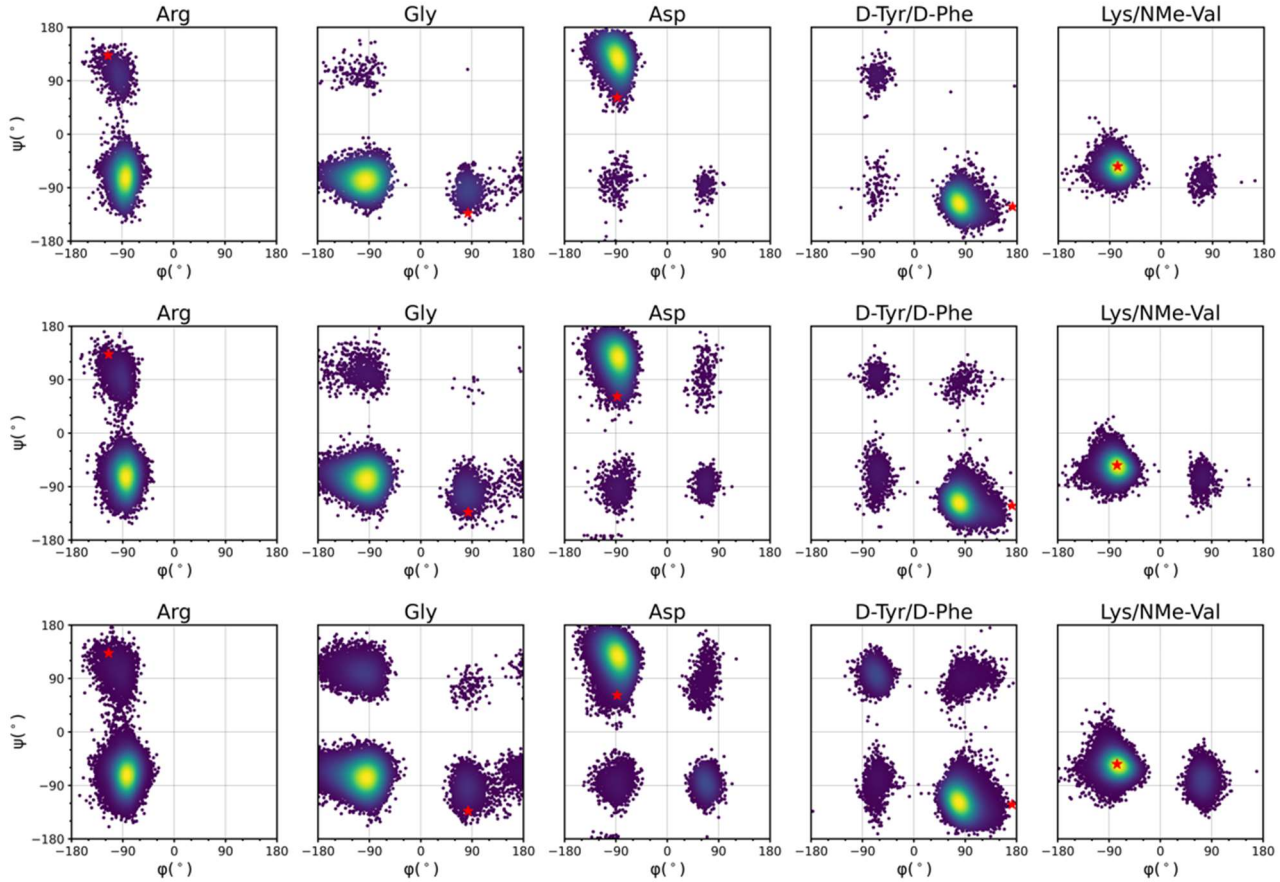

Figure S11. Ramachandran dihedral angles density plots of the other cRGDs in *NP-cRGD MD simulations* at increasing cRGD density, from top to bottom, respectively.

S3. SOM trained with only *cRGD MD simulation*

A Self-Organizing Map (SOM) was also trained with a set of intermolecular distances measured along the only *cRGD MD simulation*, as described in Section 2.2.2. Figure S12 shows the atoms which were selected to train the SOM and the trained SOM is reproduced in Figure S13B. In this case, 49 atoms were selected and 10 clusters were found.

Each neuron (hexagon) of the SOM represents a ligand/protein conformational microstate and neurons close to each other represent similar conformations. As it is evident from Figure S13A the binding mode is quite stable and conserved, only the conformations in the bottom-right corner are significantly different from the other ones. These conformations correspond to the *chelating* binding mode, as shown in the inset corresponding to Cluster C in Figure S13D. From this analysis, we can conclude that the binding mode of the aspartate side chains is the most relevant interaction in discriminating cRGD binding modes.

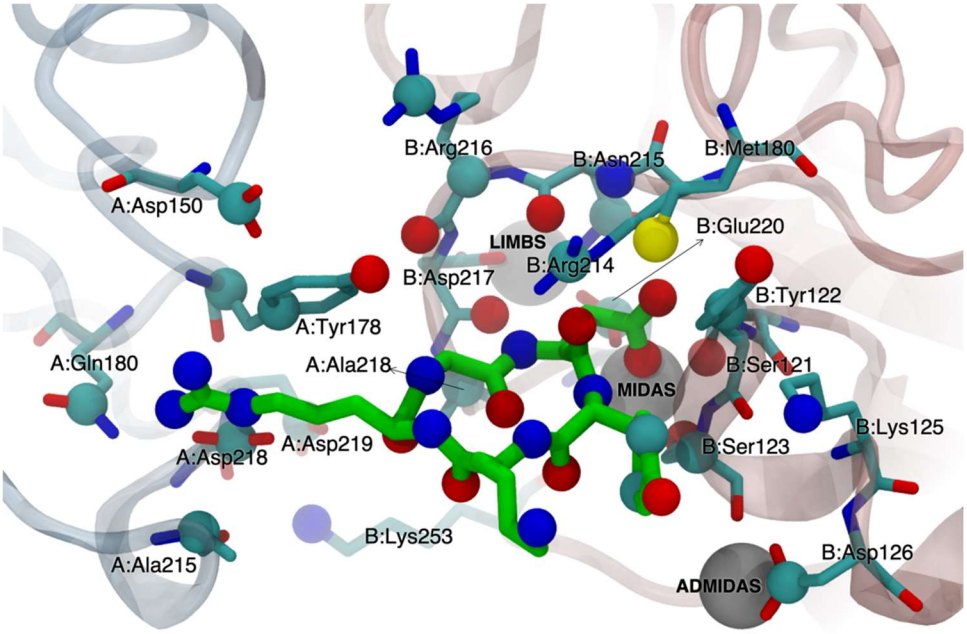

Figure S12. Atoms selected for the distance calculation used in the SOM training. Selected atoms are shown as spheres (19 atoms for the cRGD ligand and 30 atoms for the protein). The C atoms of cRGD are shown in green.

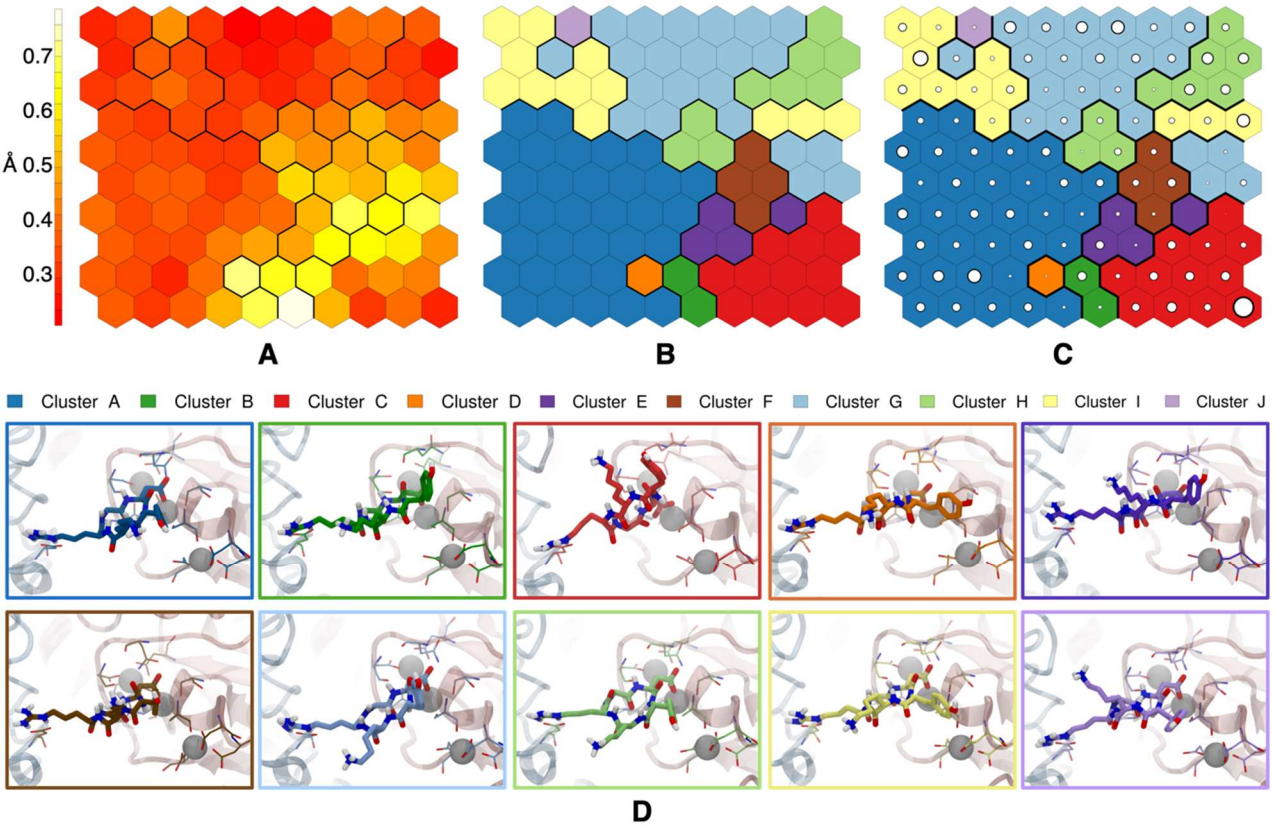

Figure S13. The SOM trained with the binding conformations assumed by the cRGD/integrin complex along the MD simulation. (A) Neighbors distance map in which each neuron is colored according to its average distance from neighboring neurons. (B) SOM with the number of each neuron and cluster of neurons represented by colors and thick black lines. (C) Population map where the size of the circle in each neuron is proportional to its population. (D) Snapshots of the representative binding mode of each cluster of the map extracted from the MD simulation.

S4. Additional results

S4.1 The binding of cRGD to integrin $\alpha_V\beta_3$: Steered MD simulations

For the *bridging* mode, the rupture of cRGD-protein binding (Figures S14A,B) in SMD simulation happens between 200 and 300 ps and requires a force of ~800 kJ mol^{-1} nm^{-1}, which corresponds to ~500 pN, much higher than the one used in experiments (100-200 pN).[15] In particular, the arginine interactions break simultaneously at 225 ps (Figure S14C), while the aspartate interactions first shift to the *chelating* binding mode for few picoseconds and then definitely break at 260 ps (Figure S14D). We, then, performed independent MD simulations starting from different configurations along the unbinding pathway and, since the *bridging* binding mode is overall maintained (Figure S14F), we calculated the free energy profile in Figure S14E. The few umbrella simulations in which OD1 and OD2 share the same distance value from Ca^{2+}_{MIDAS}, corresponding to the *chelating* binding mode, are not discarded because they reproduce the intermediate binding situation also found around 225 ps in the SMD simulation (Figure S14D).

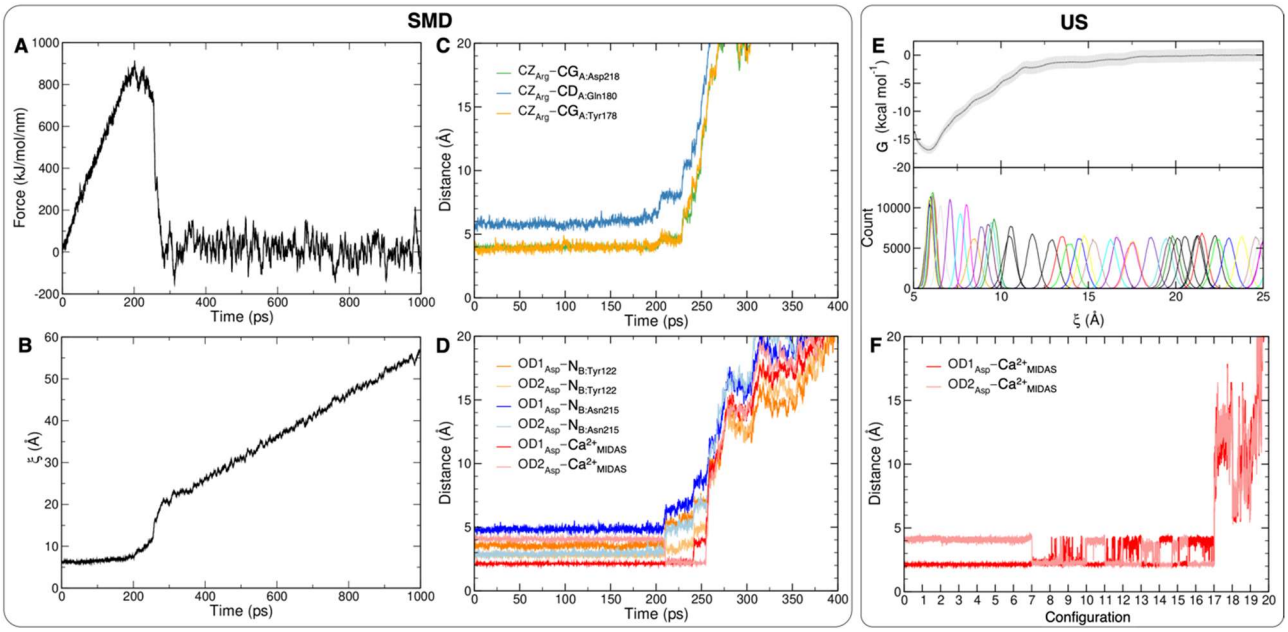

Figure S14. Steered MD simulation (A-D) and umbrella sampling (E-F) results for the *bridging* binding mode. (A) Force applied on the collective variable ξ over time in SMD simulation. (B) Value of the distance between COMs, ξ, over time in SMD simulation. Distance of relevant arginine–protein (C) and aspartate–protein (D) interactions along the SMD simulation. (E) Free energy profile and its corresponding umbrella histograms. (F) Distance between OD1 and OD2 of aspartate and Ca^{2+}_{MIDAS} for the first 20 umbrella simulations, taken in order of increasing value of the starting-point collective variable ξ.

When starting from the *chelating* binding mode, the rupture happens slightly later in time (300 ps - 350 ps) and requires a higher force (1100 kJ mol^{-1} nm^{-1}) with respect to the *bridging* binding mode (Figure S15A,B). The interactions are lost in the same order as above (Figure S15C,D): the first to break are those involving the arginine side chain, proving them to be the weakest in agreement with previous SMD simulations,[13] and, at a second stage, those involving aspartate break. In the subsequent US simulations, the aspartate *chelating* binding mode is maintained as long as the

distance is short enough for the ligand to be bound to the protein (Figure S15F) and the free energy profile obtained is reported in Figure S15E.

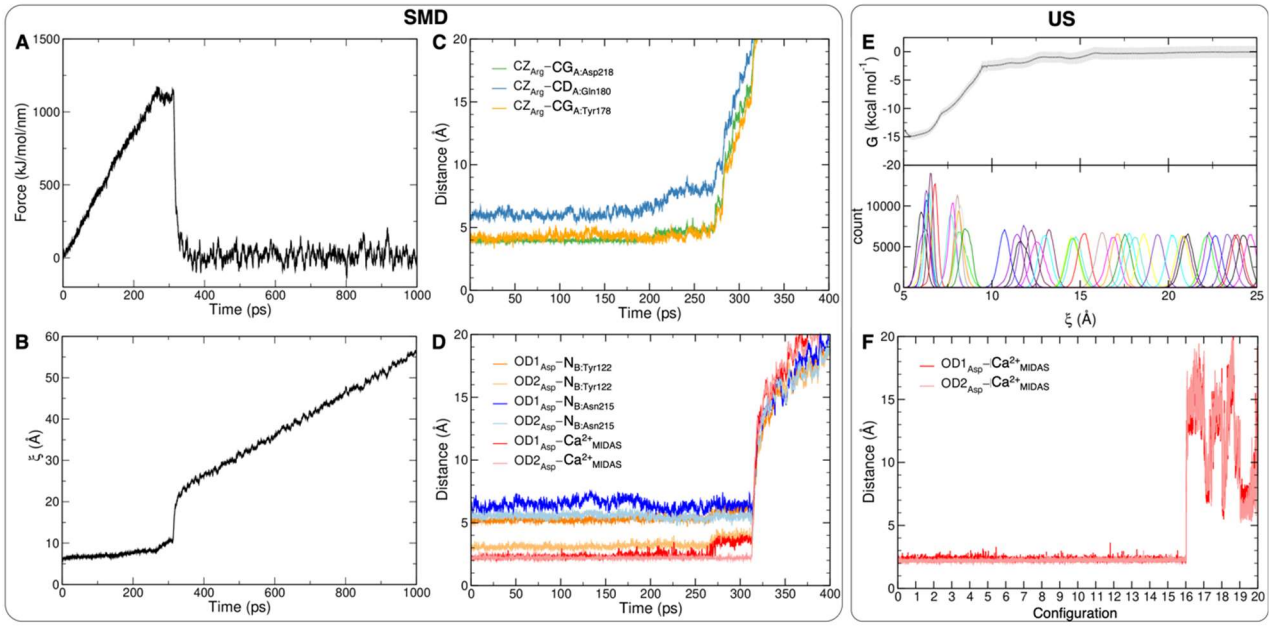

Figure S15. Steered MD simulation (A-D) and umbrella sampling (E-F) results for the *chelating* binding mode. (A) Force applied on the collective variable ξ, which is treated as a spring, over time in SMD simulation. (B) Value of the distance between COMs, ξ, over time in SMD simulation. Distance of relevant arginine–protein (C) and aspartate–protein (D) interactions along the SMD simulation. (E) Free energy profile and its corresponding umbrella histograms. (F) Distance between OD1 and OD2 of aspartate and Ca^{2+}_{MIDAS} for the first 20 umbrella simulations, taken in order of increasing value of the starting-point collective variable ξ.

S4.2 The effect of conjugation to PEGylated TiO_2 NPs on the ligand binding to integrin $\alpha_V\beta_3$

Table S3. H-bonds number between the ligand and the integrin or water for the *cRGD MD simulation* (average over the all MD simulation, for the *bridging* binding mode (50-350 ns), and for the *chelating* binding mode (450-500ns)) and for the three cRGD-conjugated PEGylated TiO_2 nanoparticle models averaged over the MD production phase.

H-bonds number	cRGD: *Bridging*	cRGD: *Chelating*	NP-PEG-cRGD10	NP-PEG-cRGD20	NP-PEG-cRGD50
Ligand–integrin	2.8(±0.9)	2.3(±0.9)	2.5(±1.0)	0.9(±0.7)	2.3(±0.9)
Ligand–water	6.0(±1.9)	6.6(±2.1)	6.1(±1.9)	7.5(±2.1)	5.7(±2.0)

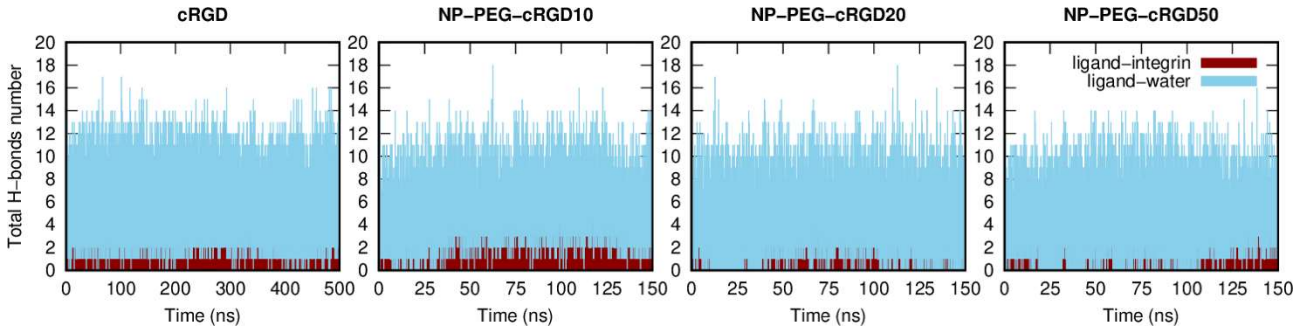

Figure S16. Total average H-bonds number of the ligand, split into ligand–integrin contribution (dark red) and ligand–water contribution (light blue).

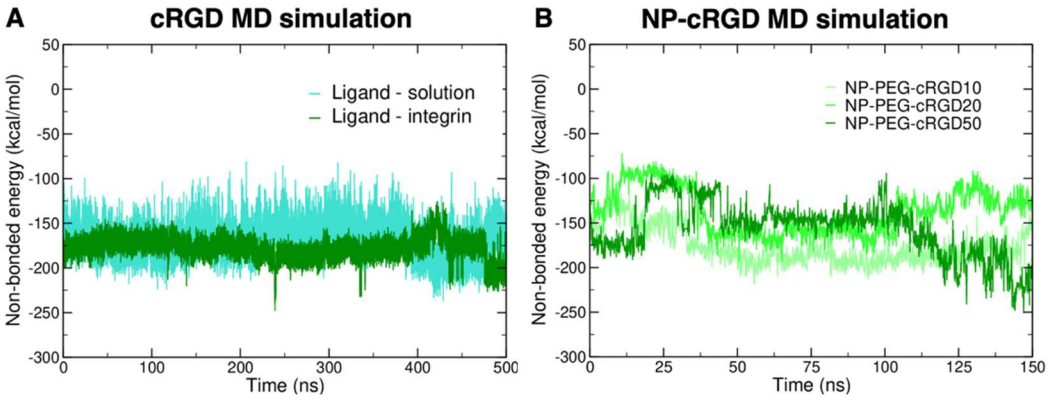

Figure S17. Non-bonded interaction energy profiles between the ligand and integrin or the solution in *cRGD MD simulation* (A) and between the ligand and integrin in *NP-cRGD MD simulation* (B).

Conformational analysis: Self-Organizing Maps

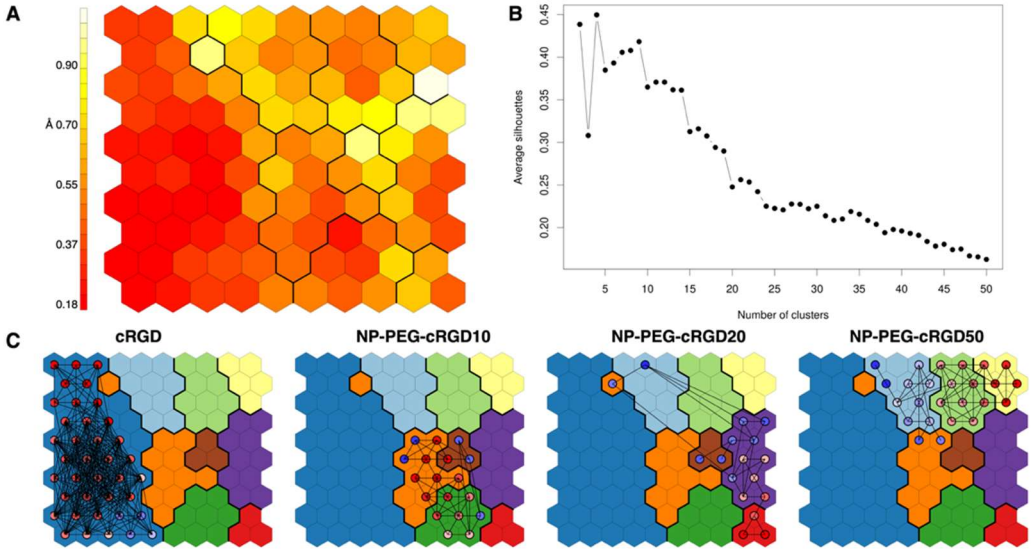

Figure S18. (A) Neighbors distance map. (B) Silhouette profile for SOM neurons clustering. (C) Transition or path maps for the four different systems. Time progress goes from blue to red along the simulations.

Based on the analysis of the population maps, one could conclude that the binding mode of cRGD ligand to integrin $\alpha_V\beta_3$ is deeply affected by the conjugation to PEGylated TiO_2 NPs. Nevertheless, the scenario is more complex and requires a deeper analysis. For instance, it is interesting to examine which features determine the clustering of the four systems in different areas of the map. To this aim, we have colored each neuron according to the value of selected descriptors in Figure 8 (manuscript), where blue and red neurons are the ones in which the descriptor has a low or a high value, respectively. In Figure 8A the per-neuron average values of the three cRGD intramolecular criteria discussed in Section 3.3.1 are reported: when the cRGD is conjugated to the PEG chains on the NP, the neurons turn to white and blue and, therefore, the conformation of the RGD portion is less extended (lower values of d_1, d_2 and ϑ) than in the *cRGD MD simulation* and in the reference CGT crystal structure (magenta-contoured neuron in Figure 7). In particular, the higher similarity with the neurons, i.e. the ligand binding conformations, in the *cRGD MD simulation* is found for NP-PEG-cRGD10, followed by NP-PEG-cRGD50.

In order to get a detailed insight into the differences in the binding mode, similar maps with the values of selected ligand–protein intermolecular distances are also reported, where blue neurons are the ones in which the selected atoms are close to each other, namely the interaction is present, and red ones are the ones in which the interaction is missing. For arginine–integrin distances (Figure 8B) it is evident that the protein residue numbers may change but not their type: in fact, arginine is always interacting with at least one aspartate group. However, it is true that the conjugation to the NP perturbs the binding in the sense that the ligand contacts different residues, probably because of the bent conformation that it assumes when exposed by the NP, as discussed above. On the aspartate side (Figure 8C), the carboxylate group is more stably found bridging the Ca^{2+}_{MIDAS} (red neurons) in the most of the neurons, apart from the few blue neurons that correspond to the *chelating* mode. However, the two H-bonds with B:Asn215 and B:Tyr122, examined in Section 3.2.1 in *cRGD MD simulation*, are sometimes missing and substituted by H-bonds with water molecules.

S4.3 Besides the ligand binding to the pocket: all the other nanodevice/integrin interactions

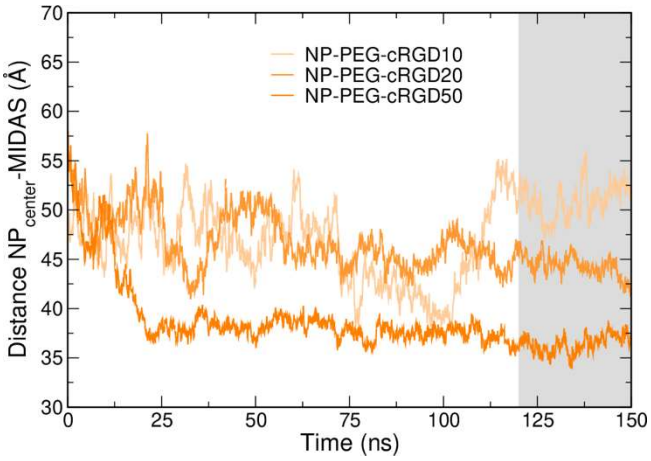

Figure S19. Distance between the NP center and the Ca^{2+}_{MIDAS} ion along the trajectory for the three cRGD densities. The grey region is the one considered for averages in *NP-cRGD MD simulations*.

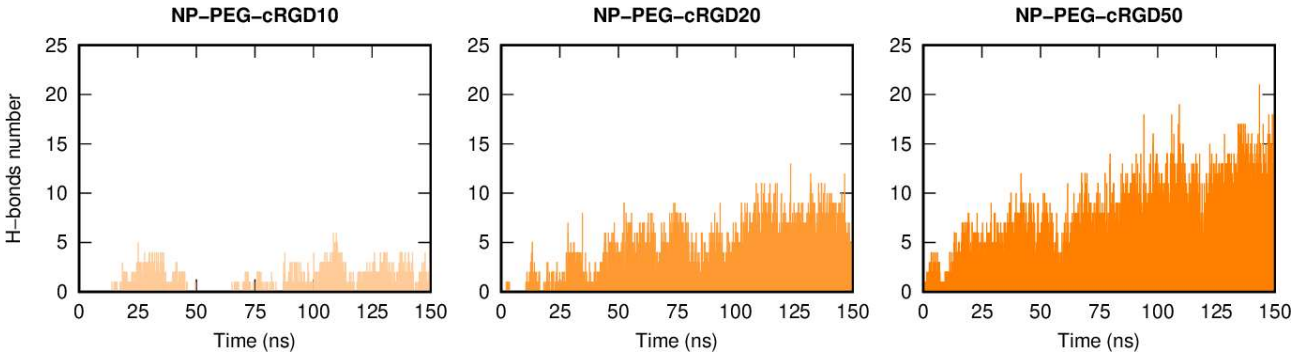

Figure S20. Average H-bonds number between the nanodevices and the integrin at different cRGD densities.

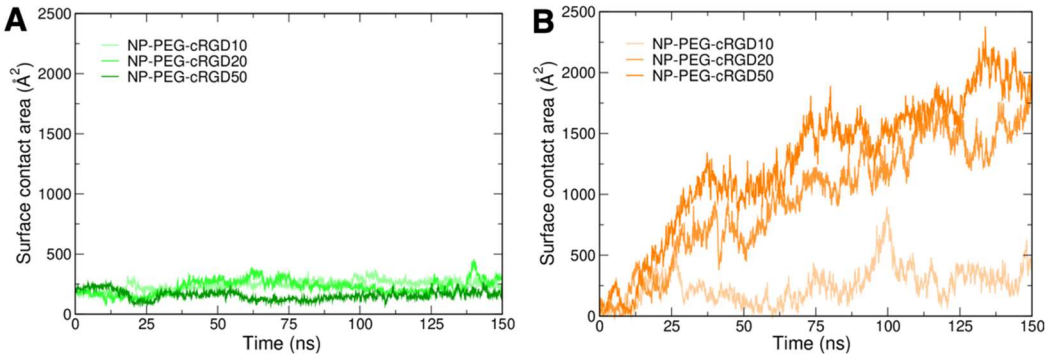

Figure S21. Surface contact area between the ligand and the integrin (A) and between the whole nanodevice, excluding the cRGD in the binding pocket, and the integrin (B) for all the three systems at the lowest, intermediate and highest cRGD density.

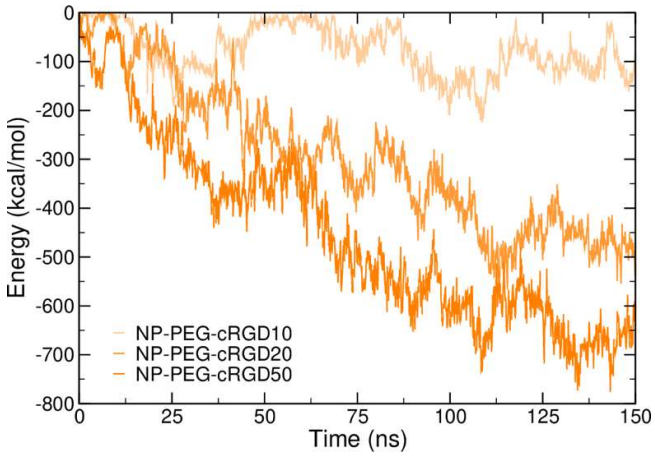

Figure S22. Non-bonded interaction energies between the protein and the nanodevice, excluding the ligand, at the lowest, intermediate and highest cRGD density.

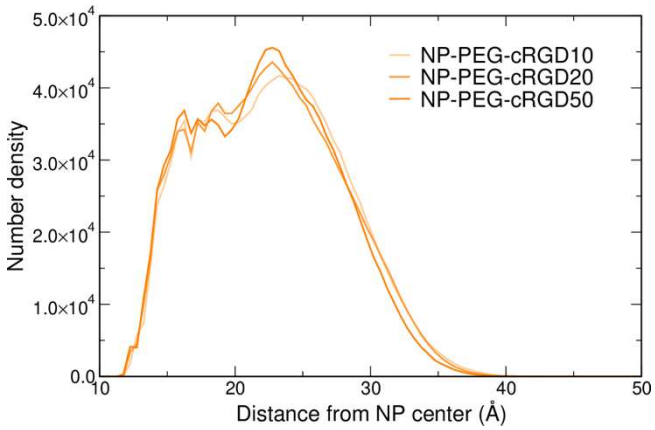

Figure S23. Number density profiles of PEG chains from the geometric center of the NP at the three cRGD densities.

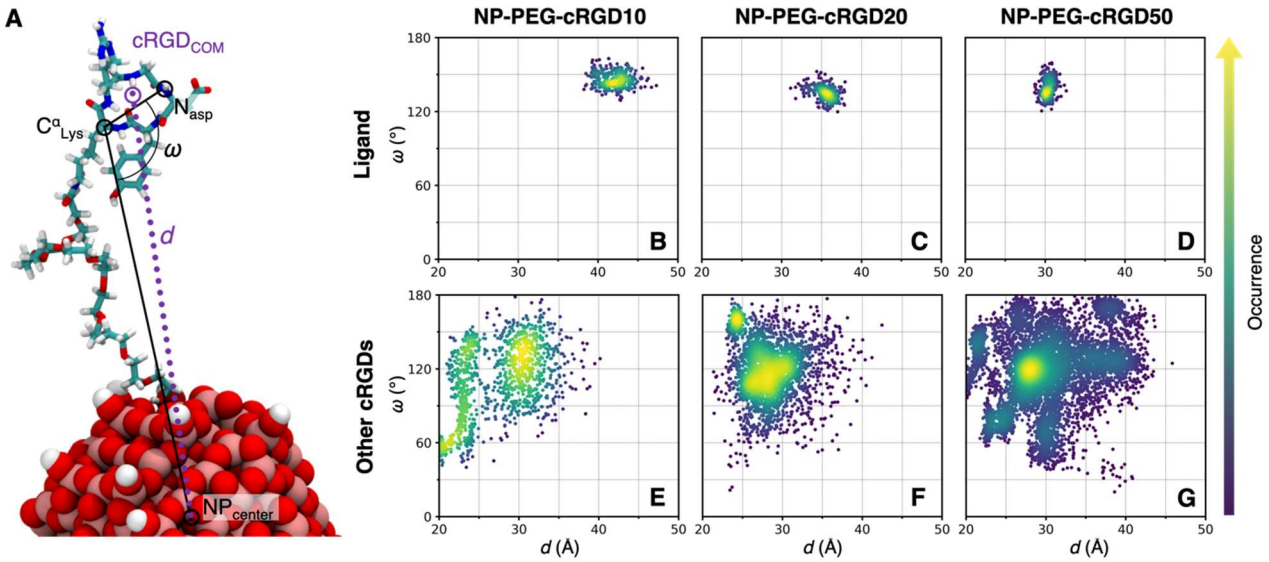

Figure S24. Heatmaps showing the relationship between the angle *ω*, which is a measure of cRGD orientation with respect to the surface, and the distance *d* of cRGD from the NP center, as defined on the left of the panel (A) for the ligand and other cRGDs at the lowest (B,E), intermediate (C,F), and highest (D,G) cRGD density. Only the production phase is considered.

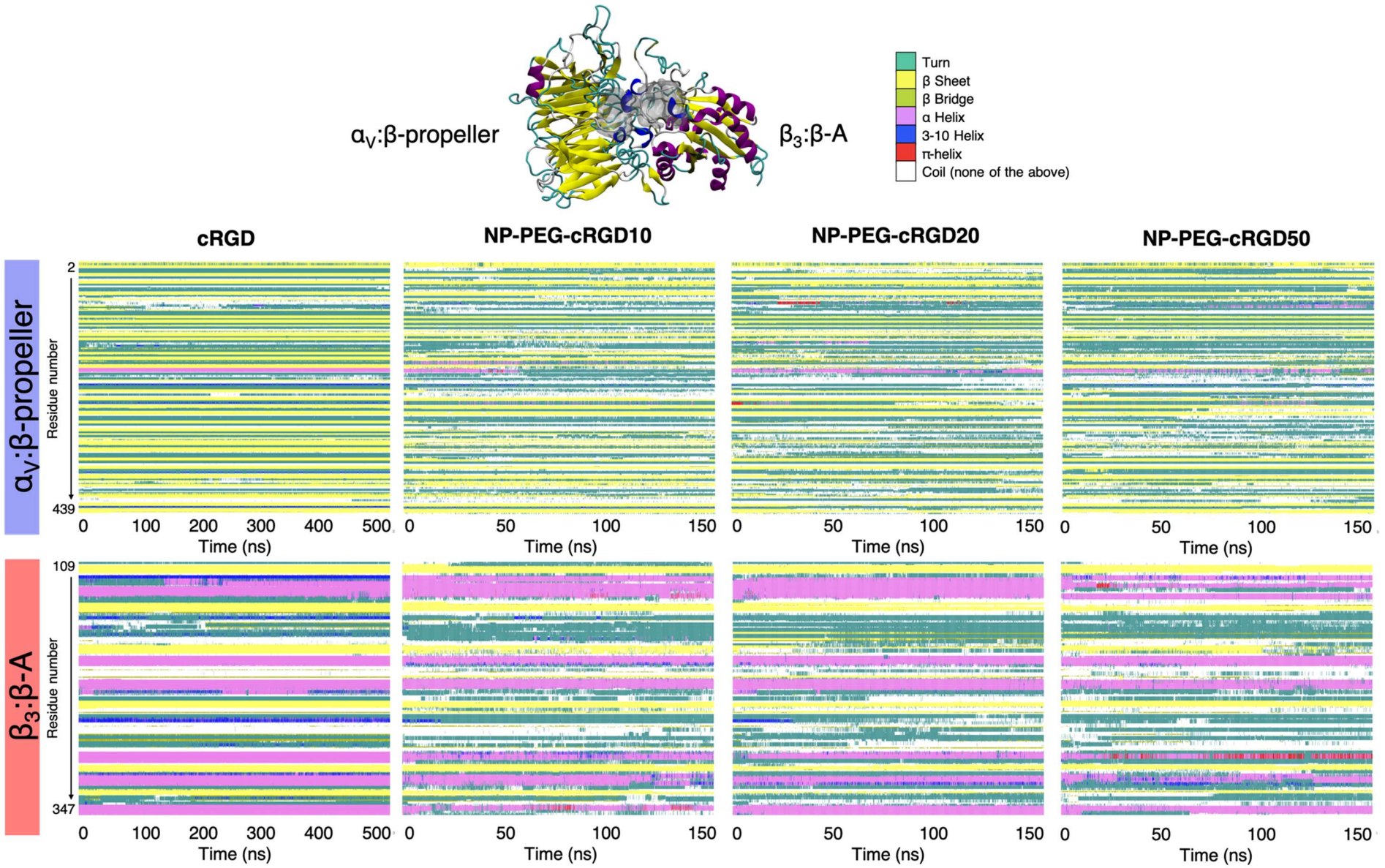

Figure S25. Top: The structure of β-propeller and β-A domains colored according to their secondary structure as found in the crystal. Bottom: Secondary structure analysis of β-propeller (subunit α_v) and β-A (subunit β_3) domains for the *cRGD MD simulation* and the *NP-cRGD MD simulations* at the three cRGD densities.

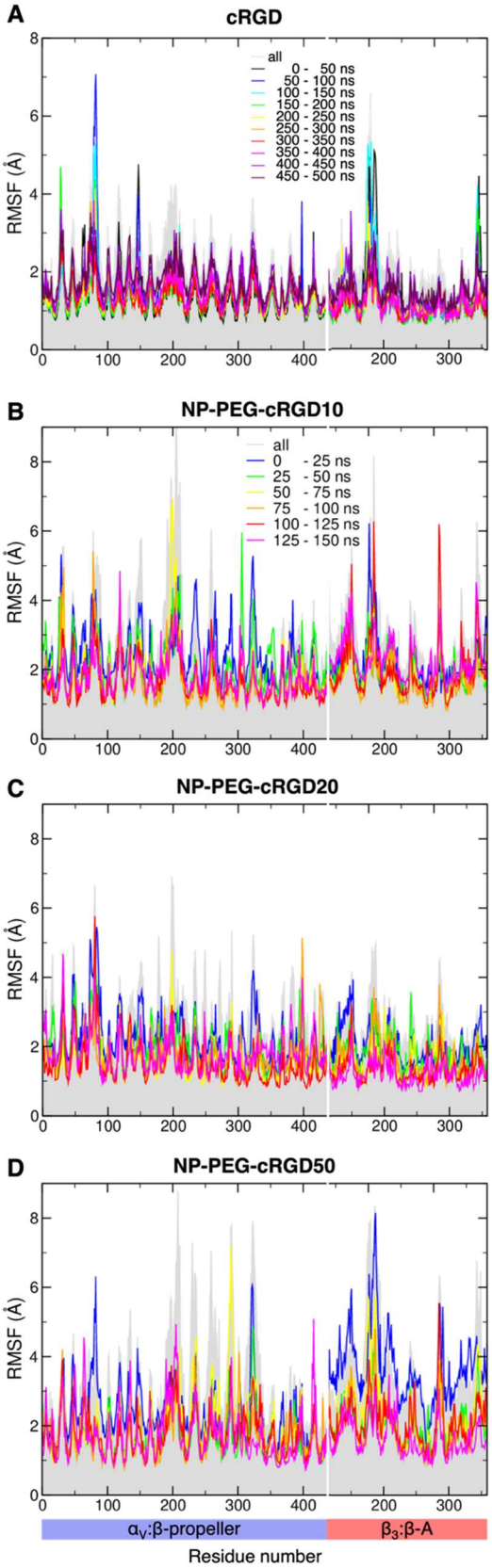

Figure S26. RMSF of the two domains involved in cRGD binding over the whole MD simulation and for different time intervals for the *cRGD MD simulation* (A) and the NP MD simulation at the lowest (B), intermediate (C) and highest (D) cRGD density.

References

1 G. Madhavi Sastry, M. Adzhigirey, T. Day, R. Annabhimoju and W. Sherman, *J Comput Aided Mol Des*, 2013, **27**, 221–234.

2 T. A. Halgren, R. B. Murphy, R. A. Friesner, H. S. Beard, L. L. Frye, W. T. Pollard and J. L. Banks, *J Med Chem*, 2004, **47**, 1750–1759.

3 R. A. Friesner, R. B. Murphy, M. P. Repasky, L. L. Frye, J. R. Greenwood, T. A. Halgren, P. C. Sanschagrin and D. T. Mainz, *J Med Chem*, 2006, **49**, 6177–6196.

4 T. A. Halgren, R. B. Murphy, R. A. Friesner, H. S. Beard, L. L. Frye, W. T. Pollard and J. L. Banks, *J Med Chem*, 2004, **47**, 1750–1759.

5 J. L. Banks, H. S. Beard, Y. Cao, A. E. Cho, W. Damm, R. Farid, A. K. Felts, T. A. Halgren, D. T. Mainz, J. R. Maple, R. Murphy, D. M. Philipp, M. P. Repasky, L. Y. Zhang, B. J. Berne, R. A. Friesner, E. Gallicchio and R. M. Levy, *J Comput Chem*, 2005, **26**, 1752–1780.

6 J.-P. Xiong, S. Thilo, R. Zhang, A. Joachimiak, M. Frech, S. L. Goodman and M. A. Arnaout, *Science (1979)*, 2002, **296**, 151–155.

7 R. B. Best, X. Zhu, J. Shim, P. E. M. Lopes, J. Mittal, M. Feig and A. D. MacKerell, *J Chem Theory Comput*, 2012, **8**, 3257–3273.

8 J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. de Groot, H. Grubmüller and A. D. MacKerell, *Nat Methods*, 2017, **14**, 71–73.

9 A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin and M. Karplus, *Journal of Physical Chemistry B*, 1998, **102**, 3586–3616.

10 J. Liao, F. Marinelli, C. Lee, Y. Huang, J. D. Faraldo-Gómez and Y. Jiang, *Nat Struct Mol Biol*, 2016, **23**, 590–599.

11 J. Kahlen, L. Salimi, M. Sulpizi, C. Peter and D. Donadio, *J Phys Chem B*, 2014, **118**, 3960–3972.

12 A. T. Church, Z. E. Hughes and T. R. Walsh, *RSC Adv*, 2015, **5**, 67820–67828.

13 D. Craig, M. Gao, K. Schulten and V. Vogel, *Structure*, 2004, **12**, 2049–2058.

14 R. Araya-Secchi, T. Perez-Acle, S. G. Kang, T. Huynh, A. Bernardin, Y. Escalona, J. A. Garate, A. D. Martínez, I. E. García, J. C. Sáez and R. Zhou, *Biophys J*, 2014, **107**, 599–612.

15 L. Alhalhooly, M. I. Confeld, S. O. Woo, B. Mamnoon, R. Jacobson, S. Ghosh, J. Kim, S. Mallik and Y. Choi, *ACS Appl Mater Interfaces*, 2022, **14**, 7671–7679.