Article

Analysis of $\alpha_4\beta_1$ Integrin Specific Antagonists Binding Modes: Structural Insights by Molecular Docking, Molecular Dynamics and Linear Interaction Energy Method for Free Energy Calculations

João Hermínio Martins da Silva,[a,b] Laurent Emmanuel Dardenne,[c] Wilson Savino[b] and Ernesto Raul Caffarena*,[a]

[a]Programa de Computação Científica - Fundação Oswaldo Cruz, Antiga Residência Oficial, Av Brasil 4365, 21045-900 Rio de Janeiro-RJ Brazil

[b]Laboratório de Pesquisa sobre o Timo, Instituto Oswaldo Cruz - Fundação Oswaldo Cruz, Av Brasil 4365, 21045-900 Rio de Janeiro-RJ Brazil

[c]Grupo de Modelagem Molecular em Sistemas Biológicos, Laboratório Nacional de Computação Científica, Av Getúlio Vargas 333, 25651-075 Petrópolis-RJ Brazil

A integrina VLA-4 (integrina $\alpha_4\beta_1$) participa da fisiopatologia de uma grande variedade de doenças, incluindo asma, esclerose múltipla, artrite reumatóide e diabetes. A seletividade de ligantes por essa integrina permanece ainda um problema sem resposta, principalmente devido ao fato das estruturas 3D da maioria delas serem ainda desconhecida. Inicialmente, construímos um modelo computacional tridimensional do sítio ativo da integrina α_4B_1, utilizando a estrutura cristalográfica da integrina $\alpha_v\beta_3$ como molde. Após, realizamos um estudo por *docking* molecular usando um conjunto de sete antagonistas de $\alpha_4\beta_1$ derivados de 4-[*N*'-(2-metilfenil)ureido]fenilacetil, avaliando os seus modos de ligação. A técnica de dinâmica molecular foi utilizada de modo a aumentar a eficiência da exploração do perfil energético obtido pelo algoritmo de *docking*. Segundo a nossa análise, os compostos foram sistematicamente agrupados em dois modos de ligação principais, sendo a conformação estendida o modo prevalente de ligação dos antagonistas no sítio. Cálculos de energia livre também confirmaram essa constatação. Esse estudo aumenta a compreensão dos mecanismos de ligação dessa família de antagonistas e pode fornecer informações úteis ao desenho racional de fármacos específicos para o VLA-4.

The VLA-4 antigen ($\alpha_4\beta_1$ integrin) is involved in the pathophysiology of a variety of diseases including asthma, multiple sclerosis, rheumatoid arthritis and diabetes. The ligand selectivity toward this integrin remains a difficult problem, mainly due to the fact that 3D structures of most integrins are still unknown. We initially built a 3D computational model of the $\alpha_4\beta_1$ ligand binding site, taking the crystal structure of the integrin $\alpha_v\beta_3$ as template. Then, we performed a computational study on a set of seven $\alpha_4\beta_1$ antagonists, evaluating the binding modes of 4-[*N*'-(2-methylphenyl)ureido]phenylacetyl and derivatives by molecular docking. Molecular dynamics simulations were used to improve the receptor-ligand energy landscape exploration by the docking algorithm. The compounds were systematically arranged in two main binding modes, and in all cases, pointed out that these antagonists preferably bind to the $\alpha_4\beta_1$ integrin active site in an extended conformation that resembles the one in solution. LIE (linear interaction energy) calculations also confirmed this statement given that the most prevailing binding mode is also the energetically most favored one. This study benefits the comprehension of the mechanism of this family of antagonists and may provide useful information for rational drug design.

Keywords: inflammation, integrins, VLA-4, docking, molecular dynamics

---

*e-mail: ernesto@fiocruz.br

Introduction

Integrins of the β_1 sub-family (formerly named very late antigens of VLAs) correspond to at least nine transmembrane heterodimers composed of a given α subunit noncovalently coupled to the β_1-integrin chain (CD29). These molecules are able to bind one or more extracellular matrix (ECM) ligands, although some of them can also interact with a cell membrane counter-receptor. Functionally, β_1 integrins are related to cell migration, both in normal and pathological conditions, such as inflammation and metastasis.[1-3]

Of particular interest is $\alpha_4\beta_1$ integrin (VLA-4 or CD49d/CD29), which is expressed by a variety of hemopoietic cell types including monocytes, T and B lymphocytes, basophils, and eosinhophils. This integrin binds the alternatively spliced type III connecting segment (CS-1) of fibronectin, as well as the vascular cell adhesion molecule-1 (VCAM-1). From a therapeutic point of view, it has been shown that blocking VLA-4/VCAM-1 interaction resulted in amelioration of distinct inflammatory diseases, such as asthma, multiple sclerosis, rheumatoid arthritis and type 1 diabetes.[4-8]

The $\alpha_4\beta_1$ integrin binds to fibronectin in the ECM and to VCAM-1 on the endothelium, through the minimum binding epitope Leu-Asp-Val (LDV), discovered by Komoriya *et al.*[9] This pattern of three residues is homologous to the Leu-Asp-Ser (LDS) motif found in VCAM-1, which has been reported as the VCAM-1 binding site of $\alpha_4\beta_1$, suggesting that VCAM-1 and fibronectin could share an identical or overlapping $\alpha_4\beta_1$ binding site.[5]

A large number of LDV mimetic moieties have been identified as potent and often subtype-selective ligands. Namely, the fibronectin-derived compound 4-[*N'*-(2-methylphenyl)ureido]phenylacetyl-Leu-Asp-Val (PUPA-LDV) has been proved to be a potent highly selective antagonist of the $\alpha_4\beta_1$.[10] A set of analogue compounds, based on the PUPA-LDV leading structure, was obtained from a computational database searching method and their IC_{50} values have been measured.[10,11]

However, detailed structural information about ligand-receptor specific interactions between this particular family of ligands and the $\alpha_4\beta_1$ integrin remains relatively poor. This is likely due to the intrinsic difficulty in obtaining the three-dimensional structure of membrane receptors such as $\alpha_4\beta_1$ through X-ray crystallography.

We applied a combined approach of homology modeling, ligand-receptor molecular docking and molecular dynamics (MD), in order to determine the main binding modes of 4-[*N'*-(2-methylphenyl)ureido] phenylacetyl-LDV and derivatives within the binding site of the $\alpha_4\beta_1$ integrin. MD simulations were performed to identify the most prevalent ligand conformations in solution and establish a connection with its binding modes within the $\alpha_4\beta_1$ integrin active site. Linear interaction energies (LIE) calculations were also carried out to identify the preferred binding modes, suggesting that there exists a preference for a specific way of binding.

Methodology

Comparative Modeling

The alignment between the integrin $\alpha_4\beta_1$ and the template $\alpha_v\beta_3$ was obtained using the Align software of EMBOSS package. Sequences were retrieved from the Entrez Protein database, under NP000876 e P05556 access numbers. The BLOSUM62 matrix was used with a gap and extension penalties of 11 and 1, respectively.[12]

A homology model of the $\beta1$ subunit $\alpha_4\beta_1$ integrin has been previously reported by You *et al.*[13] This work served us as a landmark to model the whole headpiece of human α_4 and β_1 subunits with the program Modeller (version 7.0) according to the comparative protein modeling methodology.[14, 15]

It is very unlikely that the conformational transitions that integrins undergo (*i.e.*, conversions between different conformations that represent diverse activity levels) would occur spontaneously during MD simulations using current approaches. Therefore, the structure of the $\alpha_4\beta_1$ integrin headpiece has been modeled using as template the X-ray structure of the $\alpha_v\beta_3$ integrin[16] in complex with cyclo(-RGDf[Nme]V-) (from hereafter, RGD-ligand), also known as cilengitide, (PDB entry code: 1L5G) to ensure that the active conformation of the binding site is correctly modeled. This three-dimensional structure, extensively used in previous works,[5,17] is the only available template so far for building homology model of integrins lacking I-domain in the active state. Moreover, the integrin family is a remarkably well conserved family and this information is helpful concerning the building of a homology model.[18] The human β_1 and β_3 domains are homologous, sharing 43% identity in the sequence alignment; 33.5% of identity is found between α_4 and α_v subunits. The sequence lengths of the headpiece encompass 237 and 431 residues for β and α domains, respectively. These percentages of residue identities guarantee an acceptable reliability for the generated structural homology model.

Thirty full-atoms models were built and ranked using the Modeller objective function, which is highly efficient in classifying different models calculated from the same alignment.[19] The highest-ranking model of the VLA-4

integrin headpiece was subjected to further refinement through a short minimization protocol using the Gromacs classical force field.[20] Prior to the minimization, the atomic coordinates of Mg^{2+} ions of the Metal Ion Dependent Adhesion Site (MIDAS) motif were merged into the binding site as present in the crystal structure used as template. In a previous study with the integrin alpha5beta1, the high conservation of the binding site was confirmed between alphaVbeta3 and alpha5beta1. Furthermore, according to mutagenesis data[21], cross linking[22] and X-ray structure[23] the coordination of MIDAS is conserved and indicates this region as a key element for interacting with ligands similar to RGD.

The stereochemical quality of the resulting model was checked with the program PROCHECK[24,25] and VERIFY3D.[26]

Molecular docking

Firstly, we tested the computational approach to $\alpha_v\beta_3$ integrin in complex with RGD-ligand to set basic rules and protocols of the docking procedure (*i.e.,* determination of a suitable set of parameters to reproduce experimental results) and compared the obtained results with the binding mode of RGD-ligand.

The docking of RGD ligand to the $\alpha_v\beta_3$ integrin headpiece was carried out using the Autodock program package version 3.0.5[27] allowing full flexibility to the residue side chains.

The Lamarckian-Genetic-Algorithm (LGA), as implemented in the Autodock program, was applied using a protocol of 2048 independent runs with an initial population size of 300 individuals, a maximum number of 500000 evaluations, a mutation rate of 0.02, a crossover rate of 0.80 and an elitism value of 1. For the local search, the Pseudo Solis and Wets algorithm was applied and another non-specified values were left to their respective default values. Each docking experiment was performed in a grid of 127 points *per* dimension, with a discretization of 0.275 Å, encompassing the binding site located in the headpiece of the heterodimer N-terminal.

This set of parameters resulted in an optimum way to reproduce the experimental binding modes of RGD-ligand in the $\alpha_v\beta_3$ integrin active site, which was extrapolated to docking experiments of PUPA ligands within the $\alpha_4\beta_1$ headpiece.

The molecular docking experiments were carried out allowing PUPA-LDV and derivatives (Figure 1) to rotate freely except torsion rotations in amines and peptide bonds.

The crystal structure of the extracellular domains of the $\alpha_v\beta_3$ integrin in complex with the RGD-ligand allowed us to better understand the ligand binding mode. It was observed that the acid oxygen atoms point to the central MIDAS
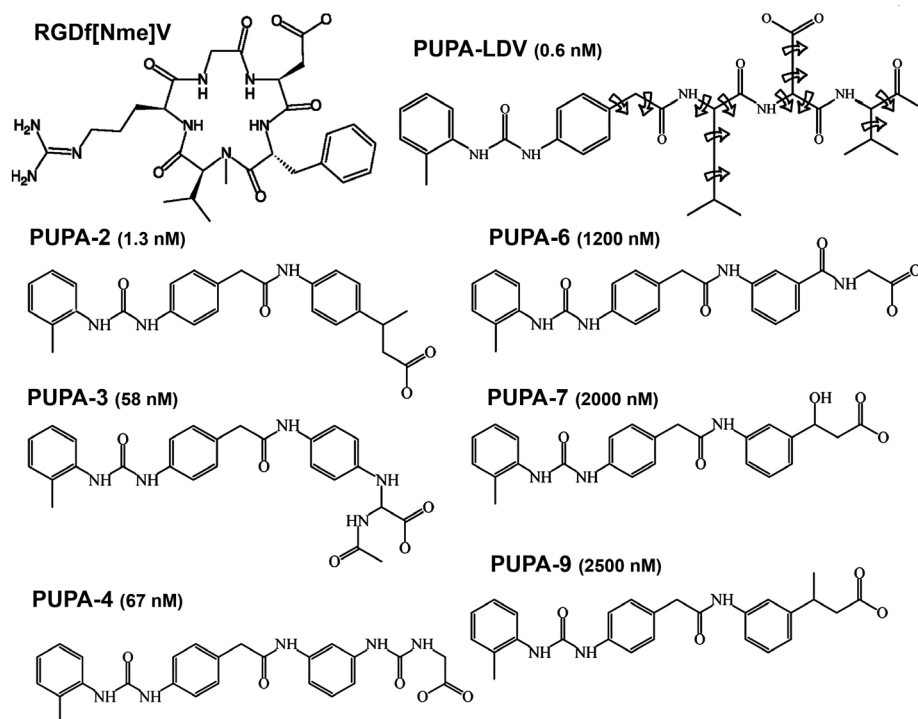

Figure 1. Scheme of PUPA-LDV and derivatives. Torsions free to rotate during docking process are marked with arrows. IC_{50}'s values are shown beside the 2D structures.

ion position. Previous reports showed the importance of the MIDAS motif for binding to carboxylic groups of acid residues.[18,28,29] This is the main reason why, out of the 2048 solutions obtained for each ligand, we have only selected those that presented any of their carbonyl oxygen atoms directed towards the central MIDAS ion using a cut-off of 3.5 Å. The exclusion criterion has been previously applied[17] based on structural information observed for RGD-ligand in the X-ray complex[16] and its importance has been discussed.[30] Due to this exclusion procedure, a larger number of independent runs were necessary to assure a better statistical analysis.

Due to the high number of degrees of freedom of the investigated ligands, the docked conformations differing by less than 3.5 Å in positional root mean square deviation (rmsd) were clustered together and represented by the outcome with the most favorable free energy of binding.

MD simulations

To find out the preferred conformations of PUPA-LDV and derived compounds isolated in aqueous solution, MD simulations were carried out in explicit solvent. Calculations were done with the GROMACS package[20] and the Gromos96 classical force field[31] in the Gibbs ensemble (300 K, 1 bar).[32] Atomic partial charges were obtained using the MMFF94 parameterization.[33]

Coordinates of the molecular complex were immersed in a cubic SPC[34] water box of volume 125 nm^3 and minimized using periodic boundary conditions to adjust the atomic positions to the force field. All interatomic bonds were constrained using the LINCS algorithm.[35] Non-bonded interactions were taken into account using the 6-12 Lennard-Jones potential using a cut-off radius of 14 Å and a Coulomb potential with reaction field with dielectric constant equals to 66 and cut-off radius of 18 Å. MD simulations evolved freely for 5 ns, saving trajectories and velocities every 1 ps to analyze structural features.

Linear interaction energy

Calculations of relatives free energies for different modes of PUPA-LDV and its most active derivatives were performed using the LIE method[36] with the software package Q[37] and the Gromos96 force field.[31] The systems were solvated with SPC water within a 20 Å sphere and non-bonded interactions across the boundary were excluded. The non-bonded interaction energies were calculated without cut-off restrictions while long-range electrostatics were treated using a multipole expansion.[38] Atoms outside

of the 18.5 Å simulation sphere were harmonically restrained to their initial positions with a 200 kcal mol^{-1} Å^{-2} force constant.

The compounds in the free state were located at the centre of the simulation sphere and restrained with a force constant of 5 kcal mol^{-1} Å^{-2} to guarantee an homogeneous solvation. All compounds were heated from 1 to 300 K using a stepwise scheme followed by an equilibration period to stabilize ligand surrounding energies before the data collection phase (500 ps). The time step used in the production phase of the simulations was 1 fs, and the temperature was set to 300 K using a weak coupling to an external bath. SHAKE[39] was used to constrain bonds and angles on solvent molecules. The total simulation time for the production period was 750 ps for the ligands in aqueous solution and 1 ns for the complex. Energies were sampled every 0.5 ps in the production part of the simulations, which were then used in equation (1) to estimate the free energy of binding. The MD run until the collected energies showed stability for a period not shorter than 1000 ps. Stability was addressed by comparing the average values of the first and second halves of the collection period. Error estimates in the calculated free energies were obtained by adding the errors in the potential energy from both the water and protein simulations, properly scaled through the use of the LIE α and β factors.

In equation (1), U_{el} and U_{LJ} are the electrostatic and Lennard-Jones interaction energies between ligand (*l*) and its surroundings (*s*) in the binding site and in aqueous solution, respectively; the brackets < >'s denote MD ensemble averages over trajectories and α, β and γ are empirical parameters.

$$\Delta G_{binding} = \alpha(\langle U_{l\text{-}s}^{LJ}\rangle_{bound} - \langle U_{l\text{-}s}^{LJ}\rangle_{free}) + \beta(\langle U_{l\text{-}s}^{el}\rangle_{bound} - \langle U_{l\text{-}s}^{el}\rangle_{free}) + \gamma \quad (1)$$

In detailed studies of ligand binding, Hansson *et al.*[40, 41] have determined that an α value of 0.181 can adequately reproduce the free energies of binding for a variety of ligand-protein systems, along with a set of β values determined by Free Energy Perturbation (FEP), depending on the chemical nature of compounds. According to these calculations, the β value is 0.5 for charged ligands.

Results and Discussion

Homology Model

As the complete three-dimensional structure of $\alpha_4\beta_1$ integrin is not available, we used the X-ray crystallographic structure of the $\alpha_v\beta_3$ as a template to generate a homology
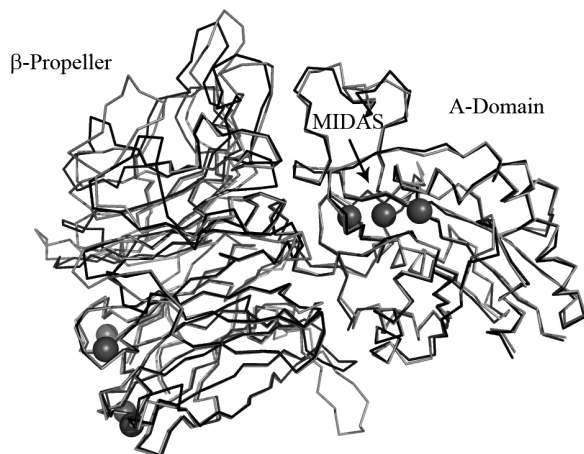
Figure 2. Three-dimensional structure of extracellular domains of $\alpha_4\beta_1$ integrin overlapped on the template. The superposition was performed considering all the backbone atoms. $\alpha_4\beta_1$ loops are shown in dark grey, while $\alpha_V\beta_3$ atoms are shown in light grey. The MIDAS ion is indicated by a black arrow. The major difference between the two complexes resides in an external loop of $\alpha_V\beta_3$ integrin that is not present $\alpha_4\beta_1$ integrin. Minor differences are shown in the conformation of the specificity determining loop (SDL). Three metal ions are visible at MIDAS, ADMIDAS, and LIMBS region as spheres. The rmsd value between β-propeller domains of crystal template and α_4 model is 0.75 Å. For β_A domains the rmsd value between alpha carbons of main chain β_1 and β_3 headpiece is 0.57 Å. The generation of all figures was done with Pymol v0.99rc6 software.[38]

model structure of the $\alpha_4\beta_1$ integrin portion containing the ligand binding site (Figure 2).

The β-Propeller and β_A domains, responsible for the ligand binding, constitute the headpiece of the $\alpha_4\beta_1$ integrin. This particular region includes 431 residues of the α subunit and 237 of the β subunit. Out of these amino acids only 48 and 37, from α and β subunits respectively, are in the interface between each other. These values were calculated using the PISA Server.[42] Only three hydrogen bonds were observed between α and β subunits in the model. (Tyr253:α-Gly286:B, Tyr364:α-Ala320:β, and Glu370:α-Lys289:β). The solvent accessible surface (SAS) was approximately 29,976 Å^2, with only 2,977 Å^2 in the interface region.

The rmsd value between the β-propeller domain of the template and the α_4 domain of the model was 0.75 Å (considering only the alpha carbons). For the β_A domains (*i.e.,* β_1 template and β_3 model) we found a rmsd of 0.57 Å. Figure 2 shows the superposition of $\alpha_4\beta_1$ and $\alpha_V\beta_3$ headpieces.

A particular attention was paid on the modeling of the β subunit amino acids that form the MIDAS and ADMIDAS regions (Asp150, Lys151, Ser152, Tyr153, Ser154, Glu249, Asp279) due to the crucial role that the divalent ions play in the binding process.

The stereochemical quality of the resulting model was evaluated with the program PROCHECK. Global G-factors

of −0.36 and −0.11 were obtained for β-propeller and β_A domains, respectively.

The majority of the residues of the modeled protein occupied the most favored regions of the Ramachandran plot, while others were located in additional allowed regions. 94.5% (alpha domain) and 96.6% (beta domain) of the residues are in the most favored regions; 4.4% (alpha domain) and 2.4% (beta domain) in additional allowed regions and only 1.1% (α domain) and 1.0% (β domain) were in disallowed regions.

The 3D-1D quality was also validated with VERIFY3D and no errors were detected. According to this software, 85.42% of the α subunit residues and 85.71% of β subunit residues presented an average 3D-1D score greater than 0.2, corroborating the good quality of the modeled structure of the binding site.

Molecular Docking

Firstly, we performed docking essays between RGD-ligand and $\alpha_V\beta_3$ as a way to determine the most suitable docking algorithm parameters to work with. The results obtained with the best parameter set reproduced the crystallographic structure (1L5G) conformation of the ligand in almost 60% of the cases, with at least 0.5 Å of rmsd.

To define the binding modes for the PUPA's set of ligands we docked these molecules within the gap of the integrin headpiece, between the β propeller and β_A domains.

After selection (*i.e.,* binding modes where the ligand directs the carboxyl group of atoms towards the MIDAS ion) and clustering of docked solutions sixty four different families could be found for the PUPA-LDV, with rmsd values less than 3.5 Å between members of the same family. For the remaining compounds, the number of families ranged from 8 to 12 excepting PUPA-2, where only three families fulfilling the selection criterion were found. Despite the large number of families, docking calculations also showed that PUPA-LDV and its derivatives were able to be systematically assembled in two main binding modes, defined by the position of the N terminal cap in the cleft, named hereafter MODE1 and MODE2 (Figure 3). A third (but less defined) binding mode, characterized by exposing the PUPA cap to the solvent, was seldom observed.

A set of common interactions was detected in the binding modes for all investigated ligands. For instance, atoms of the terminal carboxyl group (as well as the side chain oxygen atoms of aspartic acid for PUPA-LDV) were always coordinating the Mg^{2+} in the MIDAS region. Ser154, which also participates in the metal ion coordination, together with Ser247 played a significant role in the binding of the compounds setting up hydrogen bond contacts with
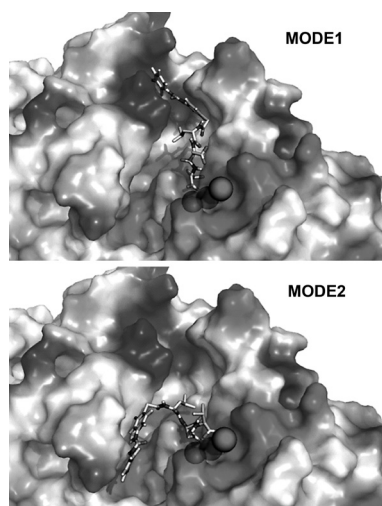
Figure 3. Binding modes of VLA-4 within the binding site represented as Connolly surface. MODE1 is seen in the upper panel, whereas MODE2 is depicted in the lower panel. The metal ions at the MIDAS region are represented as spheres.

these atoms. Main chain atoms of Tyr153 and Asn244 also contributed to stabilize the conformation of the ligand.

The main feature of the docked ligands in MODE1 was that they remained completely stretched, inside the crevice between the propeller and the β_A domain on the integrin head, in such a way that the PUPA cap mainly interacted with the residues Lys158, Lys159, Lys190 and Thr208. Cation-π interactions were apparently formed between side chains of Lys159, Lys158, Lys190 and the aromatic rings of the N-terminal cap. Although the force field has not been parameterized to take into account these special interactions, it is possible to envisage its existence due to the proximity of the intervening groups. Furthermore, the side chain of Thr208 usually hydrogen bonded to the PUPA carbonyl oxygen atom.

In MODE2, the compounds adopted a bent conformation in such a way that PUPA was located in the vicinity of Val245, Lys246 and Phe247. Cation-π interactions were also observed between Lys246 and PUPA rings. Peptide bond atoms of Val245 frequently hydrogen bonded the carbonyl or amide atoms of the ligand.

Finally, in MODE3 PUPA cap was situated outside the gap formed between the β propeller and β_A domains. This particular dizzy mode presented a less defined pattern, when compared to the previous ones, and their free energies of binding were always higher than the ones obtained in previous modes. Also, this specific mode, which was not observed for all ligands, was less populated and less probable, given that the hydrophobic PUPA rings were completely exposed to water.

Interactions between PUPA-LDV (and derivatives) and the $\alpha_4\beta_1$ integrin binding site are summarized in Table 1. The existence of interactions was determined based on distance criteria, namely, cation-π < 4.5 Å, hydrogen bond (O-O) < 3.5 Å, and electrostatic interaction < 5.0 Å.

When analyzing the results for all ligands, we were not able to find a good correlation between the calculated free energies of binding and the experimental IC_{50}'s.[11] This result was already expected because even though the current docking functions used to evaluate ligand-receptor binding affinities are suitable for ranking ligands of large databases, they have a poor capability to predict quantitatively, or even qualitatively, useful information about ligand-receptor affinities.[43] It is important to note that the lack of correlation between the experimental and the calculated ligand-receptor binding affinities are not directly associated to a poor performance of docking programs in reproducing the experimentally observed ligand-receptor binding modes. Actually, the current docking methodologies have a good performance to generate ligand conformations similar to

Table 1. Major interactions between PUPA-LDV (and derivatives) and α4β1 integrin binding site. Determination of interactions was based on distance criterion: cation-π < 4.5 Å, hydrogen Bond (O-O) < 3.5 Å, electrostatic interaction < 5.0 Å. Intermolecular interactions were mapped from docking calculation and monitored through MD simulations. Average length column shows the mean interaction distances (Å) held along the MD simulations. The asterisk means that the interaction was not maintained

	MODE1			MODE2	
AA	Interaction Type	Average length (MD)[a]	AA	Interaction Type	Average length (MD)[a]
K159(α)	Cation-π	*	Y153(β)	Hydrogen Bond	2.9 (0.1)[b]
K189(α)	Cation-π	*	S154(β)	Hydrogen Bond	2.9 (0.1)[b]
Y153(β)	Hydrogen Bond	3.3 (0.3)[b]	N244(β)	Hydrogen Bond	*
S154(β)	Hydrogen Bond	3.2 (0.2)[b]	D246(β)	Hydrogen Bond	3.6 (0.6)[b]
T208(β)	Hydrogen Bond	*	S247(β)	Hydrogen Bond	2.2 (0.4)[b]
N244(β)	Hydrogen Bond	3.4 (0.8)[b]	Mg^{2+} (MIDAS)	Electrostatic interaction	2.0 (0.05)[b]
L245(β)	Hydrogen Bond	3.3 (0.7)[b]			
S247(β)	Hydrogen Bond	*			
Mg^{2+} (MIDAS)	Electrostatic interaction	2.0 (0.07)[b]			

[a]Average length column shows the mean interaction distances (Å) held along the MD simulations; [b]The average lenght column shows the mean interaction distances (Å) plus its standard deviation value held along the MD simulations;*means that the interaction was not maintained.

crystallographically determined ligand-receptor structures. In most cases, the correct reproduction of the binding mode does not coincide with the binding affinity prediction performance.[43] This lack of correlation is likely due to a poor representation of important energy contributions, such as protein flexibility, solvation and entropy. However, PUPA-LDV ligand was correctly identified as the ligand with the best affinity for the receptor. Our results also indicated that the PUPA-LDV binding MODE1 was the preferred one.

When analyzing the results for the derived compounds, we observed a lack of correlation with experimental ligand binding affinities. However, we observed a clear preference for binding MODE1 for this particular set of ligands (Figure 4A) and also noticed that docking distances between MIDAS Mg^{2+} ion and the acid oxygen atoms were always lower for MODE1 than for MODE2 (Table 2).
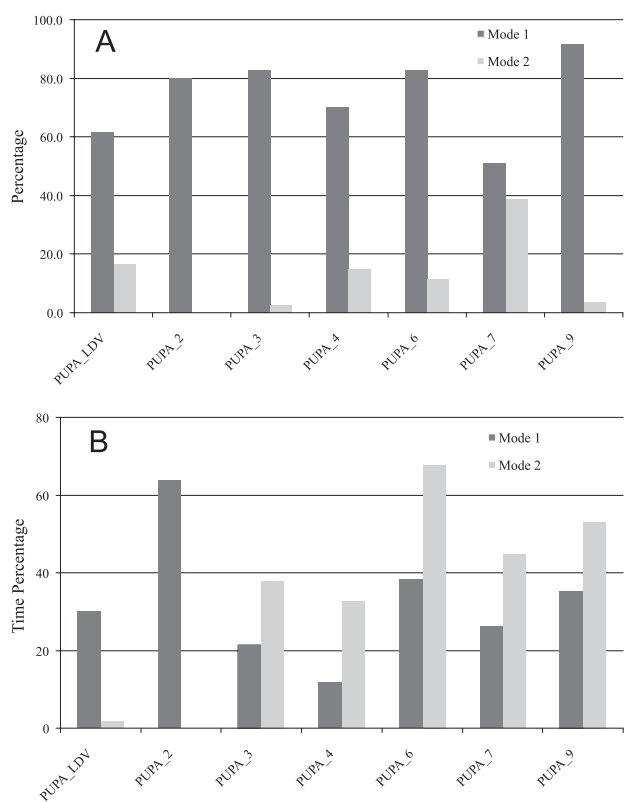

Figure 4. A) Percentage of occurrence of the binding modes obtained after molecular docking. It is clear the prevalence of MODE1 for all ligands. PUPA-2 did not exhibit the MODE2. **B)** Percentage of time that the trajectory conformations present rmsd values less than 2.0 Å when compared with MODE1 and MODE2 docked solutions. Most of ligands showed a clear preference for conformations in solution close to the associated with the MODE1.

MD simulations

The combination of molecular dynamics and docking procedures to investigate how ligands interact with receptors is well established.[44,45] Most of the studies apply molecular

Table 2. Distances (Å) between MIDAS Mg^{2+} and carboxylic oxygen atoms for both binding modes. The distances were obtained from docking calculation and 1 ns of MD simulations

Ligand	MODE1		MODE2	
	Docking	MD	Docking	MD
PUPA-LDV	2.1	1.99 (0.07)[a]		
	2.2	1.97 (0.05)[a]		
PUPA-2	2.2	1.94 (0.04)[a]	---	---
PUPA-3	1.7	1.93 (0.05)[a]		
	2.2	1.93 (0.08)[a]		
PUPA-4	1.8	1.97 (0.05)[a]	3.1	1.94 (0.05)[a]
PUPA-6	1.6	1.94 (0.05)[a]		
	1.8	1.92 (0.05)[a]		
PUPA-7	1.6	1.95 (0.05)[a]	1.8	1.95 (0.06)[a]
PUPA-9	1.7	1.94 (0.04)[a]	2.7	1.94 (0.08)[a]

[a]standard deviation values.

dynamics to improve the position of the ligand within the binding site, after a docking searching. Herein, we used molecular dynamics as an approach attempting to correlate the preferential ligand binding conformations within the $\alpha_4\beta_1$ integrin active site and their dynamical behavior when isolated in solution. Molecular dynamics simulations were performed to investigate thermodynamical (LIE calculations) and dynamical aspects (permanence time of docked solutions in water) of the ligands in the binding site and in solution. However, some structural features of the dynamical behavior of ligands within the headpiece of the integrin could be monitored. MD simulations showed that some of docking interactions were unstable, especially those forming cation-π with the PUPA cap, different from the hydrogen bonds formed with structural MIDAS amino acids (Ser152, Tyr153, Ser 154, Asn244, Leu245, Asp246 and Ser247). Average distance values and fluctuations along 1 ns of simulation are shown in Table 1.

The LIE free energy methodology indicated that the PUPA-LDV, in MODE1, resulted as the best ligand conformation for the whole series of compounds, supporting docking observations. According to the set of parameters used for LIE determination, MODE2 showed a high energy gap of almost 120 kcal mol^{-1}, (Figure 4B) with respect to the MODE1 conformation. Looking at the electrostatic energy contribution, we presumed that such interactions might have an important role in the prevalence of that particular binding mode, since this ligand presents 2 negatively charged groups, (Asp side chain and the Val C-terminal). It is important to note that during the dynamics evolution, the C terminal of PUPA-LDV drifted towards the MIDAS Mg^{2+} ion, producing a decrease in the electrostatic energy, facilitated by the high flexibility of this ligand.

For the whole set of ligands, the MODE1 showed lower electrostatic energies when compared to MODE2 (Table 3),

Table 3. Electrostatics and van der Waals energies obtained from LIE calculations. Free energy values are calculated according to LIE equation. $\Delta G_{bind} = \alpha(\langle U_{l\text{-}s}^{LJ}\rangle_{bound} - \langle U_{l\text{-}s}^{LJ}\rangle_{free}) + \beta(\langle U_{l\text{-}s}^{le}\rangle_{bound} - \langle U_{l\text{-}s}^{le}\rangle_{free}) + \gamma$ where energy units are in kcal mol^{-1}. Values in parenthesis indicate fluctuations

| | Water | | Complex | | | | | |
| | | | MODE1 | | | MODE2 | | |
	E_{el}	E_{vdW}	E_{el}	E_{vdW}	ΔG	E_{el}	E_{vdW}	ΔG
PUPA-LDV	–408.70(18.07)	–34.31(0.52)	–662.23(7.80)	–56.35(2.77)	–130.75	–397.79(7.51)	–54.26(1.43)	–7.59
PUPA-2	–181.46(1.99)	–32.48(0.40)	–285.81(2.10)	–47.15(0.88)	–54.83	-	-	-
PUPA-3	–179.21(2.71)	–35.08(0.63)	–310.10(6.48)	–50.59(1.53)	–68.25	–249.59(1.88)	–51.49(1.10)	–35.54
PUPA-4	–190.83(2.77)	–33.15(0.59)	–317.28(8.06)	–47.65(1.65)	–65.85	–248.88(5.10)	–51.85(1.79)	–31.65
PUPA-6	–190.20(2.88)	–32.87(0.65)	–350.06(3.74)	–43.84(1.60)	–81.91	–286.33(6.06)	–53.92(2.88)	–52.42
PUPA-7	–176.86(3.51)	–30.62(0.81)	–313.12(7.60)	–43.81(1.73)	–70.51	–255.76(5.48)	–45.39(1.77)	–40.87
PUPA-9	–177.47(2.94)	–32.46(0.55)	–301.46	–50.36(2.52)	–65.24	–248.48(8.28)	–53.69(1.81)	–39.48

contributing to a more favorable interaction. The energy difference between the MODE1 and MODE2 showed an average value of 35 kcal mol^{-1} for the compounds, except PUPA-LDV that presented a more accentuated gap between binding modes.

Although, the observed shorter distances between charged oxygen atoms and Mg^{2+} for MODE1 obtained by Docking procedure might explain the prevalence of MODE1 over MODE2, MD simulations showed that once in motion in the integrin active site, the charged groups of compounds maintained similar distance values for both binding modes. Apparently the proximity of the N-terminal CAP with several charged residues as Lys155, Lys159, Lys189, Lys190 and Glu157, in MODE1 appears to be responsible for the lowest electrostatic energies. On the other hand, the MODE2 showed only two charged residues near the PUPA cap: Lys246 and Glu276 with their side chains exposed to the solvent.

Molecules in solution are in constant motion and conformational transitions between different states occur frequently. During the binding process, the guest molecule experiences a conformational entropy loss ($\Delta S < 0$), resulting in a more static conformation within the active site. This effect of reduction of the ligand conformational entropy during the binding process is more important for highly flexible ligands (*i.e.,* high number of conformational degrees of freedom), which generally results in a decrease of the affinity of a flexible ligand towards a given receptor (*i.e.*, increases the free energy of binding, $\Delta G = \Delta H - T\Delta S$). For a particular flexible ligand, a decrease in the conformational entropy can be compensated by strong receptor-ligand interactions. However, it is also natural to expect that the conformational entropy loss effect can be less important if the flexible ligand binds to the receptor active site in a conformation close to that in solution.

Another important point associated with high flexible ligands is the fact that the receptor-ligand docking algorithms present a success rate of finding the correct ligand-receptor binding modes which is very sensitive to the number of the ligand conformational degrees of freedom.[44] The ability of the current docking searching procedures to explore successfully the associated ligand-receptor energy hypersurface decreases with the increasing of the ligand flexible bonds number.[45]

Furthermore, we attempted to correlate the preferential ligand binding conformations within the $\alpha_4\beta_1$ integrin active site and their dynamical behavior when isolated in solution.

In order to verify if the MODE1 and MODE2 docked conformations were feasible in solution, we compared them to the ones obtained from MD simulations trajectories of the molecules in explicit solvent, through the calculation of the rmsd values. Figure 4 shows the percentage of time the conformations obtained from the trajectories present rmsd ≤ 2.0 Å, taking MODE1 and MODE2 docked solutions as references respectively.
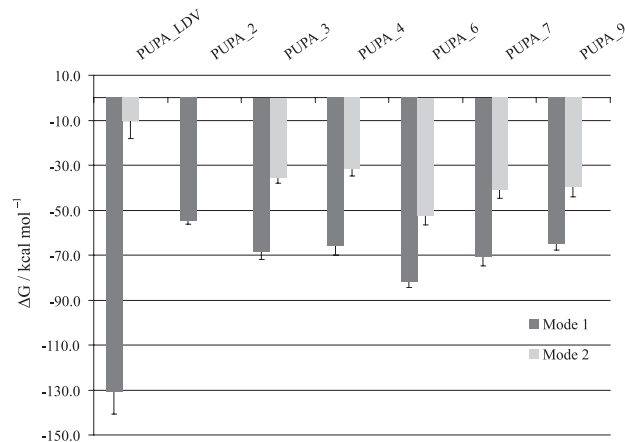

Figure 5. Calculated ΔG values for the set of ligands. The MODE1 appeared to be the more favorable one for all ligands.

The most active ligands UPA-LDV and PUPA-2 exhibited a clear preference for conformations in solution close to the more extended MODE1 docked conformation and MODE2 conformation was not observed in solution for PUPA-2 at all (Figure 5). It is also interesting that as long as the IC$_{50}$ values increases, increases the proportion of MODE2 in solution, as well.

These results provide good evidence that the conformation in solution of compounds is somehow related to their prevalence for binding the receptor. So, the existence of prevailing conformations in solution similar to the observed ones when the inhibitor is within the integrin active site, can be an indication that the energetic costs of ligand conformational relaxation on complex formation are minimized. Likewise, the entropic cost of the binding process is diminished.

Conclusions

Herein, we combined computational methodologies from MD simulations and molecular docking, to predict the binding modes of a set of ligands, derivatives of the 4-[*N*'-(2-methylphenyl)ureido]phenylacetyl-LDV compound within the gap between the α and β subunits of α$_4$β$_1$ integrin.

For these highly flexible compounds, we used MD simulations results to find the most probable ligand conformations in solution, correlating them with the potential binding modes within the binding site of α$_4$β$_1$.

Docking experiments of PUPA-LDV and derivatives in the active site of the VLA-4 antigen have identified two main binding modes, with a well defined binding pattern. These two binding modes cases present interactions between the antagonists and both subunits of the α$_4$β$_1$ integrin. In MODE1, the ligands remain completely stretched inside the crevice between the propeller and the β$_A$ domain on the integrin head. In MODE2 the ligands adopt a bent conformation in such a way that the PUPA cap is located in the vicinity of Val245, Lys246 and Phe247 of α subunit.

The amino acids Lys159, Lys189 and LysK190 of the α subunit, and Tyr153, Ser154, Thr208, Asn244 and Ser247 of the β subunit associated to the binding MODE1 and Tyr153, Ser154, Asn244 and Ser247 of the β subunit associated to the binding MODE2 were identified as essential residues for the binding process, converting them into interesting targets for site directed mutagenesis experiments.

The analysis of the free energy of binding (LIE), provided by the Autodock scoring function, has correctly identified PUPA-LDV as the ligand with the best affinity for the receptor. For its derivatives no correlation was obtained with the reported IC$_{50}$'s values. However, the LIE calculations have shown that practically all ligands have a clear preference for conformations within α$_4$β$_1$ active site associated with binding MODE1, *i.e.*, there exists a prevailing stretched conformation of PUPA-LDV and derivatives inside the active site cavity. It is very likely that the existence of the stretched conformation in solution helps the molecular recognition, at least for the two most active ones. In this sense, the predominance of an extended conformation in solution is a potential feature that would help in the binding process resulting in an enhanced affinity of the prototype compound for α$_4$β$_1$ integrin. Additionally, MD simulations of ligands in solution can be valuable approach to enable prediction of the binding modes of potentially new α$_4$β$_1$ integrin antagonists.

Acknowledgements

References

1. Hynes, R. O.; *Cell* **1992,** *69*, 11.

2. Ahmed, N.; Riley, C.; Rice, G.; Quinn, M.; *Clin. Exp. Metastasis* **2005,** *22*, 391.

3. Giannelli, G.; Astigiano, S.; Antonaci, S.; Morini, M.; Barbieri, O.; Noonan, D. M.; Albini, A.; *Clin. Exp. Metastasis* **2002,** *19*, 217.

4. Abraham, W. M.; Sielczak, M. W.; Ahmed, A.; Cortes, A.; Lauredo, I. T.; Kim, J.; Pepinsky, B.; Benjamin, C. D.; Leone, D. R.; Lobb, R. R.; *J. Clin. Invest.* **1994,** *93*, 776.

5. Macchiarulo, A.; Costantino, G.; Meniconi, M.; Pleban, K.; Ecker, G.; Bellocchi, D.; Pellicciari, R.; *J. Chem. Inf. Comput. Sci.* **2004,** *44*, 1829.

6. Hafler, D.A.; *J. Clin. Invest.* **2004,** *113*, 788.

7. Yang, X. D.; Karin, N.; Tisch, R.; Steinman, L.; McDevitt, H. O.; *Proc. Natl. Acad. Sci. U.S.A* **1993,** *90*, 10494.

8. Seiffge, D.; *J. Rheumatol.* **1996,** *23*, 2086.

9. Komoriya, A.; Green, L. J.; Mervic, M.; Yamada, S. S.; Yamada, K. M.; Humphries, M. J.; *J. Biol. Chem.* **1991,** *266*, 15075.

10. Lin, K.; Ateeq, H. S.; Hsiung, S. H.; Chong, L. T.; Zimmerman, C. N.; Castro, A.; Lee, W. C.; Hammond, C. E.; Kalkunte, S.; Chen, L. L.; Pepinsky, R. B.; Leone, D. R.; Sprague, A. G.; Abraham, W. M.; Gill, A.; Lobb, R. R.; Adams, S. P.; *J. Med. Chem.* **1999,** *42*, 920.

11. Singh, J.; Van Vlijmen, H.; Liao, Y.; Lee, W. C.; Cornebise, M.; Harris, M.; Shu, I. H.; Gill, A.; Cuervo, J. H.; Abraham, W. M.; Adams, S. P.; *J. Med. Chem.* **2002,** *45*, 2988.

12. Rice, P.; Longden, I.; Bleasby, A.; *Trends Genet.* **2000,** *16*, 276.

13. You, T. J.; Maxwell, D. S.; Kogan, T. P.; Chen, Q.; Li, J.; Kassir, J.; Holland, G. W.; Dixon, R. A.; *Biophys. J.* **2002,** *82*, 447.

14. Sali, A.; Blundell, T. L.; *J. Mol. Biol.* **1993,** *234*, 779.

15. Sali, A.; Potterton, L.; Yuan, F.; van Vlijmen, H.; Karplus, M.; *Proteins-Struct. Funct. Genet.* **1995,** *23*, 318.

16. Xiong, J. P.; Stehle, T.; Zhang, R.; Joachimiak, A.; Goodman, S. L.; Arnout, S. L.; *Science* **2002,** *296*, 151.

17. Marinelli, L.; Gottschalk, K. E.; Meyer, A.; Novellino, E.; Kessler, H.; *J. Med. Chem.* **2004,** *47*, 4166.

18. Newham, P.; Humphries, M. J.; *Mol. Med. Today* **1996,** *2*, 304.

19. Fiser, A.; Do, R. K.; Sali, A.; *Protein Sci.* **2000,** *9*, 1753.

20. Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J.; *J. Comput. Chem.* **2005,** *26*, 1701.

21. Humphries, M. J.; Symonds, E. J.; Mould, A. P.; *Curr. Opin. Struct. Biol.* **2003,** *13*, 236.

22. Yahalom, D.; Wittelsberger, A.; Mierke, D. F.; Rosenblatt, M.; Alexander, J. M.; Chorev, M.; *Biochemistry* **2002,** *41*, 8321.

23. Xiong, J. P.; Stehle, T.; Zhang, R.; Joachimiak, A.; Frech, M.; Goodman, S. L.; Arnaout, M. A.; *Science* **2002,** *296*, 151.

24. Laskowski, R. A.; Moss, D. S.; Thornton, J. M.; *J. Mol. Biol.* **1993,** *231*, 1049.

25. Laskowski, R. A.; Rullmannn, J. A.; MacArthur, M. W.; Kaptein, R.; Thornton, J. M.; *J. Biomol. NMR* **1996,** *8*, 477.

26. Eisenberg, D.; Luthy, R.; Bowie, J. U.; *Methods Enzymol.* **1997,** *277*, 396.

27. Morris, G. M.; Goodsell, D. S.; Huey, R.; Olson, A. J.; *J. Comput.-Aided Mol. Des.* **1996,** *10*, 293.

28. Humphries, M. J.; *Biochem. Soc. Trans.* **2000,** *28*, 311.

29. Springer, T. A.; *Curr. Opin. Struct. Biol.* **2002,** *12*, 802.

30. Arnaout, M. A.; Goodman, S. L.; Xiong, J. P.; *Curr. Opin. Cell. Biol.* **2002,** *14*, 641.

31. Van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hunenberger, P. H.; Kruger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G.; The GROMOS96 Manual and User Guide; *Biomolecular simulation;* Groningen, Zurich, 1996.

32. Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R.; *J. Chem. Phys.* **1984,** *81*, 3684.

33. Halgren, T. A.; *J. Comput. Chem.* **1996,** *17*, 490.

34. Berendsen, H. J. C.; Postma, J. P. M.; Gunsteren, W. F. V.; Hermans, J.; *Intermolecular Forces;* Reidel, Dordrecht, 1981.

35. Hess, B.; Bekker, H.; Berendsen, H.; Fraaije, J.; *J. Comput. Chem.* **1997,** 18, 1463.

36. Aqvist, J.; Medina, C.; Samuelsson, J. E.; *Protein Eng.* **1994,** *7*, 385.

37. Marelius, J.; Kolmodin, K.; Feierberg, I.; Aqvist, J.; *J. Mol. Graphics Modell.* **1998,** *16*, 213.

38. Lee, F. S.; Warshel, A.; *J. Chem. Phys.* **1992,** *97*, 3100.

39. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C.; *J. Comput. Phys.* **1977,** *23*, 327.

40. Hansson, T.; Marelius, J.; Aqvist, J.; *J. Comput.-Aided Mol. Des.* **1998,** *12*, 27.

41. Marelius, J.; Hansson, T.; Aqvist, J.; *Int. J. Quantum Chem.* **1998,** *69*, 77.

42. Krissinel, E.; Henrick, K. (2005); *Detection of protein assemblies in crystals*; Berthold, M. R.; Glen, R., Diederichs, K., Kohlbacher, O., and Fischer, I., eds.; Complife 2005, LNBI 3695, Berlin Heidelberg: Springer-Verlag, 2005.

43. Warren, G. L.; Andrews, C. W.; Capelli, A.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S.; *J. Med. Chem.* **2006,** *49* (20), 5912.

44. De Magalhães, C. S.; Barbosa, H. J. C.; Dardenne, L. E.; *Lecture Notes in Computer Science* **2004,** *3102*, 368.

45. De Magalhães, C. S.; Barbosa, H. J. C.; Dardenne, L. E.; *Genet. Mol. Biol.* **2004,** *27*, 605.