UNIVERSITÀ DEGLI STUDI DI MILANO
FACOLTÀ DI SCIENZE E TECNOLOGIE
DIPARTIMENTO DI CHIMICA

Doctorate school in Chemical Sciences and Technologies

Curriculum

Chemical Sciences (XXVII cycle)

Integrin and cadherin ligands: interaction studies by computational methods and bioaffinity NMR on intact cells

PhD thesis of

Ileana GUZZETTI

R09684

Tutor: Dr. Donatella Potenza

Coordinator: Prof. Emanuela Licandro

Academic year 2013/2014

Doctoral Final Oral Examination: November, 24th 2014

Examination Committee: Chairperson: Prof. F. Nicotra

Second Member: Prof. L. Lay

Third Member: Prof. F. Sansone

Fourth Member: Dr. M. Petitou

Acknowledgements

Getting through my dissertation required more than academic support, and I have many people to thank for listening to and, at times, having to tolerate me over the past three years. I am using this opportunity to express my gratitude to all these people.

First I want to thank my tutor Dr. D. Potenza for giving me the opportunity to gain experience and to be involved in several projects. She taught me that our goals have to go beyond scientific publications. We should focus, instead, on the reward given by understanding how a new system works and how to effectively apply new methods.

A special thanks goes to Dr. F. Vasile who first taught me how to use the NMR equipment, I am thankful for her invaluable constructive and friendly advice.

I express my warm thanks to Dr. L. Belvisi and Dr. M. Civera for their time, good scientific advice and many insightful discussions. Dr. L. Belvisi has been one of my primary resources for getting my science questions answered. I also have to thank Prof. C. Gennari, Prof. U. Piarulli and Prof. A. Bernardi for their suggestions, help and for listening patiently to my presentations. In addition, I want to thank my colleagues and all the people who helped me not only in recent years, but also at the beginning of my work. In particular, I want to thank Dr. E. Moroni who mentored me during my Bachelor thesis.

Table of Contents

Supplementary information for Chapter 2

Supplementary information for Chapter 3

Preface

On a molecular level protein – ligand interactions are central to a number of biological processes, but its investigation was inherently difficult due to several problems, especially for membrane proteins. Usually NMR spectroscopy is valued for its ability to shed light on molecular structure, but actually its greatest potential lies in the information that can be provided about molecular interactions at an atomic level.[1] In the past ten years, a huge amount has been published on NMR ligand interactions including numerous reviews.[2,3,4,5,6]

In this thesis, we present NMR spectroscopy as a powerful tool for the study of small molecule – biomolecule complexes, with a focus on ligand – based strategies. The analysis of this type of interactions produces a whole set of challenges, including the characterization of the dynamic behaviour and the conformational properties of the ligands in complex with the target macromolecules. Membrane proteins change their conformation if extracted from their environment. Therefore it is important to work in the biophysical neighbourhood of the membrane itself and not in an isotropic extracellular medium.[2]

Membrane proteins, in fact, cause problems for the application of NMR-based ligand-screening methods because of the need to maintain the proteins in a membrane environment. For this reason, studies were conducted directly with the proteins embedded in liposomes or micelles. This environment can interact with ligands non-specifically and can overlap with ligand signals, where the ligand is the observable of our experiments. We, in particular, worked with intact tumour cells overexpressing the proteins of interest; also in this case, many experimental problems have raised, such as maintaining the cells in suspension, verifying that the interaction is not aspecifically, verifying that the protein is in its activate form *etc.*. Even though the use of NMR spectroscopy is highly underdeveloped in the area of drug discovery for membrane proteins and

[1] Potenza, D.; Vasile, F. Seminars in Organic Synthesis, "A. Corbella" Summer School, 37th, Gargnano, Italy, June 18-22 **2012**, 225 – 248.
[2] a) Meyer, B.; Peters, T. *Angew. Chem. Int. Ed.* **2003**, 42, 864 – 890; b) Mayer, M. and Meyer, B. *J. Am. Chem. Soc.* **2001**, 123, 6108 – 6117; c) Johnson, M. A. and Pinto, B. M. *Carbohydrate research* **2004**, 339, 907 – 928.
[3] Angulo, J.; Nieto, P. M. *Eur. Biophys. J.* **2011**, 40, 1357 – 1369.
[4] Calle, L. P.; Canada, F. J.; Jimenez-Barbero, J. *Nat. Prod. Rep.* **2011**, 28, 1118 – 1125.
[5] Bhunia, A.; Bhattacharjya, S.; Chetterjee, S. *Drug Discov. Today* **2012**, 17, 505 – 513.
[6] Carlomagno, T. *Annu. Rev. Biophys. Biomol. Struct.* **2005**, 34, 245 – 266.

only a few papers treated these systems in their natural environment, these proteins constitute the most important class of drug targets, demonstrating the necessity to gain insight into their interactions with potential ligands.[7] Most of the biomedically relevant proteins as well as about 30% of the proteins in mammalian systems are membrane bound or integrated proteins and the interactions of peptides with these membranes are fundamental to a number of biological processes, such as the insertion and folding of peptides in membranes, the rupturing of membranes by toxins, the membrane-mediated mechanism of peptide-receptor interactions.[7]

A variety of biophysical methods have been developed, such as Surface Plasmon Resonance (SPR), isothermal titration calorimetry, fluorescence polarization assay, fluorescence resonance energy transfer, enzyme-linked immunosorbent assay *etc.*, but none of them gives insight into the binding epitope of the ligand at an atomic level with proteins in their natural environment and X-ray, that could theoretically provide this information, can find difficulties to crystallize this kind of receptors in an active form.[2,8,9]

NMR spectroscopic techniques have emerged as powerful methods to identify and characterize the binding activity of ligands with receptor proteins; transferred-NOE (*tr*-NOE) and Saturation Transfer Difference (STD) have gained momentum. Among the ligand-based NMR techniques, STD and transferred-NOE focus on the NMR signals of the ligand and utilize NOE effects between protein and ligand. These methods do not require labelled protein, since only the ligand NMR signals are detected and only a small amount of protein is required. These techniques are particularly useful in the medium–low affinity range, therefore they have been adopted to detect ligand interactions in various systems. In particular, they are used for: *i)* the definition of the bioactive conformation of the ligand in the bound state (*tr*-NOESY), *ii)* the identification and characterization of the binding mode of the ligand in membrane proteins with epitope mapping of the ligand itself (STD).[1,2] These techniques will be better described in Chapter 1.

In particular, our group is interested in the study of the interaction between membrane proteins and small peptidomimetic ligands. Only a few reviews[10,11] treated the study of

[7] Yanamala, N.; Dutta, B. B. et al. *Chem. Biol. Drug Des* **2010**, 75, 237 – 256.
[8] Cala, O.; Guillière, F.; Krimm, I. *Anal. Bioanal. Chem.* **2014**, 406, 943 – 956.
[9] Dalvit, C. *Drug Discov. Today* **2009**, 14, 1051 – 1057.
[10] Claasen, B.; Axmann, M.; Meinecke, R and Meyer, B. *J. Am. Chem. Soc.* **2005**, 127, 916 – 919.
[11] Meinecke, R. and Meyer, B. *J. Med. Chem.* **2001**, 44, 3059 – 3065.

peptidomimetics - membrane proteins interactions in their natural environment. In spite of the difficulties, some groups[12,13,14,15] have focused their research on biologically relevant transient protein - ligand interactions, obtaining significant results in the field and inspiring other groups to a further improvement of the methods applied to these complex systems.

During my PhD, I had the highly qualifying opportunity to grasp these new potent NMR methods, *tr*NOE and STD,[1,2] and to apply them for assessing small molecule ligand interactions with intact tumour cells and isolated proteins. Specifically, two main topics have been addressed:

1) One part of my PhD thesis was developed within the framework of a PRIN project (MIUR PRIN project 2010NRREPL "Synthesis and Biomedical Applications of Tumour-Targeting Peptidomimetics") involving different research units and aimed at the design, synthesis, structural analysis and biological evaluation of new peptidomimetic integrin ligands. Integrins are the main cell surface receptors mediating cell adhesion to extracellular matrices and one of the best known aspects for the inhibition of integrins is their ability to recognize endogenous ligands characterized by short RGD (Arg-Gly-Asp) sequences.[16] On the base of these information and due to the presence of some RGD cyclic ligands in clinical phases, our group developed a small library of peptidomimetics containing diketopiperazine scaffolds, known to be a β-turn inducer, and the recognition RGD sequence. My contribution to the project concerned the application of NMR and computational techniques aimed at obtaining a detailed conformational study of the diketopiperazine peptidomimetic ligands containing the Arg-Gly-Asp (RGD) and *iso*DGR sequence, but overall a detailed investigation of ligand-receptor interactions with live cells overexpressing integrins on their membrane, *e. g.* platelets and tumour cells. In particular, I set up live cell NMR methodologies by using human platelets and investigating the interaction of some RGD peptidomimetic ligands with the platelet-specific integrin $\alpha_{IIb}\beta_3$. The results of these experiments are discussed in the first part of Section 2-Chapter 2. Then, I expanded the live cell NMR study by

[12] Mari, S.; Serrano-Gomez, D.; Canada, F. J.; Corbì, A. L.; Jimenez-Barbero, J. *Chem . Int. Ed.* **2005**, 44, 296 – 298.
[13] Meyer, B.; Weimar, T.; Peters, T. *Eur. J. Biochem.* **1997**, 246, 705 – 709.
[14] Bernardi, A.; Arosio, D.; Manzoni, L.; Monti, D.; Posteri, H.; Potenza, D.; Mari, S.; Jimenez-Barbero, J. *Org. Biomol. Chem.* **2003**, 1, 785 – 792.
[15] Airoldi, C.; Giovannardi, S.; La Ferla, B.; Jimenz-Barbero, J.; Nicotra, F. *Chemistry - A European Journal* **2011**, 17, 13395 – 13399.
[16] a) Shimaoka, M.; Springer, T. A. *Nature Rev. Drug Discov.* **2003**, 2, 703 – 716; b) Hynes, R. O. *Cell* **2002**, 110, 673 – 687.

using human tumour cells and gaining insight into the interaction of the small library of ligands with the tumour-associated integrins $\alpha_V\beta_3$ and $\alpha_5\beta_1$ (second part of the **Section 2-Chapter 2**). The NMR results were supported by docking calculations of the ligands in the active site of the different integrins and were compared to the results of competitive receptor binding assays and competitive cell adhesion experiments.

2) The second part of my PhD was mainly focused on cadherins, a class of adhesive proteins that promotes homophilic interactions. This project, part of a FIRB project coordinated by Dr. M. Civera (Computer-aided design, synthesis and biological evaluation of peptidomimetics targeting N-cadherin as anticancer agents, MIUR-FIRB "Futuro in Ricerca" RBFR088ITV), benefits from different collaborations and aimed at obtaining a thorough understanding of the peptidomimetics - cadherin interactions and a full comprehension of their binding mechanism.[17] Our group set out to rationally design and synthesize small peptidomimetic molecules aimed at providing partial or total inhibition of the cadherin - cadherin interaction. The rational design of mimics of the tetrapeptide sequence Asp1-Trp2-Val3-Ile4 (DWVI), which is the initial part of the adhesion arm,[18] has been performed by replacing the central dipeptide Trp2-Val3 unit of the DWVI adhesive motif with several scaffolds developed in our laboratories.[19] An initial computational study led to the selection of some promising candidates to be synthesized. Due to the difficulties to crystallize small molecule - cadherin complexes, NMR has proved to be a useful method of investigation. The NMR study of the interactions was conducted with three ligands and the isolated proteins, that a recent paper demonstrates to have a role in tumour progression either if embedded in the membrane or free in solution.[20] In particular, the interactions between the EC1-EC2 domain of the two cadherins, E- (epithelial) and P- (placental) and three ligands were

[17] Doro, F.; Colombo, C.; Alberti, C.; Arosio, D.; Belvisi, L.; Casagrande, C.; Fanelli, R.; Manzoni, L.; Parisini, E.; Piarulli, U.; Luison, E.; Figini, M.; Tomassetti, A.; Civera, M. "Computational design of novel peptidomimetic inhibitors of cadherin homophilic interactions" *European Journal of Medicinal Chemistry* (submitted).

[18] Parisini, E.; Higgins, J. M. G.; Liu, J.-H.; Brenner, M. B.; Wang, J.-H. *J. Mol. Biol.* **2007**, 373, 401 – 411.

[19] a) Manzoni, L.; Arosio, D.; Belvisi, L.; Bracci, A.; Colombo, M.; Invernizzi, D.; Scolastico, C. *J. Org. Chem.* **2005**, 70, 4124 – 4132; b) Ressurreição, A. S. M.; Bordessa, A.; Civera, M.; Belvisi, L.; Gennari, C.; Piarulli, U. *J. Org. Chem.* **2008**, 73, 652 – 660; c) Arosio, D.; Belvisi, L.; Colombo, L.; Colombo, M.; Invernizzi, D.; Manzoni, L.; Potenza, D.; Serra, M.; Castorina, M.; Pisano, C.; Scolastico, C. *ChemMedChem* **2008**, 3, 1589 – 1603; d) Manzoni, L.; Belvisi, L.; Arosio, D.; Civera, M.; Pilkington-Miksa, M.; Potenza, D.; Caprini, A.; Araldi, E. M. V.; Monferini, E.; Mancino, M.; Podestà, F.; Scolastico, C. *ChemMedChem* **2009**, 4, 615 – 632; e) Marchini, M.; Mingozzi, M.; Colombo, R.; Guzzetti, I.; Belvisi, L.; Vasile, F.; Potenza, D.; Piarulli, U.; Arosio, D.; Gennari, C. *Chem. Eur. J.* **2012**, 18, 6195 – 6207.

[20] Brouxhon, S. M.; Kyrkanides, S.; Teng, X.; Athar, M.; Ghazizadeh, S.; Simon, M.; O'Banion, M. K.; Ma, L. *Oncogene* **2014**, 33, 225 – 235.

examined. The aim of the project, however, is not only the analysis of the ligand epitope, but further studies should be done. NMR studies using labeled protein constructs and molecular dynamics simulations allowing both ligand and protein flexibility will significantly improve the structural understanding of the binding of peptidomimetic ligands to the extracellular domains of the cadherins. The combined use of NMR and computational approaches will permit to increase the affinity of our ligands for specific cadherins, by optimization of the interactions with the receptor. The results obtained for this part of the project are well described in Chapter 3.

The research activities described in the thesis have led to the following publications and communications:

PUBLICATIONS

- **Determination of the binding epitope of RGD – peptidomimetics to $\alpha_V\beta_3$ and $\alpha_{IIb}\beta_3$ integrin rich – intact cells by NMR and computational studies**

  I. Guzzetti, M. Civera, F. Vasile, E. M. Araldi, L. Belvisi, C. Gennari, D. Potenza, R. Fanelli, and U. Piarulli, *Org. Biomol. Chem.* **2013**, *11*, 3886 – 3893

- **Cyclic isoDGR Peptidomimetics as Low-Nanomolar $\alpha_V\beta_3$ Integrin Ligands**

  M. Mingozzi, A. Dal Corso, M. Marchini, I. Guzzetti, M. Civera, U. Piarulli, D. Arosio, L. Belvisi, D. Potenza, L. Pignataro, and C. Gennari, *Chem. Eur. J.* **2013**, *19*, 3563 – 3567

- **Cyclic RGD-Peptidomimetics containing Bifunctional Diketopiperazine Scaffolds as New Potent integrin Ligands**

  M. Marchini, M. Mingozzi, R. Colombo, I. Guzzetti, L. Belvisi, F. Vasile, D. Potenza, U. Piarulli, D. Arosio and C. Gennari, *Chem. Eur. J.* **2012**, *18*, 6195 – 6207

COMMUNICATIONS

- EMBO Workshop, 1-6 June 2014 (Principina Terra, Italy)

Poster presentation (P16): "Insights into molecular bases of cadherin – small ligands interaction by STD-NMR"

I. Guzzetti, D. Potenza, L. Belvisi, M. Civera, F. Doro, E. Parisini, V. Nardone

- Computationally Driven Drug Discovery (CDDD) 3rd meeting, 4 – 6 March 2014 (Verona)

Poster presentation: "Targeting protein – protein interactions in cancer with peptidomimetics: insights into ligand conformation and ligand-receptor interactions"

L. Belvisi, M. Civera, C. Gennari, I. Guzzetti, U. Piarulli, D. Potenza

- **SMASH NMR Conference**, 22 -25 September 2013 (Santiago de Compostela, Spain)

Oral and poster presentation (P54): "A comprehensive NMR Analysis of the Interaction between DKP-RGD ligands and Membrane Proteins on Intact Cells"

I. Guzzetti, D. Potenza, F. Vasile, L. Belvisi, M. Civera, I. Silvestri, U. Piarulli, C. Gennari

- 35th FGMR Discussion Meeting and Joint Conference of the German, Italian and Slovenian Magnetic Resonance Societies, 9 – 12 September 2013 (Frauenwörth, Germany)

Poster presentation (P93): "STD-NMR experiments on tumor cells to investigate RGD ligand – membrane protein interaction"

I. Guzzetti, D. Potenza, F. Vasile, L. Belvisi, M. Civera, I. Silvestri, U. Piarulli, C. Gennari

Chapter 1

NMR AND COMPUTATIONAL METHODS

Section 1: NMR methods

Drug discovery is a complex and costly endeavor, where few drugs that reach the clinical testing phase make it to market. High-throughput screening (HTS) is the primary method used by the pharmaceutical industry to identify initial lead compounds. Unfortunately, HTS has a high failure rate and is not particularly efficient at identifying viable drug leads. These shortcomings have encouraged the development of alternative methods to drive the drug discovery process.[1] Specifically, NMR spectroscopy and molecular docking are routinely being employed as important components of drug discovery research; moreover, NMR has demonstrated its utility in fragment-based drug design, a novel and increasingly popular strategy for lead discovery that provides an alternative to conventional high-throughput screening.[2,3,4]

NMR spectroscopy and molecular docking are perfectly complementary techniques, and were essential for our research. Mono- and bidimensional experiments were useful for the attribution and for the conformational studies of the peptidomimetics:

- TOCSY (TOtally Correlated SpettroscopY),[5,6] a method developed specifically for peptides that permits to identify protons of the same spin system.
- NOESY (Nuclear Overhauser Effect SpectroscopY),[5,6] useful for the identification of dipolar couplings between protons.
- HSQC (Heteronuclear Single-Quantum Correlation spectroscopy),[5] heteronuclear sequence ^{1}H-^{13}C useful for the identification of proton – carbon coupling (^{1}J).

For the study of ligand-protein interactions, two are the methods used and these will be described in detail in the next paragraphs, *tr*-NOESY (transferred NOE) and STD (Saturation Transfer Difference).

For the computational studies, the programs and methods used, all based on molecular mechanics, are:

- MacroModel,[7,8] with its graphical interface Maestro, used both for the conformational analysis (MC/MM) and for molecular dynamics simulations (MC/SD).

- Glide (Grid-based ligand docking with energies),[9] used for the identification of the binding modes (docking) and for the prediction of the binding affinity (scoring) of the ligands to the receptor.

1.1.1 Protein – ligand interactions by NMR

Nuclear Magnetic Resonance (NMR) spectroscopy has long been a favourite tool of chemists interested in host-guest systems because it permits access to a wealth of information about the molecular recognition reaction. NMR has evolved dramatically in the last 20 years and, in parallel with the development of NMR methods for the determination of protein structure, a variety of tools aimed at detecting protein - ligand interactions have been proposed and are being now used both in industrial and academic laboratories as valuable tools for structure-based drug discovery.[10]

Solution-state NMR has become an accepted method for studying the structure of small proteins in solution. This has resulted in over 3000 NMR-based co-ordinate sets being deposited in the Protein Data Bank. It is becoming increasingly apparent, however, that NMR is also a very powerful tool for accessing interactions between macromolecules and various ligands. These interactions can be assessed at a wide variety of levels, *e.g.* qualitative screening of libraries of pharmaceuticals, "chemical shift mapping", dissociation constants, the complete three-dimensional structure determination of a protein-ligand complex.[11]

In the past ten years, a huge amount has been published on NMR investigations of protein - ligand interactions, including numerous reviews.[12-16] The binding process is regarded as an equilibrium condition which results from a balance between association and dissociation events. NMR spectroscopy is based on the study of simple parameters such as chemical shift and the dipole interactions, which allow to determine both the structure and conformation of the molecules examined. These same parameters may be employed to deduce key features of ligand - receptor molecular recognition processes. NMR-based methods, in addition, can be applied to weak and transient protein - protein or protein - ligand complexes that are difficult to study by other structural methods. Theoretically, all NMR spectroscopic parameters can be useful for

assessing the binding activity of a ligand to a protein. Actually, only parameters with high sensitivity are used: chemical shift, relaxation times, diffusion constants, NOEs, or saturation serve as measures of binding.[2]

The binding of a ligand to a protein is depicted below. The ligand, L, a small molecule (such as a peptide or oligosaccharide) binds reversibly to a protein P.

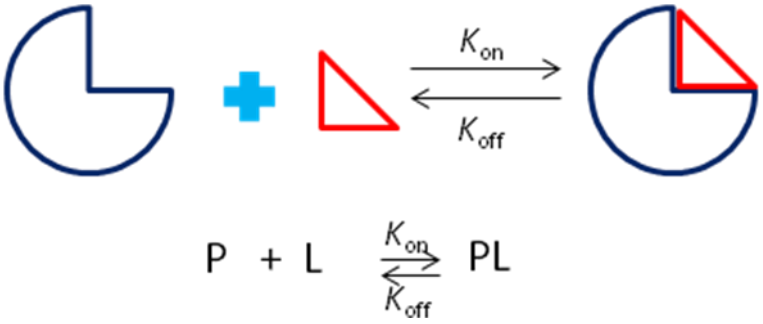


Figure 1.1.1. Ligand - protein equilibrium in a intermolecular exchange process.[2]

The binding of ligands to macromolecules is an intermolecular exchange process that can be evaluated by an association constant, K_A or by a dissociation constant, K_D, that, for a protein with a single binding site, are defined as:

$K_A = k_{on} / k_{off} = [PL]/[P][L]$ Equation 1

$K_D = [L][P]/[PL] = k_{off} / k_{on}$ Equation 2

where [L], [P] and [PL] are the equilibrium concentrations of "free" ligand, "non-bonded" protein and complexed state, respectively (**Figure 1.1.1.**).

The association constants of typical binding reactions of biomolecules range from 10^2 to 10^{12} M^{-1}. Values for protein - carbohydrate and protein - peptide interactions are typically weaker. Since the association rate is usually rapid, limited only by molecular diffusion, dissociation rates are also rapid, leading to fast exchange of the ligand between free and bound states.[17] These conditions, with a K_D between $>10^{-3}$ and $<10^{-7}$M, are the typical ones where NMR techniques can

be applied. In fact, an advantage of using NMR to measure protein - ligand interaction is that the NMR method extends the range of measurable interactions into the mM range, a region not well covered by traditional biochemical binding assays.

The NMR properties of free and bound states are different. The NMR signals from free and bound states can be in the "fast" (one averaged signal), "slow" (two non-averaged signals) or intermediate exchange regime (broad lines). If $K_D > 10^{-5}$ M the exchange regime is normally fast, while if $K_D < 10^{-6}$ M, the exchange rate is often slow. k_{off} and ($\delta_f - \delta_b$) are key parameters in deciding the exchange regime (**Figure 1.1.2.**).[11] As already mentioned, the interaction protein - ligand is a dynamic system for which the rate of exchange between free and bound forms is decisive for the quantitative NMR assessment of species in solution.

Chemical exchange in NMR refers to any process in which a nucleus exchanges between two or more environments in which its NMR parameters (chemical shift, scalar coupling, dipolar coupling, relaxation rate) differ. It is worth remembering that it is considered a very slow exchange process that which occurs within seconds, while a fast process is included in the scale between micro-and nano-seconds. **Figure 1.1.2.** shows the time scale related to the NMR parameters and to the dynamic processes of the molecules.[2]

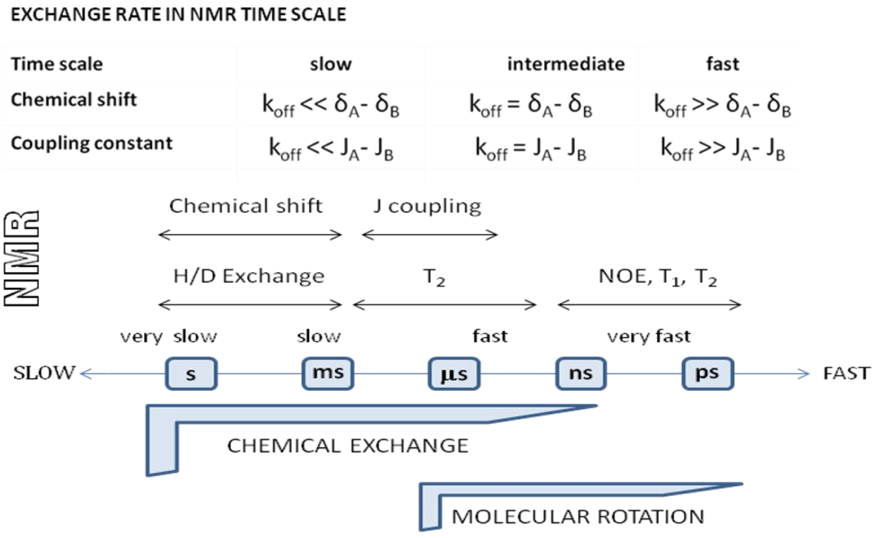


Figure 1.1.2. Exchange rate relative to chemical shift and constant coupling (top). Time scale in NMR dynamics (bottom).[2]

For a system **in slow exchange** (lifetime is long) on the chemical shift time scale, the frequency (or chemical shift) of each environment is sampled before exchange so, solved and integrated signals may be expected for the free and bound state (**Figure 1.1.2.**). In practice, a situation of this type is very rare to achieve because of the extreme difficulty to detect and integrate μM signals into a complex and crowded spectrum. More generally, systems **in fast exchange** are studied. For these systems the observed NMR response of a ligand (or protein) is the mole fractions weighted average of the NMR parameters of the free and bound states.

$M_{obs} = X_L M_L + X_{PL} M_{PL}$ Equation 3

Where M_{obs} is any NMR observable (chemical shift, scalar coupling, dipolar coupling, relaxation rate) characteristic of the equilibrium system, X_L and X_{PL} are the mole fractions of the free and bound ligand, and M_L and M_{PL} are the NMR parameters of the ligand in the free and bound state, respectively. The mole fractions are defined as: $X_L + X_{PL} = 1$ and $X_P + X_{PL} = 1$.

As mentioned above, this equation is applicable if the exchange is fast on the relaxation and chemical shift timescales of the system, namely if k_{off} is larger than the chemical shift difference between the free and the bound forms of the ligand (**Figure 1.1.3.**). Under this conditions, only one NMR line is observed (**Figure 1.1.3.**). The NMR observable M_{obs}, grown on the ligand molecule in the bound state, is transferred to the free ligand and vice versa; this transfer process is mediated by a rapid exchange between M_{PL} and M_L during the time necessary to measure M_{obs}. The individual populations X_L and X_{PL} depend on the K_D value and on the concentrations of L and P. From equation 3 it is evident that the resonances of the ligand in a ligand-target mixture can only be observed if the ligand is in excess with respect to the target.[2]

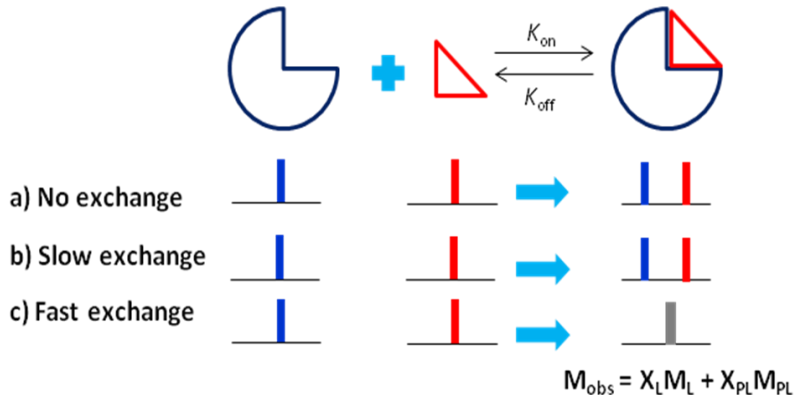


Figure 1.1.3. In the fast exchange, the frequency domain contains a single signal whose chemical shift is a weighted average of the chemical shifts of the two states.

The object of the NMR observation might be the ligand or the protein. Protein – ligand complexes are analyzed using the so-called protein-observed and ligand-observed NMR experiments in which the NMR parameters of the protein and the ligand, respectively, are compared in their free and bound states.[2,18]

In protein-observed methods, the most typical parameter and the easiest to follow is the chemical shift. Chemical shift perturbations of the protein resonances observed on ligand addition are identified to localize the ligand binding site. This enables one to immediately distinguish specific from non-specific binding. The 3D structure of the protein – ligand complex can be resolved via heteronuclear experiments performed on isotopically labeled (^{13}C, ^{15}N, ^{2}H) protein samples. The structure resolution requires molecular dynamics calculations with experimental NMR restraints resulting from chemical shifts, scalar couplings, nuclear Overhauser effects (NOEs), paramagnetic interactions or dipolar couplings. The major drawbacks of protein-observed methods are the experimental time and the need for highly stable and soluble protein. In addition, these methods are limited in routine practice to proteins with low molecular masses (less than 30 kDa) to avoid great effort with regard to both labeling strategies and resonance assignment.[19]

Ligand-based methods do not require labelled protein, since only the ligand NMR signals are detected and only a small amount of protein is required. These techniques are particularly useful in the medium–low affinity range, therefore they have been adopted to detect ligand interactions in various systems. Among the ligand-based NMR techniques, Saturation Transfer

Difference (STD) and transferred-NOE focus on the NMR signals of the ligand and utilize NOE effects between protein and ligand. Therefore, the STD technique can be used as an epitope mapping device to describe the target-ligand interactions, while the *tr*NOE offers easy detection of binding events, additionally producing information on the bound conformation of the ligands. In addition to *tr*-NOESY and STD, there are other NMR experiments, based on the Overhauser effect, able to screen potential leads from a compound library. WaterLOGSY[20] and INPHARMA[21] represent an example, even if they are less widespread NMR techniques.

1.1.2 tr-NOESY (transferred-NOE)

The fundamentals of the NOE were described very early in the history of NMR, in a classic paper by Solomon published in 1955.[22a] Solomon's work then lay almost dormant for some years until the advent of double resonance techniques led to the more widespread availability of spectrometers having a decoupler. The first paper to demonstrate the power of the NOE in structural problems was by Anet and Bourne in 1965.[22b] In 1972, Balaram et al.[22c] observed a negative NOE enhancement between a protein and a peptide binding to it, and they showed that negative enhancements are a general property of large molecules. The second major advance was the introduction of the two dimensional NOE experiment, largely by Ernst's group. These advances could not have had the impact they did without developments in instrumental techniques. The availability of the first FT spectrometers around 1970, the development of superconducting magnets that allow more stability and the advent of dedicated computer permit the diffusion of NOE experiments.[22]

In order to understand the *tr*-NOESY experiment, it is important to know the theory of the nuclear Overhauser effect (NOE). The nuclear Overhauser effect (NOE) is defined as the change in intensity of a certain resonance after perturbation of the population of a spin in close spatial proximity (by saturation or inversion). The interaction through space of the dipoles of neighbouring nuclei, or cross-relaxation, gives rise to the NOE.[17]

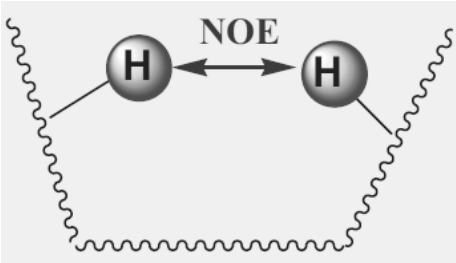


Figure 1.1.4. Nuclear Overhauser effect between two protons closed to space.

The cross-relaxation rate, σ, governs the magnitude and sign of the NOE. For an isolated two-spin system, the change in intensity of spin I after perturbation of spin S is described by Solomon (**Figure 1.1.5.**) equation:

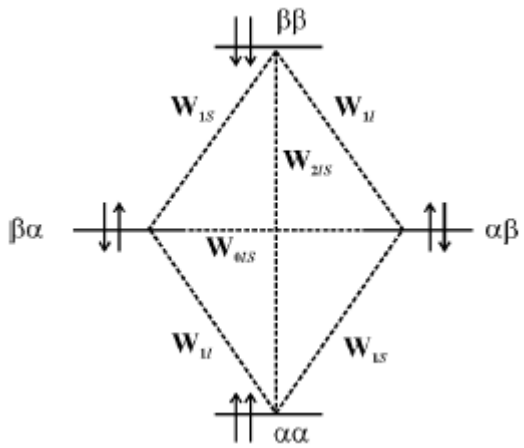


Figure 1.1.5. Solomon diagram.

$$dI_Z/dt = - (I_Z - I_{Z0}) (W_{0IS} + 2W_{1I} + W_{2IS}) - (S_Z - S_{Z0}) (W_{2IS} - W_{0IS}) \qquad \text{Equation 5}$$

where the W factor are transition probabilities (rate constants) for zero-, single-, and double-quantum transitions between energy levels of the two-spin system, as shown in **Figure 1.1.5.**, and I_Z and S_Z are the longitudinal components of the magnetization of spins I and S. Transitions between the energy levels are stimulated by transient local magnetic fields induced by the dipoles of the spins tumbling past each other. The W transition probabilities (rate constants) therefore depend on the rate of molecular tumbling. The term W_{2IS} is the rate constant for double-quantum transitions and is stimulated by local magnetic fields fluctuating at frequencies of $2\nu_0$

(the frequency difference between the $\alpha\alpha$ and $\beta\beta$ energy levels); therefore, it is greatest if the rate of molecular tumbling is rapid. This term gives rise to positive NOE enhancements. The term W_{0IS} describes zero-quantum transitions, and is stimulated by fields fluctuating at very low frequencies ($\nu_I - \nu_S \sim 0$); this term gives rise to negative NOE enhancements. The term ($W_{2IS} - W_{0IS}$) in the Solomon equation is the overall rate constant for dipole-dipole transitions giving rise to the NOE, or cross-relaxation; this term is known as the cross-relaxation rate constant, σ_{IS}. The term ($W_{0IS} + 2W_{1I} + W_{2IS}$) is the dipolar longitudinal relaxation rate constant of spin I, or ρ_{IS}. An expression for the steady-state nuclear Overhauser effect (NOE), the change in I spin intensity on saturation of spin S (abbreviated $f_I\{S\}$), may be derived by setting $dI_Z/dt = 0$ and $S_Z = 0$. Since $S_{0Z} = (\gamma_S/\gamma_I)I_{0Z}$,

$$f_I\{S\} = (I_Z - I_{0Z})/I_{0Z} = (\gamma_S/\gamma_I) * ((W_{2IS} - W_{0IS})/(W_{0IS}+2W_{1I}+W_{2IS})) = (\gamma_S/\gamma_I)*(\sigma_{IS}/\rho_{IS}) \qquad \text{Equation 6}$$

When the dependence of the W rate constants on the precession frequencies of the spins (ω) and on the molecular tumbling rate (defined by the rotational correlation time, τ_C), is incorporated into the above expression, the result is,

$$f_I\{S\} = (\gamma_S/\gamma_I) * ((5 + \omega^2\tau_c^2 - 4\,\omega^4\tau_c^4)/(10 + 23\omega^2\tau_c^2 + 4\omega^4\tau_c^4)) \qquad \text{Equation 7}$$

which reduces to $0.5 *(\gamma_S/\gamma_I)$ for small molecules tumbling rapidly ($\omega\tau_C << 1$); and to $-1 * (\gamma_S/\gamma_I)$ for large molecules tumbling slowly ($\omega\tau_C >> 1$). NOE enhancements are expected to be positive for small molecules, and negative for large molecules. In practice, the most useful information is obtained from transient NOE experiments, rather than from steady-state NOE experiments. From transient NOE experiments, NOE buildup rates, governed by σ_{IS}, are measured. The initial rate of NOE buildup is directly proportional to σ_{IS}. In addition, the relative values of σ_{IS} provide a measure of internuclear distances, because the dipole-dipole interaction, which gives rise to cross-relaxation, has a strong dependence on distance (r^{-6}).[17]

Transferred NOE is a transient NOE experiments.[12,23-26] In this experiment, the system is excited by high frequency pulses. During this mixing time the intra- and intermolecular transferred-NOEs build up where the intermolecular *tr*NOEs are usually much larger than intramolecular effects. Tipically, the 2D-NOESY experiment is used to detect transient NOEs and consequently the *tr*-NOESY. *tr*NOE has the potential to offer easy detection of binding events, additionally producing information on the bound conformation of the ligands. This approach provides key information on the ligands' binding mode in the natural environment and allows to deduce information on the structural requirements of the ligand in the bound state. Transferred intramolecular NOEs are well established as unique sources of structural information on bound ligand.

The observation of *tr*NOEs relies on different tumbling times τ_c of free and bound molecules. Low- or medium-molecular-weight molecules (MW<1000–2000) have a short correlation time τ_c and, as a consequence, such molecules exhibit positive NOEs, no NOEs, or very small negative NOEs depending on their molecular weight, shape, and the field strength. Large molecules, however, exhibit strongly negative NOEs.[27] When a small molecule (ligand) is bound to a large-molecular weight protein (the protein receptor molecule) it behaves as a part of the large molecule and adopts the corresponding NOE behavior, that is, it shows strong negative NOEs, so-called *tr*NOEs. These *tr*NOEs reflect the bound conformation of the ligand. Binding of a ligand to a receptor protein can thus easily be distinguished by looking at the sign and size of the observed NOEs. Furthermore, the discrimination between *tr*NOEs originating from the bound state and NOEs of the ligand in solution can also be achieved by the build-up rate, that is, the time required to achieve maximum intensity, which for *tr*NOEs is in the range of 50 to 100 ms, whereas for nonbinding molecules it is four- to ten-times as long (**Figure 1.1.6.**).

Critical parameters for this experiment are:

a) Mixing time. The mixing time must be short enough so that the contribution of the free ligand is negligible and long enough to allow visualization of the signal in the spectrum.

b) The molar ratio of ligand to receptor. It should be emphasized that the *tr*-NOESY experiment works well for ligands that have K_D in the μM - mM range.

c) Spin diffusion. Although the intermolecular spin diffusion is present in every system, it is smaller in magnitudo than intramolecular spin diffusion and should not affect the determination of peptide ligand structure.[28] In addition, spin-diffusion-mediated cross-peaks can be detected by employing *tr*-ROESY experiments, due to the alternation of sign between direct and indirect effects.[29]

tr-NOESY is a robust complement to the STD method. Initially, *tr*-NOESY experiment was used for screening mixtures of carbohydrates that had potential binding affinity for a lectin, *Aleuria aurantia agglutinin* (AAA).[23] Then, the *tr*-NOESY method was also successfully applied to characterize the conformation of natural products bound to different proteins.[30]

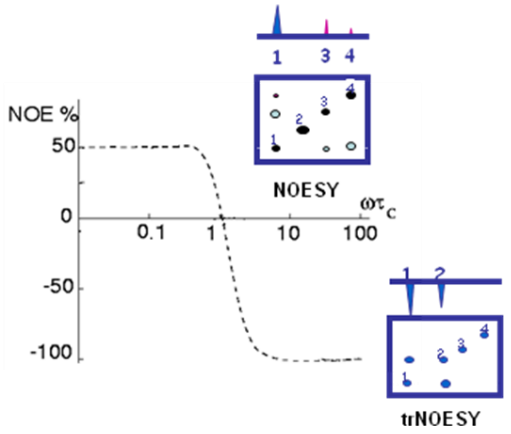


Figure 1.1.6. Schematic view of the *tr*-NOESY experiment. NOESY technique shows slow NOE build-up with weak positive maxima for free small ligands (left, cross-peaks with the opposite sign of the diagonal). *tr*-NOESY shows strong negative maxima for the bound state (medium, cross-peaks with the same sign of the diagonal). The analysis of these negative cross-peaks provides information on the bound geometry of the ligand. The *tr*-NOESY has the potential to offer easy detection of binding events, additionally producing information on the bound conformation of the ligands. On the right, the graphic of NOE % as a function of the correlation time is reported.

In conclusion, for a small molecule ligand binding reversibly to a protein, fast exchange on the "relaxation time scale" implies that k_{off} is greater than s, in both the free and bound states of

the ligand. This condition leads to averaging of the cross-relaxation rate, and therefore of the observed NOE, between the ligand free and bound states:

$\sigma_{OBS} = p_B{*}\sigma_B + p_F{*}\sigma_F$, Equation 8

where p_B and p_F are the fractions of bound and free ligand. This leads to the phenomenon of the transferred nuclear Overhauser effect, which allows the determination of bound-ligand conformations. The transferred nuclear Overhauser effect (*tr*NOE) refers to the observation of negative NOE effects, representative of the bound conformation, for a small-molecule ligand in solution with a protein, in contrast to the positive NOE effects normally expected for a small molecule in the absence of protein. Since σ_B is much greater than σ_F, σ_{OBS} is dominated by σ_B even with an excess of free ligand:

$p_B\sigma_B >> p_F\sigma F$

$\sigma_{OBS} \approx p_B\sigma_B$

The large, negative NOE effects which develop when the ligand is bound to the protein are transferred from the bound to the free state of the ligand by chemical exchange. This allows transferred NOE effects to be measured on smaples containing a mixture of the protein and an excess of free ligand.[17]

The sequence used for our experiments is the so-called “noesyesgpph”; it is a gradient-based phase-sensitive 2D NOESY with excitation sculpting.[31]

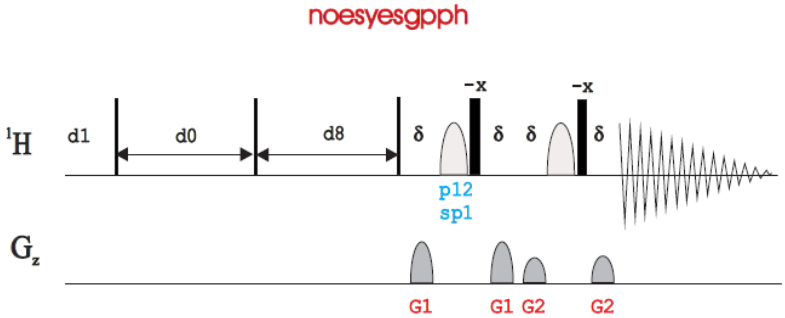


Figure 1.1.7. noesyesgpph sequence used for our experiments; d1 corresponds to the relaxation delay (typically few second, 1- 5 times than T1), d0 is an increment delay, d8 is the mixing time (usually fixed at 200 ms or 700 ms), sp1 is the shape pulse for selective excitation.

1.1.3 STD (Saturation Transfer Difference)[2]

Saturation transfer NMR was a pioneering technique developed in 1979 by Wuthrich's group[32] to obtain resonance assignments of the heme group in oxidized cytochrome c-557 from *Crithidia oncopelli*. After twenty years, in 1999, Mayer and Meyer[33] proposed the same STD technique to elucidate ligand binding to the receptor. STD-NMR is one of the most widespread NMR methods to study the interactions between small ligands and macromolecular receptors.

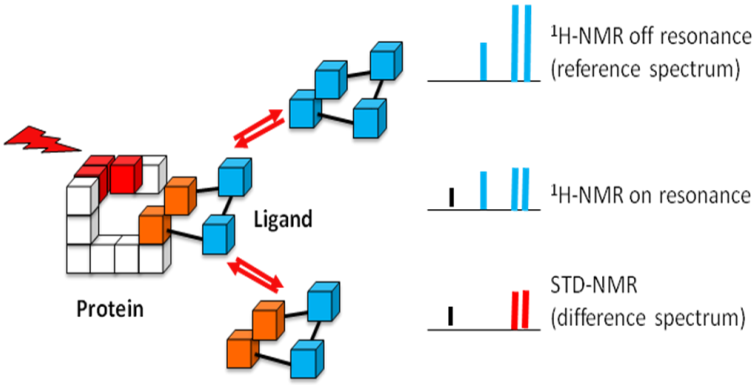


Figure 1.1.8. Illustration of the 1D STD-NMR experiment applied to a sample containing a protein receptor in the presence of small ligands in molar excess. The experiment provides for the acquisition of two 1D NMR spectra: a standard ^{1}H-NMR (off resonance) spectrum (top), and a ^{1}H-NMR (on resonance) spectrum with selective saturation of receptor proton signals (medium). If protein–ligand ^{1}H–^{1}H distances are short enough, the magnetization of protein proton can be transferred to a small molecule by intermolecular NOE. The difference spectrum (STD) will only show the signals of the ligand protons which have been in intimate contact with the receptor (bottom).

Originally proposed as a technique for the rapid screening of compound libraries, its scope has been extended to determine the ligand regions in contact with the receptor (epitope mapping).[34] The method is based on the transfer of saturation from the protein to the bound ligand which in turn, by exchange, is moved into solution where it is detected (**Figure 1.1.8.**).

During the period of saturation, the magnetization gradually moves from the protein to the protons of the ligand when the ligand binds to the target. This saturation process is very efficient,

so the modulation of the ligand signal induced by the protein is readily detected, even in the presence of a large excess of ligand. The ligand protons nearest to the protein are most likely to be saturated to the highest degree, and therefore have the strongest signal in the mono-dimensional STD spectrum. The ligand protons located further away are saturated to a lower degree, and their STD intensities are weaker. Therefore, the degree of saturation of individual ligand protons (expressed as absolute-STD percent) reflects their proximity to the protein surface and can be used as an epitope-mapping method to describe the target-ligand interactions (**Figure 1.1.8.**). Moreover, the sensitivity to discriminate between binders and non binders is excellent, and this experiment has been successfully used to identify ligands with binding activity from multi-component mixtures.[12] The STD experiment is implemented as a difference experiment (**Figure 1.1.8.**). Two mono-dimensional spectra are acquired: one with selective excitation of the protein turned on, and one where the selective excitation is moved to an empty spectral region. If the irradiation conditions are adequately chosen, it is possible to saturate very efficiently only protons of the protein, but not those of the ligand. In this situation, if binding occurs, magnetization from the receptor protons will be transferred to ligand protons that are close in space in the bound state. For transient interactions with rapid exchange, the ligand polarization in the bound state is transferred to the free state where the saturation accumulates during the irradiation time of the experiment. In fact, the enthalpic relaxation of a molecule of low molecular weight, is much slower than the kinetic constant k_{off} of the binding process. The difference spectrum will show only the signals of ligands that were at some time associated with the protein. For non binding molecules, the intensity of the two spectra (on- and off-resonance) is the same and their difference spectrum cancels out.

Figure 1.1.9. shows the STD pulse sequence associated with the sequence Watergate for water suppression. The subtraction of the spectra can be obtained by the phase cycling after every scan. The saturation of the protein is obtained by a selective pulse, usually a Gaussian pulse is used because it reduces the side band problem and excites uniformly the region of signals. In addition, it is preferred to apply a train of pulses of the duration of 50 ms each with intervals of 1 ms, instead of using a long selective pulse. The times of saturation are in the range of 1-3 seconds.

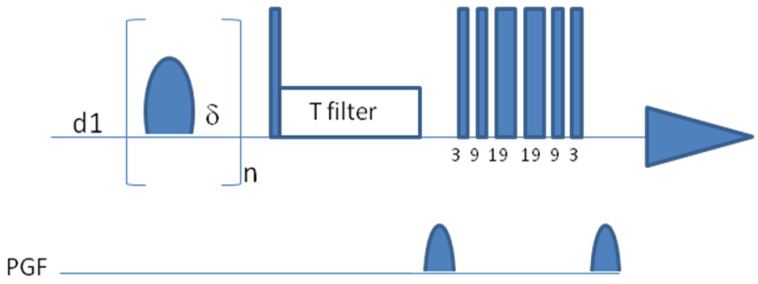


Figure 1.1.9. Pulse sequence for 1D-STD experiment associate with Watergate sequence (water suppression).[2]

In STD NMR experiments, a sample should contain the ligand (in a large molar excess, from 10:1 to 100:1) and the receptor (large molecules with MW > 15 KDa). Generally, the sample containing the protein is irradiated in the area of the aromatic protons or in the area of the spectrum between -2 and 1 ppm (valine, leucine and isoleucine residues), then the radiation is spread on all the protein residues exploiting the spin diffusion phenomenon. Once chosen the frequency of radiation, the protocol provides for the acquisition of an STD experiment on a sample containing only the ligand. In this way, the presence of artifacts due to Overhauser effects between the protons of the ligand can be excluded.

STD experiments have enormous potential, they have been used to select, with success, libraries of both natural and synthetic compounds, also using immobilized receptors.[35] The STD NMR technique may be applied to a large variety of systems, from medium-size receptors to large entities, such as viruses,[36] bacteria and even intact cells.[37-39]

In the first STD experiments, the interactions between a protein receptor and carbohydrates were studied.[23,29,38] These systems are characterized by low binding affinities (the dissociation rate constant K_D was in the micro-molar range) and generate a substantial saturation transfer. Among others, the groups of Jimenez-Barbero and Meyer[23,24,25,33,34] have focused their research on biologically relevant transient protein – carbohydrate interactions.

The STD technique was also applied to study the interaction between protein receptors and peptide ligands.[34,39] Usually, a set of compounds is screened with a target receptor, in other cases the ability of a ligand to interact with similar binding sites within a single protein is studied.

Often, peptide ligands show a dissociation constant (K_D) in the nM range. Then, the study of these systems becomes difficult with NMR-STD technique because it is the limits of the detection. Furthermore, in our cases the proteins of interest are proteins whose function is strongly related to their conformation (for example, the proteins expressed on the cell membrane). So, the STD experiments must be directly recorded on samples containing a suspension of living cells without having to isolate the receptor protein. In this respect, a second generation of STD, termed STDD (saturation transfer double difference) has become a tool in the epitope mapping of drug molecules. In STDD, a STD spectrum of a cell plus ligand system is recorded. Then another spectrum containing only the cell is recorded at the same concentration as the previous spectrum. Next, a difference spectrum of these two systems is computed to perform the epitope mapping of drug, bound to the cell.

The intensity of the STD signals depends, among other things, on the irradiation time/saturation time and on the excess of ligand molecules used. The more ligand that is used and the longer the irradiation time the stronger the STD signal is. The ligand protons nearest to the protein are most likely to be saturated to the highest degree, and therefore have the strongest signal in the STD spectrum; whereas the ligand protons located further away are saturated to a lower degree, and their STD intensities are weaker. Therefore, the degree of saturation of individual ligand protons (expressed as absolute-STD percent)[12,33] reflects their proximity to the protein surface and can be used as an epitope-mapping method to describe the target–ligand interactions.

All STD intensities depend on the duration of radiofrequency irradiation of the protein (saturation time), and the extent of saturation transferred to the ligand protons generally increases with this time.[33] The ligand group epitope mapping by STD intensities is typically obtained for a given saturation time (normally of the order of seconds). Ideally, epitope information would not depend on the chosen saturation time. However different relaxation rates of the ligand protons can produce artifacts in the epitope definition. Because all artifacts are consequences of differences in the ability to accumulate saturation in the free state, a way to

cancel them is to derive STD intensities close to zero saturation time, or to calculate the initial slope of a curve builded up at different saturation time.[13]

Determination of binding constants

The dissociation constant (K_D) (Equation 2) which describe protein–ligand interaction can be obtained from STD NMR spectra. At equilibrium, the fraction of proteins bound to the ligand $f_{p'}$ is given by

$f_{p'} = [PL]/([PL]+[P])$ Equation 9

The combination of equations (2) and (9) yields

$f_{p'} = [L]/([L]+K_D)$ Equation 10

Equation 10 is one of the fundamental equations that elucidate the outcome of the NMR experiment. It shows that increasing the ligand concentration [L] increases the fractional occupation $f_{p'}$ of the receptor-binding site according to a hyperbolic function of [L]. In the limit $[L] << K_D$, the fractional population of the ligand is directly proportional to its concentration. In the opposite limit $[L] >> K_D$, the protein receptor is completely saturated by the ligand ($f_{p'} = 1$). The interesting situation occurs when $[L] \approx K_D$ and the receptor is half-saturated; that is, half of the receptor molecules exist in a one-to-one complex with the ligand. Therefore, ligands with weaker affinity have larger K_D and require the addition of more ligand to saturate the receptor-binding site. Thus, by adjusting the receptor and ligand concentrations, it is possible to "select" the maximum K_D to obtain an optimum observable NMR signal. Such a tuning of the detection threshold is very useful in ligand-based screening experiments and in particular in STD experiments for screening large compound libraries. At equilibrium, the transient presence of the ligand molecule at the protein site is responsible for chemical exchange, a parameter that is used only by NMR methods in ligand-screening experiments.

STD can be used for K_D ranging from 10^{-3} to 10^{-8}M. Operatively, the value of K_D can be obtained from the initial growth rate of the STD amplification factor. Since the STD-NMR intensity reflects the concentration of ligand-receptor complex in solution,[12,33] Mayer and Meyer[34] proposed a conversion rule from the observed experimental intensities (I_0-I_{sat}/I_0) to the STD amplification factors (STD-AF), that is by multiplying the observed STD by the molar excess of ligand over protein (ε).

$$\text{STD-AF} = \varepsilon(I_0-I_{sat})/I_0=\varepsilon\eta_{STD} \quad \text{Equation 11}$$

This converts the STD intensity, which depends on the fraction of bound ligand, into a factor (STD-AF) that is function of the fraction of bound protein. Thus, the evolution of the STD-AF along a ligand titration series enables construction of a saturation curve in the form of an association isotherm. These curves can be adjusted by means of Equation 10 and the value of K_D results from the mathematical fit of the experimental curve. In various cases, however, apparent K_D values obtained from STD NMR titrations have been shown to depend strongly on experimental conditions (e.g., saturation time, relative concentration of species, monitored proton).

Angulo *et al*.[40] showed that the origin of such discrepancies can be found to be the fast ligand rebinding in solution, where a previously saturated ligand molecule returns to the complex (bound state) before its perturbed polarization (*i.e.* saturation) has completely relaxed back to equilibrium. This means that the strongest effect will be observed for low ligand-to-protein ratios, as typically found during the initial steps of a titration series (starting at the lowest ligand concentration). This dependence on the fraction of bound ligand also explains why such artifacts in K_D determination are stronger for systems with larger affinity (the biases typically increase from mM to nM interactions), and for titrations with larger protein concentrations. After careful study of the artifacts introduced by the different experimental conditions of the STD NMR experiment, Angulo *et al*.[40] concluded that:

1. STD data overestimate K_D, and so its value is always grater (or equal) than true thermodynamic value.

2. the protons with largest STD intensities give greater discrepancy.

3. the discrepancy is higher for largest protein concentration (low ligand/protein ratio)

The accurate determination of dissociation constants requires the removal of all effects from fast protein – ligand rebinding in solution. Because these effects are related to the accumulation of saturated ligand in the bulk solution, they are minimized as τ_{sat} approaches 0. As for determination of ligand epitopes, this requires determination of the initial slopes of the build-up curves of STD-AF values with the saturation time. The procedure for exact K_D determination[30] involves the initial slopes of a curve recorded with τ_{sat}-dependent STD-AF for each ligand concentration (instead of just one STD NMR experiment at a single saturation time, as suffices for epitope mapping).

1.1.4 Water suppression sequences

The experiments conducted during my thesis were performed both in a H_2O/D_2O 9:1 and in a phosphate buffer solutions. So, we need pulse sequences that permit to eliminate the effects of the huge water resonance on the quality of the spectrum.

The solvent signal suppression technique can be divided into:

a) Saturation based methods: discrimination by frequency or by relaxation times.

b) Methods avoiding solvent saturation

c) Magnetization Destruction-Based Methods

d) Coherence Selection

e) Postacquisitional Methods

The proton signals of the water solvent fall at around 4.7 ppm, overlapping at all the signals that fall in the range between 4.5 and 4.9 ppm. For the STD experiments we used the suppression sequence called WATERGATE (Water suppression by GrAdient Tailored Excitation),[41] while for the mono- and bidimensional experiments (TOCSY, NOESY) we used the Excitation Sculpting technique (DPFGSE, Double Pulse Field Gradient Spin Echo);[42] this last is an improvement of the WATERGATE sequence where a second spin echo sequence with gradients is added to cancel phase errors.

One of the simplest and widely used ways to eliminate a strong water signal is to use presaturation.[43]

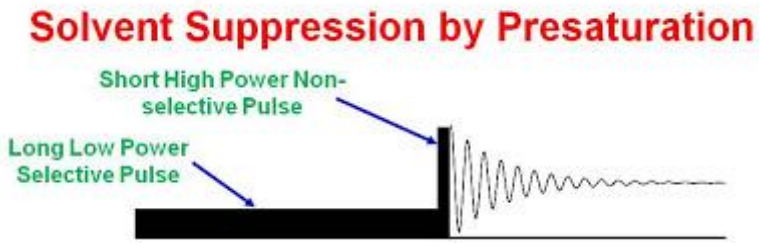


Figure 1.1.10. Solvent suppression by presaturation.

In this technique, the transmitter is set to the water resonance; a very long (seconds) low power (mW) pulse is given. The excitation profile of this pulse is very narrow due to its length and it saturates the water resonance at the transmitter frequency. A non-selective hard 90° pulse is then given to place all remaining spin in the transverse plane for detection.

WATERGATE[43] is widely used in many complicated pulse sequences. Unlike presaturation which irradiates the water resonance with a long low power pulse, this method is based on the gradient spin echo technique.

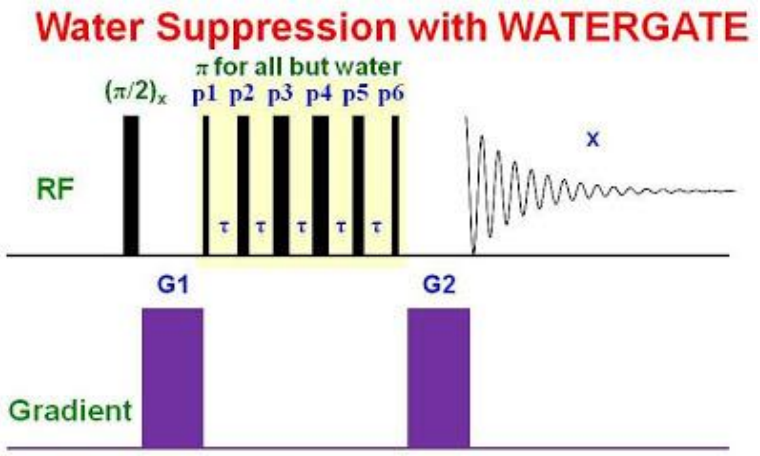


Figure 1.1.11. Water suppression with WATERGATE.

The transmitter frequency is set on the water resonance. A non-selective hard 90° pulse is applied followed by a 1-2 msec gradient pulse. The gradient pulse dephases all of the resonances. A composite pulse (consisting of 6 hard pulses separated by a delay, τ) is then applied which acts as a 180° pulse for everything except peaks on resonance (*i.e.* water) and any peaks at frequencies n/τ away from the transmitter, where n is an integer. t is chosen such that $1/\tau$ lies outside of the

spectral width (typically several hundred μsec). The second gradient pulse (equal in magnitude, duration and sign, to the first) further dephases the water resonance at the center of the spectrum which was unaffected by the composite pulse, but rephases everything else which was inverted by the composite pulse. The gradients and composite pulse act as a gradient spin echo for all. The FID is then collected with the water resonance suppressed by the two dephasing gradients.

Section 2: Computational methods

1.2.1 Molecular Mechanics (MM)

In a quantum mechanical description of a molecule, electrons need to be included. Thus a very large number of particles must be considered in order to fully describe the system. Calculations performed at a quantum mechanical level are very time consuming and a different approach is needed when dealing with molecules bigger than a few tens of atoms. Molecular mechanics is then invariably used in this case. Molecular mechanics is based on the validity of two main assumptions:

- the Born-Oppenheimer approximation and the Potential Energy Surface
- Force fields

While the Born-Oppenheimer approximation enables the possibility of writing the potential energy of the system as a function of the nuclear coordinates, discarding the electrons, the force fields give a functional form to the description of the potential energy.

The Born-Oppenheimer approximation is the basis of all methods of molecular modeling. It states that the electronic and nuclear motions can be separated, because the mass of the electrons is much smaller and hence they can rapidly adjust to any change in the nuclear position. This allows to calculate the energy of a molecule as a function of the nuclear coordinates only.

Energy changes that are associated to a variation of the nuclear position can be considered as movements on a multidimensional surface, which is called the Potential Energy Surface (PES).[44] Each point on the PES represents a molecular conformation characterized by its structure and energy. Stationary points on the PES are points where the first derivative of the energy is zero with

respect to the coordinates. Minimum points are one type of stationary point and correspond to stable conformations of the molecule. Defining the shape of a molecule implies locating all the accessible minima and assigning them a relative energy, and hence a probability factor.

A molecular mechanics (MM) program uses a Force Field (FF) to define a PES for each molecule and to calculate an energy value for every combination of the system coordinates.

In the approach of molecular mechanics (MM) a molecule is seen as a collection of atoms, held together by forces. The forces acting between atoms have the classical form of a potential energy function and contain parameters which are determined empirically. The collective action of the interatomic forces generates the force field which describes the molecule's potential energy.

The overall potential energy of a molecule, its steric energy, is obtained as a linear combination of independent terms originating from interatomic interactions, and its qualitative formulation is as follows:

$$\text{Steric Energy} = E_{stretch} + E_{bend} + E_{torsion} + E_{VdW} + E_{electrostatic} + \ldots$$

Equation 1

The various terms reflect the origin of the model in vibrational spectroscopy and are defined as:

- Estretch = Stretch energy over all bonds
- Ebend = Bending energy over all angles
- Etorsion = Torsional (dihedral) energy (over all dihedral angles)
- EVdW = Van der Waals energy (over all atom pairs ≥ 1,4)
- Eelectrostatic = Electrostatic energy (over all charged atom pairs ≥ 1,4)

Mixed terms, such as stretch-bend or torsion-stretch energies may also be included. The VdW term and the electrostatic term act between atom pairs which are separated by at least three bonds, and hence are often called the non-bonded interaction terms. Thus, a MM force field is a collection of parametric potential energy functions that describe interatomic interactions. The different force fields are defined by the functional form chosen for the energy functions and by the values of their parameters. It must be stressed that there is no such thing as "the best form" of a FF: force fields are empirical, and are evaluated exclusively on the basis of their performance.

Furthermore, different force fields have often been developed with different molecular properties or different types of molecules in mind, and their performance will vary accordingly.

A typical form of a force field is shown in Equation 2, where the various terms represent the different forces acting between pairs of nuclei:[44]

$$V(r^N) = \sum_{bonds} \frac{k_i}{2}(l_i - l_{i,0})^2 + \sum_{angles} \frac{k_i}{2}(\theta_i - \theta_{i,0})^2 + \sum_{torsions} \frac{V_n}{2}(1 + \cos(n\omega - \gamma)) + \sum_{i=1}^{N}\sum_{j=i+1}^{N}\left(4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}\right) + \text{cross terms}$$

Equation 2

The first two terms, the stretching and bending forces, are the simplest to visualize. They correspond to the familiar IR stretching and bending vibrations and act to return the related structural parameters, bond length and bond angles, to their equilibrium values $l_0$and Θ_0. Thus, the equilibrium value of the structural parameter, for instance the equilibrium bond length l_0, is seen as the "natural" value for that specific bond and each deviation from this ideal state, caused by perturbations from the rest of the molecule, will increase the overall energy of the molecule. The size of the increase will be determined by the value of the associated force constant k and will be returned by the MM calculations as the stretch energy for the conformation been calculated. Both the force constant k and the equilibrium values of bond length and angles, l_0 and Θ_0, are parameters of the FF. Depending on the FF, the values of l_0 and Θ_0 can be derived from experimental data (X-ray, electron diffraction, MW spectroscopy etc.) or from high-level molecular orbital calculations and the values of the force constants can be optimized by trial and error and by least square methods.

The fourth term is the non-bonded term and is calculated between all pair of non-bonded atoms, *i.e.* all the atoms that are separated by more than two bonds. This term is usually modeled using a Lennard-Jones potential for van der Waals interactions and a Coulomb potential term for the electrostatic interaction. As expected for a model which ignores the existence of electrons in molecules, the electrostatic term is particularly difficult to handle. In general, a Coulomb potential

is used and is calculated by assigning to each atom or atom type a partial atomic charge, which is often obtained by fitting to the electrostatic molecular potential obtained from QM calculations.

The third term is a torsional potential, that models how the energy changes as a bond rotates. Physically, the rotation barriers about chemical bonds are generated by attractive or repulsive interactions which, in principle, could be accounted for by the non-bonded interaction terms. However, most FF do include explicit torsional potential terms, which are used to model correctly the known energy profiles for rotation about a given type of bond.

The different FF may vary in the way the parameters are obtained and optimized, as well as in the form of some of the potential terms. In all cases, it is assumed that the value of the parameters is transferable, that is that a given atom or combination of atoms will behave similarly in different molecules.

The best known FF are MM2, MM3 and MM4 (Allinger et al.), parameterized for many functional groups and for modeling hydrocarbons, AMBER (Kollman),[45] OPLS_AA (or OPLS_2001, Jorgensen)[46] and CHARMM, parameterized for proteins and nucleic acids and used to model molecular dynamics of biomolecules. They mainly differ in:

- Force Field of the first class, mainly developed for the treatment of extended systems (proteins or DNA). Armonic functions are used for the stretching or bending terms and the cross-interaction terms were implicitly kept into consideration.
- Force Field of the second class, mainly developed for the treatment of small/medium molecules with high accuracy. These FF explicit the cross-interaction terms, use exponential terms for the Van der Waals energy and expansion of third or fourth degree for the bending and stretching terms.

In particular, OPLS_2001 and OPLS_2005, both belonging to the first class Force Field, were used in the thesis.

1.2.2 Solvation Models: the implicit water GB/SA solvation model[44]

Most chemical processes occur in solution, and, although many molecular modeling studies are still carried out without including solvation models, the solvent medium can have a major effect on molecular structures and energies. This is particularly important for the study of biological interactions which take place between polar molecules in the presence of water.

The simplest and crudest way of including solvent-related effects in molecular mechanics calculation is to increase the dielectric constant in the coulombic electrostatic term of the potential energy. In vacuum, the value of ε_0 is 1. Polar solvents, like water, dampen the electrostatic interactions and their effect can be modeled by assigning to ε_0 the value of the solvent's dielectric constant (*e.g.* for water ε_0= 80). More realistic solutions are provided by so-called **explicit** or **implicit** solvation models.

In the most detailed microscopic approach, solvent molecules are treated explicitly, and the electrostatic properties of both solvent and solute are obtained by averaging over a very large number of configurations of the system. In the most widely used explicit model, TIP3P, the water molecule is considered as a rigid molecule having three interaction sites, corresponding to its three atoms. The partial positive charges on the hydrogen atoms are exactly balanced by an appropriate negative charge located on the oxygen atom. The van der Waals interaction between two water molecules is computed using a Lennard-Jones function with just a single interaction point per molecule centered on the oxygen atom; no van der Waals interactions involving the hydrogen atoms are calculated.

Explicit solvation calculations of this kind run slowly because hundreds of explicit solvent molecules need to be included. Furthermore, convergence can be a problem, since different solvent starting configurations often give different final energies, and hence simple energy minimization is not useful in an explicit solvent.

This prompted interest in models which incorporate the influence of the solvent in an implicit fashion.[47] There are two main types of implicit solvation models: empirical models and models based on continuum electrostatics.

Solvent models based on continuum electrostatics define the solute interior and the solvent as regions with different dielectric constants, and the electrostatic solvation free energy is computed by solving the Poisson-Boltzmann equations.[48] These methods represent a rigorous treatment of continuum electrostatics, which takes into account solvation of single charges as well as screening effects (charge-charge and charge-dipole interactions). However, the calculations are too time-consuming to be performed routinely. Often one also encounters convergence problems, which may depend on the resolution of the solute-solvent boundary, on the partial charge representation, and on the difficulty in mapping forces related to the dielectric boundary onto individual atoms.

The implicit model GB/SA (Generalized Born/Solvent-Accessible Surface Area) was used in the calculations of the thesis. As for the other empirical models, GB/SA dissects the solvation free energy, that is the free energy change to transfer a molecule from vacuum to solvent, in three components, an electrostatic component, a van der Waals component and one component associated with creating the solute cavity within the solvent (Eq. 3). The last two terms depend on the solvent-accessible surface area of the solute, the first term is usually derived either by the Onsager or the Born model.

$$\mathbf{G_{sol} = G_{cav} + G_{VdW} + G_{pol}}$$ Equation 3

The contribution of each group to the ΔG_{VDW} and ΔG_{cav} terms is often approximated by a linear function of its solvent-accessible surface area. The GB/SA system does not show convergence problems and allows simple minimizations and conformational searches to be performed in all the solvents that have been parameterized (water and chloroform).

1.2.3 Conformational analysis

Conformational search methods are used to sample the PES, *i.e.* to locate all the accessible minima and their associated relative energies. Distance geometry methods can be used to derive molecular conformations from experimental data; Metropolis Monte Carlo and molecular

dynamics simulation methods can be used to explore the entire conformational space of molecules. Methods which use a search algorithm coupled with an energy minimization are used to identify the preferred conformations of a molecule, *i.e.* the conformations localized at minimum points on the energy surface. Such conformational searches can be performed using either **systematic search** or **random search** algorithms.

The simplest type of **systematic search** is called a grid search, and it contains various steps. First, all rotatable bonds are identified and systematically rotated through 360° using a fixed increment. Every conformation so generated is subjected to energy minimization to locate the closest (downhill) minimum conformation. The search stops when all the possible combination of torsion angles have been generated and minimized. After elimination of duplicates (duplicates are defined as molecular structures that do not differ by more than a defined threshold), the resulting minima are yielded as the search output. It can readily be seen that the number of conformations generated by a grid search is much larger than the number of minima on the surface and that many starting conformations will end up in the same minimum energy structure. Many of the initial conformations will be very high in energy, and some of them, for instance those containing major steric clashes of atoms, may be discarded even before the minimization. Nonetheless, the number of structures to be generated and minimized increases combinatorially with the number of rotatable bonds.

A **random search** method moves from one region of the PES to a completely unconnected region in a single step by performing a random change of coordinates. Although the change can be made both in Cartesian coordinates or in internal coordinates, it has been shown that the latter method is much more efficient at exploring the conformational space of molecules, because it greatly reduces the number of degrees of freedom to be.[49] After the random change has been made, the structure is refined using energy minimization. If the minimized conformation has not been found before, and falls within a predefined energy window above the instant global minimum of the search, it is stored. A conformation is then selected to be used as the starting point for the next iteration, and the cycle starts again. The procedure continues until a given number of iterations have been performed or until the user decides that no new conformations

can be found. A flow chart of the steps is given in **Figure 1.2.1.** There are many ways in which the structure for input to the next iteration can be selected, and this choice will influence the efficiency of the method. It has been shown that low-energy final conformers are favored by selecting low-energy starting geometries at each Monte Carlo step.[50] The rational for such an empirical finding is that low-energy full structures are composed of low-energy substructures and a limited number of random variations of internal coordinates leaves many of the energetically favorable substructures intact.

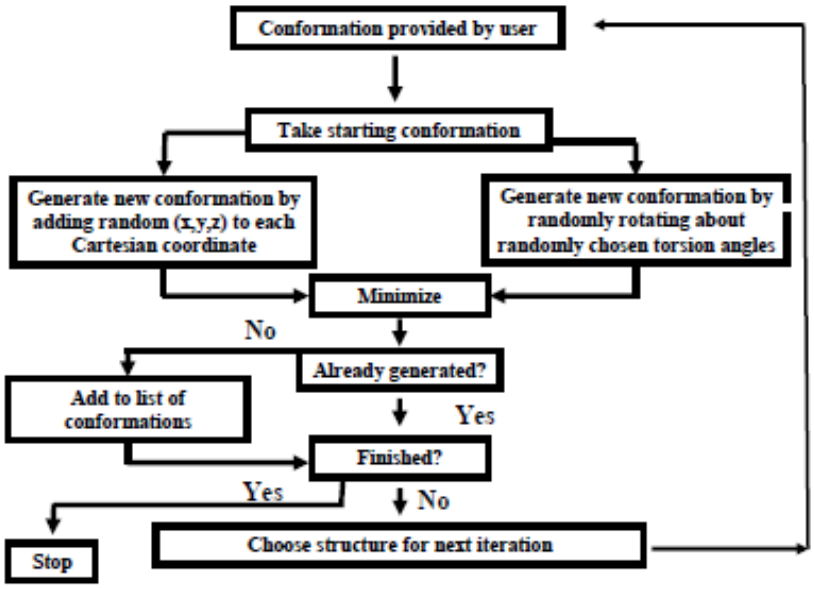


Figure 1.2.1. Flow chart of the processes conducted by an MC/MM simulation.

There is not a "natural" endpoint of a random search procedure and the choice has to be made by the user. Generally the search is stopped after a predefined number of steps (1000 steps for rotatable bond is cited in the literature as a reasonable initial guess) and the results are analyzed for convergence. The usual strategy is to generate conformations until no new structures, or no low-energy new structures, are obtained. This generally requires each structure to be generated many times, so the number of times each minimum has been found is also used as a convergence criterion. However, one can never be absolutely sure that all of the minimum energy conformations have been found. If a run shows poor convergence, a common strategy is to

restart a new search from the instant minimum. The output of the two independent searches can then be merged to improve the coverage of the PES.

An alternative method of conformational analysis is constituted by the simulation of the motion of the molecule.

Although most of these simulations are conducted using the technique of molecular dynamics, it is possible use stochastic simulation methods (such as Metropolis Monte Carlo algorithm) or mixed methods (such as MC/SD, Metropolis Monte Carlo/stochastic Dynamics).

1.2.4 Molecular dynamics simulations

Molecular Dynamics (MD)[51] simulations are based on the calculation of the time dependent behaviour of molecular systems, giving a detailed description of the variation from one conformation to another of the system studied. Simulations generate ensembles of representative configurations in such a way that accurate values of thermodynamic and structural properties can be obtained with a reasonable amount of computation. In particular, statistical analysis links microscopic and macroscopic properties providing the fundamental principles for the description of biomolecular systems.

In the molecular dynamics it is not possible/necessary to specify the internal coordinates because the entire molecule is constantly modified and monitored. In the molecular dynamics approach a trajectory (conformations in function of the time) of the molecular system is generated through the simultaneous integration of the Newton equations of the motion; it goes to simulate the thermal motion of the molecule along the PES.

$$\frac{d^2 r_i(t)}{dt^2} = m_i^{-1} F_i \qquad F_i = -\frac{\partial V(r_i)}{\partial r_i} \qquad \text{Equation 4}$$

with

r_i = position of the i atom

t = time

F_i = force acting on the i atom.

The integration is done in a reduced time step, generally between 1 and 10 femtoseconds. In the stochastic dynamics (SD), instead, the Langevin equation[52,53,54] of motion for the Brownian motion is integrated in time, thus simulating the coupling of the atomic motions to a thermal bath or to the effect of a solvent (Equation 5).

$$\frac{d^2 r_i(t)}{dt^2} = m_i^{-1} F_i + m_i^{-1} R_i - \frac{\gamma_i dr_i(t)}{dt} \qquad \text{Equation 5}$$

with:

R_I = stochastic force;

γ_I = frictional coefficients.

γi represents how much the system is coupled to the heat bath; it assumes an universal value, in function of the Force Field, to represent the effect of the solvent. The stochastic term introduces energy, while the frictional term removes kinetic energy from the system.

At the end of a dynamic simulation is obtained:

- A number of intermediate conformations (if required) from which you can obtain information about the dynamics of the system
- Molecular data specifically required averaged over time
- Thermodynamic data (potential and kinetic energy) averaged over time
- Geometric information about the PES explored
- kinetic information

The fundamental problem of molecular dynamics simulations is their convergence. The simulation should be conducted for a period long enough to cover the PES thermally accessible so as to allow its efficient sampling and reliable results. The convergence of a simulation is evaluated through the analysis of convergence of the parameters obtained during the simulation itself, checking if the parameters converge to the desired accuracy at a fixed value, or carrying out the same simulation starting from different conformers and verifying that the results are similar.

The problem of convergence is related to the difficulty of molecular dynamics to overcome the energy barriers. This means that if the PES contains holes separated by relatively high energy

barriers (higher than RT), these may never be overcome during the simulation. In this case, the convergence is apparently good because the potential hole is well sampled, but the data cannot be considered reliable because only part of the PES was sampled.

A method to overcome the latter problem is the coupling of the molecular dynamics, or better of the stochastic dynamics, with the Metropolis Monte Carlo method, which, through a random variation of molecular coordinates, allows a more complete sampling of the conformational space.

The combined method is called Mixed-Mode Metropolis Monte Carlo/Stochastic Dynamics (MC/SD) and it was used for all calculations in the thesis. This approach combines the method Metropolis Monte Carlo (MC), which has the ability to interconvert conformations separated by large energy barriers, with the stochastic dynamics (SD) that samples effectively local conformational space. The molecular dynamics is not used since it generates a microcanonical ensemble (at constant energy), while Metropolis Monte Carlo and stochastic dynamics generate canonical ensembles of Gibbs (at constant temperature) and, therefore, can be used in combination.

In particular, the Monte Carlo Metropolis algorithm compares the energy of each new conformation, obtained by a random change of a coordinate of the previous conformation, with the energy of its predecessor. The method consists of the following steps:

1. Random selection of a coordinate of the molecule and its random variation

2. Calculation of the resulting change in the potential energy ΔV

3. If $\Delta V < 0$ then the new configuration is accepted; if $\Delta V \geq 0$ then the new configuration is accepted if the Boltzmann factor $e^{-[E(new) - E(old)/kT]}$ is greater than a random number chosen between 0 and 1, otherwise the previous configuration is taken. The term $e^{-[E(new) - E(old)/kT]}$ is the probability to accept or reject the new conformation generated.

4. It iterates through this cycle until it reaches a number N_{trial} of iterations.

In this way a large number of energetically accessible system configurations, called states, are explored; this set of states (ensemble) allows the calculation of the thermodynamic properties of the system, including the entropic term. The method is less efficient than the molecular

dynamics for the calculation of thermodynamic properties and obviously does not produce direct information on the dynamics of the system.

The combined method MC/SD differs from other procedures designed for the same purpose, in two points:

- ✓ It assigns equal importance to the dynamics and to the Metropolis Monte Carlo.
- ✓ During the simulation, there is a continuous alternation between the two methods.

It is a protocol that allows calculation through a dynamic simulation to evolve the conformation generated by a Monte Carlo step. The latter introduces large changes in a few degrees of freedom, making the sampling of the PES faster.

The algorithm used by the combined method is the following:

1. It assigns a starting structure (already minimized), the duration of the simulation, the torsion angles to vary, the maximum number of these to be changed at every step.

2 It carries out a step of Metropolis Monte Carlo changing the torsion angles (randomly from one to the maximum number specified in point 1).

3 The structure obtained is evaluated according to the criteria of Metropolis Monte Carlo; if it satisfies the cycle continues, otherwise it is discarded, taking the previous structure, and skipping to step 5.

4 A short equilibration through SD is made, to minimize the error of numerical integration which results from the discontinuity inserted in the trajectory by Metropolis Monte Carlo.

5. It performs the dynamics for the equivalent of a time step, during which the desired data are collected and mediated.

6 The intermediate structure and the data are passed to step 2, until it reaches the end of the simulation.

A problem that can plague even this type of method is the dependence on the starting structure and, in general, the convergence, which, as for the dynamics simulations, must be carefully evaluated. In the thesis the convergence has been verified by running different MC/SD simulations for each compound with different starting conformations of the molecule.

1.2.5 Docking calculations (Glide)

Docking is a computational procedure extremely important for virtual screening and identification of new drugs. Technically, the docking is the placement of a molecule in a region of interest, such as the binding site of a macromolecule, and the assessment of this pose with a "score", which theoretically should reflect the affinity between the two interacting partners, in terms of complementarity of shape and properties. The protein - ligand interaction is dictated by complementary features between the two molecules, mainly based on Van der Waals forces, hydrogen bonds and electrostatic forces. A good complementarity indicates that the molecule is potentially a good binder.

The docking programs on the market are numerous and differ in the type of algorithm used for the generation of new poses, for the function used for the calculation of the "score", or for the evaluation of the flexibility of the protein and the ligand.

Routinely a default region of the macromolecule is selected; in this region, the docking program places the ligand in all possible positions, until reaching a maximum value of favourable interactions, which corresponds to the best "fit".

The program Glide (Grid-based Ligand Docking with Energies) used in this research builds a grid (equally spaced points) of the entire binding region (bounding box) and defines a larger volume containing all the atoms, called enclosing box.

Glide uses a series of filters to search for possible locations of the ligand in the active site, and then to generate the best ligand binding poses through a coarse screening. The filter examines steric complementarity of the ligand to the protein and evaluates various ligand – protein interactions with the empirical Glide Score function. Next, the ligand binding poses selected by the initial screening are minimized in situ with the OPLS force field. Finally the score is used to rank the resulting ligand binding poses.[56]

The empirical score function calculated is:

$$Gscore = 0.065*vdW + 0.130*Coul + Lipo + Hbond + Metal + BuryP + RotB + Site \quad \text{Equation 6}$$

with

vdW = Van der Waals interaction energy

Coul = Coulomb interaction energy

Lipo = favourable hydrophobic interaction energy

Hbond = term relative to the hydrogen bond

Metal = term relative to metal interactions

BuryP = penalty terms for bulky polar groups

RotB = penalty terms for freezing rotatable bonds

Site = polar interactions in the binding site.

The "Glide score" is given, then, by the sum of different contributions related to non-covalent interactions and the weight assigned to each of these terms is derived from the "fitting" to known binding energies obtained experimentally in a variety of different receptor - ligand complexes.

∗Bibliography∗

[1] Stark, J.; Powers, R. *opics in Current Chemistry* **2012**, 326, 1 – 34.

[2] Potenza, D.; Vasile, F. Seminars in Organic Synthesis, "A. Corbella" Summer School, 37th, Gargnano, Italy, June 18 – 22 2012, 225 – 248.

[3] Pellecchia, M.; Sem, D. S.; Wuthrich, K. *Nat. Rev. Drug Discov.* **2002**, 1, 211 – 219.

[4] Kessler, H. et al. *Mol. BioSyst.* **2006**, 2, 318 – 333.

[5] "*Basic one- and two-dimensional NMR spectroscopy*" Friebolin-Horst, Wiley & Sons. Ltd..

[6] "*Modern NMR Techniques for Chemistry research*" Derome, A. E. **1987**.

[7] http://www.schrodinger.com

[8] Macro Model, version 9.5, Schrödinger, LLC, New York, NY.

[9] Glide, version 4.5 and 5.7, Schrödinger, LLC, New York, NY.

[10]Salvatella, X.; Giralt, E. *Chem. Soc. Rev.* **2003**, 32, 365 – 372.

[11]Clarkson, J.; Campbell, I. D. *Biochemical Society Transactions* **2003**, 31, 1006 – 1009.

[12]Meyer, B.; Peters, T *Angew. Chem. Int. Ed.* **2003**, 42, 864 – 890.

[13]Angulo, J.; Nieto, P. M. *Eur. Biophys. J.* **2011**, 40, 1357 – 1369.

[14]Calle, L. P.; Canada, F. J.; Jimenez-Barbero, J. *Nat. Prod. Rep.* **2011**, 28, 1118 – 1125.

[15]Bhunia, A.; Bhattacharjya, S.; Chetterjee, S. *Drug Discov. Today* **2012**, doi 1016/j.

[16]Carlomagno, T. *Annu. Rev. Biophys. Biomol. Struct.* **2005**, 34, 245 – 266.

[17]Johnson, M. A.; Pinto, B. M. *Carbohydrate Research* **2004**, 339, 907 – 928.

[18]Skimer, A. L.; Laurence, J. S. *Journal of Pharmaceutical Sciences* **2008**, 97, 4670 – 4695.

[19]Cala, O.; Guillière, F.; Krimm, I. *Anal. Bioanal. Chem.* **2014**, 406, 943 – 956.

[20]Dalvit, C.; Fogliatto, G.; Steward, A.; Veronesi, M.; Stockman, B. *J. Biomol. NMR* **2001**, 21, 349 – 359.

[21]Sanchez-Pedregal, V. M.; Reese, M.; Meiler, J.; Blommers, M. J. J.; Griesinger, C.; Carlomagno, T. *Angew. Chem. Int. Ed.* **2005**, 44, 4172 – 4175.

[22] a) Solomon, I. *Phys. Rev.* **1955**, 99, 559 - 565; b) Anet, F. A.; Bourn, A. J. R. *J. Am. Chem. Soc.* **1965**, 87, 5250 - 5251; c) Balaram, P.; Bothner-By, A. A.; Dadok, J. *J. Am. Chem. Soc.* **1972**, 94, 4015 - 4017; d) Neuhaus, D.; Williamson, M. P. *"The Nuclear Overhauser Effect in Structural and Conformational analysis"* Wiley-VCH **2000**.

[23] Meyer, B.; Weimer, T.; Peters, T. *Eur. J. Biochem.* **1997**, 246, 705 – 709.

[24]Haselhorst, T.; Espinosa, J. F.; Jimenez-Barbero, J.; Sokolowski, T.; Kosma, P.; Brade, H.; Brade, L.; Peters, T. *Biochemistry* **1999**, 38, 6449 – 6459.

[25]Mayer, M.; Meyer, B. *J. Med. Chem.* **2000**, 43, 2093 – 2099.

[26]Post, C. B. *Curr. Opin. Struct. Biol.* **2003**, 13, 581 – 588.

[27]Neuhaus, D.; Williamson, M. P. "The nuclear Overhauser effect in structural and conformational analysis" **2000**, Wiley-VCH.

[28]Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, 235 – 242.

[29]Asensio, J. L.; Canada, F. J.; Jimenez-Barbero, J. *Eur. J. Biochem.* **1995**, 233, 618 – 630.

[30]Bhunia, A.; Domadia, P. N.; Torres, J.; Hallock, K. J.; Ramamoorthy, A.; Bhattacharjya, S. *J. Biol. Chem.* **2010**, 285, 3883 – 3895.

[31]"Pulse Program Catalogue – NMR Guide 4.0" – Bruker.

[32]Keller, R. M.; Wuthrich, K. *Biochim. Biophys. Acta* **1978**, 533, 195 – 208.

[33]Mayer, M. and Meyer, B. *Angew. Chem. Int. Ed.* **1999**, 38, 1784 – 1788.

[34]Mayer, M. and Meyer, B. *J. Am. Chem. Soc.* **2001**, 123, 6108 – 6117.

[35]Benie, A. J.; Moser, R.; Bauml, E.; Blaas, D.; Peters, T. *JACS* **2003**, 125, 14 – 15.

[36]Klein, J.; Meinecke, R.; Mayer, M. and Meyer, B. *JACS* 1999, 121, 5336 – 5337.

[37]Claasen, B.; Axmann, B.; Meinecke, R.; Meyer, B. *JACS* **2005**, 127, 916 – 919.

[38] a) Mari, S.; Serrano-Gomez, D.; Canada, F. J.; Corbi, A. L.; Jimenez-Barbero, J. *Angew. Chem. Int. Ed.* **2005**, 44, 296 – 298; b) Airoldi, C.; Giovannardi, S.; La Ferla, B.; Jimenz-Barbero, J.; Nicotra, F. *Chemistry - A European Journal* **2011**, 17, 13395 – 13399.

[39]Potenza, D.; Vasile, F.; Belvisi, L.; Civera, M.; Araldi, E. M. V. *ChemBioChem* **2011**, 12, 695 – 699.

[40]Angulo, J.; Enriquez-Navas, P. M.; Nieto, P. M. *Chem. Eur. J.* **2010**, 16, 7803 – 7812.

[41]a) Piotto, M.; Saudek, V.; Sklenar, V. *J. Biomol. NMR* **1992**, 2, 661; b) Hwang, T.-L.; Shaka, J. *Journal of Magnetic Resonance* **1995**, 112, 275 – 279.

[42]Sahu, S. C.; Majumdar, A. *Current Science* **1998**, 74 n°5.

[43]u-of-o-nmr-facility.blogspot.it/2010/03/Watergate-vs-presaturation.html

[44]"Lecture Notes of Computational Organic chemistry" Bernardi, A.; Belvisi, L.; Olivucci, M.; Sinicropi, A.; Colombo, G.; Barone, V.; Crescenzi, O. Cap. 3 "3D Models of oligopeptides by Molecular Mechanics and Conformational Analysis" **2006**.

[45]Quentin, D. M.; Still, W. C. *Tetrahedron Letters* **1992**, 33, 7743 - 7746

[46]Jorgensen, W. L.; Maxwell, D. S.; Tirado, J. *Rives-Journal of the American Chemical Society* **1996**.

[47]Roux, B.; Simonson, T. *Biophys. Chem.* **1999**, 78, 1 – 20.

[48]Honig, B.; Nicholls, A. *Science* **1995**, 268, 1144 – 1149.

[49]Saunders, M.; Houk, K. N.; Wu, Y. D.; Still, C. et al. *J. Am. Chem. Soc.* **1990**, 112, 1419 – 1427.

[50]Chang, G.; Guida, W. C.; Still, W. C. *J. Am. Chem. Soc.* **1989**, 111, 4379 – 4386.

[51]"Molecular Modelling-Principles and Applications" A. R. Leach, Longman.

[52]Van Kampen, N. G. *North Holland Personal Library* **1992** "Stochastic Processes in Physics and Chemistry".

[53]Kottalam, J.; Case, D. A. *Biopolymers* **1990**, 29, 1409 – 1421.

[54]Lamm, G. and Szabo, A. *J. Chem. Phys.* **1986**, 85, 7334 – 7348.

[55]Eldrige, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. *Journal of Computer – Aided Molecular Design* **1997**, 11, 425 – 445.

[56] Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. *J. Comput. Aided. Mol. Des.* **1997**, 11, 425–445.

CHAPTER 2

ANALYSIS OF LIGAND - INTEGRIN INTERACTIONS

Section 1: Introduction

In recent years, the enormous achievements made in modern molecular and cellular biology have enabled the complex network of biological targets and their different functional properties to be unveiled, along with the possible pathogenic deviations. Of the possible biological targets, surface cell receptors (*e.g.* integrins) constitute points of intervention in the pharmacology of many pathologies.[1,2] In addition, fast progresses have been achieved in the study of protein-protein and ligand-protein interaction, that became possible also on an atomic level due to the development of new potent NMR spectroscopic techniques (**Chapter 1**).[3]

This part of my PhD thesis, that was the main focus of the three-year project, was developed within the framework of a PRIN project (MIUR PRIN project 2010NRREPL "Synthesis and Biomedical Applications of Tumor-Targeting Peptidomimetics") involving different research units and aimed at the design, synthesis, structural analysis and biological evaluation of new peptidomimetic integrin ligands.

My PhD project benefits from collaborations with the research groups of Prof. Gennari and Dr. Belvisi from University of Milan (Department of Chemistry) and of Prof. Piarulli (University of Insubria) for the design and development of new peptidomimetics targeting integrins, particularly with the group of Dr. Belvisi for the computational part, and with Prof. Gennari's and Piarulli's groups for the synthesis. The project also benefits from the collaboration with Dr. Arosio (Consiglio Nazionale delle Ricerche) for the *in vitro* binding assays and with the research group of Dr. Tringali for the cell cultures and the *in vitro* adhesion assays.

My contribution to the project concerned the application of NMR and computational techniques aimed at obtaining:

- a detailed conformational study of diketopiperazine peptidomimetic ligands containing the Arg-Gly-Asp (RGD) and isoDGR sequence by NMR techniques (VT-NMR and NOESY) in acqueous solution and by computational methods, such as Monte Carlo-

Stochastic Dynamics (MC/SD) simulations. The results of this task are discussed in this Section 1.

- a detailed investigation of ligand-receptor interactions by new potent NMR spectroscopic techniques[3] applied to live cells overexpressing integrins on their membrane, *e. g.* platelets and tumour cells.[5,6] In particular, I set up live cell NMR methodologies by using human platelets and investigating the interaction of some RGD peptidomimetic ligands with the platelet-specific integrin $\alpha_{IIb}\beta_3$. The results of these experiments are discussed in the first part of Section 2. Then, I expanded the live cell NMR study by using human tumour cells and gaining insight into the interaction of the small library of ligands with the tumour-associated integrins $\alpha_V\beta_3$ and $\alpha_5\beta_1$ (**"Interaction with ECV-304"** and **"Interaction with MDA-MB-231"**). The NMR results were supported by docking calculations of the ligands in the active site of the different integrins and were compared to the results of competitive receptor binding assays and competitive cell adhesion experiments.

In this section an overview on integrins and on their role in many common diseases, in particular thrombosis and cancer, will be shortly presented. A more detailed discussion can be found in some published reviews.[8,9]

2.1.1 Integrins

The "integrin" terminology was applied in a 1987 review article (Hynes, R. O. *Cell*) to describe a family of structurally, immunochemically and functionally related cell-surface heterodimeric receptors, which integrated the extracellular matrix with the intracellular cytoskeleton to mediate cell migration and adhesion.[9b] In fact, integrins are a large family of transmembrane non-covalently linked heterodimeric receptors composed of an α- and β-subunit. Until now at least 24 such heterodimers have been identified, comprising 18 α-subunits and 8 β-subunits that formed specific, but diverse heterodimers (**Figure 2.1.1.**).[8]

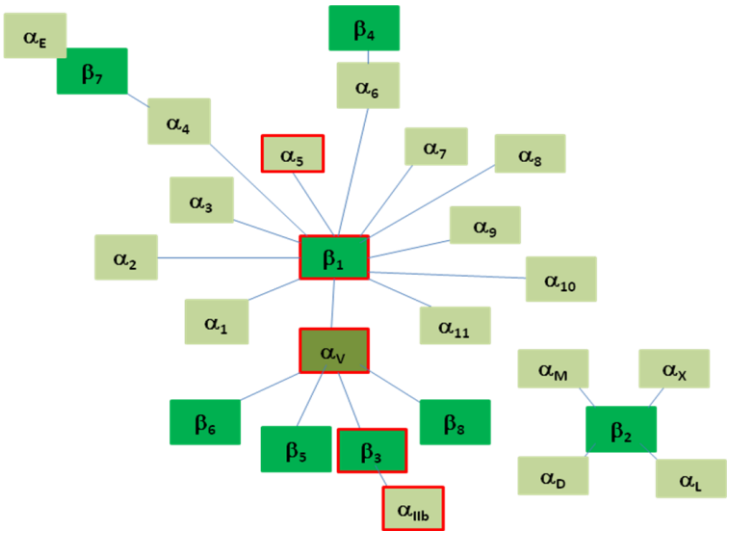


Figure 2.1.1. Integrin subfamilies; in green are reported different types of β-chains, in light green α-chains except for α_V, in fact it can associate with many different β-subunits, whereas other α-subunits are singular in their choice of β-subunit partner.

Initially, integrin subfamilies were defined by their β-subunit, between them the most important are β_1, β_2, and β_3; however, the α-subunit seemed to be important in defining the ligand binding properties of the integrin, and, consequently, defining integrin subfamilies required a more flexible system that takes into consideration also integrin ligand-binding properties and functions.[8]

The subunits are type-I transmembrane proteins equipped with an extensive extracellular domain (amino-terminal) and a small cytoplasmic tail (carboxy-terminal) connected by a single transmembrane chain (**Figure 2.1.2.**).

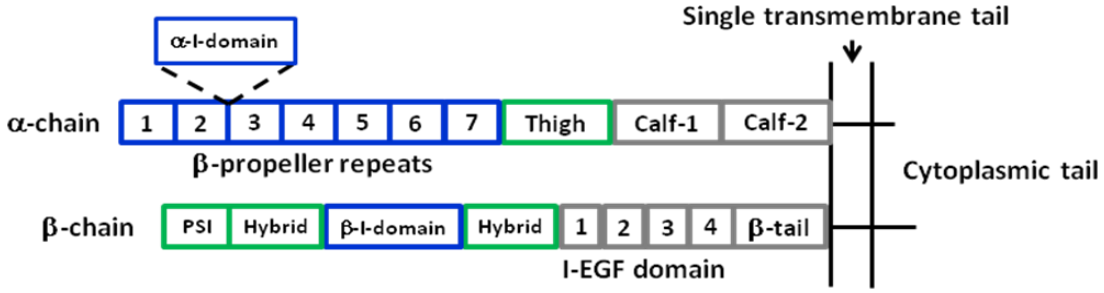


Figure 2.1.2. Subunits structure; α-I domain = Interactive Domain, I-EGF domain = Integrin-Epidermal Growth Factor-like domain, PSI = Plexin Semaphorin Integrin.

α chain has a molecular weight varying between 120 and 200 kDa and consists of about 1100 amino acids, while β chain varies between 90 and 100 kDa (800 amino acids).

The extracellular part of the α subunit is characterized by three or four regions close to the terminal nitrogen, which contain from 12 to 15 amino acids capable of binding divalent cations such as Ca^{2+} or Mg^{2+}. Both subunits have a transmembrane region of about 20 - 30 residues and an intracellular region comprising from 15 to 80 residues. While the amino-terminal portion of both chains form a globular head that contributes to their association and formation of the ligand pocket, the carboxy-terminal portion of β subunit connects integrins with actin filaments of the cytoskeleton through intracellular proteins such as talin, vinculin and α-actin.[10] The various components of the two subunits and their interaction partners are shown schematically in **Figure 2.1.3.**.

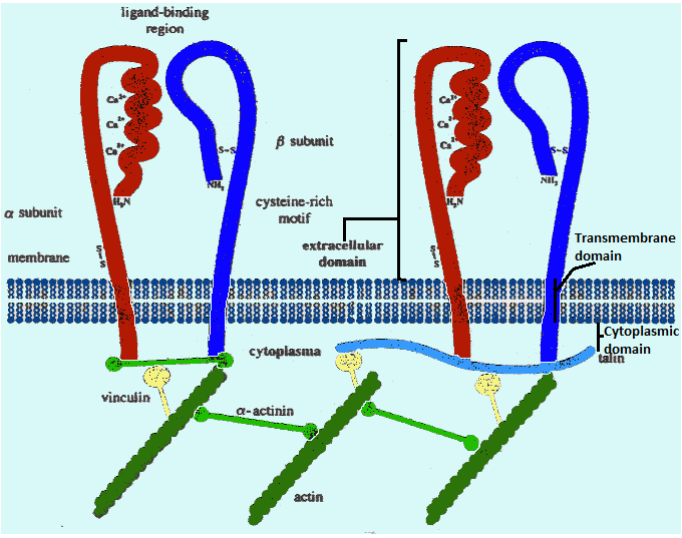


Figure 2.1.3. Interactions with extracellular matrix proteins (Ligand-binding region) and with cytoskeleton (cytoplasm).

The cytoplasmic domain interacts with components of the cytoskeleton, allowing the adhesion of the extracellular domains to different ligands, such as glycoproteins of the extracellular matrix (*i.e.,* fibronectin, fibrinogen, collagen and vitronectin) or proteins on the surface of other cells.

Therefore, integrins are constituted by a complex "integral" transmembrane which has the task of modulating the association between the extracellular matrix and the cytoskeleton.[11] This

connection regulates the activity of all the cells, leading integrins to play an essential role in cellular functions.

Both extracellular domains of the α and β chains contain sites with metal ions that allow interaction with the ligand; the most important of these sites is called "Metal-Ion-Dependent Adhesion Site" (MIDAS) (**Figure 2.1.4.**).[8]

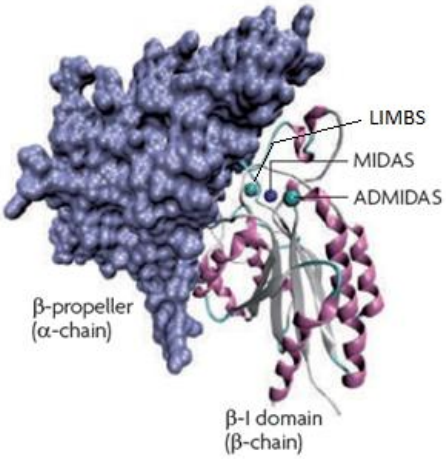


Figure 2.1.4. Receptor site with the three metal ions (Ca^{2+} or Mg^{2+}); MIDAS, Adjacent to MIDAS = ADMIDAS, Ligand-associated-Metal-Binding-Site=LIMBS.[8]

The two main functions of integrins are:

1) Structural functions: attachment of the cell to other cells and to the ECM (Extracellular Matrix); in fact, integrins are often described as the “glue of life”. This means that the major function of these receptors is to maintain the proper structure of tissues by organization of cell-cell and cell-ECM tethering.

2) Signal transduction from the ECM to the cell and *vice versa*. The signals involve: cell growth, cell migration, cell division, cell survival, cellular differentiation, apoptosis (programmed cell death). Integrins transduce information from the ECM (extracellular matrix) to the cell as well as reveal the status of the cell to the outside, allowing rapid and flexible responses to changes in the environment.[8,12]

Integrins display three distinct conformations: a resting inverted V-shaped conformation; an intermediate partially activated extended state; and a fully activated open, ligand-binding conformation (**Figure 2.1.5.**).

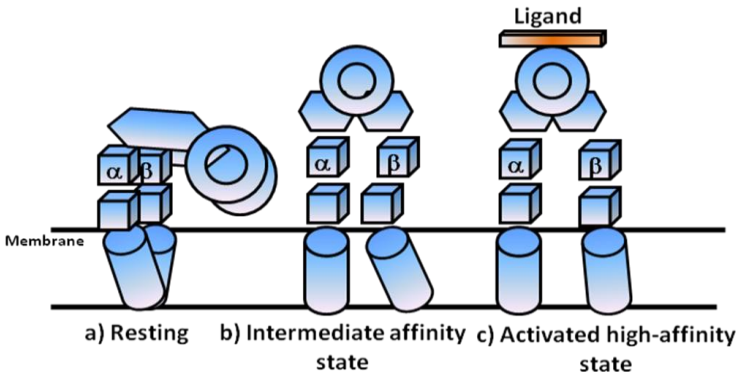


Figure 2.1.5. Swichtblade like opening of integrins. Three major integrin states: a) bent (with a closed headpiece), b) extended with a closed headpiece, and c) extended with an open headpiece. A number of intermediates exists between these three main activation states due to outside-in signaling and inside-out signaling ("switchable model").

Some integrins can bind their ligands in a resting state, whereas others require activation before binding. Activation is mediated by signaling events leading to alterations in the intracellular domains, also known as "inside-out signaling". Ligand binding results in transmission of a signal into the cell, which is known as "outside-in signaling" (**Figure 2.1.5.**). Their participation in signal transduction results in cytoskeleton reorganization (shape change, adhesion, migration), regulation of cell proliferation, cell survival and apoptosis, as previously said. On the whole, the ability of integrins in mediating stable adhesion to the ECM or to other cells, in participating in the assembly of an ECM, in controlling migration, and in regulating signaling cascades, explains their prominent role for the correct development and function of various tissues. Absence of functional integrins is associated with aberrant development of tissues and their functions, ultimately resulting in the initiation and prolongation of many diseases such as thrombosis and related pathologies, tumor formation and metastasis spread, tumor angiogenesis, vascular damage brought about by osteoporosis.[13,14,15]

However, although integrins are known to be involved in various disorders, identifying the specific integrin involved and their precise role is difficult because many diseases are multifactorial and integrins are only one of many types of receptors involved. Moreover, many cells express multiple integrins that exhibit complementary binding properties.[8]

A brief overview of the two disorders in which integrins are known to have an important role and that are the main focus of this thesis is presented.

Thrombosis: the first-disease-related process to be clearly associated with integrins was thrombosis - the formation of a blood clot inside a blood vessel, which occurs when platelets adhere to damaged blood vessels and become activated. These activated platelets recruit other platelets, resulting in the formation of a haemostatic plug. This is an essential mechanism for preventing blood loss, but inappropriate thrombus formation can lead to a stroke or to a heart attack. In the 1970s, it was noted that some patients with a severe bleeding disorder known as Glanzmann's thrombasthenia lacked two functional glycoproteins (GPIIb and GPIIIa) on their blood platelets, which are now known as the integrin $\alpha_{IIb}\beta_3$. It was subsequently found that the platelet – platelet interaction that mediates thrombus formation is facilitated by fibrinogen binding to the platelet-specific integrin $\alpha_{IIb}\beta_3$ following platelet activation caused by thrombotic stimuli. This knowledge provided the basis for the development of the first three integrin-targeted drugs to be approved, which were introduced in the 1990s (see "Integrin ligands").[8,16]

Cancer: During carcinogenesis, the growth of the tumour and its subsequent metastasis is highly dependent on tumor cells being able to regulate their attachment to the extracellular matrix and to adjacent cells. Given that integrins have important roles in cell attachment, survival, migration, invasion and angiogenesis, which are all critical for carcinogenesis, they have attracted considerable attention as potential anticancer targets. Many different integrins have been implicated in carcinogenesis, in particular $\alpha_V\beta_3$, $\alpha_V\beta_5$ and $\alpha_5\beta_1$ play a key role in modulating cancer development. These may be important in angiogenesis, in which endothelial cell survival is crucial.[8] They also have a key role in metastasis, and its expression in melanoma is associated with highly metastatic disease. Integrin $\alpha_5\beta_1$ is a major cellular receptor for the extracellular matrix protein fibronectin and plays a fundamental role during mammalian development.[17] Potentially, blocking integrin function could interfere with tumor growth and metastasis.

The ectodomain of the $\alpha_V\beta_3$ and of $\alpha_5\beta_1$ receptors has been crystallized in the resting state, while the ligand-binding domain of $\alpha_{IIb}\beta_3$ has been crystallized in both the inactive and activated conformations.[16,17,18,19]

In addition, from a structural point of view, $\alpha_V\beta_3$, $\alpha_5\beta_1$ and $\alpha_{IIb}\beta_3$ are all characterized by the lack of the α-I-domain.

Compared with other cell adhesion receptor classes, integrin's ligand recognition mechanism is highly unique in three aspects. First, the ligand recognition specificity of each integrin heterodimer is determined combinatorially, in that both the α and the β subunits contribute to the selective ligand binding of the resultant heterodimeric receptor, and even the same β (or α) subunit will bind different ligands when paired with a different α (or β) subunit. Second, in contrast to other divalent cation-dependent cell adhesion molecules such as cadherins in which metals do not directly bridge two molecules across the cell-cell junction, the core mechanism of integrin-ligand recognition involves a direct coordination bond between a Mg^{2+} ion bound on the integrin (called MIDAS) and a carboxylate oxygen from the ligand. Lastly, the ligand-binding affinity of integrins can be modulated allosterically via conformational changes that take place outside the binding pocket.[17]

2.1.2 Integrin ligands

Over the last decade, it has been identified a wide variety of potent integrin antagonists, with low molecular weight, whose biological targets are $\alpha_V\beta_3$, $\alpha_{IIb}\beta_3$ or $\alpha_5\beta_1$ integrins. One of the best known aspects for the inhibition of α-I-domain lacking integrins is their ability to recognize endogenous ligands characterized by short RGD (Arg-Gly-Asp) sequences (**Figure 2.1.6.**).

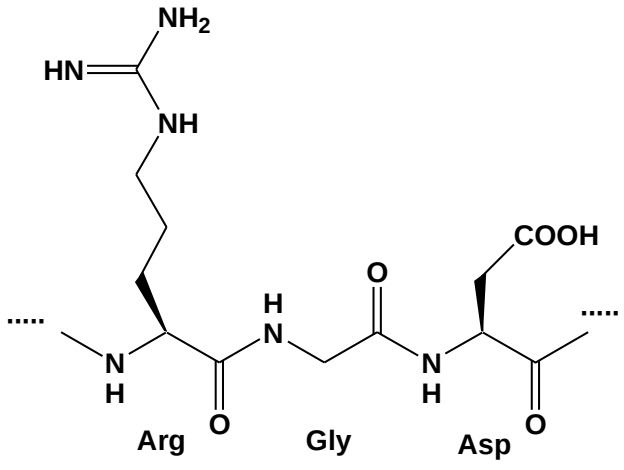


Figure 2.1.6. RGD sequence of integrin endogeneous ligands.

In fact, despite the great variability of the receptors, only a few integrins recognize a single endogenous ligand; most of them is able to bind to a wide variety of ligands, which are also recognized by other adhesion receptors of the cell surface.[20] Among the natural ligands of integrins we find laminin, fibronectin, vitronectin, fibrinogen and fibrin. Therefore, the first phase in the planning and development of synthetic antagonists of these receptors was focused on creating a series of low molecular weight peptides where various amino acid residues flank the RGD consensus sequence. This sequence was used as template for the development of potent and selective monomeric ligands, both linear and conformationally constrained by cyclization, characterized by a peptide, pseudopeptide or not peptide sequence. It was shown that linear ligands containing the RGD sequence have, however, low specificity toward integrins, while homologous peptides having a cyclic structure generally show higher activity and selectivity. These efforts leads to the design and synthesis of effective, conformationally constrained cyclic (pseudo) peptides, such as the inhibitor Cilengitide (**Figure 2.1.7.**), which proved to be one of the most potent and selective antagonists of the integrin $\alpha_V\beta_3$,[21] reached phase III clinical trials for the treatment of glioblastoma multiforme and now is in phase II studies for head and neck cancer as well as lung cancer.

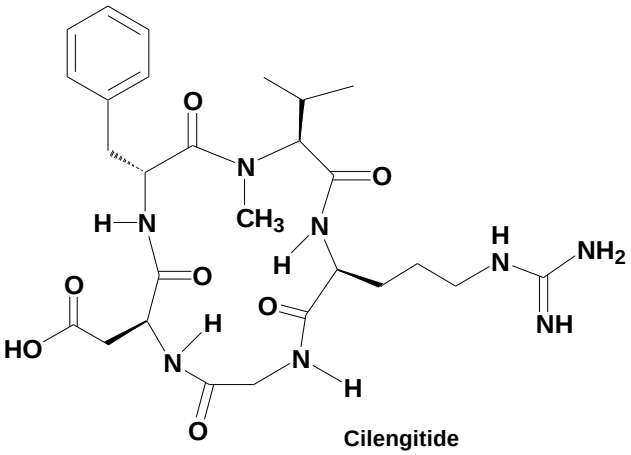


Figure 2.1.7. Structure of the inhibitor Cilengitide (cyclo-[Arg-Gly-Asp-D-Phe-N(Me)-Val]) developed by Kessler in the early 90s.

In addition, a crucial step forward in this important area arose from the crystal structure determination of the extracellular segment of vitronectin receptor $\alpha_V\beta_3$ and of its complex with

Cilengitide,[18] and from the resolution of the crystallographic structure of integrin $\alpha_{IIb}\beta_3$[19] in complex with a synthetic antagonist in use as antithrombotic, called Eptifibatide or Integrilin, whose structure is shown in **Figure 2.1.8.**.

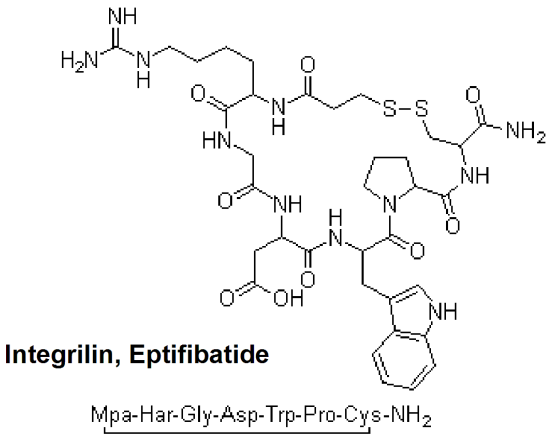


Figure 2.1.8. Structure of the inhibitor Integrilin developed by Robert M. Scarborough e David Phillips (COR Therapeutics).

Recently, in 2012, the crystal structure of the $\alpha_5\beta_1$ integrin headpiece fragment bound by an allosteric inhibitory antibody was determined both in the absence and presence of a ligand peptide containing the RGD sequence.[17] The structure, solved as a complex with a Fab fragment of the anti-β1 inhibitory antibody SG/19, revealed high similarity to the ligand-unbound form of $\alpha_V\beta_3$ and $\alpha_{IIb}\beta_3$ integrins. Like integrin $\alpha_V\beta_3$, $\alpha_5\beta_1$ integrin was crystallized in its resting conformation. It was particularly important to obtain structural information about β_1 integrins because they constitute the largest and probably the most ancient integrin subclass.

The specific recognition of RGD peptide by three different integrins is reported in **Figure 2.1.9.**. A more detailed explanation of the single domain will be given in the next Sections.

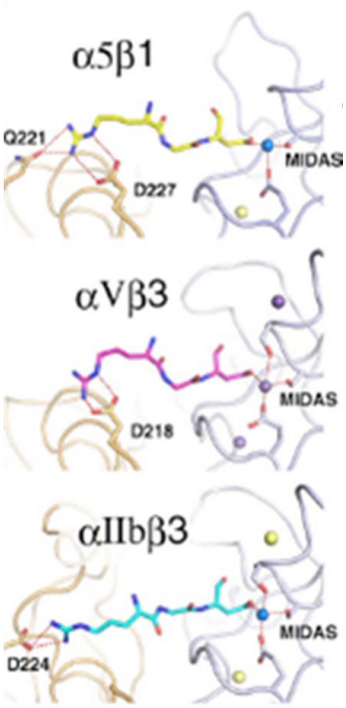


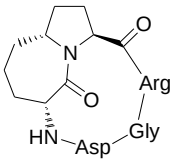
Figure 2.1.9. Specific recognition of RGD peptide by three different integrins: $\alpha_5\beta_1$, $\alpha_V\beta_3$ and $\alpha_{IIb}\beta_3$. Hydrogen and coordination bonds are shown in clashed lines. Metal ions in the β–I-domain are shown in spheres.[17]

Capitalizing on these fundamental results, numerous cyclic peptidomimetics containing the RGD sequence and pseudopeptide scaffolds of different nature were introduced.

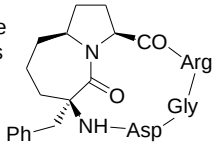
These include ligands incorporating unnatural 6,5- and 7,5- fused 1-aza-2-oxobicycloalkane amino acids (ST1646, ST2605, DB58) (**Figure 2.1.10.**),[22,23,24] or diketopiperazinic scaffold-containing β-amino acids (AV98),[25] RGD peptidomimetics containing γ-aminocyclopentane carboxylic acid[26] (PRGC008) or 4-aminoproline residue[27] (PRGC021)[28] and RGD-Morpholine peptides (c-[RGDf-D-Morph]).[29]

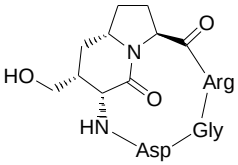
Fifteen-membered cyclic RGD pseudopentapeptides

University of Milano, CNR-ISTM | University of Milano and Pavia, CNR-ISTM | University of Milano, CNR-ISTM



azabicycloalkane dipeptide mimics





ST1646

$\alpha_V\beta_3$ IC_{50} 3.8±1.2 nM
$\alpha_V\beta_5$ IC_{50} 1.4±0.2 nM

inhibition of ^{125}I-echistatin binding

C. Scolastico et al.
Mol. Cancer Ther. 2005, 4, 1670

ST2605

$\alpha_V\beta_3$ IC_{50} 6.4±0.1 nM
$\alpha_V\beta_5$ IC_{50} 7.7±0.04 nM

L. Colombo, C. Scolastico et al.
ChemMedChem 2008, 3, 1589

DB58

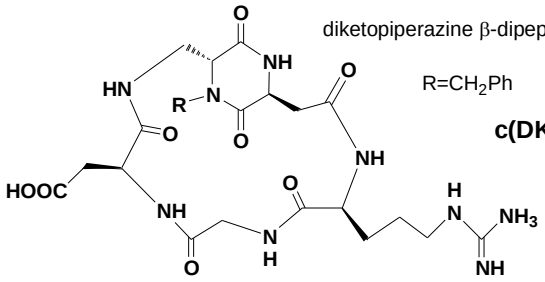
$\alpha_V\beta_3$ IC_{50} 25.8±4.0 nM
$\alpha_V\beta_5$ IC_{50} 205±33.5 nM

inhibition of biotinylated vitronectin binding

L. Manzoni, L. Belvisi et al.
ChemMedChem 2009, 4, 615

Seventeen-membered cyclic RGD pseudopentapeptides

University of Milano and Insubria



diketopiperazine β-dipeptide mimics

R=CH_2Ph

c(DKP2-RGD)

$\alpha_V\beta_3$ IC_{50} 3.2±2.7 nM
$\alpha_V\beta_5$ IC_{50} 114±99 nM

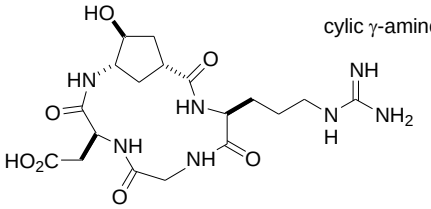
inhibition of biotinylated vitronectin binding

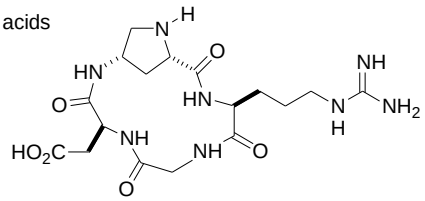
C. Gennari, U. Piarulli et al.
Chem. Eur. J. 2009, 15, 12184

Fourteen-membered cyclic RGD pseudotetrapeptides

University of Parma

cyclic γ-amino acids





PRGC008

$\alpha_V\beta_3$ IC_{50} 5.6±0.1 nM
$\alpha_V\beta_5$ IC_{50} 4.56±0.13 nM

inhibition of ^{125}I-echistatin binding

G. Casiraghi et al.
J. Med. Chem. **2005**, 48, 7675

PRGC021

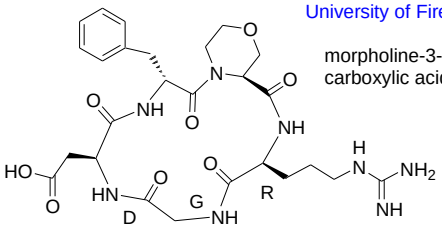
$\alpha_V\beta_3$ IC_{50} 0.47±0.20 nM
$\alpha_V\beta_5$ IC_{50} 30±11 nM

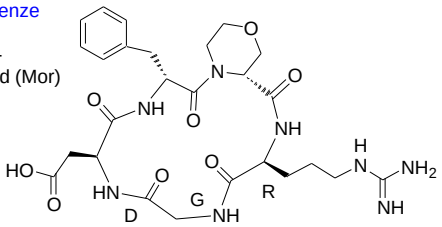
F. Zanardi et al.
J. Med. Chem. **2008**, 51, 1771

Fifteen-membered cyclic RGD pseudopentapeptides

University of Firenze

morpholine-3-carboxylic acid (Mor)





c[RGDf-(L)-Mor]

$\alpha_V\beta_3$ IC_{50} 157±0.9 nM
$\alpha_V\beta_5$ IC_{50} 15.1±2.3 nM

inhibition of ^{125}I-echistatin binding

A. Guarna et al.
Bioorg. Med. Chem. 2009, 17, 1542

c[RGDf-(D)-Mor]

$\alpha_V\beta_3$ IC_{50h} 6.5±2.0 nM (high affinity state)
$\alpha_V\beta_5$ IC_{50} 21.0±2.1 nM

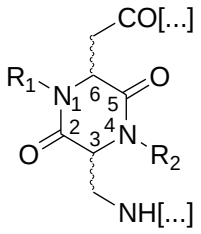
Figure 2.1.10. Structure of some cyclic ligands based on a RGD sequence and various scaffolds.

However, up to now, a large number of linear or cyclic peptidic and peptidomimetic ligands have been developed, which are all related to the common recognition motif RGD[28,30] and a few potent ligands are presently in different stages of clinical trials for cancer therapy, or in clinical use for antithrombotic therapy. Notwithstanding these recent achievements, the discovery of new ligands displaying high activity and selectivity together with an optimal pharmacological profile still remains a challenge.

During my thesis work, studies have been focused on hybrid cyclic ligands containing a diketopiperazinic scaffold and the recognition sequence RGD (see AV98 in **Figure 2.1.10.**). These scaffolds affect the spatial arrangement of the pattern, locking it in various conformations, described in the paragraph below.

2.1.3 DKP-RGD peptidomimetic ligands

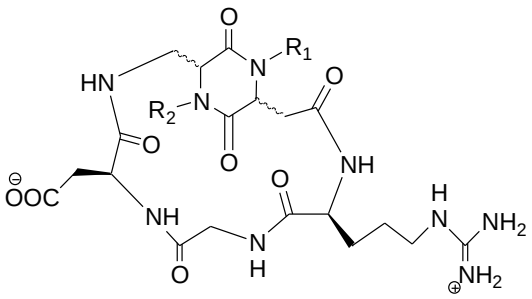
New bifunctional diketopiperazine (DKP) scaffolds [**DKP-1** (*cis*) and **DKP-2/DKP-7** (*trans*)], which are formally derived from aspartic acid and 2,3-diaminopropionic acid and bear a carboxylic acid and an amino functionality, were recently developed and their stereochemistry and substitution patterns were varied (**Figure 2.1.11.**). In particular the scaffolds differ in: 1) the relative stereochemistry, namely *cis* (**DKP-1**) or *trans* (**DKP-2** – **DKP-7**); 2) the absolute stereochemistry of the *trans* scaffolds [3R,6S (**DKP-2**, **DKP-4**, **DKP-5**) or 3S, 6R (**DKP-3**, **DKP-6**, **DKP-7**)]; 3) the substitution at the endocyclic nitrogen atoms, which can be either hydrogen or benzyl (**DKP-2**,**DKP-3**, **DKP-4**, **DKP-6**) or dibenzyl (**DKP-5**, **DKP-7**).



DKP-1 = 3*S*, 6*S*, R^1 = H, R^2 = CH_2Ph
DKP-2 = 3*R*, 6*S*, R^1 = H, R^2 = CH_2Ph
DKP-3 = 3*S*, 6*R*, R^1 = H, R^2 = CH_2Ph
DKP-4 = 3*R*, 6*S*, R^1 = CH_2Ph, R^2 = H
DKP-5 = 3*R*, 6*S*, R^1 = CH_2Ph, R^2 = CH_2Ph
DKP-6 = 3*S*, 6*R*, R^1 = CH_2Ph, R^2 = H
DKP-7 = 3*S*, 6*R*, R^1 = CH_2Ph, R^2 = CH_2Ph

Figure 2.1.11. Bifunctional DKP scaffolds.

These bifunctional diketopiperazines were introduced into the 17-membered cyclic RGD peptidomimetics **Cyclo-[DKP-1-RGD] – Cyclo-[DKP-7-RGD]** (**Figure 2.1.12.**).



(1) *cyclo*(**DKP-1**-RGD) = 3*S*, 6*S*, R^1 = H, R^2 = CH_2Ph
(2) *cyclo*(**DKP-2**-RGD) = 3*R*, 6*S*, R^1 = H, R^2 = CH_2Ph
(3) *cyclo*(**DKP-3**-RGD) = 3*S*, 6*R*, R^1 = H, R^2 = CH_2Ph
(4) *cyclo*(**DKP-4**-RGD) = 3*R*, 6*S*, R^1 = CH_2Ph, R^2 = H
(5) *cyclo*(**DKP-5**-RGD) = 3*R*, 6*S*, R^1 = CH_2Ph, R^2 = CH_2Ph
(6) *cyclo*(**DKP-6**-RGD) = 3*S*, 6*R*, R^1 = CH_2Ph, R^2 = H
(7) *cyclo*(**DKP-7**-RGD) = 3*S*, 6*R*, R^1 = CH_2Ph, R^2 = CH_2Ph

Figure 2.1.12. Cyclic RGD peptidomimetics containing bifunctional DKP scaffold.

The configuration of the two amino acids may affect the spatial arrangement of the RGD motif, while the different nature of the nitrogen substitutes can influence the affinity for the receptor for steric repulsion or lipophilic interactions. The context of the RGD sequence (flanking residues, 3D presentation, and individual features of the integrin binding pockets) determines the specificity and efficacy of interaction.

In the case of the N-dibenzyl derivative **Cyclo-[DKP-7-RGD]**, the molecule can exist as two different separable conformers (diastereomers) due to hindered rotation of one ring around the other, in a way reminiscent of the ansa-cyclopeptides (*i.e.*, the DKP N-benzyl group cannot pass inside the macrolactam ring). So, for this compound two diastereomers (P1 – P2) were formed in the macrolactamization step and were isolated in a 2(A):1(B) ratio.

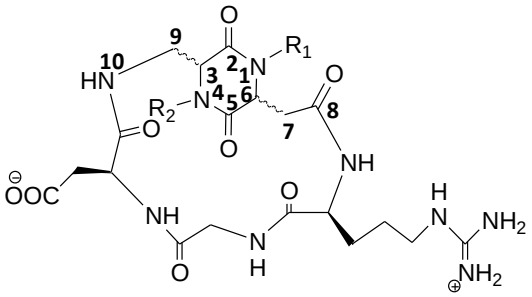
2.1.4 NMR characterization and conformational studies[4]

During the period of my thesis, I have contributed to the characterization of the conformations adopted by these RGD ligands, where diketopiperazine scaffolds were introduced in order to induce β-turn motifs and stabilize the structure of the cycle. Natural peptides or proteins most frequently form α-helices, β-sheets, as well as reverse-turns.[27] A turn is defined as a site where the peptide backbone reverses the direction of propagation by adopting a U-shaped conformation.

The terms β-turns and γ-turns have more restricted definitions and describe turns of four or three consecutive residues, respectively. These turns can be stabilized by intramolecular H-bonding. In the β-turn, the carbonyl of the first residue *i* may be hydrogen bonded to the NH of the fourth residue (*i+3*), producing a ten-membered cycle. In γ-turns, the C=O of the first residue (*i*) may be hydrogen bonded to the NH of the third residue (*i+2*), giving rise to a seven-membered ring.

For this reason, conformational studies in the free-state of these cyclic conformationally constrained RGD peptidomimetics were conducted by ^{1}H-NMR experiments (VT-NMR and NOESY) in aqueous solution and by computational methods, such as mixed-mode Metropolis Monte Carlo/Stochastic Dynamics (MC/SD) simulations. Monodimensional ^{1}H-NMR experiments were conducted to detect intramolecular hydrogen bonds, by measuring the chemical shift of the N–H protons and their temperature coefficients ($\Delta\delta/\Delta T$), while NOESY spectra were recorded to investigate both sequential and long-range NOE's that provide evidences of preferred conformations. These NOE contacts were used as constraints (see ∗Supplementary Information∗) for the MC/SD simulations,[31] performed using the implicit water GB/SA solvation model[32] and the OPLS_2001 force field,[33] chosen for its capability to better parameterize the diketopiperazine scaffolds of our compounds.

The structure and connectivity of ligands **Cyclo-[DKP-1-RGD]** and **Cyclo-[DKP-2-RGD]** to **Cyclo[DKP-7-RGD]** were unambiguously assigned by means of mono- and bidimensional ^{1}H- and ^{13}C-NMR spectra (see ∗Supplementary Information∗). The relevant NMR data are summarized in **Table 2.1.1.**.



Structure	δ (ppm)						Δδ/ΔT (ppb/K)						Significant NOE contacts
	NH_1	NH_4	NH_{10}	NH_{Arg}	NH_{Gly}	NH_{Asp}	NH_1	NH_4	NH_{10}	NH_{Arg}	NH_{Gly}	NH_{Asp}	
cyclo[**DKP-1**-RGD]	8.35	-	7.46	8.40	8.75	8.10	-7.3	-	-2.0	-7.0	-8.0	-3.7	NH_{Asp}-NH_{10}; NH_{Asp}-NH_{Gly}
cyclo[**DKP-2**-RGD]	8.35	-	8.78	8.57	8.18	8.29	-8.7	-	-10.7	-7.0	-5.7	-7.7	NH_{Arg}-NH_{Gly}
cyclo[**DKP-3**-RGD]	8.10	-	8.28	8.80	8.00	7.85	-5.7	-	-8.5	-6.0	-4.5	-3.5	NH_{Arg}-NH_{Gly}
cyclo[**DKP-4**-RGD]	-	8.17	7.59	8.29	8.27	8.88	-	-9.1	-0.7	-9.3	-8.2	-9.3	--
cyclo[**DKP-5**-RGD]	-	-	8.58	8.48	8.23	8.42	-	-	-11.0	-7.5	-4.7	-8.2	NH_{Arg}-NH_{Gly}
cyclo[**DKP-6**-RGD]	-	8.07	7.90	8.32	8.35	8.80	-	-4.9	-5.1	-7.6	-6.7	-8.0	NH_{Asp}-NH_{10}; NH_4-NH_{10}
cyclo[**DKP-7**-RGD]-A	-	-	8.04	8.66	7.93	7.76	-	-	-7.5	-5.0	-3.0	-1.0	NH_{Arg}-NH_{Gly}; NH_{Asp}-NH_{Gly}
cyclo[**DKP-7**-RGD]-B	-	-	7.72	8.34	8.45	8.55	-	-	-4.0	-7.0	-7.0	-5.0	NH_{Asp}-NH_{10}

Table 2.1.1. ^{1}H-NMR and NOE data of cyclic RGD peptidomimetics in water.

On the basis of spectroscopic data and computational simulations, different preferred conformations featuring specific intramolecular H-bonding patterns and turn motifs were identified for the different ligands, depending on configuration and substitution of the DKP scaffold. In particular, five distinct types of conformations were recognized and classified as follows: Type I-cis (β-turn at Gly-Asp), Type I-trans (β-turn at Gly-Asp), Type II (β-turn at Arg-Gly), Type III (β-turn at DKP-Arg) and Type IV (β-turn at Asp-DKP). As shown below, these conformations are characterized by various values of the distance between the Cb atoms of Asp and Arg residues, a key parameter for the RGD fitting into the active site of integrins.[18] In particular, an extended conformation of the RGD motif displaying a distance of about 9 Å is required to allow the carboxylate and guanidinium groups to properly interact with the receptor acting as an electrostatic clamp.

Compound **Cyclo-[DKP-1-RGD]**[25] exists as an equilibrium of two different preferred conformations. The NOESY spectrum shows two mutually exclusive long-range NOE contacts. The cross peak between DKP-NH_{10} and NH_{Asp} (strong) is indicative of a β-turn conformation at Gly-Asp stabilized by a hydrogen bond between DKP-NH_{10} and Arg-C=O (referred as Type I H-bonding pattern, **Figure 2.1.13.A**). The chemical shift value (7.46 ppm) and the $\Delta\delta/\Delta T$ value (-2 ppb/K) of the amide proton DKP-NH_{10} indicate that this proton is strongly locked in an intramolecularly H-bonded state. The cross peak between NH_{Gly} and NH_{Asp} (medium) is indicative of an alternative β-turn conformation at Arg-Gly, stabilized by a hydrogen bond between Asp-NH and C(8)=O (referred as Type II H-bonding pattern, **Figure 2.1.13.B**).

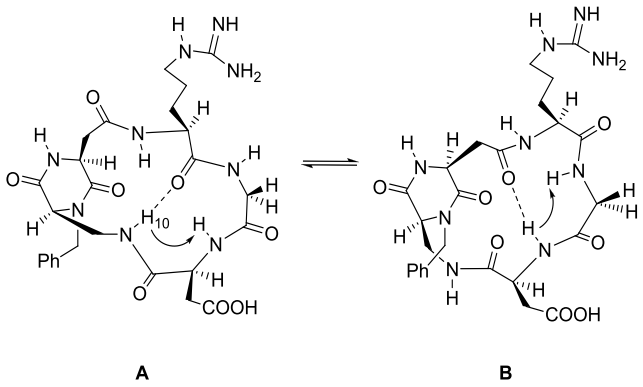


Figure 2.1.13. Preferred intramolecular hydrogen-bonded patterns proposed for compound **DKP-1** on the basis of spectroscopic data. The arrows indicate significant NOE contacts. A) Type I H-bonding pattern, Gly-Asp β-turn motif. B) Type II H-bonding pattern, Arg-Gly β-turn motif.

Three-dimensional structures satisfying long-range NOE contacts were generated for compound **Cyclo-[DKP-1-RGD]**[25] performing two 10 ns restrained MC/SD simulations and applying the DKP-NH_{10}/NH_{Asp} or the NH_{Asp}/NH_{Gly} distance restraint derived from the NOESY spectra. More than 90% of the conformations sampled during the first simulation adopted a non-extended arrangement of the RGD sequence characterized by a β-turn at Gly-Asp and the presence of the corresponding hydrogen bond between DKP-NH_{10} and Arg-C=O. In addition, the formation of a γ-turn at Gly stabilized by the hydrogen bond between NH_{Asp} and Arg-C=O was observed for 40% of the simulation. A Cβ(Arg)–Cβ(Asp) average distance of 7.4 Å was obtained during this MC/SD

calculation. A representative energy minimized conformation selected by cluster analysis and featuring both H-bonds is shown in **Figure 2.1.14.A** (Type I-*cis* H-bonding pattern).

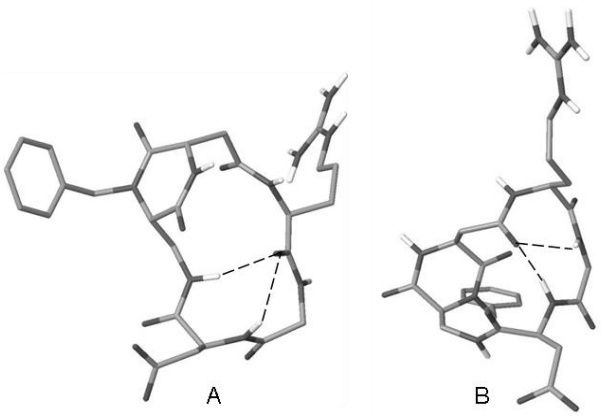


Figure 2.1.14. Structures of **Cyclo-[DKP-1-RGD]** as obtained by restrained MC/SD simulations based on experimental distance information, after energy minimization. A) Type I-*cis* H-bonding pattern, γ-turn at Gly and βII'-turn at Gly-Asp [CβArg)–Cβ(Asp)=7.9 Å]. B) Type II H-bonding pattern, γ-turn at Arg and βII'-turn at Arg-Gly [Cβ(Arg)–Cβ(Asp)=6.6 Å]. Tube representation: O, N in dark gray, C in light gray and N-H hydrogens in white; for the sake of clarity, all H atoms bound to carbon are omitted.

Approximately 60% of the conformations sampled during the simulation of **Cyclo-[DKP-1-RGD]** featuring the NH_{Asp}/NH_{Gly} distance restraint, adopted a non-extended arrangement of the RGD sequence characterized by a β-turn at Arg-Gly and the corresponding hydrogen bond between NH_{Asp} and C(8)=O. In addition, the formation of a γ-turn at Arg stabilized by the hydrogen bond between NH_{Gly} and C(8)=O was observed for 40% of the simulation. The Cβ(Arg)–Cβ(Asp) average distance in this MC/SD calculation was 6.8 Å. A representative energy minimized conformation selected by cluster analysis and featuring both H-bonds is shown in **Figure 2.1.14.B** (Type II H-bonding pattern).

The similarity of the NMR data of compounds **Cyclo-[DKP-2-RGD]** and **Cyclo-[DKP-3-RGD]** and, hence, of the conformation of these two ligands is quite surprising, considering the opposite configuration of the diketopiperazine scaffold [**DKP-2** (3*R*,6*S*); **DKP-3** (3*S*,6*R*)] which should impart

a different stereochemical orientation to the two side arms of the diketopiperazine. This conformational similarity can be interpreted in terms of a quasi-enantiomeric structure of the two ligands (excluding the configuration of the remote RD amino acid side chains, **Figure 2.1.15.**). Both the compounds are apparently characterized by conformational mobility, as suggested by the values of chemical shifts and $\Delta\delta/\Delta T$ reported in **Table 2.1.1.**. The only exception is proton NH-Asp of **Cyclo-[DKP-3-RGD]** (7.85 ppm, $\Delta\delta/\Delta T$ = -3.5 ppb/K), which might be involved in a Type II H-bonding pattern (for definition of Type II H-bonding pattern, see **Figure 2.1.13.B** above). This last conformation, however, is less important due to the lack of an associated NOE contact and we will not report in the other analysis. On the other hand, the presence in both cases of a NOE contact between NH_{Gly} and NH_{Arg} suggests the formation of a β-turn motif at DKP-Arg, stabilized by a hydrogen bond between NH_{Gly} and C(5)=O (referred to as Type III H-bonding pattern, **Figure 2.1.15.**).

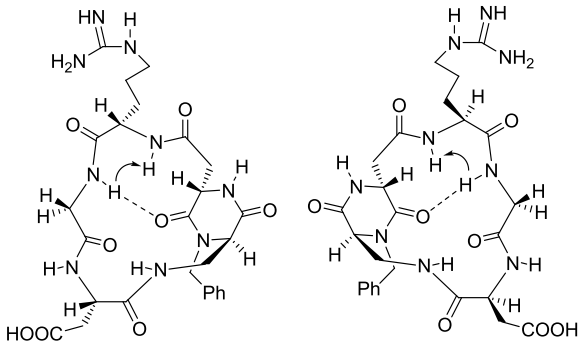


Figure 2.1.15. Preferred intramolecular hydrogen-bonded pattern proposed for compound **DKP-2** (left) and **DKP-3** (right) on the basis of spectroscopic data. The arrows indicate significant NOE contacts. The DKP-Arg β-turn motif is referred to as Type III H-bonding pattern.

The presence of this H-bond is also supported by the rather upfield chemical shift value of NH_{Gly} in these two ligands (8.18 and 8.00 ppm for **DKP-2** and **DKP-3**, respectively) and the relatively low temperature dependence (-5.7 and -4.5 ppb/K, respectively).

Ligand **Cyclo-[DKP-5-RGD]**, with the dibenzylated diketopiperazine, shows NMR features similar to ligand **Cyclo-[DKP-2-RGD]** (for definition of Type III H-bonding pattern, see **Figure 2.1.15.** above): a NOE contact between NH_{Gly} and NH_{Arg} and a rather shielded NH_{Gly} (8.23 ppm) with a relatively low temperature coefficient (-4.7 ppb/K).

The distance restraint corresponding to the NOE contact between NH_{Gly} and NH_{Arg} was applied in the 10 ns MC/SD simulations of compounds **Cyclo-[DKP-2-RGD]**, **Cyclo-[DKP-3-RGD]** and **Cyclo-[DKP-5-RGD]**. More than 90% of the conformations sampled during each of these simulations adopted an extended arrangement of the RGD sequence characterized by a pseudo β-turn at DKP-Arg and the formation of the corresponding hydrogen bond between the NH_{Gly} and C(5)=O. Interestingly, only for compound **Cyclo-[DKP-3-RGD]**, the additional formation of a β-turn at Arg-Gly stabilized by the hydrogen bond between NH_{Asp} and C(8)=O was observed for 15% of the simulation. These results and the NMR data (showing δ 7.85 ppm and $\Delta\delta/\Delta T$ -3.5 ppb/K for NH_{Asp} of **Cyclo-[DKP-3-RGD]**) suggest the contribution of a Type II/Type III H-bonding pattern to the conformational equilibrium of **Cyclo-[DKP-3-RGD]** (mainly populated by a Type III H-bonding pattern). Cβ(Arg)-Cβ(Asp) average distances of 9.3, 8.8, and 9.1 Å were obtained during the MC/SD calculations of **Cyclo-[DKP-2-RGD]**, **Cyclo-[DKP-3-RGD]** and **Cyclo-[DKP-5-RGD]**, respectively. A representative energy minimized conformation selected by cluster analysis and featuring the H-bond between the Gly-NH and C(5)=O (Type III H-bonding pattern) is shown in **Figure 2.1.16** for RGD peptidomimetic **Cyclo-[DKP-2-RGD]**.

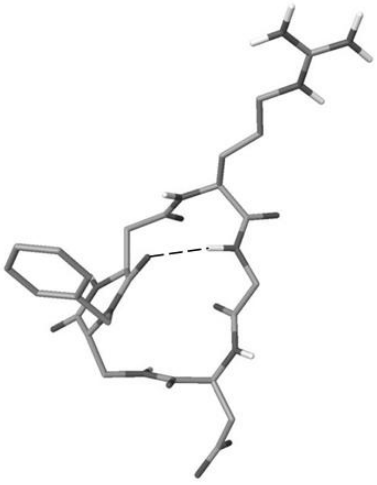


Figure 2.1.16. Structures of **Cyclo-[DKP-2-RGD]** as obtained by restrained MC/SD simulations based on experimental distance information, after energy minimization. Compound **Cyclo-[DKP-2-RGD]**, Type III H-bonding pattern, distorted inverse γ-turn at Asp and pseudo β-turn at DKP-Arg [Cβ(Arg)–Cβ(Asp)=9.4 Å]. Tube representation: O, N in dark gray, C in light gray and N-H hydrogens in white; for the sake of clarity, all H atoms bound to carbon are omitted.

Ligands **Cyclo-[DKP-4-RGD]** and **Cyclo-[DKP-6-RGD]**, featuring the diketopiperazine scaffolds **DKP-4** (3*R*,6*S*) and **DKP-6** (3*S*,6*R*) respectively (with the benzyl substitution at the endocyclic nitrogen N-1, instead of N-4), show a different NMR pattern.

Ligand **Cyclo-[DKP-4-RGD]** is characterized by the absence of relevant NOE contacts, a very low temperature dependence (-0.7 ppb/K) and a quite upfield chemical shift value (7.59 ppm) for proton NH_{10}. These two features suggest a Type I H-bonding pattern (for definition of Type I H-bonding pattern, see **Figure 2.1.18.A** below), notwithstanding the apparent lack of NOE contact between NH_{Asp} and NH_{10}.

Due to the absence of relevant long-range NOE contacts, several 10 ns runs of unconstrained MC/SD simulations were performed for high affinity ligand **Cyclo-[DKP-4-RGD]** (containing *N*-1-benzylated **DKP-4**, 3*R*,6*S*) starting from different 3D structures. Most of the conformations sampled during these simulations adopted an extended arrangement of the RGD sequence [Cβ(Arg)-Cβ(Asp) average distance of 8.8 Å] and approximately 40% of them are characterized by a β-turn at Gly-Asp and the presence of the corresponding hydrogen bond between DKP-NH_{10} and Arg-C=O. These results provide a structural model in agreement with NMR data showing a low temperature dependence (-0.7 ppb/K) and an upfield chemical shift value (7.59 ppm) for proton NH_{10}.

A representative energy minimized conformation selected by cluster analysis and featuring the H-bond between DKP-NH_{10} and Arg-C=O (Type I-*trans* H-bonding pattern) is shown in **Figure 2.1.17** for RGD peptidomimetic **Cyclo-[DKP-4-RGD]**.

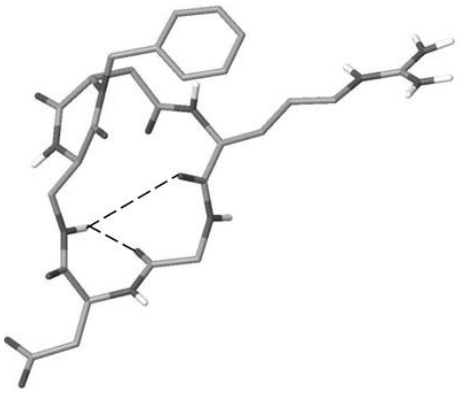


Figure 2.1.17. Structure of **Cyclo-[DKP-4-RGD]** as obtained by MC/SD simulations, after energy minimization. Compound **Cyclo-[DKP-4-RGD]**, Type I-*trans* H-bonding pattern, inverse γ-turn at

Asp and distorted βII'-turn at Gly-Asp [Cβ(Arg)–Cβ(Asp)=9.0 Å]. Tube representation: O, N in dark gray, C in light gray and N-H hydrogens in white; for the sake of clarity, all H atoms bound to carbon are omitted.

It is worth noting how the combination of the *trans* **DKP-4** scaffold with the Gly-Asp β-turn occurs by generating an extended RGD arrangement, while the combination of the *cis* **DKP-1** scaffold with the same secondary motif resulted in a non-extended RGD disposition (see above, **Figure 2.1.14.A**). Accordingly, two Type I H-bonding patterns have been defined (Type I-*cis* and Type I-*trans*), depending on the *cis* or *trans* relative stereochemistry of the diketopiperazine scaffold.

On the other hand, ligand **Cyclo-[DKP-6-RGD]** is characterized by a rather strong NOE contact between NH_{Asp} and NH_{10} and a moderate one involving NH_4 and NH_{10}. These two contacts are mutually exclusive and hence indicative of an equilibrium between two different conformations, respectively Type I and Type IV binding modes (**Figure 2.1.18.A** and **B**). The hydrogen bonded status of the two amide protons NH_4 and NH_{10}, as indicated by their rather low temperature dependence (-4.9 and -5.1 ppb/K, respectively) and quite upfield chemical shift values (8.07 and 7.90 ppm, respectively), corroborate this assumption. The Type IV H-bonding pattern could feature a pseudo β-turn at Asp-DKP stabilized by a hydrogen bond between NH_4 and Gly-C=O (NOE contact between NH_4 and NH_{10}).

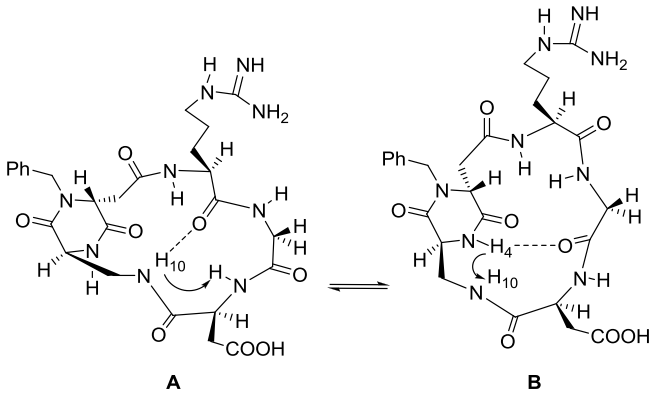


Figure 2.1.18. Preferred intramolecular hydrogen-bonded pattern proposed for compound **Cyclo-[DKP-6-RGD]** on the basis of spectroscopic data. A) Type I H-bonding pattern, Gly-Asp β-turn motif. B) Type IV H-bonding pattern, pseudo β-turn at Asp-DKP. The arrows indicate significant NOE contacts.

Three-dimensional structures satisfying long-range NOE contacts were generated for compound **Cyclo-[DKP-6-RGD]** (containing *N*-1-benzylated **DKP-6**, 3*S*,6*R*) performing two 10 ns restrained MC/SD simulations and applying the DKP-NH_{10}/NH_{Asp} or the NH_4/NH_{10} distance restraint derived from NOESY spectra (**Table 2.1.1.**, **Figure 2.1.18.**).

Most of the conformations sampled during the first simulation adopted an extended arrangement of the RGD sequence [Cβ(Arg)-Cβ(Asp) average distance of 9.0 Å] and approximately 40% of them are characterized by a β-turn at Gly-Asp and the corresponding hydrogen bond between DKP-NH_{10} and Arg-C=O. A representative energy minimized conformation selected by cluster analysis and featuring this H-bond is shown in **Figure 2.1.19.A** (Type I-*trans* H-bonding pattern). Approximately 70% of the conformations sampled during the simulation of **Cyclo-[DKP-6-RGD]** featuring the NH_4/NH_{10} distance restraint adopted an extended arrangement of the RGD sequence [Cβ(Arg)-Cβ(Asp) average distance of 8.8 Å] characterized by a pseudo β-turn at Asp-DKP and the corresponding hydrogen bond between NH_4 and Gly-C=O. In addition, the formation of a γ-turn at Asp stabilized by the hydrogen bond between NH_{10} and Gly-C=O was observed for 50% of the simulation. A representative energy minimized conformation selected by cluster analysis and featuring these H-bonds is shown in **Figure 2.1.19.B** (Type IV H-bonding pattern).

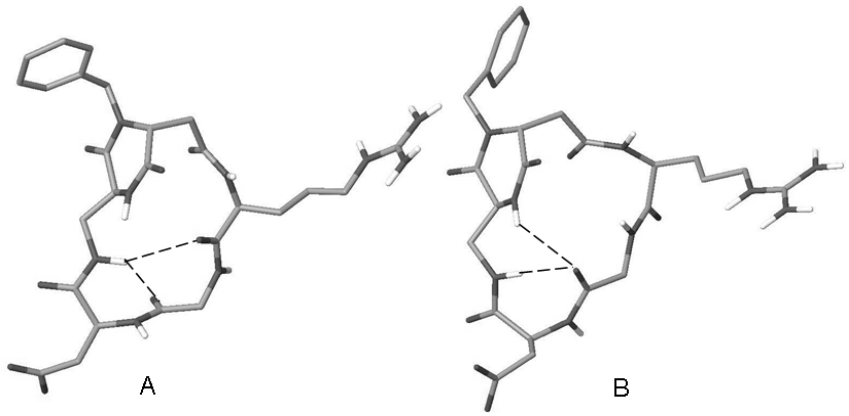


Figure 2.1.19. Structures of **Cyclo-[DKP-6-RGD]** as obtained by restrained MC/SD simulations based on experimental distance information, after energy minimization. A) Type I-*trans* H-bonding pattern, inverse γ-turn at Asp and distorted βII'-turn at Gly-Asp [Cβ(Arg)–Cβ(Asp)=9.0 Å]. B) Type IV H-bonding pattern, inverse γ-turn at Asp and pseudo β-turn at Asp-DKP [Cβ(Arg)–Cβ(Asp)=8.8 Å]. Tube representation: O, N in dark gray, C in light gray and N-H hydrogens in white; for the sake of clarity, all H atoms bound to carbon are omitted.

The dibenzylated diketopiperazine containing peptidomimetic **Cyclo-[DKP-7-RGD]** was eventually studied. Compound **Cyclo-[DKP-7-RGD]** was obtained as a mixture of two diastereomers **Cyclo-[DKP-7-RGD]-A** and **Cyclo-[DKP-7-RGD]-B** due to hindered rotation of one ring around the other (see **Figure 2.1.20.**), whose conformations in solution were studied separately.

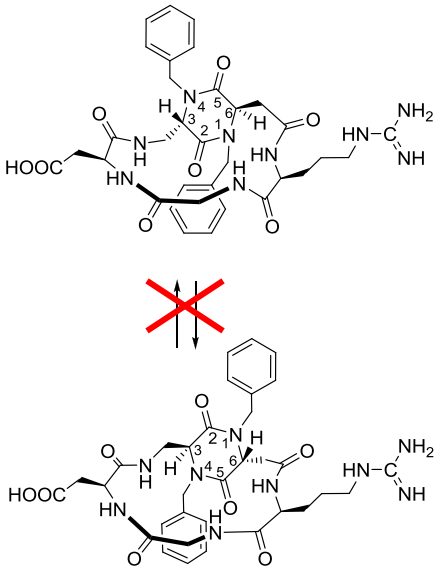


Figure 2.1.20. Compounds **Cyclo-[DKP-7-RGD]-A** and **Cyclo-[DKP-7-RGD]-B**, diastereomers of **Cyclo-[DKP-7-RGD]** due to hindered rotation.

In particular, compound **Cyclo-[DKP-7-RGD]-A** displays two mutually exclusive NOE contacts between NH_{Arg} and NH_{Gly} and between NH_{Asp} and NH_{Gly}. These three protons, on the other hand, show also a rather strong hydrogen bonded status, as indicated by their low temperature dependence and, at least for NH_{Asp} and NH_{Gly}, their upfield chemical shift (**Table 2.1.1.**). These data indicate an equilibrium between two different conformations: one displaying a Type III H-bonding pattern (β-turn at DKP-Arg, see **Figure 2.1.15.**) and a second one showing a Type II H-bonding pattern (β-turn at Arg-Gly, see **Figure 2.1.14.**), like **Cyclo-[DKP-1-RGD]**. Finally, ligand **Cyclo-[DKP-7-RGD]-B** shows a single NOE contact between NH_{Asp} and NH_{10} and a hydrogen bonded status for NH_{10} (δ 7.72 ppm and $\Delta\delta/\Delta T$ -4 ppb/K, see **Table 2.1.1.**). These values are indicative of a Type I H-bonding pattern (see **Figure 2.1.17.A** above).

Three-dimensional structures satisfying long-range NOE contacts were generated for RGD peptidomimetic **Cyclo-[DKP-7-RGD]** performing three 10 ns restrained MC/SD simulations and applying the distance restraints derived from NOESY spectra of diastereoisomers **Cyclo-[DKP-7-RGD]**-**A** and **Cyclo-[DKP-7-RGD]**-**B** (**Table 2.1.1.**): in the first simulation NH_{Arg}/NH_{Gly} relevant in **Cyclo-[DKP-7-RGD]**-**A**, in the second simulation NH_{Asp}/NH_{Gly} also relevant in **Cyclo-[DKP-7-RGD]**-**A**, and in the third simulation DKP-NH_{10}/NH_{Asp} relevant in **Cyclo-[DKP-7-RGD]**-**B**.

All the conformations sampled during the first two simulations adopted a non-extended arrangement of the RGD sequence [Cβ(Arg)-Cβ(Asp) average distance of 6.6 Å] characterized by the simultaneous presence of different turn motifs (pseudo β-turn at DKP-Arg, γ-turn at Gly and pseudo β-turn centered at the DKP unit). The structural models provided by these restrained MC/SD simulations differ from the conformations hypothesized on the basis of NMR data of **Cyclo-[DKP-7-RGD]**-**A** [equilibrium between Type III (pseudo β-turn at DKP-Arg) and Type II (β-turn at Arg-Gly) H-bonding patterns, see the previous section]. However, also the calculated structures are able to provide an explanation for the NOE contacts and the NMR temperature coefficients observed for compound **Cyclo-[DKP-7-RGD]**-**A**.

The distance restraint corresponding to the NOE contact between DKP-NH_{10} and NH_{Asp} (observed in the NOESY spectrum of high affinity ligand **Cyclo-[DKP-7-RGD]**-**B**) was applied in the third 10 ns MC/SD simulation of compound **Cyclo-[DKP-7-RGD]**. Most of the conformations sampled during this simulation adopted an extended arrangement of the RGD sequence [Cβ(Arg)-Cβ(Asp) average distance of 9.0 Å] and approximately 50% of them are characterized by a β-turn at Gly-Asp and the corresponding hydrogen bond between DKP-NH_{10} and Arg-C=O (Type I-*trans* H-bonding pattern).

Contrary to what observed for the other cyclic RGD peptidomimetics containing DKP scaffolds, rotation of the DKP ring cannot be observed during the simulations performed on compound **Cyclo-[DKP-7-RGD]**, confirming **Cyclo-[DKP-7-RGD]**-**A** and **Cyclo-[DKP-7-RGD]**-**B** as two different separable conformers (diastereomers) due to hindered rotation of one ring around the other.

In conclusion, we identified five types of conformation, namely Type I-*cis* (β-turn at Gly-Asp), Type I-*trans* (β-turn at Gly-Asp), Type II (β-turn at Arg-Gly), Type III (β-turn at DKP-Arg) and

Type IV (β-turn at Asp-DKP). The most populated ones are the Type III, adopted by the compounds **Cyclo-[DKP-2-RGD]**, **Cyclo-[DKP-3-RGD]**, **Cyclo-[DKP-5-RGD]**, and **Cyclo-[DKP-7-RGD]-A** and the Type I-*trans*, adopted by **Cyclo-[DKP-4-RGD]**, **Cyclo-[DKP-6-RGD]** and **Cyclo-[DKP-7-RGD]-B** (**Figure 2.1.21.**).

A) **B)**

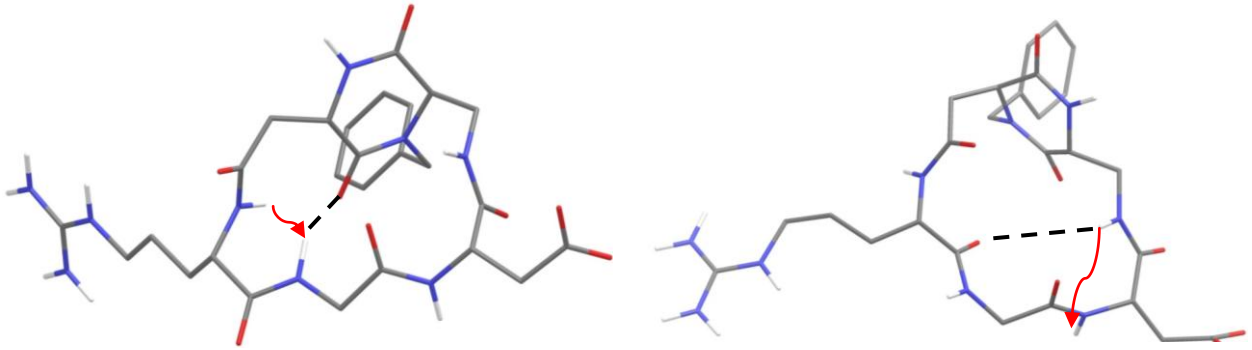


Figure 2.1.21. Representation of the two most populated conformations; A) Example of a Type III H-bonding pattern, B) Example of a Type I-*trans* H-bonding pattern. Red arrows indicate NOE contacts, while dashed black line the H-bonds.

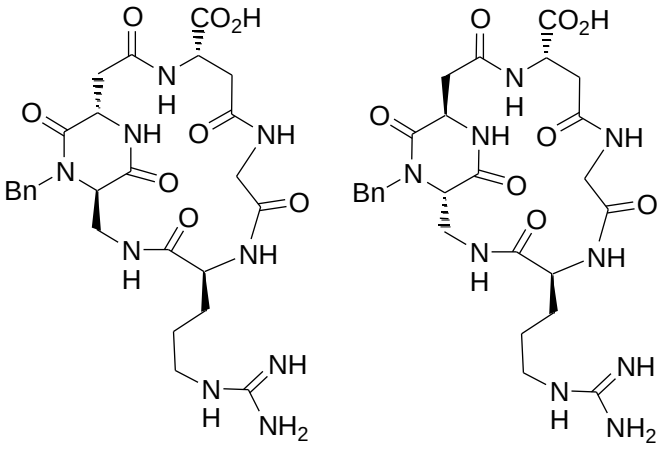
2.1.5 *iso*DGR peptidomimetics

Recent biochemical studies have shown that a spontaneous post-translational modification, occurring at the Asn – Gly – Arg (NGR) motif of the extracellular matrix protein, fibronectin, leads to the *iso*Asp-Gly-Arg (*iso*DGR) sequence.[34] This rearrangement, which involves the attack of the protein backbone NH on the primary amide side chain of asparagines and ring opening of the succinimide, is a known side reaction normally leading to loss of biological activity.[35] In this case, however, the result is a gain of protein function and the creation of a new adhesion binding site for integrins. Subsequent biochemical, spectroscopic and computational investigations have shown that the *iso*DGR sequence can fit into the RGD-binding pocket of $\alpha_V\beta_3$ integrin, establishing the same electrostatic clamp as well as additional polar interactions. Then, rapid NGR- to *iso*DGR- sequence transition in fibronectin generates $\alpha_V\beta_3$ inhibitors (named "isonectins") that competitively bind RGD binding sites. Given the structural and functional similarities between RGD and *iso*DGR, it is conceivable that peptides containing the *iso*DGR motif could also be exploited for the generation of integrin antagonists.[36,37,38]

In fact, in 2010, Kessler *et al.* created different libraries of head-to-tail-cyclized pentapeptides containing the *iso*DGR amino acid motif (*i.e.*, c(G*iso*DGRG), c(F*iso*DGRG)), and studied these constrained *iso*DGR peptides for their affinities for integrin subtypes. The first approach was to synthesize a small library of peptides based on the retrosequence of the highly active integrin-binding cyclic peptide c(-RGDfV-), for example, c(-Vf*iso*DGR-).[38]

Based on these observations, a few conformationally constrained cyclopeptides containing the *iso*DGR sequence have been synthesized by our group.

Two cyclic *iso*DGR peptidomimetics were synthesized containing the bifunctional diketopiperazine scaffolds mentioned above (**Figure 2.1.22.**) and we investigated their conformation in solution.[37]



8 = 3*R*, 6*S* **9** = 3*S*, 6*R*

Figure 2.1.22. Cyclic *iso*DGR peptidomimetics containing the bifunctional diketopiperazine scaffolds.

Although *iso*DGR- and RGD- containing ligands can share the same binding site on integrins, their effects on integrin function are not necessarily identical for the following reasons. It is well established that the interaction between integrins and ligands is regulated by conformational changes. Conformational changes and integrin activation can be transmitted bidirectionally across the membrane through interactions with intracellular molecules (“inside-out signaling”) and with extracellular ligands (outside-in signaling”). It has been suggested that

different RGD-containing ligands induce different signaling pathways. Thus, it is possible that RGD and *iso*DGR ligands induce differential conformational changes and signaling in vivo.[34] Notably, *iso*DGR docks onto $\alpha_V\beta_3$-integrin in an inverted orientation compared with RGD docking to $\alpha_V\beta_3$-integrin. This orientation allows *iso*DGR to bind the integrin α- and β-subunits through the *iso*Asp and Arg residues. The acidic and basic side chains of these residues are at the correct distance and orientation to engage stabilizing interactions with the polar regions of the integrin, thus reproducing the canonical interactions between RGD and $\alpha_V\beta_3$-integrin. Therefore, *iso*DGR is a natural fit for the RGD-binding pocket of $\alpha_V\beta_3$ integrin.[36]

It appears, therefore, that *iso*DGR, like RGD, can differentially recognize integrins depending on its molecular scaffold.

2.1.6 NMR characterization and conformational studies of *iso*DGR compounds[37]

The structure and connectivity of compounds **8** and **9** were unambiguously assigned from the mono- and bidimensional ^{1}H and ^{13}C NMR spectra (see ∗**Supplementary Information**∗). The preferred conformations of **8** and **9** were investigated by ^{1}H NMR spectroscopy of dilute H_2O/D_2O 9:1 solutions and by computational methods. One-dimensional ^{1}H-NMR spectroscopy experiments were conducted to detect intramolecular hydrogen bonds, by measuring the chemical shift of the amide N-H protons, their temperature coefficients ($\Delta\delta/\Delta T$) and their chemical exchange rate in the presence of D_2O. The results are reported in the **Table 2.1.2.**.

Ligand	**8**			**9**		
Structure	Cyclo[DKP2-isoDGR]			Cyclo[DKP3-isoDGR]		
	δ(ppm)	$\Delta\delta/\Delta$T (ppb/K)	Chemical Exchange rates	δ(ppm)	$\Delta\delta/\Delta$T (ppb/K)	Chemical Exchange rates
NH_1	8.34	-6.99	0 min	7.60	-1.70	0 min
NH_{10}	8.43	-8.57	20 min	8.46	-8.02	1h27min
NH_{Asp}	8.23	-4.90	0 min	8.63	-7.03	12 min
NH_{Gly}	7.88	-4.76	0 min	8.16	-7.20	7 min
NH_{Arg}	8.26	-6.87	5-10 min	8.27	-6.92	12-17 min

Table 2.1.2. List of the temperature coefficients, chemical shifts and chemical exchange rates of the two *cyclo*[DKP-*iso*DGR] peptidomimetics.

NOESY spectra were also recorded to investigate both sequential and long-range NOE signals that provide evidences of preferred conformations. Three-dimensional structures satisfying long-range NOE contacts were generated by restrained mixed-mode Metropolis Monte Carlo/stochastic dynamics (MC/SD) simulations,[31] using the implicit water GB/SA solvation model.[32] The data obtained from the NMR experiments show that compound **8** (**Figure 2.1.23.**) is characterized by a high conformational flexibility. In fact, the $\Delta\delta/\Delta T$ values range from −4.7 ppb K^{-1} (Gly-NH) to −8.5 ppb K^{-1} (DKP-NH10) and the N-H chemical shifts indicate that the amide protons of **8** are all solvent exposed. However, the strong NOE contact between *iso*Asp-NH and Gly-NH (**Figure 2.1.23.**, structure on the left) indicates a preferred folding through the formation of a pseudo-β-turn stabilized by an intramolecular hydrogen bond between Gly-NH and DKP-C(5)=O.

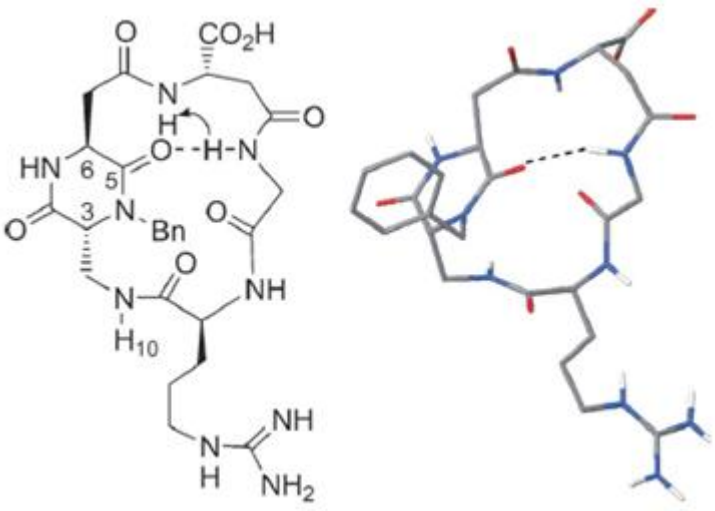


Figure 2.1.23. Left: preferred intramolecular hydrogen-bonded pattern proposed for the cyclic *iso*DGR peptidomimetic **8**, on the basis of spectroscopic data. The arrow indicates the NOE contact. Right: 3D structure of **8** as obtained by restrained MC/SD simulations based on experimental distance information, after energy minimization [pseudo-β-turn at DKP-*iso*Asp, Cβ(Arg)-COO^-(Asp) distance=8.8 Å].

The distance restraint corresponding to the NOE contact between *iso*Asp-NH and Gly-NH was applied in the MC/SD simulations of Compound **8**. Approximately 50 % of the conformations sampled during the simulations adopted an extended arrangement of the *iso*DGR sequence characterized by a pseudo-β-turn at DKP-*iso*Asp and the formation of the corresponding hydrogen

bond between Gly-NH and C(5)=O. A Cβ(Arg)-COO^-(Asp) average distance of 8.6 Å was obtained during the MC/SD calculations. A representative energy minimized conformation featuring the 11-membered ring hydrogen bond between Gly-NH and C(5)=O is shown in **Figure 2.1.23.** (structure on the right).

NMR spectroscopy experiments performed on the *iso*DGR peptidomimetic **9** suggest a different conformational preference. In fact, compound **9** exists in two preferred conformations (**Figure 2.1.24.a** and **b**). The NOESY cross-peak between DKP-NH10 and Arg-NH (strong) is indicative of a β-turn conformation stabilized by a hydrogen bond between DKP-NH10 and *iso*Asp-C=O (**Figure 2.1.24.a**), whereas the NOESY cross-peak between Gly-NH and *iso*Asp-NH (weak) is indicative of a pseudo-β-turn motif (**Figure 2.1.24.b**) similar to that observed for compound **8**.

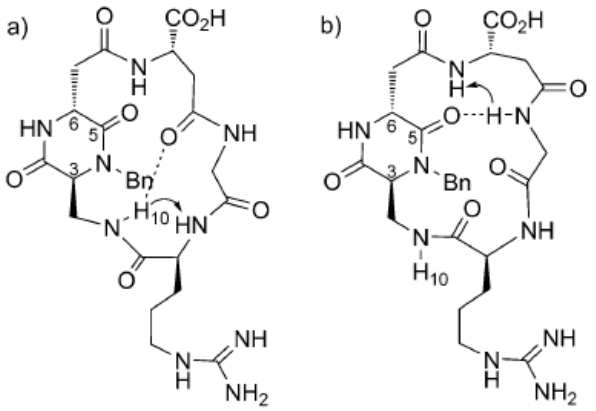


Figure 2.1.24. Preferred intramolecular hydrogen-bonded patterns proposed for the cyclic *iso*DGR peptidomimetic **9**, on the basis of spectroscopic data. The arrows indicate the NOE contacts. a) β-turn at Gly-Arg; b) pseudo-β-turn at DKP-*iso*Asp.

Only the distance restraint corresponding to the strong NOE contact between DKP-NH10 and Arg-NH (**Figure 2.1.24.b**) was applied in the MC/SD simulations of compound **9**. It is worth noting that approximately 80 % of the conformations sampled during the simulations adopted an extended arrangement of the *iso*DGR sequence, characterized by a pseudo-β-turn at DKP-*iso*Asp and the formation of the corresponding hydrogen bond between Gly-NH and C(5)=O (**Figure 2.1.25.**). In addition, the formation of a distorted β-turn at Gly-Arg (**Figure 2.1.25.**), characterized by hydrogen bond distances of about 3.5 Å between DKP-NH10 and *iso*Asp-C=O, was observed for 10 % of the simulations. A Cβ(Arg)-COO^-(Asp) average distance of 10.7 Å was obtained during the

MC/SD calculations of compound **9**. Two representative energy-minimized conformations featuring the two hydrogen bond patterns compatible with the strong NOE contact between DKP-NH10 and Arg-NH are shown in **Figure 2.1.25.**.

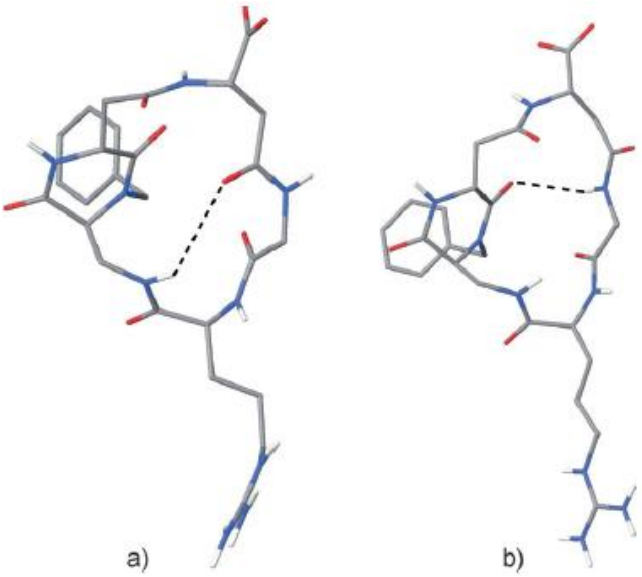


Figure 2.1.25. The 3D structures of compound **9** as obtained by restrained MC/SD simulations based on experimental distance information, after energy minimization. a) Distorted β-turn at Gly-Arg, populated for 10 % of the simulation, Cβ(Arg)-COO^-(Asp) distance=10.7 Å; b) pseudo-β-turn at DKP-*iso*Asp, populated for 80 % of the simulation, Cβ(Arg)-COO^-(Asp) distance=10.8 Å.

Cyclic *iso*DGR peptidomimetics **8** and **9** were examined in vitro for their abilities to compete with biotinylated vitronectin for binding to the purified $\alpha_V\beta_3$ receptor (**Table 2.1.3.**). Screening assays were performed by incubating the immobilized integrin $\alpha_V\beta_3$ with increasing concentrations (10^{-12} – 10^{-5} M) of the two *iso*DGR ligands in the presence of biotinylated vitronectin (1 mg mL^{-1}), and measuring the concentration of bound vitronectin in the presence of the competitive ligands.

Ligands	$\alpha_V\beta_3$ IC_{50} [nM]
8	46.7 ± 18.2
9	9.2 ± 1.1
cyclo[G*iso*DGRphg]	89 ± 19[b]
cyclo[(*3S,6R*)DKP-RGD]	4.5 ± 1.1
cyclo[RGDfV]	3.2 ± 1.3
cyclo[RGDf(N-Me)V] (Cilengitide)	0.6 ± 0.1[c]

Table 2.1.3. Inhibition of biotinylated vitronectin binding to $\alpha_V\beta_3$ receptor. [a] IC_{50} values were calculated as the concentration of compound required for 50% inhibition of biotinylated vitronectin binding as estimated by GraphPad Prism software; all values are the arithmetic mean ± SD of triplicate determinations. [b] Determined by a solid phase binding assay by using supported vitronectin, soluble $\alpha_V\beta_3$ integrin, specific primary and secondary antibodies. [c] Calculated as the concentration of compound required for 50 % inhibition of biotinylated vitronectin binding.

The IC_{50} values of compounds **8** and **9** reported in **Table 2.1.3.** compare favorably with other *iso*DGR ligands, such as the cyclopentapeptide *cyclo*[G*iso*DGRphg][38a] recently published by Kessler and co-workers, as well as with known RGD ligands, such as the cyclic peptidomimetic *cyclo*[(*3S, 6R*)DKP-RGD],[4,25] cyclopentapeptides *cyclo*(RGDfV) and Cilengitide.[21b] In particular, ligand **9** is ten-times more potent in binding integrin $\alpha_V\beta_3$ than *cyclo*[G*iso*DGRphg], only 2-3 times less potent than the RGD ligands *cyclo*[(*3S,6R*)DKP-RGD] and *cyclo*(RGDfV), and fifteen-times less potent than the powerful RGD ligand Cilengitide.

In order to rationalize, on a molecular basis, the affinity of compounds **8** and **9** for the $\alpha_v\beta_3$ receptor, docking studies were performed starting from the representative conformations obtained from the restrained MC/SD simulations. The crystal structure of the extracellular segment of integrin $\alpha_v\beta_3$ complexed with the cyclic pentapeptide Cilengitide (PDB ID: 1L5G) was taken as a reference model for the interpretation of the docking results in terms of ligand – protein interactions. In the X-ray complex, Cilengitide binds to the interface of the α and β units forming specific electrostatic interactions. The acid and basic pharmacophoric groups and their orientation are essential for binding to $\alpha_v\beta_3$, because they act like an electrostatic clamp and interact with charged regions of the receptor binding site.

Docking calculations starting from the pseudo-β-turn DKP-*iso*Asp geometries of compound **8** (**Figure 2.1.23.**, right) and from the distorted β-turn Gly-Arg conformations of compound **9** (**Figure 2.1.25.**, left), produced top-ranked poses conserving the key electrostatic interactions but lacking further stabilizing hydrogen bond interactions with the β subunit (*e.g.*, hydrogen bonds of ligand carboxylate oxygen with the backbone amides of Asn215 and Tyr122, and hydrogen bonds between ligand backbone N-H and protein backbone carbonyl groups).

On the contrary, docking calculations starting from the pseudo-β-turn DKP-*iso*Asp geometries of Compound **9** (**Figure 2.1.25.**, right) produced top-ranked poses conserving all the important interactions of the X-ray complex (**Figure 2.1.26.**).

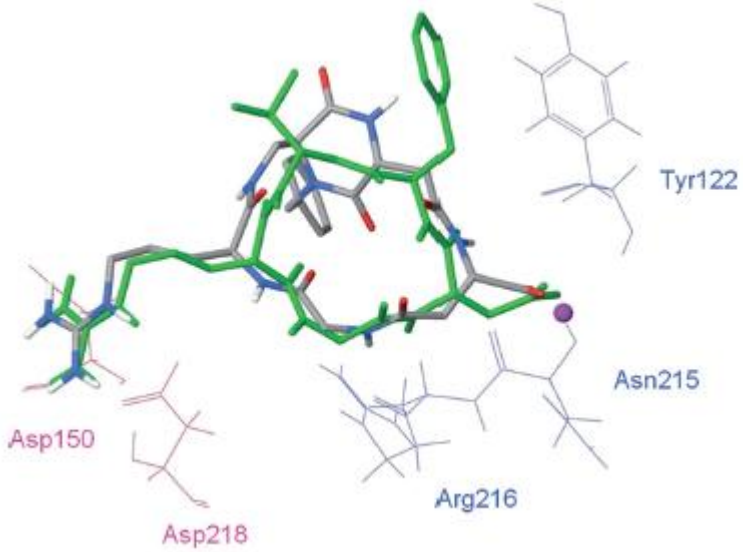


Figure 2.1.26. Best pose of compound **9** (atom color tube representation, pseudo-β-turn at DKP-*iso*Asp) into the crystal structure of the extracellular domain of $\alpha_v\beta_3$ integrin (α unit red and β unit blue wire representation) overlaid on the bound conformation of Cilengitide (green tube representation). Only selected integrin residues involved in the interactions with the ligand are shown. The metal ion in the MIDAS region is represented by a magenta sphere. Nonpolar hydrogen atoms were removed for clarity.

The positively charged Arg guanidinium group of the ligand interacts with the negatively charged carboxylates of Asp218 and Asp150 in the α unit, one carboxylate oxygen of the ligand is coordinated to the metal cation in the metal-ion-dependent adhesion site (MIDAS) region of the β unit, whereas the second carboxylate oxygen forms hydrogen bonds with the backbone amides of

Asn215 and Tyr122 in the β unit. A further stabilizing interaction involves the formation of a hydrogen bond between the ligand backbone NH of the Gly residue and the backbone carbonyl group of Arg216 in the β unit.

These docking calculations were reported here, while the docking calculations for the RGD compounds are better explained in the Section 2, for more clarity.

2.1.7 Cell lines expressing integrins

Integrins are membrane proteins and, in contrast to soluble proteins, there is little information available about ligand-receptor interactions that occur at membrane surfaces. In fact, the aggregation behaviour of these proteins is unpredictable, the production of the isolated, soluble membrane proteins is difficult and, when a truncated, soluble form of the receptor is accessible, the binding affinity of this form may differ from that of the native receptor embedded into a membrane. So, it is crucial to characterize this type of interactions working in the biophysical neighborhood of the membrane itself and not in an isotropic extracellular medium.

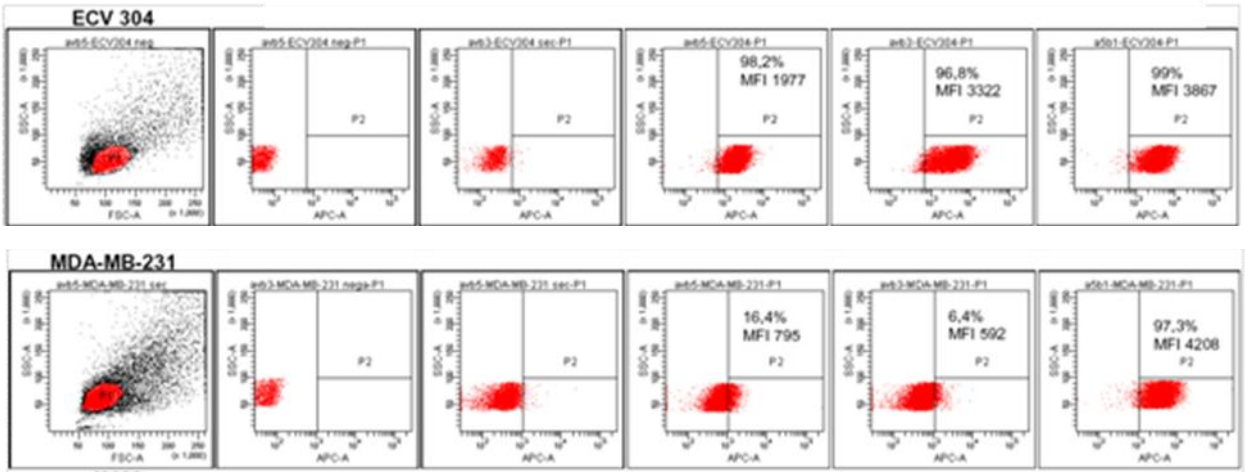
Many studies showed the application of different NMR techniques to study the interaction between some RGD-based ligands and integrins embedded in planar lipid bilayers or liposome.[39,40,41]

However, more recent studies demonstrate many differences in the interaction if evaluated into liposomes or directly in the native intact cells overexpressing integrins,[42] supporting the importance to study membrane proteins in their natural environment.

Before my thesis, numerous NMR experiments were conducted using transfected cells without giving results due to the fact that the number of the receptors expressed on the cell surface is largely variable (unpublished results). However, also the work with cells that naturally overexpressed integrins has numerous problems; in fact, the initiation and progression of cancer is accompanied by alterations in the expression profile of various adhesion receptors, including integrins. Endothelial cells in the angiogenic vessels within solid tumours express several proteins that are absent or hardly detectable in established blood vessels, including integrins and receptors for certain angiogenic growth factors. The $\alpha_V\beta_3$ integrin, for instance, is expressed at low levels on epithelial cells and mature endothelial cells; however, it is highly expressed on activated

endothelial cells in the neovasculature of tumours, including osteosarcomas, neuroblastomas, glioblastomas, melanomas, lung carcinomas, and breast cancer.[43]

Notwithstanding these problems, we found different available cell lines that naturally express integrins in different ratio. At the beginning, in particular, I set up live cell NMR methodologies by using human platelets, that express on their surface only integrin $\alpha_{IIb}\beta_3$. Platelets are cells without a nucleus, but with an active energy metabolism. As reported in the paragraphs above, in the activated state, the integrin binds to fibrinogen and then mediates thrombus formation by platelet aggregation.[42] Then, we used ECV-304 bladder and MDA-MB-231 breast cancer cell lines. In the **Figure 2.1.27.** and in **Table 2.1.2.** is reported the expression of different types of integrins in different cell lines obtained by fluorescence analysis (FACS) method, that exploits fluorescent signals emitted from individual proteins express on the cell surface.



Cell lines	MFI			Expression $\alpha_5\beta_1/\alpha_V\beta_3$
	$\alpha_V\beta_5$ integrin	$\alpha_V\beta_3$ integrin	$\alpha_5\beta_1$ integrin	
ECV-304	1977	3322	3867	1.16
MDA-MB-231	795	592	4208	7.11

Figure 2.1.27. and **Table 2.1.2.** Cytofluorimetric analysis of $\alpha_V\beta_3$, $\alpha_V\beta_5$ and $\alpha_5\beta_1$ integrin expression on human epithelial cancer cell panel; MFI= Mean Fluorescence Intensity.

Despite their epithelial origin, ECV-304 cells share many characteristics with the human endothelium, and they are widely used as endothelial-like cells. ECV-304 bladder cancer cell line,

derived from solid tumours, highly expresses $\alpha_V\beta_3$ integrin. The average number of $\alpha_V\beta_3$ receptors per ECV-304 cell was estimated to be about $4*10^3$, as well as the average number of $\alpha_5\beta_1$ receptors. Instead, for MDA-MB-231 the average number of $\alpha_5\beta_1$ receptors is about $4*10^3$, but expresses seven-fold excess of $\alpha_5\beta_1$ with respect to $\alpha_V\beta_3$ (ratio = 7.11). T98G cells express approximately a three-fold excess of $\alpha_5\beta_1$ with respect to $\alpha_V\beta_3$ (ratio = 2.75). This suggests that a comparison among these two cell lines should be a good way to investigate the selectivity of test compounds for integrins $\alpha_5\beta_1$ and $\alpha_V\beta_3$. In our next discussion, however, to evaluate the obtained NMR results we had to take into consideration the ratio of expression rather than the absolute values. In the **Figure 2.1.28.** the optical microscopy analysis of the two cell lines are reported.

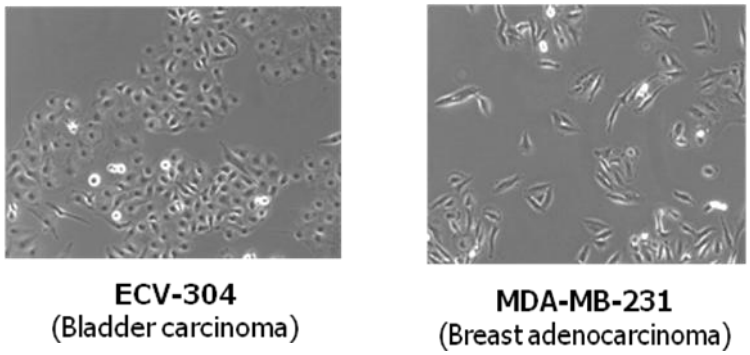


Figure 2.1.28. Optical microscopy analysis of ECV-304 and MDA-MB-231 human cell lines; Magnification 20x.

Section 2: Results and Discussion

2.2.1 INTERACTIONS WITH PLATELETS

In this part of the work, we have used NMR methodologies to shed light on the interaction of the peptidomimetic integrin ligands Cyclo-[DKP-1-RGD] (**1**), Cyclo[DKP-2-RGD] (**2**) and Cyclo[DKP-3-RGD] (**3**) with human platelets, known to express on their membrane only integrin $\alpha_{IIb}\beta_3$. Previous work demonstrated that the binding of selected RGD ligands to the integrin $\alpha_{IIb}\beta_3$ of intact human platelets can be studied by using saturation transfer double difference NMR (STDD)[42] and *tr*-NOESY[5] NMR techniques. In this thesis, STD and *tr*-NOESY NMR experiments were conducted in the presence of platelet suspension with the aim to identify the epitope and the bound conformation of three DKP-RGD ligands, and to investigate the effect of scaffold stereochemistry on ligand binding mode.

In fact, the selected DKP-RGD ligands differ only for the configuration at the two stereocenters of the DKP scaffold. The STD-NMR results were compared to docking studies into the X-ray structure of $\alpha_{IIb}\beta_3$. For compounds **2** and **3**, docking calculations showed a good disposition of the ligands into the binding site, conserving all the important interactions observed in the X-ray complex.

2.2.1.1 The docking model of $\alpha_{IIb}\beta_3$

Starting from the ligand representative conformations obtained from the MC/SD simulations, docking calculations were performed using GLIDE v 4.5.[47] Computational models for the interaction of RGD peptidomimetic ligands with the $\alpha_{IIb}\beta_3$ integrin were built using a previously developed docking approach,[4,25] starting from the crystal structure of the integrin $\alpha_{IIb}\beta_3$ headpiece bound to the cyclic heptapeptide Eptifibatide or Integrilin (Protein Data Bank entry 2VDN, see "**Experimental Section**"), which is currently in clinical use to prevent thrombosis (**Figure 2.2.1.**).[19] Integrilin is a cyclic heptapeptide mimic derived from a disintegrin containing the Lys-Gly-Asp-Trp sequence with the Lys residue converted to homoarginine. The homoArg with its longer aliphatic chain than Arg confers selectivity to integrilin for $\alpha_{IIb}\beta_3$, whereas Trp confers high affinity.

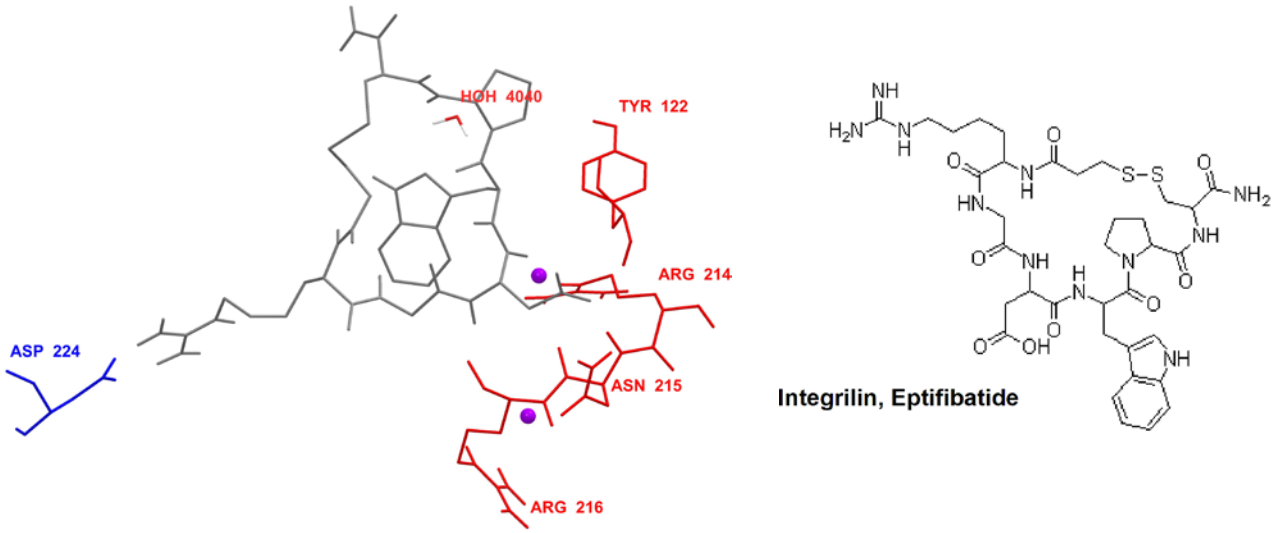


Figure 2.2.1. Left: representation of the binding site of the integrin $\alpha_{IIb}\beta_3$ (red and blue) complexed with Integrilin (gray) from PDB 2VDN. Right: structure of Eptifibatide.

In the X-ray complex, Eptifibatide binds to the interface of the α and β units forming specific electrostatic interactions. The acid and basic pharmacophoric groups and their orientation are essential for the binding because they act like an electrostatic clamp, interacting with charged regions of the receptor binding site.

Eptifibatide is disposed in an extended conformation in the receptor site (with C_β(Arg) – C_β(Asp) = 9.14 Å) placing the carboxylic and guanidine groups on the positive Ca^{2+} ion (MIDAS) in the β subunit and on the negative Asp224 residue side chain of the α subunit, respectively. Further stabilizing interactions occur between the ligand and the receptor residues. The guanidine group of the Integrilin homoArg forms a charged hydrogen bond to α_{IIb} residue Asp224 as well as a hydrogen bond to an α_{IIb} backbone carbonyl (Phe160). The carboxylate of the Integrilin Asp side chain directly coordinates to the MIDAS cation and is further stabilized by three hydrogen bonds to the β_3 backbone (Asn215, Tyr122 and Ser123). A further stabilizing interaction involves the formation of a hydrogen bond between the ligand backbone NH of the Asp residue and the backbone carbonyl group of Arg216 in the β_3 unit. In addition, there is a π-stacking interaction between the Trp moiety and the Phe160 residues of the α chain. All these interactions were used as a guide for the description and for the evaluation of the binding mode of the studied compounds.

Docking results showed that the two geometries adopted by compound **1**, namely Type I-cis and Type II H-bonding pattern, due to their short Cβ(Arg) – Cβ(Asp) distances (values lower than 8 Å) fail in forming all the interaction pattern observed for Eptifibatide. Probably the bent conformations prevent the guanidine and carboxylic groups from achieving the required distance for binding to the $\alpha_{IIb}\beta_3$. In fact, docking calculations starting from the RGD extended conformations featuring the Type III H-bonding pattern for compounds **2** and **3** (Cβ(Arg) – Cβ(Asp) distance values ≥ 8.8 Å), produced top-ranked binding modes conserving all the important interactions of the X-ray complex.

2.2.1.2 Control experiments and *tr*-NOESY results of compounds 1 – 3[4]

Compounds **1** – **3** were thus analyzed in the presence of human platelets for the first time using STD and *tr*-NOESY NMR techniques.[3,39,44] All the experiments with platelets were conducted at 298 K in a non deuterated buffer.

Negative control experiments were performed using mannose that was added to a platelet suspension containing a previously developed peptidomimetic ligand (ST1646).[5,22,23,24] Under these conditions a NOESY spectrum was recorded. The platelet membrane does not contain mannose receptors and the NOESY spectrum showed negative cross-peaks (*tr*NOEs) for the tested compound, that are with the same negative sign as the diagonal, indicating a binding situation, while positive cross-peaks were present for mannose, indicating absence of binding. This suggests that the interaction observed for the tested compound is not aspecific, but rather depends on a specific receptor on the platelet membrane. In addition, no transferred NOEs were detected in the control experiments for the free ligands at identical conditions without the presence of platelets.[5]

In addition, as many studies suggest the prominent role played by divalent cations, Ca^{2+}, as a key element in receptor activation, further negative controls were conducted in the presence and absence of Ca^{2+} ions (unpublished results). In the absence of Ca^{2+} ions, the integrin receptors are not activated and we did not have any evidence of binding both from the *tr*-NOESY and from STD experiments. In fact, *tr*-NOESY experiments, conducted in the presence of ST1646 compound and without Ca^{2+} ions indicated the absence of binding, with positive cross-peaks, and we did not observe STD signals in the STD experiment.

Transferred intramolecular NOEs are well established as unique sources of structural information on bound ligand. The technique, fully described in the **Chapter 1**, relies on the size-dependence of the intramolecular NOE, which shows slow NOE build-up with weak positive maxima for free ligands and rapid build-up with strong negative maxima for the bound state. If dissociation of the ligand occurs quickly enough ($K_d > 10^{-7}$ M), a sufficient percentage of the observable free ligand will retain intense negative *tr*NOE and thus binding of a ligand to a receptor protein can be easily detected by looking at the sign and size of the observed NOEs.

Compound **1** in the free-state revealed the presence of two main conformations (see paragraph "**NMR Characterization and Conformational Analysis**", **Chapter 1**): a Gly-Asp β-turn conformation (Type I-*cis*) stabilized by a hydrogen bond between DKP-NH10 and Arg-C=O (**Figure 2.2.2.a**), and an Arg-Gly β-turn conformation (Type II) stabilized by a hydrogen bond between NH_{Asp} and C(8)=O (**Figure 2.2.2.b**).

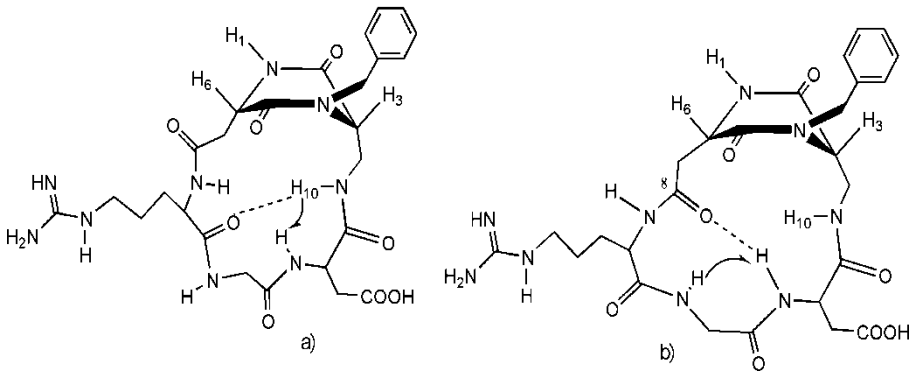


Figure 2.2.2. Preferred free state conformations of *cis*-DKP-RGD-peptidomimetic. Arrows indicate significant NOE contacts and dotted lines the hydrogen bonding patterns stabilizing β-turn motifs: (a) Gly-Asp β-turn; (b) Arg-Gly β-turn.

The weakness of the NH- NOE contacts in the *tr*-NOESY experiments does not allow considerations about the conformational equilibrium that characterized the compound in the bound-state. However, the *tr*-NOESY experiment of compound **1** in the presence of platelets at 298 K showed negative cross-peaks (*i.e.*, with the same negative sign as the diagonal) in contrast to the positive ones found in the free-state, which gave evidence of binding. Probably, the conformational equilibrium of compound **1** is unalterated by the binding to platelets.

Compound **2** is characterized by a DKP-Arg β-turn conformation (Type III) stabilized by a hydrogen bond between NH_{Gly} and C(5)=O (**Figure 2.2.3.**), as suggested by the relevant long-range NOE interaction between NH_{Gly} and NH_{Arg} (positive at 298 K).

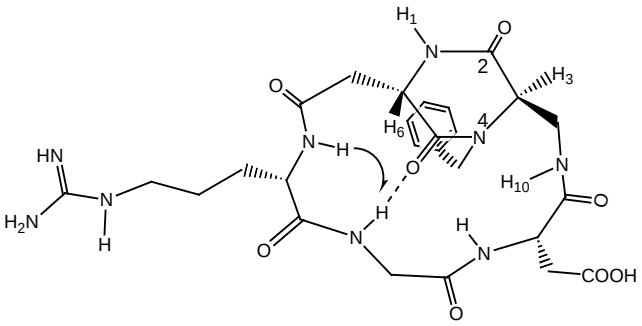
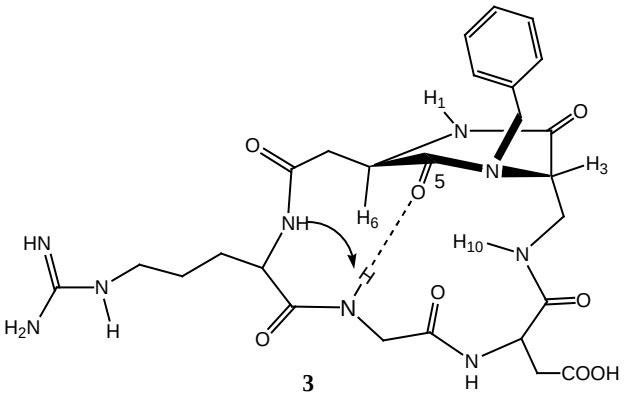


Figure 2.2.3. Free and bound conformation of *trans*-DKP-RGD-peptidomimetic **2**. Hydrogen bond (dotted line) and NOE contact (arrow) are indicated.

Also in this case, the weakness of the NH- NOE cross-peaks does not allow considerations about the conformation adopted by compound **2** in the bound-state; however, the presence of negative cross-peaks in the *tr*-NOESY experiment gave evidence of binding. In addition, the cross-peaks between the aromatic protons and the Arg-side chain protons, which can be observed in the NOESY experiments (free ligand), are lacking in the *tr*-NOESY spectrum. The absence of these cross-peaks in the bound state suggests an extended conformation of the Arg-side chain, suitable for the constitution of the electrostatic clamp inside the binding pocket.

Also compound **3** adopts, in the free state, mainly one conformation characterized by the presence of a β-turn stabilized by a hydrogen bond between NH_{Gly} and C(5)=O (Type III, **Figure 2.2.4.**), and by an extended arrangement of the RGD sequence [C_β(Arg) – C_β(Asp) distance of about 9 Å]. NOESY experiments showed a long range contact (positive) between NH_{Gly} and NH_{Arg}, confirming this conformation.



3

Figure 2.2.4. Free and bound conformation of *trans*-DKP-RGD-peptidomimetic **3**. Hydrogen bond (dotted line) and NOE contact (arrow) are indicated.

The *tr*-NOESY experiment does not permit to evaluate the bound conformation because the cross-peaks intensity is close to zero and the only detectable cross-peaks were related to the platelet membrane. However, in the free-state the cross-peaks are positive and not close to zero, this variation suggests that a binding event happened, even if the cross-peaks do not have the same negative sign as the diagonal.

2.2.1.3 STD-NMR and molecular docking results of compounds 1 - 3

Then, STD experiments[5] were performed in the presence of the platelet suspension for all the three compounds at 298 K. The STD sequence helps identifying the group epitope, revealing which moieties of the ligand molecule are closest to the receptor in the bound state. The method, fully described in Chapter 1, is based on the transfer of saturation from the protein to the bound ligand which, by exchange, is released into solution where it is detected. The degree of saturation of individual ligand protons (expressed as absolute-STD percent) reflects their proximity to the protein surface and can be used to describe the ligand-target interactions.

Compound 1

The STD-NMR experiment on compound **1** displayed signals for the guanidine NH, DKP-NH1 and the aromatic protons (1.0, 0.6 and 0.3% STD absolute intensity, **Figure 2.2.5.**, blue bars). These

results highlight that this compound interacts through its basic moiety with the α_{IIb} subunit of integrin $\alpha_{IIb}\beta_3$.

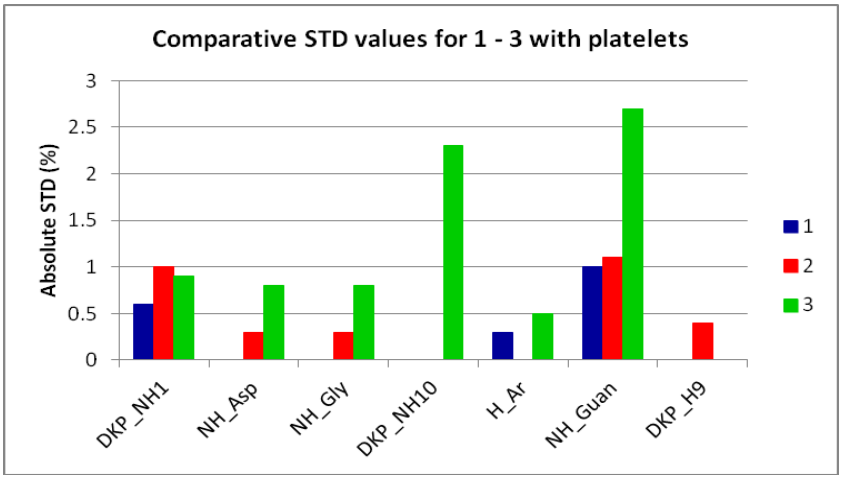


Figure 2.2.5. Relevant absolute STD percentages for the protons of cyclic DKP-RGD **1** - **3** obtained at T = 298 K in the presence of a platelet suspension: compound **1** (blue bars), compound **2** (red bars), compound **3** (green bars).

Docking calculations into the $\alpha_{IIb}\beta_3$ integrin starting from the two conformations of compound **1**, namely Type I-cis and Type II, showed poses failing in forming all the interaction pattern observed for Eptifibatide. Probably these bent conformations (Cβ(Arg) – Cβ(Asp) distances values lower than 8 Å) prevent the guanidine and carboxylic groups to achieve the required distance for binding to the $\alpha_{IIb}\beta_3$. In fact, for compound **1** the interaction mainly occurs through its basic moiety (guanidine NH) with the α_{IIb} subunit, for which we observed the largest STD effect. We observed STD effects also for aromatic and DKP-NH_1 protons of the scaffolds but not for other protons of the RGD sequence, confirming a non-optimal or weak interaction of **1** with the receptor.

Compound 2

In the STD experiment of compound **2** in the presence of platelets, the protons involved in the interaction belong to the DKP scaffold (DKP-NH_1 and DKP-H9, 1.0 and 0.4% STD absolute intensity, respectively) as well as to the RGD sequence (NH_{Gly}, NH_{Asp} and the guanidine NH, 0.3, 0.3

and 1.1% STD absolute intensity, respectively), while the aromatic protons do not appear in contact with the protein (**Figure 2.2.5.**, red bars).

Docking calculations, starting from the preferred RGD extended conformation of compound **2** (**Figure 2.2.3.**) obtained from previous MC/SD simulations, produced top-ranked binding poses conserving all the important polar interactions of the X-ray complex with Eptifibatide. The guanidine group of the ligand Arg forms a charged hydrogen bond to α_{IIb} residue Asp224 as well as a hydrogen bond to an α_{IIb} backbone carbonyl (Phe160 or Tyr189). The carboxylate of the ligand Asp side chain directly coordinates to the MIDAS cation and is further stabilized by three hydrogen bonds to the β_3 backbone (Asn215, Tyr122 and Ser123) and one to the β_3 Asn215 side chain. A further stabilizing interaction involves the formation of a hydrogen bond between the ligand backbone NH of the Asp residue and the backbone carbonyl group of Arg216 in the β_3 unit. Moreover, favourable van der Waals contacts can be detected between DKP-NH_1 and the Hβ of Ser123, between DKP-H9 and the aromatic protons of β_3-Tyr122, and between NH_{Gly} and the methyl protons of β_3-Ala218. So, docking results are consistent with the STD data showing the RGD sequence in close contact to the receptor.

Compound 3

In the case of compound **3**, the largest STD effect is again observed for the guanidine NH (2.7% STD absolute intensity, see **Figure 2.2.5.**, green bars); hence this proton closely approaches the α_{IIb} subunit of the protein. Remarkably, also DKP-NH10, DKP-NH_1, NH_{Gly}, NH_{Asp} and the aromatic protons of this ligand show quite intense STD effects (2.3, 0.9, 0.8, 0.8 and 0.5%, respectively) and, as observed for **2**, also compound **3** can be considered as lying close to the protein surface. The epitopes of the compounds **2** and **3** appear to be quite similar, except for the aromatic protons that are not present for compound **2**.

Also for compound **3**, the STD-NMR results in the presence of platelets can be nicely interpreted by the docking poses of the ligand into the $\alpha_{IIb}\beta_3$ integrin binding site. In fact, docking studies starting from the extended conformation (**Figure 2.2.4.**) obtained from MC/SD simulations, produced top-ranked binding modes conserving all the key RGD interactions of the X-ray complex between $\alpha_{IIb}\beta_3$ integrin and Eptifibatide (**Figure 2.2.6.**).

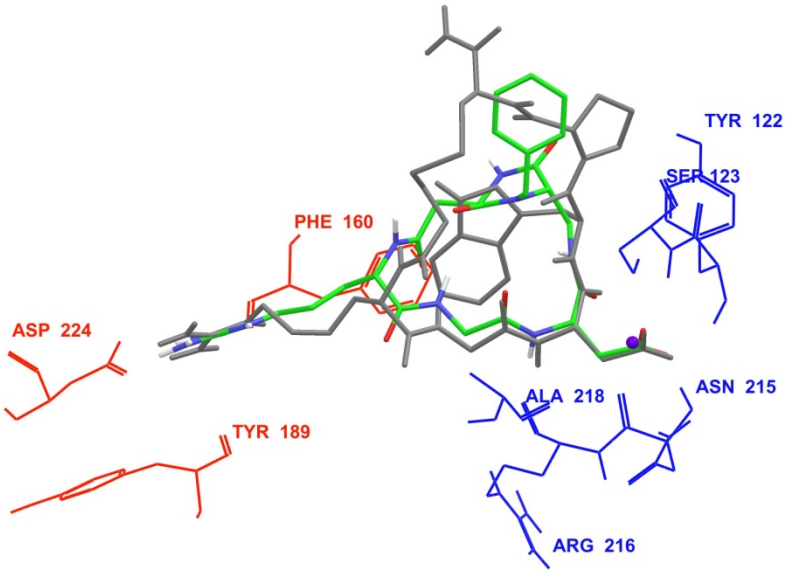


Figure 2.2.6. Docking best pose of *trans*-DKP-RGD-peptidomimetic **3** (tube representation, C atoms in green, N in blue, O in red, H in white) into the crystal structure of the integrin $\alpha_{IIb}\beta_3$ headpiece (α unit red and β unit blue wire representation) overlaid on the bound conformation of Eptifibatide (gray tube representation). Only selected integrin residues involved in the interactions with the ligand are shown. The metal ion at MIDAS is shown as a magenta CPK sphere. For the sake of clarity, all H atoms bound to carbon are omitted.

Compared to compound **2**, for compound **3** favourable van der Waals contacts can be detected between the ligand benzyl group and the aromatic moiety of β_3-Tyr122 and the Hα of Ser123 (**Figure 2.2.7.**).

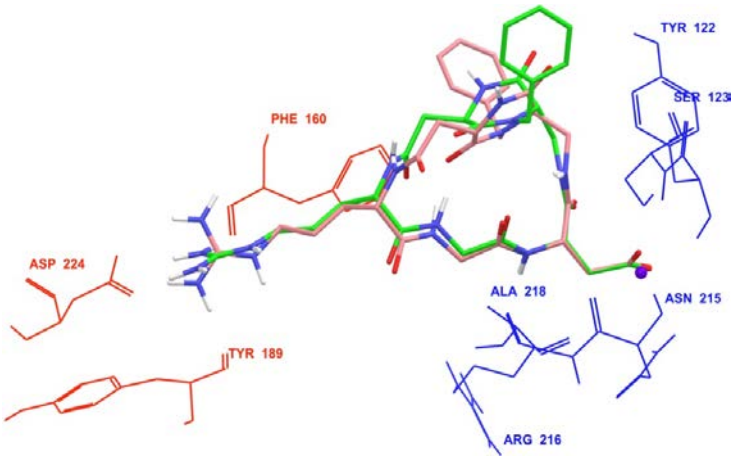


Figure 2.2.7. Docking best poses of *trans*-DKP-RGD-peptidomimetics **2** and **3** (pink and green atom colour tube representation, respectively) into the crystal structure of the integrin $\alpha_{IIb}\beta_3$ headpiece (α unit red and β unit blue wire representation).

The superimposition of the best poses of **2** and **3** into the binding site showed a similar disposition of the two compounds, perfectly overlapped in the RGD sequence, but different in the benzyl group which points into opposite direction (**Figure 2.2.7.**). According to the absence of the STD effect for the aromatic protons of compound **2**, in the docking poses no residue of the protein is in contact with the ligand benzyl group.

2.2.2 INTERACTIONS WITH ECV-304 BLADDER CANCER CELLS

Starting from the encouraging results obtained with platelets, we applied the same NMR techniques to study the interaction of DKP-RGD ligands **1** - **3** with integrin $\alpha_V\beta_3$. To this end, we performed our experiments on the ECV-304 bladder cancer cells, known to overexpress on their membrane integrin $\alpha_V\beta_3$ (see paragraph "**Cell lines expressing integrins**"). We then expanded our investigation considering also the DKP-RGD ligands **4** - **7**. Aware of the fact that the ECV-304 cell line expresses a panel of different integrins, namely $\alpha_V\beta_3$, $\alpha_5\beta_1$ and $\alpha_V\beta_5$, we also performed experiments on MDA-MB-231 breast cancer cells, known to overexpress mainly integrin $\alpha_5\beta_1$.

For clarity, I will discuss first the results obtained with ECV-304 cells for compounds **1** - **3** and I will compare them with those obtained with platelets. Then, I will show the results of ligands **4** - **7** with ECV-304 cells and the comparison with the data achieved with MDA-MB-231 cells.

2.2.2.1 The docking model of $\alpha_V\beta_3$

A computational model for the interaction with the $\alpha_V\beta_3$ integrin was built starting from the X-ray structure of the extracellular segment of integrin $\alpha_V\beta_3$ complexed with the cyclic pentapeptide ligand Cilengitide (Protein Data Bank entry 1L5G) (**Figure 2.2.8.**).[18] In the X-ray complex, Cilengitide binds to the interface of the α and β units forming specific electrostatic interactions. The acid and basic pharmacophoric groups and their orientation are essential for the binding because they act like an electrostatic clamp, interacting with charged regions of the receptor binding site.

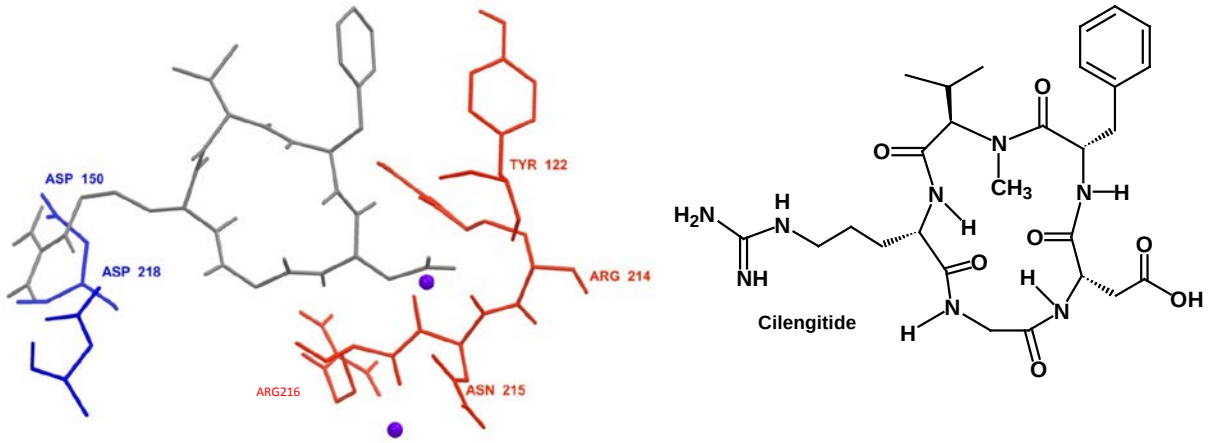


Figure 2.2.8. Left: representation of the binding site of the integrin $\alpha_V\beta_3$ (red and blue) complexed with Cilengitide (gray) from PDB 1L5G. Right: structure of Cilengitide.

Cilengitide is disposed in an extended conformation (C_β(Arg) – C_β(Asp) = 8.87 Å) in the receptor site placing the carboxylic and guanidine groups respectively on the β (Ca^{2+}) and α chain (residues Asp150 and Asp218). The stabilization of the compound inside the receptor site occurs also with other types of interactions, such as H-bonding (with the residues Arg216, Asn215 and Tyr122 of the β subunit) and hydrophobic interactions (for example with Tyr122 of the β subunit). All these interactions were used as a guide for the description and for the evaluation of the binding mode of the studied compounds. In particular, the positively charged Arg guanidinium group of the ligand interacts with the negatively charged side chains of Asp218 and Asp150 in the α unit, one carboxylate oxygen of the ligand Asp side chain is coordinated to the metal cation in the metal-ion-dependent adhesion site (MIDAS) region of the β unit, whereas the second

carboxylate oxygen forms hydrogen bonds with the backbone amides of Asn215 and Tyr122 in the β unit. Further stabilizing interaction involves the formation of a hydrogen bond between the ligand backbone NH of the Asp residue and the backbone carbonyl group of Arg216 in the β unit. Favourable aromatic ring interactions between the ligand benzylic groups and the β_3-Tyr122 side chain could also be observed. The X-ray structures of $\alpha_V\beta_3$ and $\alpha_{IIb}\beta_3$ complexed with RGD ligands have revealed an identical atomic basis for the interaction: RGD binds to the interface of the α and β subunits, the Arg residue fitting into a cleft in the α subunit, and the Asp coordinating a cation in the β subunit. The RGD tripeptide adopts a highly extended conformation across the $\alpha_V\beta_3$ and $\alpha_{IIb}\beta_3$ integrin intersubunit interface. The Arg and Asp side chains extend in opposite directions, and the backbone in between is also extended. Looking at the distinguishing features, the basic ligand side chain must reach further into the deeper β propeller pocket of α_{IIb} to interact with Asp224, whereas the residues Asp150 and Asp218 are nearer in the shallower α_V binding pocket.

2.2.2.2 In vitro biological assays of compounds 1 - 3

We can take advantage also of the presence of biological assays both on the isolated receptors ($\alpha_V\beta_3$ and $\alpha_V\beta_5$) and on the cell line.

Integrin receptor competitive binding assays

The cyclic RGD peptidomimetics **1** - **3** were examined *in vitro* for their ability to inhibit biotinylated vitronectin binding to the purified $\alpha_v\beta_3$ and $\alpha_v\beta_5$ receptors (**Table 2.2.1.**).

Structures	**$\alpha_v\beta_3$ IC$_{50}$ [nM][a]**	$\alpha_v\beta_5$ IC$_{50}$ [nM][a]
cyclo[**DKP-1**-RGD]	**3898 ± 418**	> 10^4
cyclo[**DKP-2**-RGD]	**3.2 ± 2.7**	114 ± 99
cyclo[**DKP-3**-RGD]	**4.5 ± 1.1**	149 ± 25
ST1646[22,24]	**1.0 ± 0.5**	1.4 ± 0.8
cyclo[RGDfV][45]	**3.2 ± 1.3**	7.5 ± 4.8
cyclo[RGDf(N-Me)V][21]	**0.6 ± 0.1**	11.7 ± 1.5

Table 2.2.1. Inhibition of biotinylated vitronectin binding to $\alpha_v\beta_3$ and $\alpha_v\beta_5$ receptors. [a] IC$_{50}$ values were calculated as the concentration of compound required for 50% inhibition of biotinylated

vitronectin binding as estimated by GraphPad Prism software; all values are the arithmetic mean ± SD of triplicate determinations.

Screening assays were performed incubating the immobilized integrin receptors with various concentrations (10^{-12}–10^{-5} M) of the RGD ligands **1**, **2** and **3** in the presence of biotinylated vitronectin (1 μg/mL), and measuring the concentration of bound vitronectin in the presence of the competitive ligands (see "Experimental Section"). The ability of the new compounds to inhibit the binding of vitronectin to the isolated $\alpha_v\beta_3$ receptor was compared with that of the reference compounds c(RGDfV) and ST1646. The results are collected in **Table 2.2.1.**. Low nanomolar IC_{50} values were obtained for all the ligands containing *trans* DKP scaffolds, and a micromolar IC_{50} value was obtained for ligand **1** derived from the *cis* DKP1 scaffold. These data further demonstrate that the extended RGD sequence forced by the *trans* diketopiperazine scaffolds of compounds **2** and **3** (compared to the folded RGD sequence due to the *cis* scaffold of compound **1**) gives rise to a better pre-organization of the cyclic peptidomimetics for binding to the integrin receptors. Interestingly, unlike reference compounds, the RGD peptidomimetics were about 10 – 500-fold more selective for the $\alpha_V\beta_3$ integrin with respect to the $\alpha_V\beta_5$ in this kind of assay.

Cell adhesion assays

To assess the activity of the compounds as integrin antagonists on a cell model, cell adhesion experiments were performed using ECV-304 bladder cancer cells. Cells were allowed to adhere to the immobilized physiological ligand, vitronectin (VN, the highest affinity ligand for $\alpha_V\beta_3$ integrin) or fibronectin (FN, the highest affinity ligand for $\alpha_5\beta_1$ integrin), in the presence of increasing concentration of the tested compound (see "Experimental Section"). Adhesion assay is a cell biology test extensively used to determine the ability of compounds to compete with the adhesion of a cell line to physiological ligand.

Ligand	ECV-304	
	Vitronectin IC_{50} [μM][a]	Fibronectin IC_{50} [μM][a]
cyclo[DKP-1-RGD]	> 200	> 200
cyclo[DKP-2-RGD]	13.8 ± 0.2	2.2 ± 0.4
cyclo[DKP-3-RGD]	49.5 ± 3.0	15.0 ± 3.9
cyclo[RGDfV]	0.22 ± 0.00	0.36 ± 0.09

Table 2.2.2. Effect of DKP-RGD compounds on ECV-304 cell adhesion to vitronectin and fibronectin. [a] Each data point was performed in triplicate in two independent experiments.

As shown in **Table 2.2.2.**, compound **1** was not able to impair the adhesion of the ECV-304 cell line on either vitronectin or fibronectin, thus confirming its low affinity for the $\alpha_V\beta_3$ integrin. In contrast, both cyclic peptidomimetics **2** and **3** (containing *trans* DKP scaffolds) significantly inhibited cell adhesion to either fibronectin or vitronectin, exhibiting a micromolar anti-adhesive activity.

The micromolar activities obtained in the adhesion assays represent an ideal range for testing interactions between the ligands and the integral membrane protein of intact cells by means of *tr*-NOE and STD techniques.

2.2.2.3 *tr*-NOESY results of compounds 1 - 3

The NOE experiments for compound **1** in phosphate buffer at 298 K showed positive NOE contacts. The *tr*-NOESY experiment of compound **1** in the presence of ECV-304 cells at 298 K also showed positive cross-peaks, which indicate absence of binding. Lowering the temperature to 282 K, the NOESY spectrum of the free ligand **1** showed cross-peaks with an inverted phase, *i.e.* with the same negative sign as the diagonal. The *tr*-NOESY spectrum performed at 282 K showed a negative cross-peak between NH_{Gly} (8.82 ppm) and NH_{Asp} (8.13 ppm); this NOE contact is consistent with the β-turn conformation at Arg-Gly stabilized by the hydrogen bond between NH_{Asp} and C(8)=O mentioned above and depicted in **Figure 2.2.9.**. This would indicate that, upon binding to cells, the conformational equilibrium present in the free state is shifted towards one conformation, and precisely to the less abundant in the free state.

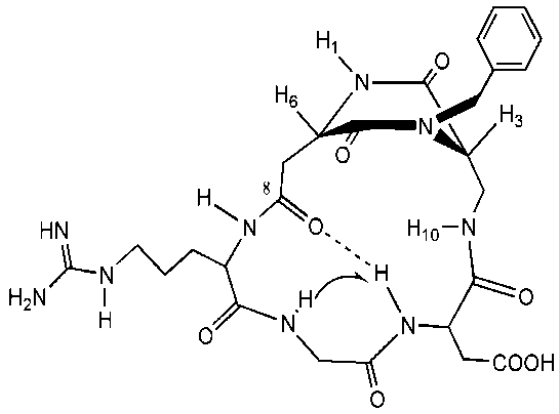


Figure 2.2.9. Preferred bound state conformation of *cis*-DKP1-RGD-peptidomimetic. Arrows indicate significant NOE contacts and dotted lines the hydrogen bonding patterns stabilizing β-turn motifs: Arg-Gly β-turn (Type II, bound conformation).

In the *tr*-NOESY experiments of compound **2** at 298 K the cross-peak intensity is close to zero, while at 282 K the contacts are negative like in the free state. However, as described above, the NOESY spectrum at 298 K showed positive cross-peaks in contrast to what observed in the *tr*-NOESY experiment, where the cross-peak intensity is close to zero; this variation is enough to demonstrate the binding event, even if the cross-peaks did not adopt in the *tr*-NOESY an inverted phase. In addition, the weakness of the NH_{Gly}/NH_{Arg} cross-peak does not allow considerations about the bound conformation, but all the other cross-peaks are overlapping with those of the free state, suggesting the maintenance of the conformation characterized by a β-turn at DKP-Arg also in the bound state (Type III, **Figure 2.2.10.**).

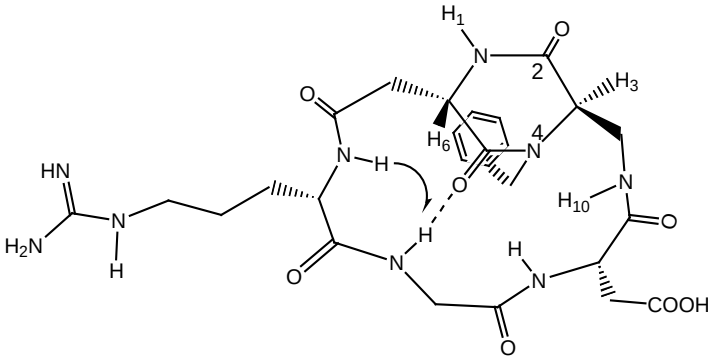
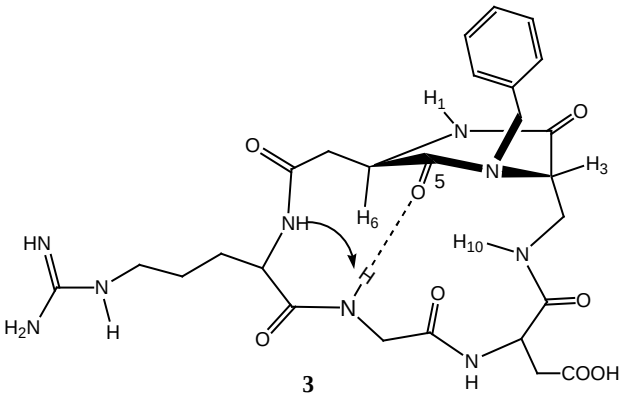


Figure 2.2.10. Bound conformation of trans-DKP2-RGD-peptidomimetic **2**. Hydrogen bond (dotted line) and NOE contact (arrow) are indicated.

For compound **3**, the NOESY experiments at 282 K in the free-state showed a long range contact (positive) between NH_{Gly} and NH_{Arg}, suggesting a DKP-Arg β-turn conformation (Type III). The *tr*-NOESY experiments of compound **3** at 282 K showed negative cross-peaks (*i.e.*, with the same phase as the diagonal), indicating a binding situation. However, only a very weak NOE contact between NH_{Gly} and NH_{Arg} is observed in the *tr*-NOESY spectrum at 282 K, indicating the maintenance of the free state conformation, Type III (**Figure 2.2.11.**), also in the bound state.



3

Figure 2.2.11. Bound conformation of *trans*-DKP3-RGD-peptidomimetic **3**. Hydrogen bond (dotted line) and NOE contact (arrow) are indicated.

Moreover, the cross-peaks between the aromatic protons and the Arg side chain protons, which can be observed in the NOESY experiments (free ligand), are lacking in the *tr*-NOESY spectrum (**Figure 2.2.12.**). These cross-peaks suggest folding of the Arg-side chain, which does not occur in the bound ligand. In fact, the formation of the electrostatic clamp inside the binding pocket requires that the side chains of Arg and Asp lie in an extended conformation.

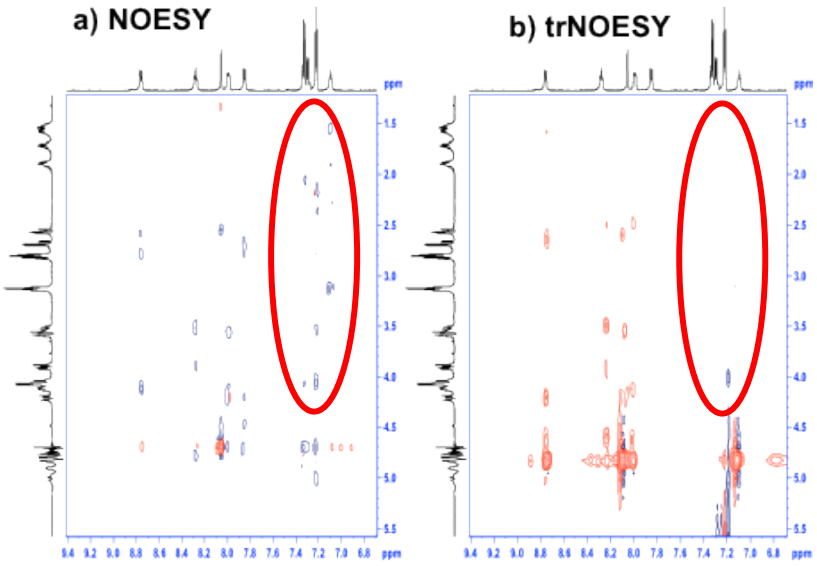


Figure 2.2.12. Left: NOESY spectrum of compound **3** in non-deuterated buffer. Right: *tr*-NOESY spectrum of compound **3** in a cell suspension. Experiments performed at T = 282 K. The aromatic region is highlighted with a red circle.

STD-NMR experiments were performed in the presence of ECV-304 cells at 298 K and at 282 K. Temperature is an important parameter for the sensitivity of STD spectroscopy; it strongly influences the kinetics and the affinity constant of protein-ligand complex formation and, doing so, it affects the observed STD signals.

2.2.2.4 STD-NMR and molecular docking results of compounds 1 – 3

Compound 1

For compound **1**, STD-NMR experiments performed at 298 K did not give satisfactory results, whereas at 282 K the region of the spectrum corresponding to the aromatic and amide protons is well resolved and can be used to measure the absolute intensity of STD (**Figure 2.2.13.**, blue bars). The largest STD effects (1.3% STD absolute intensity) were observed for the guanidine NH protons (7.26 ppm), which are expected to be in contact with the protein. Also the aromatic protons and the DKP-NH_1 showed a STD effect (0.4 and 0.8% STD absolute intensity, respectively) and consequently can be considered lying close to the protein surface. A very small STD effect is observed for NH_{Asp} (0.2% STD absolute intensity).

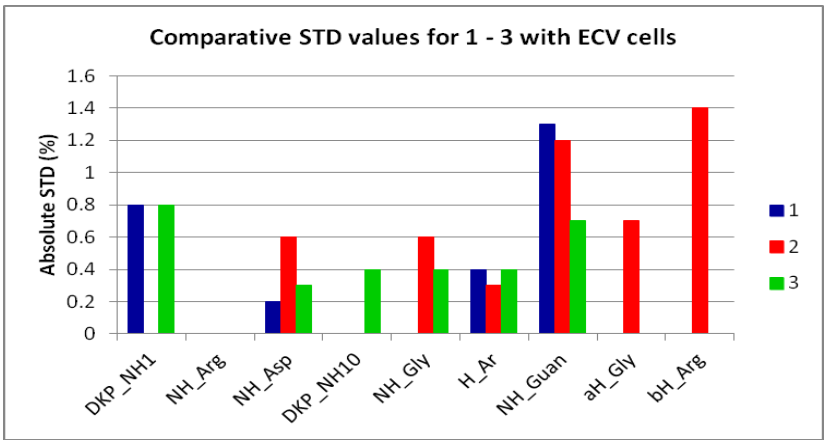


Figure 2.2.13. Relevant absolute STD percentages for the protons of cyclic DKP-RGD **1** – **3** obtained at T = 282 K in the presence of ECV-304 cells: compound **1** (blue bars), compound **2** (red bars), compound **3** (green bars).

As previously reported for the integrin $\alpha_{IIb}\beta_3$, NMR data were compared with docking results into the $\alpha_v\beta_3$ model. Docking studies starting from the bound conformation of compound **1** (Type II, **Figure 2.2.9.**) produced top-ranked poses conserving optimal interactions only between the positively charged Arg guanidinium group of the ligand and the negatively charged side chains of Asp218 and Asp150 in the α unit (**Figure 2.2.14.**).

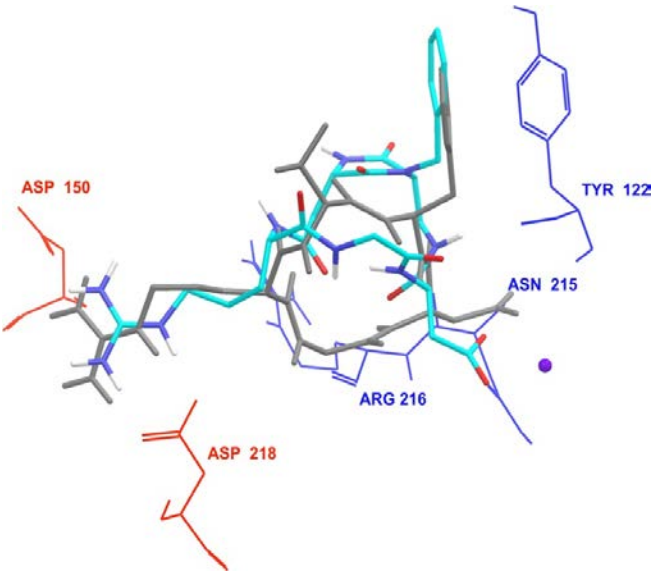


Figure 2.2.14. Docking best pose of *cis*-DKP1-RGD-peptidomimetic **1** (tube representation, C atoms in light blue, N in blue, O in red, H in white) into the crystal structure of the extracellular domain of $\alpha_v\beta_3$ integrin (α unit red and β unit blue wire representation) overlaid on the bound conformation

of Cilengitide (gray tube representation). Only selected integrin residues involved in the interactions with the ligand are shown. The metal ion at MIDAS is shown as a magenta CPK sphere. For the sake of clarity, all H atoms bound to carbon are omitted.

Probably, the short Cβ(Arg)-Cβ(Asp) distance (< 8 Å) of this geometry does not allow the simultaneous superimposition of the guanidine and carboxylic groups of the ligand with the corresponding moieties of Cilengitide. In particular, the stabilizing hydrogen bond interactions of Cilengitide carboxylate oxygen with the backbone amides of Asn215 and Tyr122 in the β unit, and the hydrogen bonds between Cilengitide backbone N-H and protein backbone carbonyl groups in the β unit are missing in the poses of compound **1**. The STD data indicated that the RGD sequence of compound **1** is not in close contact with the protein residues. Experimental and docking results are in agreement showing a non-optimal disposition of the RGD sequence into the binding site. Furthermore, also the STD effects observed for the aromatic protons and the DKP-NH_1 can be nicely interpreted on the basis of docking poses. In fact, the ligand benzyl group is suitably positioned to interact with the aromatic moiety of the β_3-Tyr122, and the DKP-NH_1 shows favourable van der Waals contacts with the β_3-Arg214 side chain (**Figure 2.2.14.**).

Compound 2

STD-NMR experiments for compound **2** were performed at two temperatures, 298 and 282 K. The experiment at 298 K showed quite large STD effects for the guanidine NH, NH_{Asp} and NH_{Gly} (2.8, 2.0 and 2.1 % STD absolute intensity, respectively), also some aliphatic protons are present, but their contributions are not quantifiable. When the temperature was lowered to 282 K, a decreased intensity of STD signals was observed for the guanidine NH, NH_{Asp} and NH_{Gly} protons (1.2, 0.6 and 0.6 % STD absolute intensity, respectively), but we can quantify the contributions of aliphatic protons. Also DKP-NH_1 is observable, but its contribution was not considered because the STD signal is present also in the control experiment. Additionally, at 282 K, the signals for the aromatic protons, $H\alpha_{Gly}$ and $H\beta_{Arg}$ (0.3, 0.7 and 1.4 % STD absolute intensity, respectively) appeared in the STD spectrum (**Figure 2.2.13.**, red bars). The absence of the aromatic protons at 298 K as well as their low STD effect at 282 K are worth noting .

Docking calculations into the $\alpha_v\beta_3$ integrin binding site, starting from the RGD extended conformation [Cβ(Arg)-Cβ(Asp) distance of about 9 Å] of compound **2** (Type III, **Figure 2.2.10.**) produced top-ranked binding modes conserving a good correspondence with the receptor-bound structure of Cilengitide. The positively charged guanidinium group of the ligand interacts with the negatively charged carboxylates of Asp218 and Asp150 in the α unit; one carboxylate oxygen of the ligand is coordinated to the metal cation in the metal-ion-dependent adhesion site (MIDAS) region of the β unit, while the second carboxylate oxygen forms hydrogen bonds with the backbone amides of Asn215 and Tyr122 in the β unit. A further stabilizing interaction involves the formation of a hydrogen bond between the ligand backbone NH of the Asp residue and the backbone carbonyl group of Arg216 in the β unit (**Figure 2.2.15.**). Even if all the important polar interactions of the X-ray complex are maintained, the phenyl ring does not interact with the aromatic moiety of β_3-Tyr122, but is placed between the guanidinium group of the β_3-Arg214 and the aromatic moiety of α_V-Tyr178.

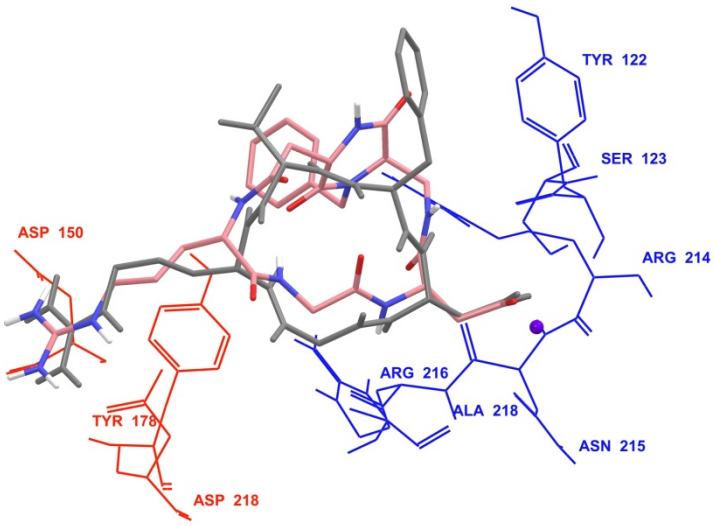


Figure 2.2.15. Docking best pose of *trans*-DKP2-RGD-peptidomimetic **2** (tube representation, C atoms in pink, N in blue, O in red, H in white) into the crystal structure of the extracellular domain of $\alpha_v\beta_3$ integrin (α unit red and β unit blue wire representation) overlaid on the bound conformation of Cilengitide (gray tube representation). Only selected integrin residues involved in the interactions with the ligand are shown. The metal ion at MIDAS is shown as a magenta CPK sphere. For the sake of clarity, all H atoms bound to carbon are omitted.

Compound 3

In the STD experiment of compound **3** at 298 K the signals of DKP-NH_1, NH_{Asp} and NH_{Gly} overlap, and therefore it is difficult to assess their involvement in the binding process (**Figure 2.2.16.**). The largest STD effect is observed for the guanidine NH protons (7.15 ppm, 2.3% STD absolute intensity).

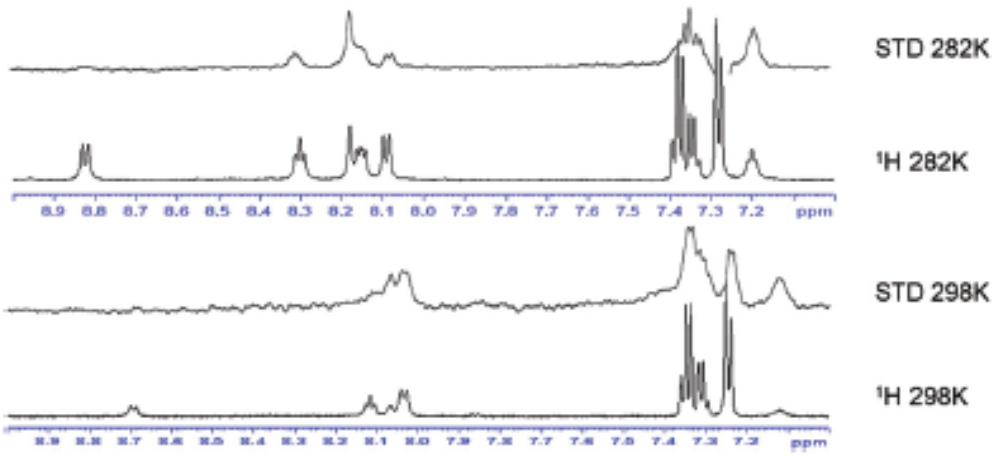


Figure 2.2.16. STD spectra and ^{1}H-NMR spectra (selected regions) of ligand **3** at 282 and 298 K. At T = 298 K, DKP-NH_1, NH_{Asp} and NH_{Gly} overlap.

At 282 K (see **Figure 2.2.13.**, green bars) the largest STD effects are observed for DKP-NH_1 (0.8% STD absolute intensity) and for the guanidine NH protons (7.26 ppm, 0.7% STD absolute intensity). The aromatic protons, DKP-NH_{10}, NH_{Gly} and NH_{Asp} show similar STD effects (0.4, 0.4, 0.4 and 0.3%, respectively) and can also be considered as lying close to the protein surface. The signals observed in the STD spectrum suggest that ligand **3** is in an extended conformation when binding to cells, and it interacts closely with both the α_v (by the guanidine NHs) and β_3 (by NH_{Asp}, NH_{Gly}, DKP-NH_{10}, DKP-NH_1 and the aromatic protons) subunits.

Docking studies starting from the bound conformation of compound **3** (Type III, **Figure 2.2.11.**) produced top-ranked binding modes conserving all the key polar interactions of the X-ray complex, as described above. Moreover, favourable van der Waals contacts can be detected between the ligand benzyl group and the aromatic moiety of β_3-Tyr122 and the Hα of Ser123, between DKP-NH_1 and the terminal methyl protons of β_3-Met180, between DKP-NH_{10} and the Hβ

of Ser123 and the carbonyl group of Tyr122, and between NH_{Gly} and the methyl protons of β_3-Ala218 (**Figure 2.2.17.**).

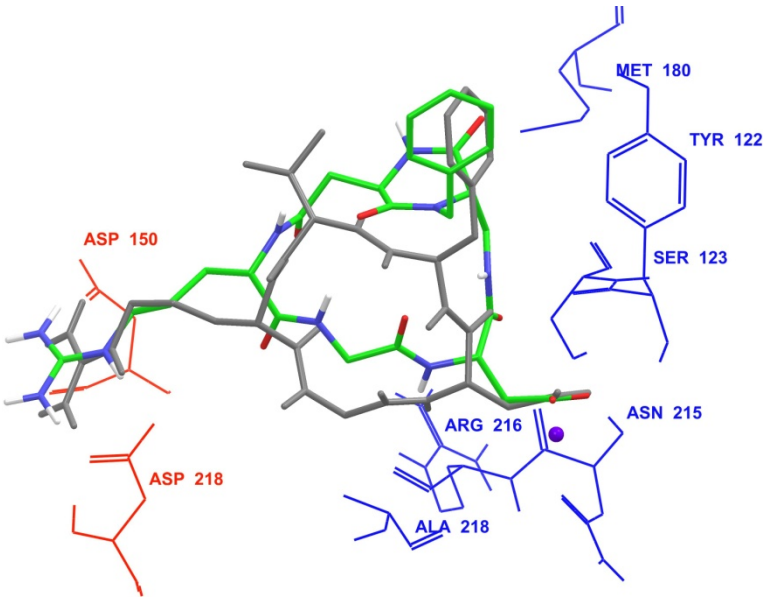


Figure 2.2.17. Docking best pose of *trans*-DKP3-RGD-peptidomimetic **3** (tube representation, C atoms in green, N in blue, O in red, H in white) into the crystal structure of the extracellular domain of $\alpha_v\beta_3$ integrin (α unit red and β unit blue wire representation) overlaid on the bound conformation of Cilengitide (gray tube representation). Only selected integrin residues involved in the interactions with the ligand are shown. The metal ion at MIDAS is shown as a magenta CPK sphere. For the sake of clarity, all H atoms bound to carbon are omitted.

2.2.2.5 Platelets and ECV-304 cells: a comparative discussion

The *tr*-NOE and STD experiments were run in the presence of ECV-304 bladder cancer cells, expressing $\alpha_v\beta_3$, $\alpha_v\beta_5$, and $\alpha_5\beta_1$ integrins, and platelets, expressing integrin $\alpha_{IIb}\beta_3$. These experiments helped identifying the binding epitopes of the peptidomimetics and defining the ligand-integrin bound conformations and the ligand-receptor interactions at a molecular level.

The ligands **1 - 3** are cyclic peptidomimetics containing the RGD sequence and chiral bifunctional diketopiperazine scaffolds, differing for the configuration at the C-3 and C-6 stereogenic centres. The scaffold stereochemistry influences the conformation of the ligands both in the free and in the cell-bound states. In particular, the *cis* (3S, 6S)-configuration of DKP1 scaffold in ligand **1** imparts a kinked conformation to the RGD sequence which implies relatively short

Cβ(Arg)-Cβ(Asp) distance values (< 8 Å) and a low binding affinity to integrin $\alpha_v\beta_3$ as proved by the high IC_{50} values in the competitive receptor binding and cell adhesion assays (**Table 2.2.1.** and **2.2.2.**). On the other hand, the *trans* stereochemistry of the DKP scaffolds in ligands **2** and **3** ensures an extended arrangement of the RGD sequence [Cβ(Arg)-Cβ(Asp) distance values of about 9 Å] which is well suited for the $\alpha_v\beta_3$ and $\alpha_{IIb}\beta_3$ integrin receptors. This is reflected in the high binding affinity to the $\alpha_v\beta_3$ integrin receptor and in the efficient ECV-304 cell adhesion inhibition displayed by these compounds (**Table 2.2.1.** and **2.2.2.**).

In particular, for compound **1**, the binding epitope is conceivably the same with few interactions for the two cell lines, the only difference is the presence of NH_{Asp} (0.2% STD absolute intesnsity) in the interaction with ECV-304 (**Figure 2.2.18.**). This could be because the basic ligand side chain must reach further into the deeper β-propeller pocket of α_{IIb} to interact with Asp224, whereas in the α_V chain the residues Asp150 and Asp218 are nearer in its shallower pocket.

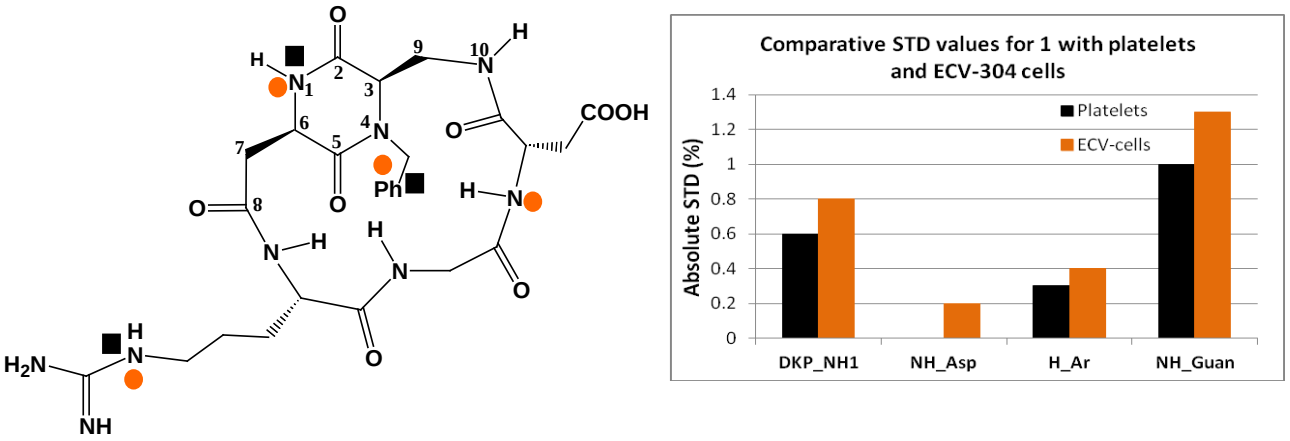


Figure 2.3.18. Binding epitopes: map of the STD signals of ligand **1** in the presence of platelets (black squares) with respect to ECV-304 cells (orange circles).

For compound **2**, the binding epitope of the RGD sequence remains for the two cell lines, but in general the intensity and the number of the STD effects suggest a greater interaction with ECV-304. In fact, some differences are observed for the DKP scaffold. STD-NMR experiments with ECV-304 cells show that the aromatic protons of the ligand are in contact with the protein, while NH1 and H9 of the DKP scaffold interact with the receptor in the experiments with platelets (**Figure 2.2.19.**).

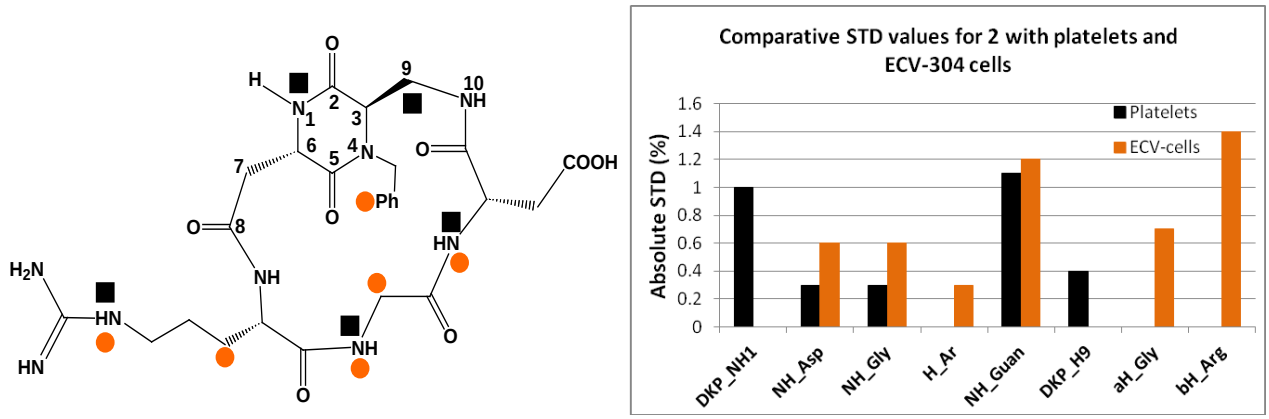


Figure 2.2.19. Binding epitopes: map of the STD signals of ligand **2** in the presence of platelets (black squares) with respect to ECV-304 cells (orange circles).

For compound **3**, the same protons belonging to both the RGD sequence and the DKP scaffold provide significant STD percentages (higher for platelets) and, as a consequence, the epitope is conceivably the same with the two cell lines (**Figure 2.2.20.**).

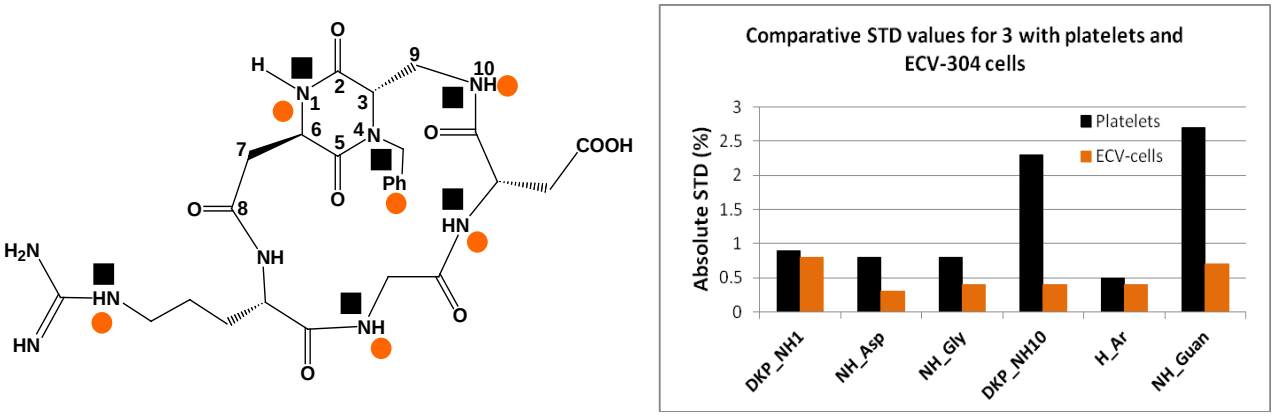


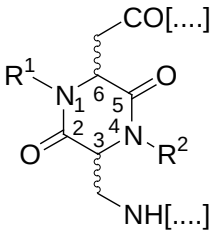
Figure 2.2.20. Binding epitopes: map of the STD signals of ligand **3** in the presence of platelets (black squares) with respect to ECV-304 cells (orange circles).

In conclusion, for both cell lines we observed that: (*i*) the primary interaction between ligand and receptor is performed by the guanidine residue of the Arginine side chain (see behaviour of peptidomimetic **1**), (*ii*) the extended conformation of the *trans* ligands **2** and **3** ensures the formation of the electrostatic clamp (STD effects on protons of the Arg and Asp residues), (*iii*) the opposite configuration at C-3 and C-6 stereogenic centres of the DKP scaffold of

the *trans* ligands **2** and **3** induces a different binding epitope for the non-RGD moiety of the ligands (diketopiperazine and benzyl substituent). Finally, these results are well modelled by docking analysis of **1** - **3** in the active sites of $\alpha_v\beta_3$ and $\alpha_{IIb}\beta_3$ integrin receptors.

2.2.2.6 In vitro biological assays of compounds 4 - 7

We completed our analysis with the study of the peptidomimetic ligands **4** - **7** containing the two enantiomeric *trans* DKP scaffolds with different substitution at the endocyclic nitrogen atoms (**Figure 2.2.21.**).



DKP-4 = 3*R*, 6*S*, $R^1 = CH_2Ph$, $R^2 = H$
DKP-5 = 3*R*, 6*S*, $R^1 = CH_2Ph$, $R^2 = CH_2Ph$
DKP-6 = 3*S*, 6*R*, $R^1 = CH_2Ph$, $R^2 = H$
DKP-7 = 3*S*, 6*R*, $R^1 = CH_2Ph$, $R^2 = CH_2Ph$

Figure 2.2.21. Bifunctional DKP scaffolds.

Integrin receptor competitive binding assays

The cyclic RGD peptidomimetics were examined *in vitro* for their ability to inhibit biotinylated vitronectin binding to the purified $\alpha_v\beta_3$ and $\alpha_v\beta_5$ receptors (**Table 2.2.3.,** see "Experimental Section" for details).

Structures	$\alpha_v\beta_3$ **IC_{50} [nM]**[a]	$\alpha_v\beta_5$ IC_{50} [nM][a]
cyclo[**DKP-4**-RGD]	**7.6 ± 4.3**	216 ± 5
cyclo[**DKP-5**-RGD]	**12.2 ± 5.0**	131 ± 29
cyclo[**DKP-6**-RGD]	**2.1 ± 0.6**	79 ± 3
cyclo[**DKP-7**-RGD]-A	**220.2 ± 82.3**	$> 10^4$
cyclo[**DKP-7**-RGD]-B	**0.2 ± 0.09**	109 ± 15
cyclo[RGDfV]	**3.2 ± 1.3**	7.5 ± 4.8

Table 2.2.3. Inhibition of biotinylated vitronectin binding to $\alpha_v\beta_3$ and $\alpha_v\beta_5$ receptors. [a] IC_{50} values were calculated as the concentration of compound required for 50% inhibition of biotinylated vitronectin binding as estimated by GraphPad Prism software; all values are the arithmetic mean ± SD of triplicate determinations.

Low nanomolar values were obtained for all the ligands except for compound **7-A**. The behavior of this ligand is peculiar, considering that the diastereomeric compound **7-B** is the most potent ligand of this series, effectively inhibiting the binding of vitronectin to the isolated $\alpha_v\beta_3$ receptor in sub-nanomolar concentration. Interestingly, unlike reference compound, the RGD peptidomimetics were about 10 – 500-fold more selective for the $\alpha_V\beta_3$ integrin with respect to the $\alpha_V\beta_5$ in this kind of assay.

Cell adhesion assays

To assess the activity of the compounds as integrin antagonists on a cell model, as already reported for compounds **1** - **3**, cell adhesion experiments were performed using ECV-304 bladder cancer cells. Cells were allowed to adhere to immobilized fibronectin or vitronectin in the presence of increasing concentration of the tested compound (**Table 2.2.4.**, see "Experimental Section" for details).

Ligands	ECV-304	
	Vitronectin IC_{50} [μM][a]	Fibronectin IC_{50} [μM][a]
cyclo[DKP-4-RGD]	>70	0.58 ± 0.21
cyclo[DKP-5-RGD]	17.3 ± 7.6	N. R.
cyclo[DKP-6-RGD]	6.8 ± 2.0	>90
cyclo[DKP-7-RGD]-A	N.D.	N.D.
cyclo[DKP-7-RGD]-B	6.6 ± 0.8	N.D.
cyclo[RGDfV]	2.00 ± 0.35	0.24 ± 0.03

Table 2.2.4. Effect of DKP-RGD compounds on ECV-304 cell adhesion to vitronectin and fibronectin. [a] Each data point was performed in triplicate in two independent experiments. N.D. = Not Detectable, N.R. = Not Reproducible.

Except for compound **7-A**, whose activity is too low to be detected, ligands **4** - **7** showed a micromolar activity in inhibiting the VN-mediated adhesion. Surprisingly, compound **4**, which has a nanomolar affinity for the isolated $\alpha_V\beta_3$ and $\alpha_V\beta_5$ receptors comparable to the other ligands, resulted 10 time less active in inihibiting the VN-ECV-304 adhesion. Moreover, compound 4 is the only compound able of inhibiting the FN-mediated adhesion at sub-micromolar concentration.

The micromolar activities obtained for compounds **4 – 7-B** in the ECV-304 cell adhesion assay represent an ideal range for testing interactions between the ligands and the integral membrane protein of intact cells by means of *tr*-NOE and STD experiments, as for the other three compounds.

2.2.2.7 *tr*-NOESY, STD-NMR and docking results of compounds 4 – 5

Compound 4

The RGD peptidomimetic ligands **4** and **5**, as well as ligand **2**, contain the (3R, 6S) DKP scaffold. The three ligands differ for the substitution at the endocyclic nitrogen atoms of the DKP scaffold. In particular, compound **4** incorporates a benzyl group at N-1 and a hydrogen atom at N-4. Investigation of the conformational preferences of ligand **4** by NMR and MC/SD simulations suggested a preferred solution conformation characterized by the type I-trans extended geometry (*i.e.* β-turn at Gly-Asp) (**Figure 2.2.22.**). This conformation, apparently unaltered upon binding to ECV-304 cells, is featured by the orientation of the DKP scaffold shown in **Figure 2.2.22.** in comparison to the different orientation adopted by the scaffold in the type III conformation.

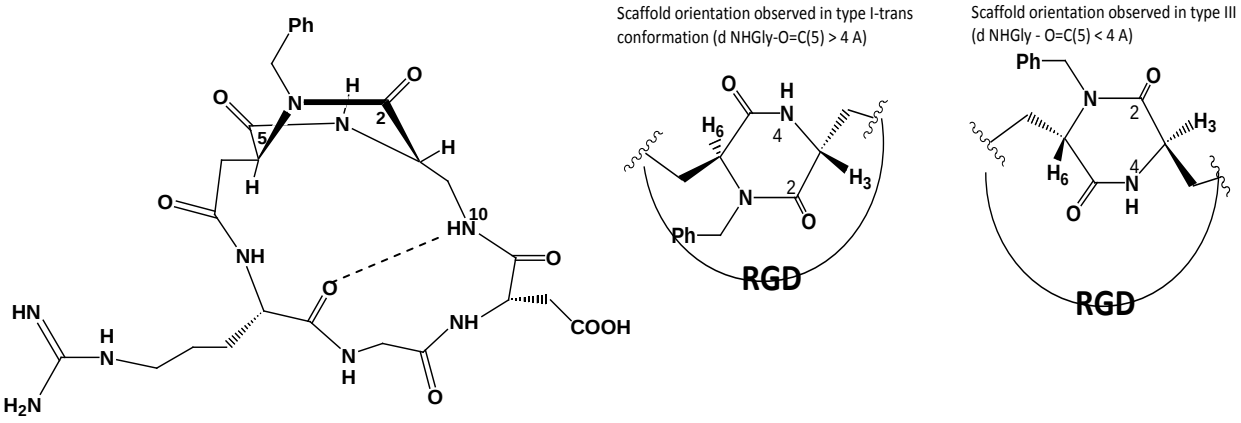


Figure 2.2.22. Left: Free and bound conformation of *trans*-DKP-RGD-peptidomimetic **4**. Hydrogen bond (dotted line). Right: representation of the rotated DKP scaffold characterizing the conformation.

As shown by the docking results described below, the ligand conformation coupled with scaffold substitution and orientation strongly influenced the orientation of the benzyl group inside the binding pocket and, as a result, the binding mode of the entire compound.

In the *tr*-NOESY experiment of compound **4** in the presence of ECV-304 cells, at 298 K the cross-peak intensity is close to zero as in the free state, at 282 K the contacts are negative (*i. e.*, with the same phase as the diagonal), while in the free state are close to zero.

STD-NMR experiments were performed at two temperatures, 298 and 282 K. The experiment at 298 K showed quite large STD effects for the protons belonging to the RGD sequence (guanidine NH, Hδ of Arg and Hβ of Asp, 1.1, 0.2 and 0.3% STD absolute intensity, respectively) as well as to the DKP scaffold (DKP-NH4 and aromatic protons, 1.0 and 0.6% STD absolute intensity, respectively). When the temperature was lowered to 282 K, additional STD signals were detected. In fact, at 282 K, the signals for the guanidine NH, Hδ of Arg, Hβ of Asp, DKP-NH4 and aromatic protons showed values of 0.4, 0.3, 0.5, 0.8 and 1.2% STD absolute intensity, respectively. Additionally, at 282 K, the signals of Arg Hβ and Hγ (0.4 and 0.4% STD absolute intensity, respectively), NH_{Arg}, DKP-H7 and DKP-NH10 (1.3, 0.9 and 1.1 % STD absolute intensity, respectively) appeared in the STD spectrum (**Figure 2.2.23.**, orange bars). For comparison, the STD data of ligands **2**, **4** and **5** sharing the (3R, 6S) DKP scaffold, are shown in the **Figure 2.2.23.**.

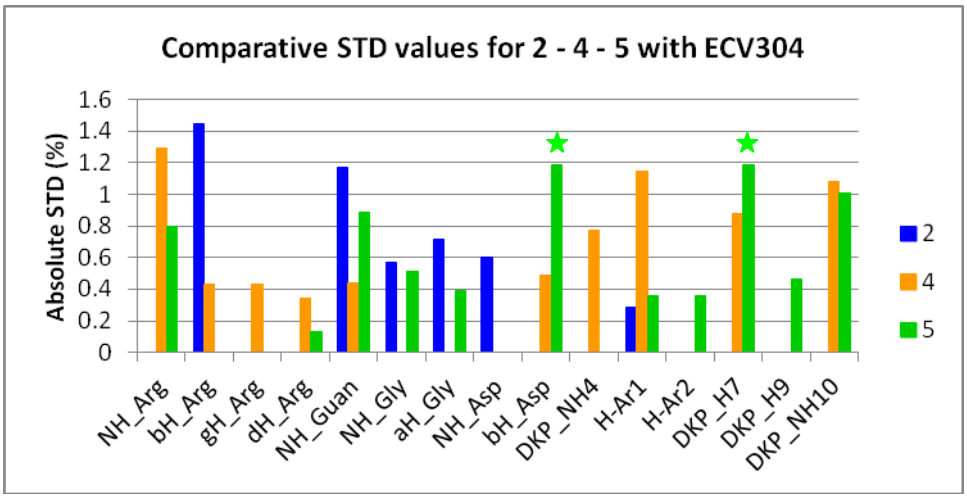


Figure 2.2.23. Relevant absolute STD percentages for the protons of cyclic DKP-RGD obtained at 282 K in the presence of ECV-304 cells: compound **2** (blue bars), compound **4** (orange bars), compound **5** (green bars). For compound **5**, signal overlaps are marked with lines above each bar.

The binding epitope of compound **4** is characterized by Arg and Asp pharmacophoric moieties (guanidine NH proton and Hβ of Asp), by many protons of the DKP scaffold and of the Arg-side chain and by the lack of Gly and Asp backbone atoms (**Figure 2.2.24.**).

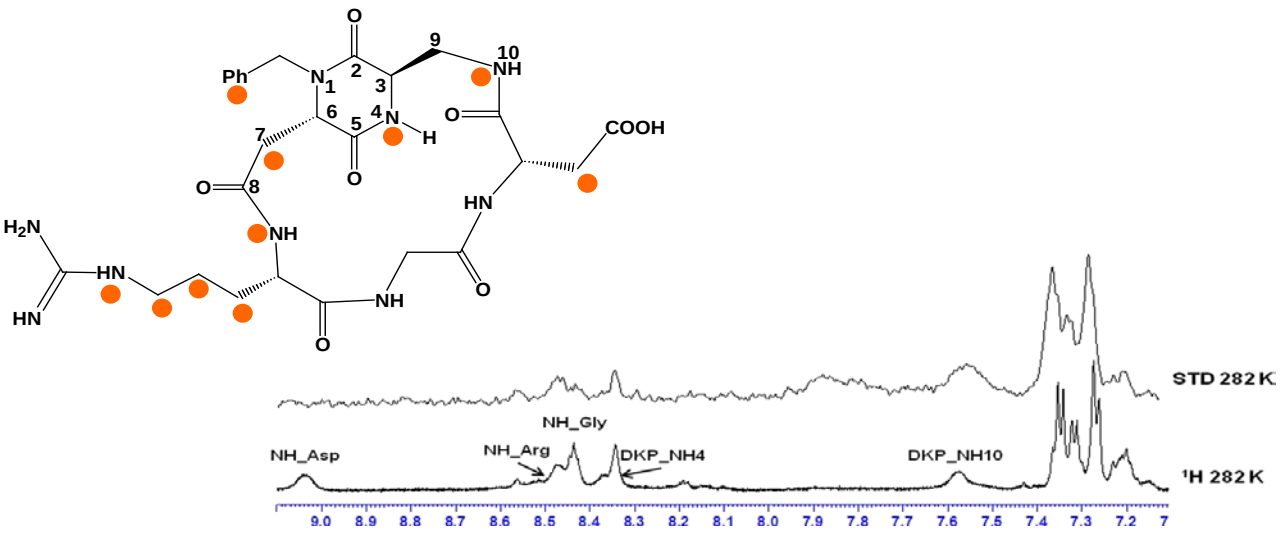


Figure 2.2.24. Binding epitope: map of the STD signals (left) of ligand **4** in the presence of ECV cells (orange circles), STD and ^{1}H-NMR spectra (right). Only the amidic and the aromatic regions are shown.

Docking calculations into the $\alpha_V\beta_3$ integrin binding site, starting from the preferred RGD extended conformation [C_β(Arg)-C_β(Asp) distance of about 9 Å] of ligand **4** (**Figure 2.2.22.**), produced top-ranked binding modes conserving a good correspondence with the receptor-bound structure of Cilengitide (**Figure 2.2.25.**, left). The analysis of the docking poses identified a different possible disposition of the molecule within the binding pocket, where the ligand still maintains the typical electrostatic clamp, but does not overlay to the Cilengitide cyclic backbone. The two types of binding mode, called A and B (their superposition is shown in **Figure 2.2.25.**, right), are very similar in terms of contacts of the RGD sequence with the receptor (average distance values for the relevant interactions identified by STD experiment are reported in **Table 2.2.5.**), but significantly different as regards the contacts of the scaffold protons. In fact, only for poses of type B binding mode favourable van der Waals contacts between ligand DKP-NH4, DKP-H7, DKP-NH10 and receptor β-chain atoms are detected (**Table 2.2.5.**).

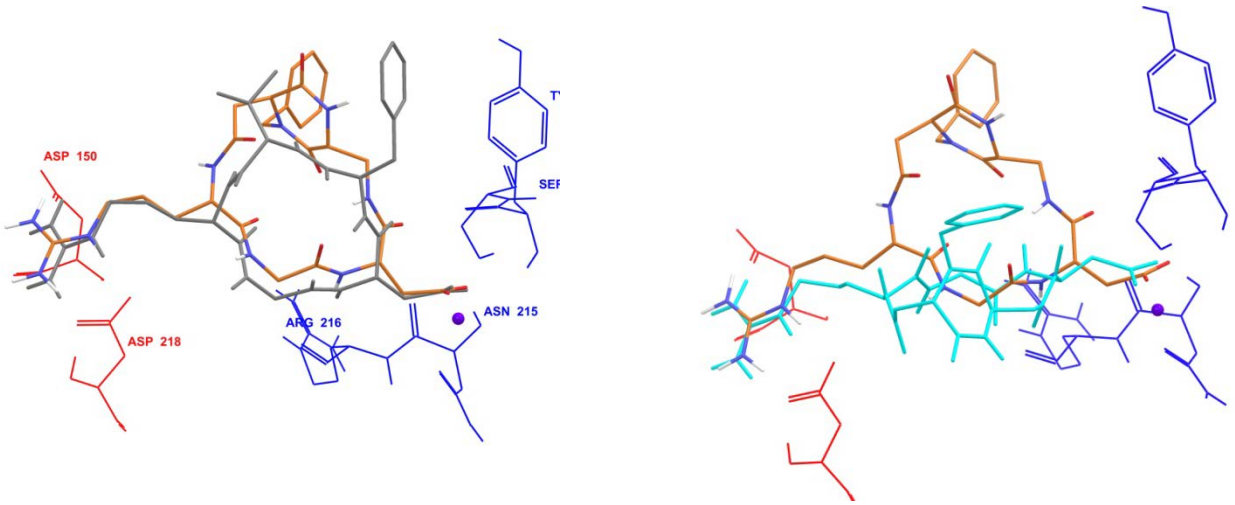


Figure 2.2.25. Left: docking best pose (binding mode A) of compound **4** (tube representation, C atoms in orange, N in blue, O in red, H in white) into the crystal structure of the extracellular domain of $\alpha_v\beta_3$ integrin (α unit red and β unit blue wire representation) overlaid on the bound conformation of Cilengitide (gray tube representation); the metal ion at MIDAS is shown as a magenta CPK sphere. Right: Superposition of A (orange) and B (cyan) binding modes identified among the docking poses of ligand **4** into the $\alpha_v\beta_3$ integrin binding site.

Ligand protons	Binding mode A		Binding mode B	
	Residues	Mean distance	Residues	Mean distance
NH-Arg	HO-Tyr178 (α)	4.67 Å	CH_3-Ala215 (α)	3.54 Å
βH-Arg	CAr-Tyr178 (α)	2.68 Å	CH_3-Ala215 (α)	2.41 Å
γH-Arg	CH_3-Ala215 (α)	2.57 Å	O_2=C-Asn218 (α)	2.88 Å
δH-Arg	HAr-Tyr178 (α)	2.90 Å	HAr-Tyr178 (α)	2.65 Å
NH_Guan	O_2=C-Asp218 (α)	2.42 Å	O_2=C-Asp218 (α)	2.20 Å
βH_Asp	O=C-Asn215 (α)	2.38 Å	O=C-Asn215 (α)	3.00 Å
DKP-NH4	HAr-Tyr122 (β)	6.88 Å (Min=4.69 Å)	CH_3-Ala218 (β)	2.38 Å
H_Ar	$CH_2\beta$-Met180 (β)	5.61 Å (Min=2.21 Å)	O=C-Asp251 (β)	4.60 Å (Min=2.48 Å)
DKP-H7	HO-Tyr178 (α)	6.07 Å (Min=3.51 Å)	NH_3-Lys253 (β)	3.63 Å
DKP-NH10	O=C-Tyr122 (β)	4.48 Å (Min=3.98 Å)	αCH-Ala218 (β)	1.84 Å

Table 2.2.5. Average values of the distances between ligand **4** protons (selected from STD experiment) and $\alpha_v\beta_3$ residues calculated from docking poses for the two binding modes A and B.

The ligand epitope disclosed by STD-NMR results nicely agree with the calculated docking poses, though type B binding mode provides the best explanation for the protons of the DKP scaffold. Therefore, we can assume that the epitope observed by NMR depends on the contribution of both the binding dispositions.

Moreover, our analysis pointed out the stereochemistry and the substitution of the DKP scaffold, that strongly influenced the type of conformation adopted by compound **4** (see paragraph "**NMR characterization and conformational analysis**"), also affected the disposition of the compound in the receptor binding pocket. NMR and docking results well agree with the nanomolar IC_{50} value obtained in the isolated $\alpha_v\beta_3$ receptor binding assay (7.6 ± 4.3 nM).

Compound 5

Similarly to compound **2**, the free state conformation of ligand **5** (featuring a benzyl group at both the endocyclic nitrogens N1 and N4) is characterized by a β-turn motif at DKP-Arg stabilized by a hydrogen bond between NH_{Gly} and C(5)=O (Type III conformation, **Figure 2.2.26.**), as suggested by the relevant long-range NOE interaction between the NH_{Gly} and NH_{Arg} (positive at 298 K and negative at 282 K).

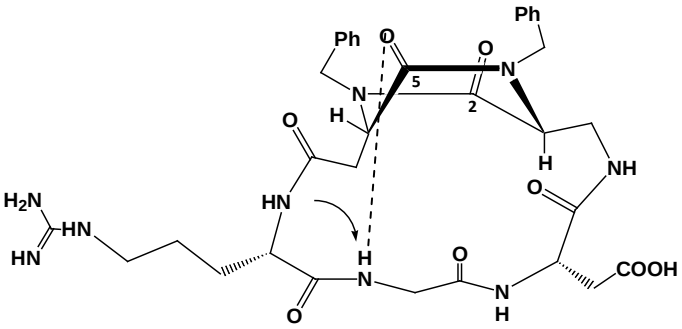


Figure 2.2.26. Free and bound type III conformation of *trans*-DKP-RGD-peptidomimetic **5**. Hydrogen bond (dotted line) and NOE contact (arrow) are indicated.

The *tr*-NOESY experiment of compound **5** in the presence of ECV-304 cells at 298 K showed negative cross-peaks (*i.e.*, with the same phase as the diagonal) in contrast with the free state, indicating a binding situation; instead, at 282 K the contacts are negative like in the free state. In addition, the *tr*-NOESY spectrum performed at 298 K showed a negative cross-peak between NH_{Arg} (8.39 ppm) and NH_{Gly} (8.08 ppm); this NOE contact (**Figure 2.2.27.**) is consistent with the β-turn conformation at DKP-Arg stabilized by the hydrogen bond between NH_{Gly} and C(5)=O mentioned above (**Figure 2.2.26.**).

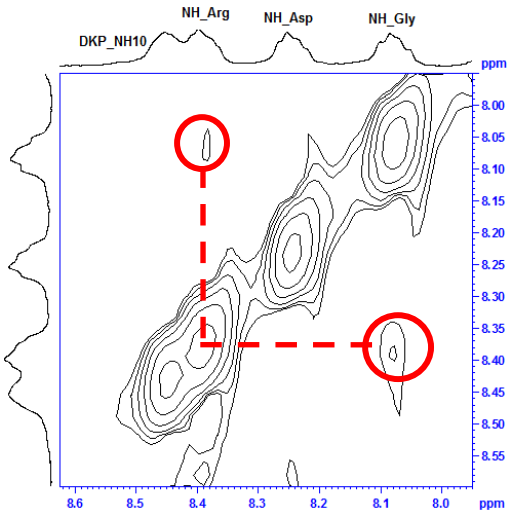


Figure 2.2.27. NOE contact between NH_{Arg} and NH_{Gly} indicating a Type III bound conformation for Compound **5**.

STD-NMR experiments were performed in the presence of ECV-304 cells at 298 K and 282 K. In these spectra, the region corresponding to the aromatic and amide protons is better resolved at 282 K with respect to 298 K and can be used to measure the absolute intensity of STD (**Figure 2.2.23.**, green bars). The largest STD effects are observed for NH_{10} (1.0% STD absolute intensity) and for the guanidine NH protons (6.95 ppm, 0.9% STD absolute intensity). Also the protons DKP-H7 and βH_{Asp} show high STD effects (1.2% STD absolute intensity), but their overlapping prevents a right evaluation of the contribution. Additionally, spectra show protons belonging to the DKP scaffold, *i.e.* the aromatic protons and DKP-H9 (0.4 and 0.5% STD absolute intensity, respectively) as well as to the RGD sequence (NH_{Arg}, δH_{Arg}, NH_{Gly} and αH_{Gly}, 0.8, 0.1, 0.5 and 0.4% STD absolute intensity, respectively) in contact with the receptor (**Figure 2.2.28.**). However, as the signals of the protons of the two aromatic rings overlapped and it was not possible to distinguish them, the STD intensity reported in the graph (**Figure 2.2.23.**, green bars) was halved for each aromatic ring.

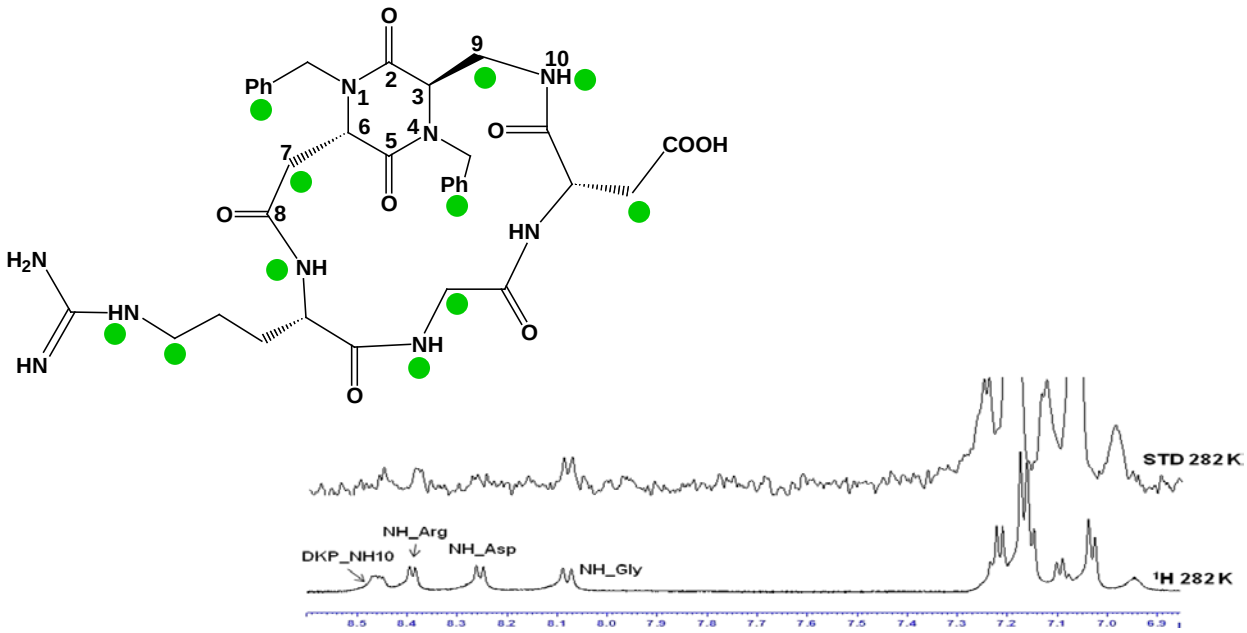


Figure 2.2.28. Binding epitope: map of the STD signals (left) of ligand **5** in the presence of ECV cells (green dots), STD and ^{1}H-NMR spectra (right). Only the amidic and the aromatic regions are reported.

Docking calculations into the $\alpha_V\beta_3$ integrin binding site, starting from the preferred RGD extended conformation of compound **5** produced top-ranked binding modes conserving a good correspondence with the receptor-bound structure of Cilengitide and all the important polar interactions of the X-ray complex. For this compound, we observe a single binding mode which provides average distances between ligand epitope protons and binding site residues in agreement with the NMR data (**Figure 2.2.29.** and **Table 2.2.6.**).

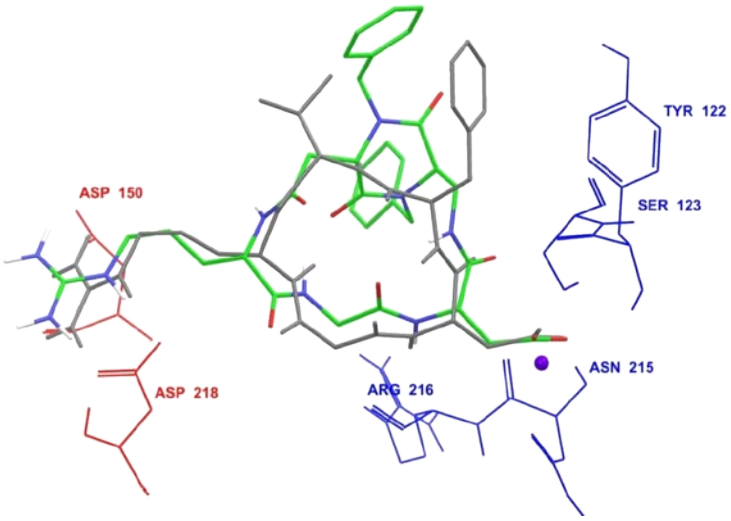


Figure 2.2.29. Docking best pose of compound **5** (tube representation, C atoms in green, N in blue, O in red, H in white) into the crystal structure of the extracellular domain of $\alpha_v\beta_3$ integrin (α unit red and β unit blue wire representation) overlaid on the bound conformation of Cilengitide (gray tube representation); the metal ion at MIDAS is shown as a magenta CPK sphere.

Ligand protons	Protein residues	Mean Distance
NH_{Arg}	CAr-Tyr178	3.61 Å
δH_{Arg}	HAr-Tyr178	2.71 Å
NH_Guan	O_2=C-Asp218	3.05 Å
NH_{Gly}	HAr-Tyr178	3.24 Å
βH_{Asp}	O=C-Asn215	2.53 Å
H_Ar	NHGuan_Arg214	3.40 Å
DKP-H7	HO-Tyr178	4.32 Å (Min=3.01 Å)
DKP_H9	HAr-Tyr122	2.71 Å
DKP-NH10	CH_2-Ser123	4.18 Å (Min=3.59 Å)
αH_Gly	CH_3-Ala218	2.42 Å

Table 2.2.6. Average values of the distances between ligand **5** protons (selected from STD experiment) and $\alpha_v\beta_3$ residues calculated from docking poses.

The stereochemistry of the DKP moiety and the dibenzylation at the two nitrogen atoms, strongly influence the conformation adopted by compound **5** (see paragraph "**NMR characterization and conformational analysis**") and its disposition inside the receptor binding

pocket. In the docking poses, the ligand is well overlaid to the experimental binding mode of Cilengitide, one aromatic moiety points outwards and the other lies within the binding site (**Figure 2.2.29.**). NMR and docking results are consistent with the IC_{50} value obtained in binding ($\alpha_v\beta_3$ IC_{50} 12.2 ± 5.0 nM) and adhesion assays (ECV-304 IC_{50} (VN) = 17.3 ± 7.6 μM).

A comparative analysis of the results obtained for the series of active ligands containing the (3R, 6S) DKP scaffold, *i.e.* ligands **2**, **4** and **5**, highlights that: *i)* the conformational preferences depend on both the scaffold stereochemistry and substitution; *ii)* all the compounds adopt extended conformations in solution that produce top-ranked docking poses fitting the experimental binding mode of Cilengitide; *iii)* both ligand conformation and scaffold substitution strongly affect the binding epitope detected by STD-NMR; (iv) a good agreement between docking models and STD data has been obtained.

In particular, compound **2** (**Figure 2.2.23.**, blue bars) mainly involves in the interaction the RGD sequence (NH_{Gly}, αH_{Gly}, NH_Guan and NH_{Asp}), compound **4** (**Figure 2.2.23.**, orange bars) mostly involves Arg and Asp pharmacophoric moieties and scaffold protons (DKP-NH_4, DKP-H7 and DKP-NH_{10}), and compound **5** (**Figure 2.2.23.**, green bars) involves in the interaction both the RGD sequence and the DKP scaffold, strengthening the idea that the double benzylation helps to stabilize the interaction with the receptor binding pocket.

2.2.2.8 *tr*-NOESY, STD-NMR and docking results of compounds 6 -7

Compound 6

The RGD peptidomimetic ligands **6** and **7**, as well as ligand **3**, contain the (3S, 6R) DKP scaffold. The three ligands differ for the substitution at the endocyclic nitrogen atoms of the DKP scaffold. In particular, compound **6**, which incorporates a benzyl group at N-1 and a hydrogen atom at N-4, is characterized by a rather strong NOE contact between NH_{Asp} and NH_{10} and a moderate NOE involving NH_4 and NH_{10}. These two contacts, as discussed in the paragraph "**NMR characterization and conformational analysis**", are mutually exclusive and indicative of an equilibrium between two different conformations, type I-trans (β-turn at Gly-Asp, **Figure 2.2.30.a**) and type IV (β-turn at Asp-DKP, **Figure 2.2.30.b**).

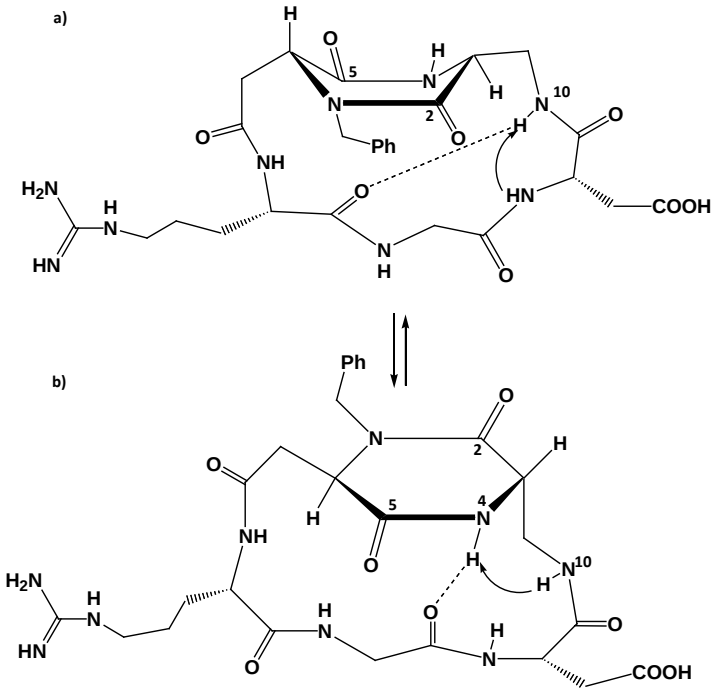


Figure 2.2.30. Free state conformations of *trans*-DKP-RGD-peptidomimetic **6**. a) Type I-trans H-bonding pattern (β-turn at Gly-Asp); b) Type IV H-bonding pattern (β-turn at Asp-DKP). Hydrogen bond (dotted line) and NOE contact (arrow) are indicated.

The two conformations are characterized by different orientations of the DKP scaffold, as shown for the Type I-trans geometry of compound **4** in comparison to the Type III conformation (**Figure 2.2.22**.).

In the *tr*-NOESY experiments at 298 K the cross-peak intensity is close to zero, while at 282 K the contacts are negative like in the free state. However, at 298 K, even if we did not observe a complete inversion of phase (from positive of the free state to negative in the bound state), the fact that the cross-peaks are not positive (like in the free state), but close to zero, suggests a binding event.

In addition, the *tr*-NOESY performed at 282 K showed a very weak negative cross-peak (**Figure 2.2.31.**) between NH_{Asp} (8.92 ppm) and NH_{10} (7.99 ppm); this NOE contact is consistent with the Type I-trans conformation (β-turn at Gly-Asp) (**Figure 2.2.30.**). This could indicate that, upon binding to cells, the conformational equilibrium present in the free state is shifted towards one conformation and, precisely, to the most populated in the free state.

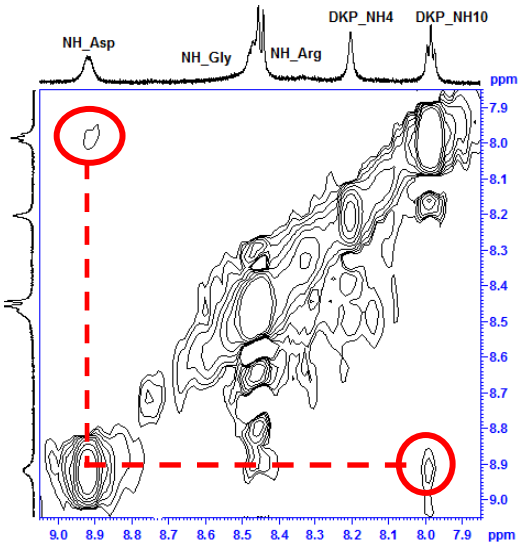


Figure 2.2.31. NOE contact between NH_{Asp} and NH_{10} indicating a Type I-trans bound conformation for compound **6**.

STD-NMR experiments in the presence of the cell suspension were performed both at 298 K and 282 K. At 298 K some overlaps make difficult to evaluate the contribution of several protons in the binding event; the experiment showed quite large STD effects for the guanidine NH, NH_4 and the aromatic protons (1.3, 1.1 and 0.6% STD absolute intensity, respectively) and STD absolute intensity below 0.2% for other aliphatic protons. At 282 K (see **Figure 2.2.32.**, orange bars), the signals for the guanidine NH, NH_4 and the aromatic protons showed values of 1.1, 1.0 and 0.7% STD absolute intensity, respectively. Additionally, at 282 K, the signals for the DKP-H9 and $H_{\alpha Gly}$ (1.4 and 0.8% STD absolute intensity, respectively) appeared with an increased intensity. However, also at 282 K the signals of NH_{Arg} and NH_{Gly} (0.3% STD absolute intensity) overlap as well as the signals of $H_{\beta Asp}$ and DKP-H7 (0.4% STD absolute intensity), making difficult to assess their contribution in the binding process.

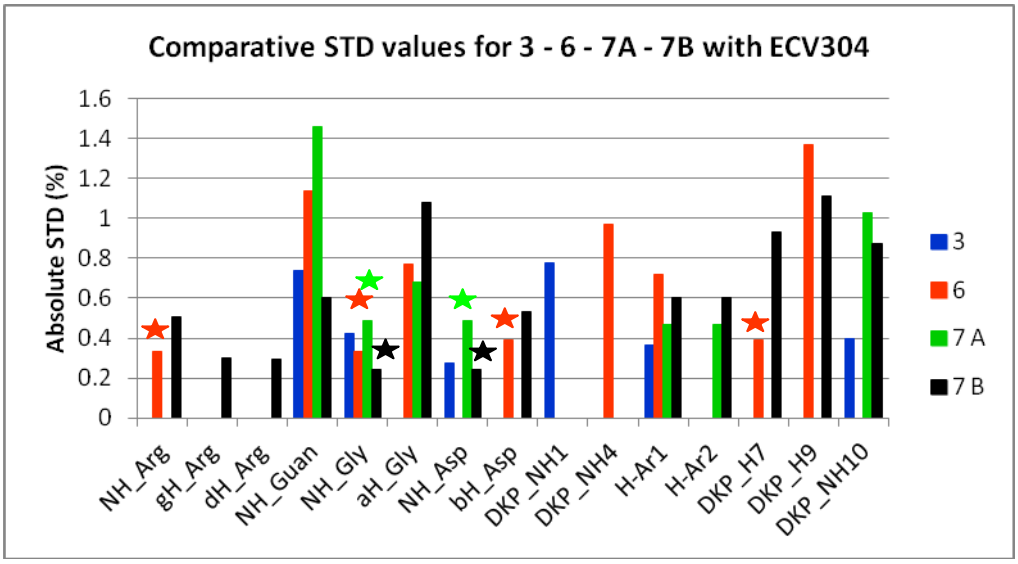


Figure 2.2.32. Relevant absolute STD percentages for the protons of cyclic DKP-RGD obtained at T = 282 K in the presence of ECV-304 cells: compound **3** (blue bars), compound **6** (orange bars), compound **7-A** (green bars), compound **7-B** (black bars). The overlaps are marked with lines above each bar.

In the **Figure 2.2.33.** the amidic and the aromatic regions of the STD spectrum at 282 K are reported with the structure, on the left, of compound **6** where the protons involved in the epitope are highlighted with orange dots.

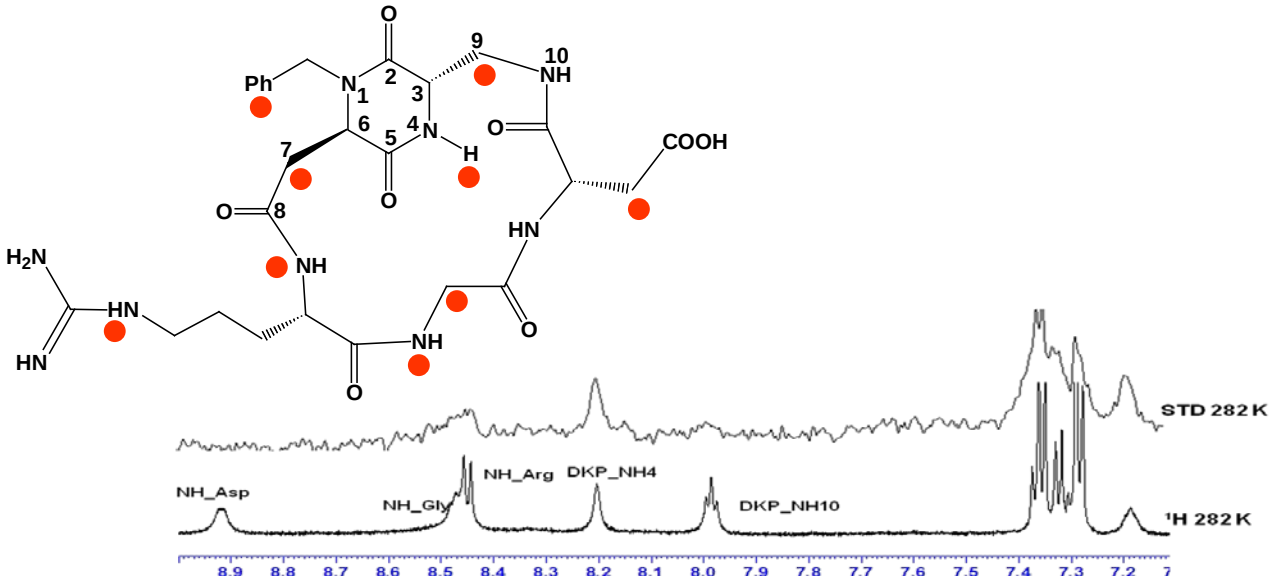


Figure 2.2.33. Binding epitope: map of the STD signals (left) of ligand **6** in the presence of ECV cells (orange circles), STD and ^{1}H-NMR spectra (right). The amidic and the aromatic regions are reported.

Docking calculations into the $\alpha_V\beta_3$ integrin binding site, starting from the bound RGD extended conformation of compound **6** (Type I-trans) produced top-ranked binding modes conserving a good correspondence with the receptor-bound structure of Cilengitide and all the important polar interactions of the X-ray complex already described (**Figure 2.2.34.**). The analysis of the docking poses in terms of ligand epitope protons that could be in contact with the receptor binding site highlights some important considerations (**Table 2.2.7.**).

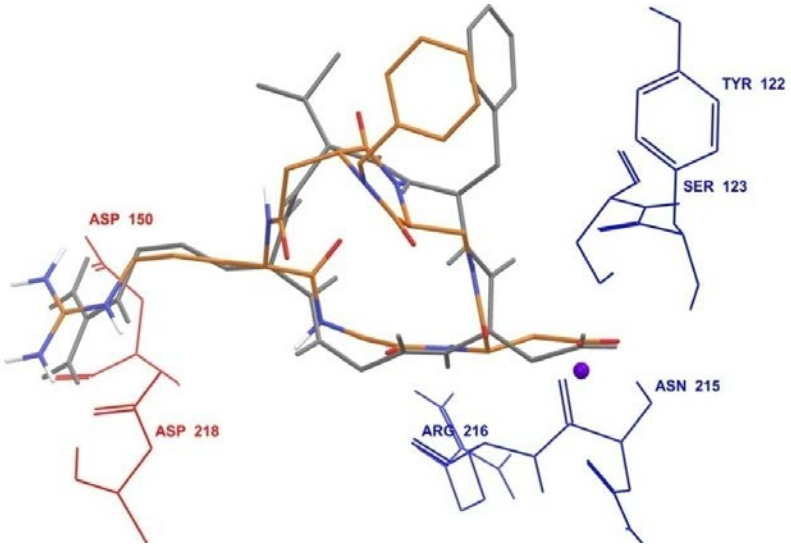


Figure 2.2.34. Docking best pose of compound **6** (tube representation, C atoms in orange, N in blue, O in red, H in white) into the crystal structure of the extracellular domain of $\alpha_v\beta_3$ integrin (α unit red and β unit blue wire representation) overlaid on the bound conformation of Cilengitide (gray tube representation); the metal ion at MIDAS is shown as a magenta CPK sphere.

Ligand protons	Protein residues	Mean Distances
•NH_{Arg}	CH_3-Ala215 (α)	4.87 Å
NH_Guan	O_2=C-Asp218 (α)	2.02 Å
•NH_{Gly}	HAr-Tyr178 (α)	2.83 Å
αH_Gly	HAr-Tyr178 (α)	3.04 Å
*βH_Asp	O=C-Asn215 (β)	2.35 Å
DKP-NH4	βCH_2-Ser123 (β)	5.00 Å (Min=4.83 Å)
H_Ar	CH_3-Ala215 (α)	4.59 Å (Min=2.79 Å)
*DKP-H7	CH_3-Ala215 (α)	5.18 Å (Min=4.55 Å)
DKP-H9	βCH_2-Ser123 (β)	2.34 Å

Table 2.2.7. Average values of the distances between ligand **6** protons (selected from STD experiment) and $\alpha_v\beta_3$ residues calculated from docking poses. Asterisks and dots indicate overlapping proton pairs.

For each overlapping proton pair of the NMR spectrum (NH_{Arg}/NH_{Gly} and $H_{\beta Asp}$/DKP-H7) there is only one proton that shows a distance compatible with the interaction *i.e.* NH_{Gly} and $H_{\beta Asp}$.

In the docking poses, the aromatic protons show a mean distance of 4.59 Å from the protein residues, but such distance can decrease up to a minimum value of 2.79 Å.

The compound shows a low nanomolar IC_{50} value in the binding assays ($\alpha_v\beta_3$ IC_{50} 2.1 ± 0.6 nM) and a low micromolar IC_{50} value in the adhesion assays (ECV-304 IC_{50}(VN) = 6.8 ± 2.0 μM).

Compound 7-A

As described in the paragraph "**NMR Characterization and Conformational Analysis**", compound **7** was obtained as a mixture of two separable diastereomers, **DKP-7 A** and **DKP-7 B**.

The low affinity isomer **7-A** ($\alpha_v\beta_3$ IC_{50} 220.2 ± 82.3 nM) displayed an equilibrium between two different conformations: one featuring a type III H-bonding pattern and an extended RGD arrangement (β-turn at DKP-Arg, **Figure 2.2.35.a**) and one showing a type II H-bonding pattern and a bent RGD arrangement (β-turn at Arg-Gly, **Figure 2.2.35.b**).

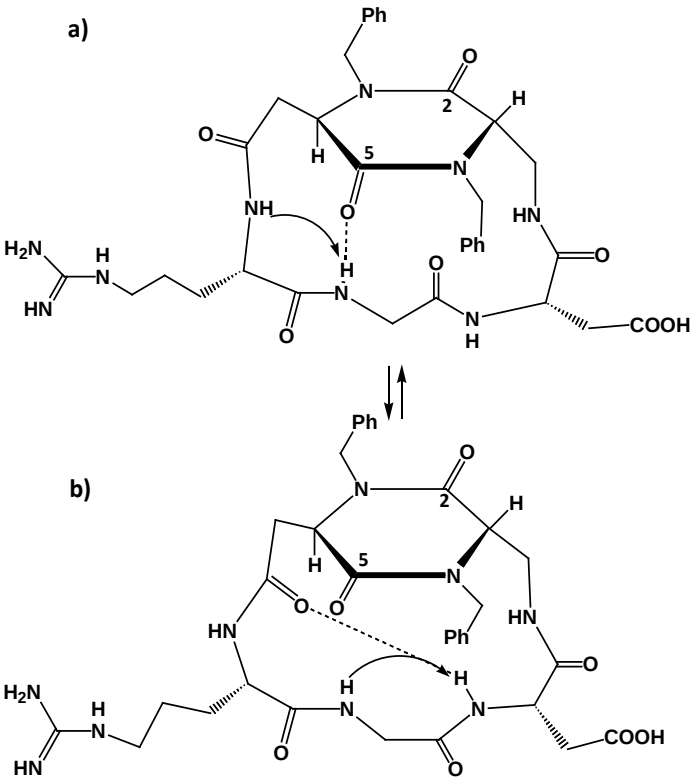


Figure 2.2.35. Free state and bound conformations of *trans*-DKP-RGD-peptidomimetic **7-A**. a) Type III H-bonding pattern (β-turn at DKP-Arg); b) Type II H-bonding pattern (β-turn at Arg-Gly). Hydrogen bond (dotted line) and NOE contact (arrow) are indicated.

In the *tr*-NOESY experiment at 298 K the cross-peak intensity was close to zero, while in the free state the cross-peaks are positive indicating a binding event, as well as in the *tr*-NOESY at 282 K where the cross-peaks had the same phase as the diagonal in contrast with the NOESY in the free state where the cross-peak intensity is close to zero. The conformational equilibrium of the compound in the free state is apparently unaltered upon binding to ECV-304; at 282 K a very weak NOE contact between NH_{Arg} (8.65 ppm) and NH_{Gly} (8.05 ppm) appears in the *tr*-NOESY spectrum, the presence of the other NOE contact between NH_{Asp} (8.04 ppm) and NH_{Gly} (8.05 ppm) is not evaluable due to the overlap between the two amidic protons.

STD-NMR experiments in the presence of the cell suspension were performed both at 298 and 282 K. However, at 298 K several overlaps made difficult to assess the involvement of the protons in the binding process. At 282 K, the largest STD effects (**Figure 2.2.32.**, green bars) are observed for NH_{10} (1.0% STD absolute intensity) and for the guanidine NH (1.5% STD absolute intensity). The aromatic protons and $H_{\alpha Gly}$ show similar STD effects (0.5 and 0.7% STD absolute intensity, respectively). Additionally, at 282 K, the signals of NH_{Gly} and NH_{Asp} overlap (0.5% STD absolute intensity) and therefore it is difficult to assess their involvement in the binding process (**Figure 2.2.36.**).

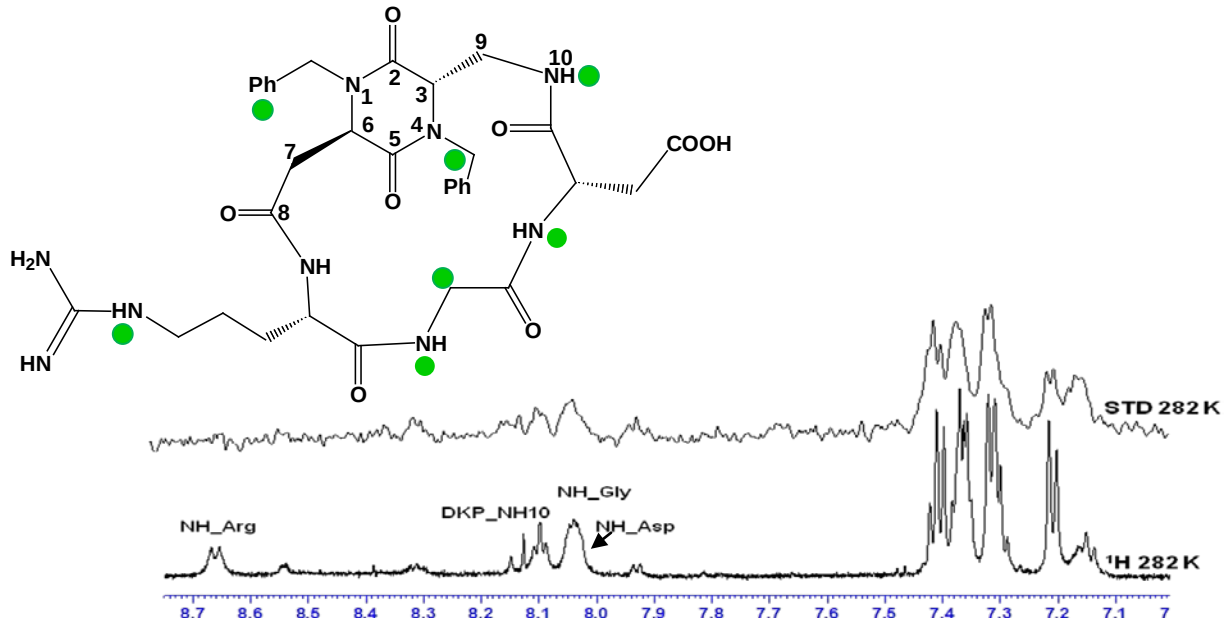


Figure 2.2.36. Binding epitope: map of the STD signals (left) of ligand **7-A** in the presence of ECV cells (green circles), STD and ^{1}H-NMR spectra (right). Only the amidic and the aromatic regions are reported.

In docking calculations, the exended Type III conformation produced top-ranked binding modes conserving a quite good correspondence with the receptor-bound structure of Cilengitide, while the non-extended Type II geometry was not able to properly fit into the $\alpha_v\beta_3$ binding pocket. For this isomer, the presence of an equilibrium between an extended (β-turn DKP-Arg) and a closed geometry (β-turn Arg-Gly) could explain the high nanomolar IC_{50} value in the $\alpha_v\beta_3$ binding assay ($\alpha_v\beta_3$ IC_{50} 220.2 ± 82.3 nM) and the poor inhibition displayed in the adhesion assays (ECV-304 IC_{50}(VN) not detectable at 100 μM).

Compound 7-B

The diastereomer **7-B** showed in the free state a single NOE contact between NH_{Asp} and NH_{10}, indicative of a type I-trans H-bonding pattern (β-turn at Gly-Asp, **Figure 2.2.37.**).

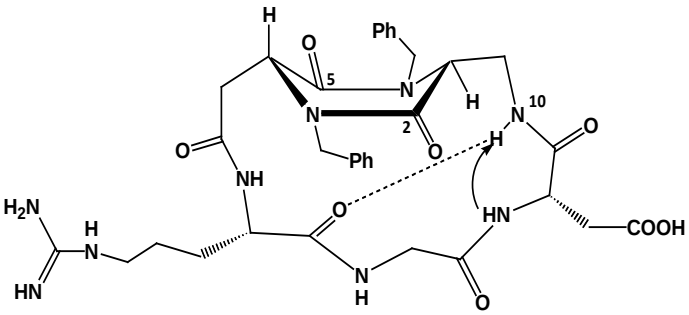


Figure 2.2.37. Free state and bound conformation of *trans*-DKP-RGD-peptidomimetic **7-B**. Type I-trans H-bonding pattern (β-turn at Gly-Asp). Hydrogen bond (dotted line) and NOE contact (arrow) are indicated.

As discussed for ligands **4** and **6**, the Type I-trans geometry adopted by compound **7-B** is characterized by a specific orientation of the DKP scaffold that differs from the orientation displayed by the Type III conformation.

The *tr*-NOESY experiment of compound **7-B** in the presence of ECV-304 cells at 298 K showed negative cross-peaks (*i.e.*, with the same phase as the diagonal) indicating a binding event.

In addition, the *tr*-NOESY spectrum performed at 282 K showed a negative cross-peak between NH_{Asp} (8.56 ppm) and NH_{10} (7.72 ppm); this NOE contact (**Figure 2.2.38.**) is consistent with the β-turn conformation at Gly-Asp (type I-trans H-bonding pattern).

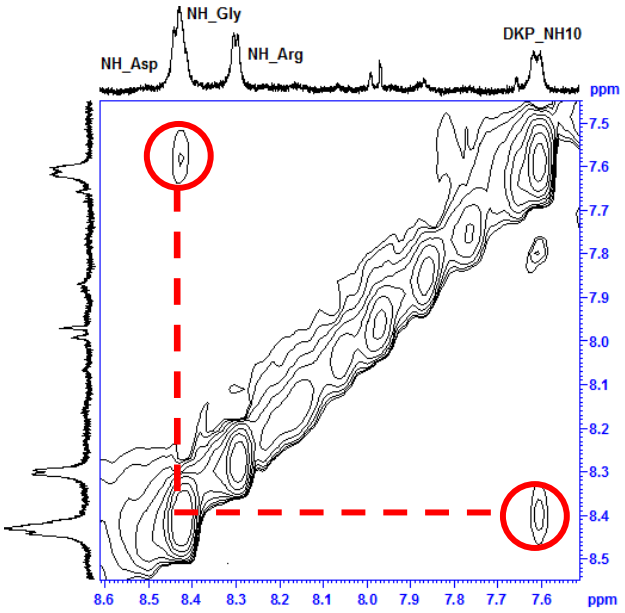


Figure 2.3.38. NOE contact between NH_{Asp} and NH_{10} indicating a Type I-trans bound conformation for compound **7-B**.

STD-NMR experiments were performed in the presence of the cell suspension both at 298 and 282 K. At 298 K STD effect is observed for the aromatic protons, the guanidine NH, the DKP-H7 and NH_{10} (0.3, 0.3, 0.7 and 0.3% STD absolute intensity). Also the signals of some aliphatic protons ($H\delta_{Arg}$, $H\beta_{Asp}$ and $H\gamma_{Arg}$, 0.2, 0.5 and 0.3% STD absolute intensity, respectively) appeared in the spectrum. However, when the temperature was lowered to 282 K, the ligand remains in the receptor site for a longer time and an increased intensity of these STD signals was observed making easier the analysis. In fact, at 282 K (**Figure 2.2.32.**, black bars), the signals for the guanidine NH and the aromatic protons showed values higher than those obtained at 298 K (0.6% STD absolute intensity), and we observed signals for NH_{Arg}, $H\gamma_{Arg}$ and $H\delta_{Arg}$ (0.5, 0.3 and 0.3% STD absolute intensity, respectively), for $H\alpha_{Gly}$, $H\beta_{Asp}$, DKP-H7, DKP-H9 and NH_{10} (1.1, 0.5, 0.9, 1.1 and 0.9% STD absolute intensity, respectively). Additionally, the signals of NH_{Gly} and NH_{Asp} overlap (0.2% STD absolute intensity) and it is difficult to assess their involvement in the binding process (**Figure 2.2.39.**).

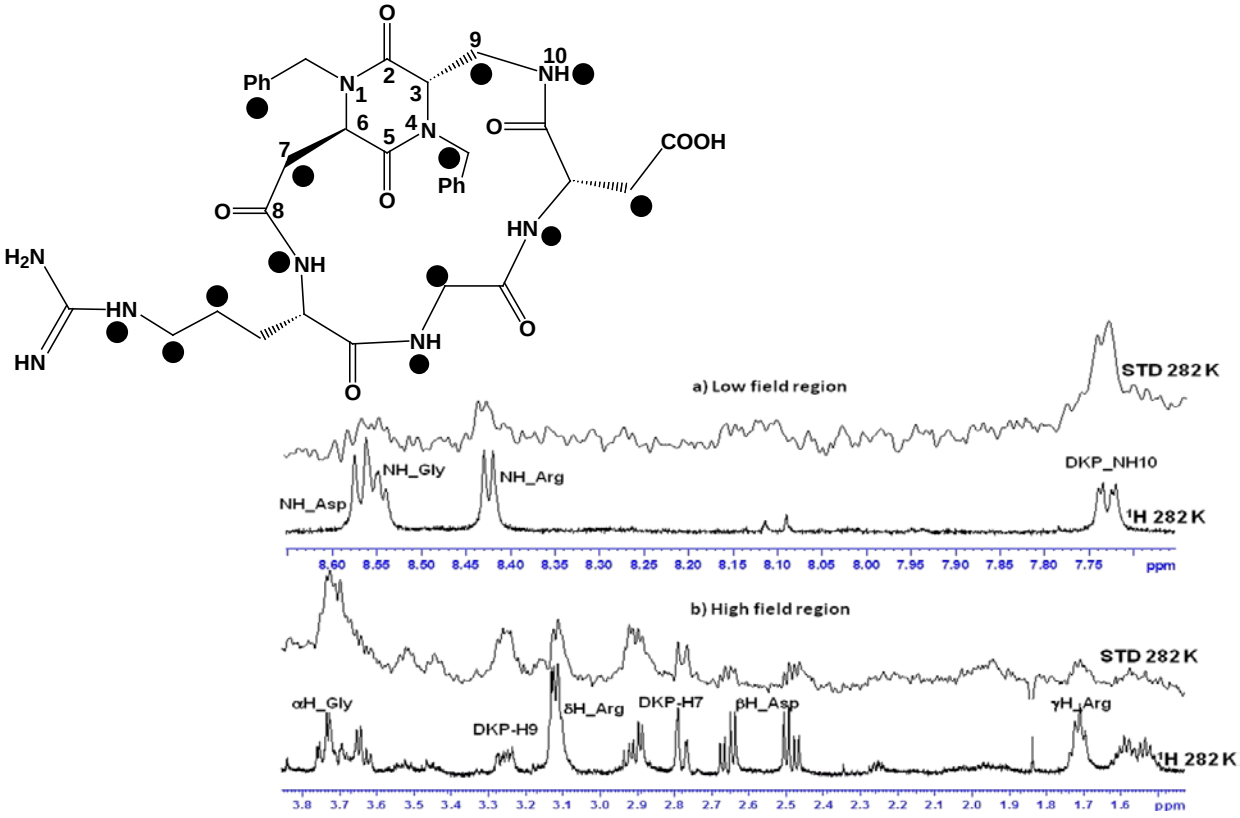


Figure 2.2.39. Binding epitope: map of the STD signals (left) of ligand **7-B** in the presence of ECV cells (black circles), STD and ^{1}H-NMR spectra (right) performed at 282 K, *a)* Amidic protons region (the aromatic zone was not reported due to the high intensity of STD signals), *b)* Aliphatic region.

Docking studies starting from the Type I-trans bound conformation of compound **7-B** (**Figure 2.2.37.**) produced two top-ranked binding modes (A and B), both conserving all the key polar interactions of the X-ray complex. The two binding modes (**Figure 2.2.40.**) are characterized by a different disposition of the ligand in the binding pocket. The binding mode A showed a good superposition of the whole molecule to the Cilengitide reference structure (**Table 2.2.8.**). In particular, the protons belonging to the RGD portion are in close contact with the receptor and the two benzyl moieties point toward the aromatic side chain of Tyr122, similarly to the (D)-Phe residue of Cilengitide. In the binding mode B, the ligand does not overlay to the RGD backbone of Cilengitide and only the electrostatic clamp is preserved. The comparison between the docking poses and the NMR data does not allow to choose one of the two binding modes, that probably both contribute to an optimal interaction.

a) b)

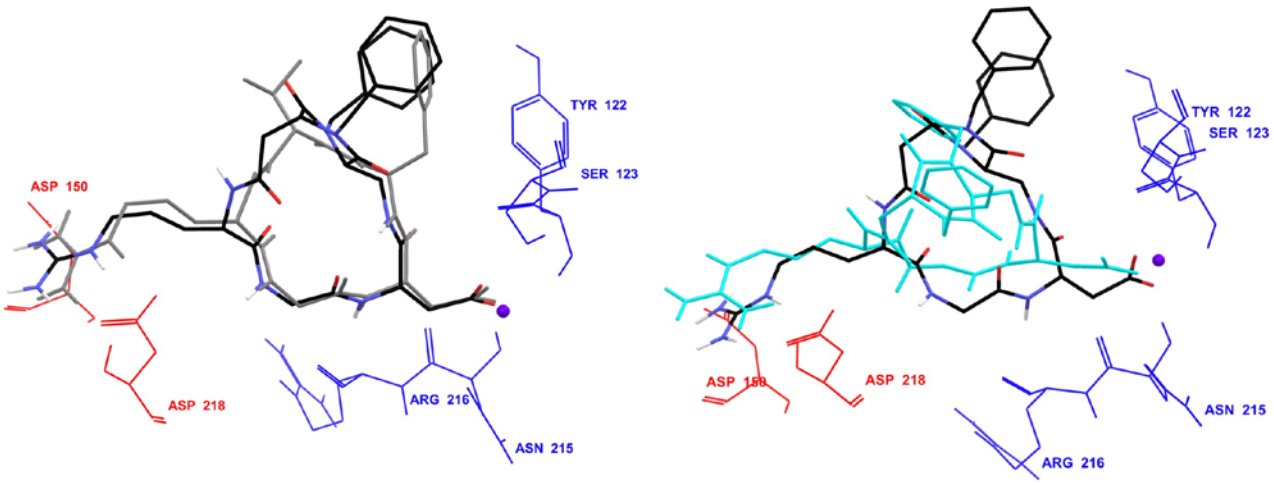


Figure 2.2.40. Left: Docking best pose (binding mode A) of compound **7-B** (tube representation, C atoms in black, N in blue, O in red, H in white) into the crystal structure of the extracellular domain of $\alpha_v\beta_3$ integrin (α unit red and β unit blue wire representation) overlaid on the bound conformation of Cilengitide (gray tube representation); the metal ion at MIDAS is shown as amagenta CPK sphere. Right: Superposition of A (black) and B (cyan) binding modes identified among the docking poses of ligand **7-B** into the $\alpha_v\beta_3$ integrin binding site.

Ligand protons	Binding mode A		Binding mode B	
	Residues	Mean distances	Residues	Mean distances
NH-Arg	CH_3-Ala215 (α)	3.84 Å	CH_3-Ala215 (α)	2.48 Å
γH-Arg	O_2=C-Asp218 (α)	3.07 Å	O_2=C-Asp218 (α)	3.66 Å
δH-Arg	HAr-Tyr178 (α)	2.79 Å	HAr-Tyr178 (α)	3.84 Å
NH_Guan	O_2=C-Asp218 (α)	2.56 Å	O_2=C-Asp150 (α)	3.87 Å
*NH-Gly	HAr-Tyr178 (α)	3.31 Å	HAr-Tyr178 (α)	3.27 Å
αH-Gly	HAr-Tyr178 (α)	2.74 Å	HAr-Tyr178 (α)	2.30 Å
*NH-Asp	αCH-Arg216 (β)	3.37 Å	αCH-Arg216 (β)	4.66 Å (Min=2.70 Å)
βH-Asp	O=C-Asn215 (β)	3.59 Å	O=C-Asn215 (β)	3.70 Å
H_Ar	HAr-Tyr122 (β)	6.97 Å (Min=3.24 Å)	NH_Guan-Arg248 (α)	3.24 Å
DKP-H7	CH_3-Ala215 (α)	4.34 Å (Min=2.92 Á)	CH_3-Ala215 (α)	2.50 Å
DKP-H9	βCH_2-Ser123 (β)	3.53 Å	CH_3-Ala252 (β)	3.28 Å
DKP-NH10	βCH_2-Ser123 (β)	4.27 Å (Min=2.84 Å)	CH_3-Ala218 (β)	3.94 Å (Min=2.53 Å)

Table 2.2.8. Average values of the distances between ligand protons (selected from STD experiment) and $\alpha_v\beta_3$ residues calculated from docking poses for the two binding modes A and B. Asterisks indicate overlapping protons.

These data well support the sub-nanomolar IC_{50} value in the $\alpha_v\beta_3$ binding assay of this compound (IC_{50} 0.2 ± 0.09 nM) and its low micromolar IC_{50} value in the adhesion assays (ECV-304 IC_{50}(VN) = 6.6 ± 0.8 μM).

As discussed above for compounds **2**, **4** and **5** derived from the (3R, 6S) DKP scaffold, also for the set of compounds **3**, **6** and **7** derived from the (3S, 6R) DKP scaffold both the bound conformation and the scaffold substitution strongly influenced the disposition of the ligand in the binding pocket. For this set of ligands, the identification of the experimental binding epitope of compounds **6** and **7** was complicated by the presence of several overlapping signals into STD spectra. The most active compound of the DKP series, the diastereomer **7-B** (**Figure 2.2.32.**, black bars), is characterized by high STD absolute percentages and lots of interactions with the receptor. On the other hand, the diastereomer **7-A** (**Figure 2.2.32.**, green bars), showing the lowest affinity for the $\alpha_V\beta_3$ integrin, displayed a STD pattern of signals intermediate between those of the compounds **3** (**Figure 2.2.32.**, blu bars) and **7-B**. Entropic reasons, *i.e.* the presence of both extended and non-extended geometries in the free state conformational equilibrium, can account for the lower affinity for the $\alpha_V\beta_3$ integrin of the isomer **7-A** compared to the other *trans* DKP-RGD ligands.

2.2.3 INTERACTIONS WITH MDA-MB-231 BREAST CANCER CELLS

The ECV-304 cell line expresses different integrins on the membrane (see "**Cell lines expressing integrins**"). In particular, integrins $\alpha_V\beta_3$ and $\alpha_5\beta_1$ show the same high level of expression ($\alpha_5\beta_1/\alpha_V\beta_3$ expression of 1.16). Since also integrin $\alpha_5\beta_1$ recognizes the RGD sequence, we wondered if the NMR results obtained on ECV-304 cells might be affected by the presence of both the integrins. For this reason, we performed *tr*-NOESY and STD NMR experiments using the MDA-MB-231 breast cancer cell line, known to overexpress only integrin $\alpha_5\beta_1$. MDA-MB-231 was then considered the reference cell line to investigate the interaction of the compounds with the $\alpha_5\beta_1$ integrin. In parallel, molecular docking calculations starting from the crystal structure of the extracellular segment of $\alpha_5\beta_1$ in complex with a linear RGD peptide[17] were implemented with the aim to obtain computational models for the interaction of the DKP-RGD ligands with the $\alpha_5\beta_1$

integrin. In addition, competitive binding assays on the isolated $\alpha_5\beta_1$ integrin and adhesion assays using MDA-MB-231 cells were carried out.

As previously discussed, the preferred free state conformations strongly influence the ligand binding to integrin receptors. Ligands **2** – **6** and **7-B** mainly adopt two types of extended conformations, regardless of the stereochemistry of the DKP scaffold: the type I-trans conformation (compounds **4**, **6** and **7-B**) and the type III conformation (compounds **2**, **3** and **5**). Accordingly, NMR results on MDA-MB-231 cells and docking calculations in the $\alpha_5\beta_1$ model will be discussed on the basis of the ligand bound conformation. First, the results obtained for the compounds adopting the type I-trans geometry and showing the highest affinities for the receptor $\alpha_5\beta_1$ (compounds **7-B**, **6** and **4**) will be discussed, than those obtained for the compounds adopting the type III geometry (compounds **2**, **3** and **5**). Ligands featuring non-extended RGD conformations (**1** and **7-A**) were not included in the study focused on the $\alpha_5\beta_1$ integrin, as this receptor is known to recognize an extended RGD arrangement almost identical to that recognized by the $\alpha_V\beta_3$ integrin.[18]

2.2.3.1 The docking model of $\alpha_5\beta_1$

β_1 integrins constitute the largest integrin subfamily and recognize a wide range of ligands depending on the partner α subunit. $\alpha_5\beta_1$ was one of the first integrin heterodimers to be identified, and its function as fibronectin receptor has been extensively studied.[46] However, the X-ray structure of this integrin closed headpiece in complex with SG/19 Fab (an allosteric inhibitory antibody) was only recently solved both in the absence and presence of a linear peptide ligand containing the Arg-Gly-Asp (RGD) sequence (PDB entry 3VI4).[17] This structure revealed high similarity to the ligand unbound form of $\alpha_V\beta_3$ and $\alpha_{IIb}\beta_3$ integrins. The most important differences between the receptor binding pockets of $\alpha_V\beta_3$ and $\alpha_5\beta_1$ are shown in the **Figures 2.2.41.** and **2.2.42.**. The mutation of Asp150 (α_V) into Ala159 (α_5) produces a less acidic region, while the mutation of Arg214 and Arg216 (β_3) into the hydrophobic residues Gly223 and Leu225 (β_1) both expands the site and reduces the polar character of the region.

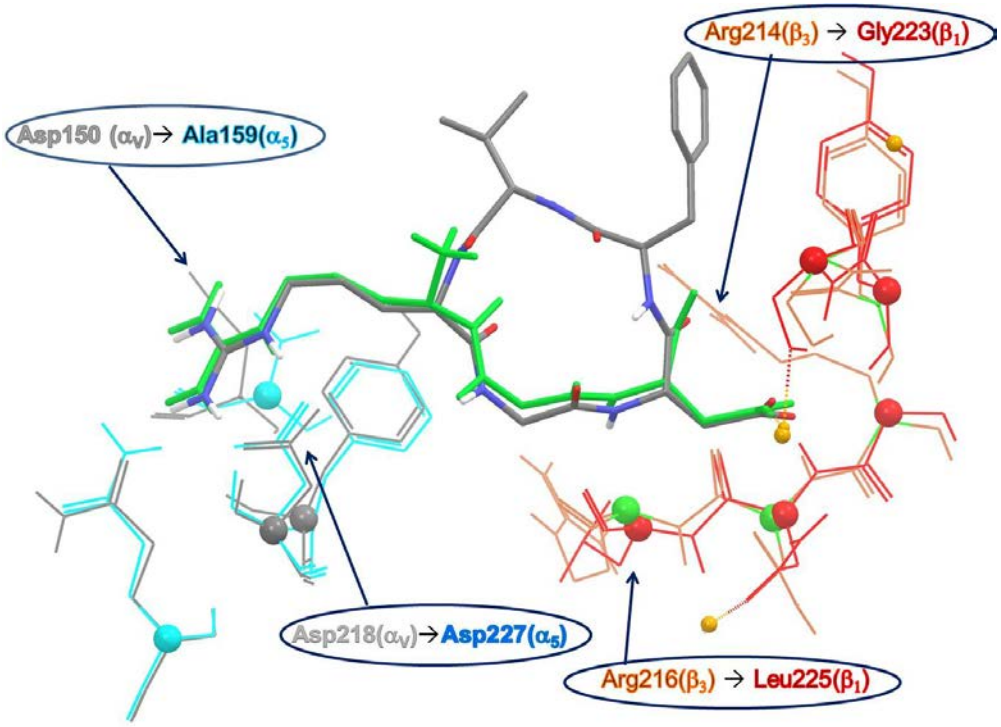


Figure 2.2.41. Schematic representation of the differences between the binding sites of $\alpha_V\beta_3$ and $\alpha_5\beta_1$ receptors; Cilengitide X-ray structure in $\alpha_V\beta_3$ (gray), α_V chain (gray) and β_3 chain (orange), linear RGD sequence in $\alpha_5\beta_1$ (green), α_5 chain (light blue) and β_1 chain (red).

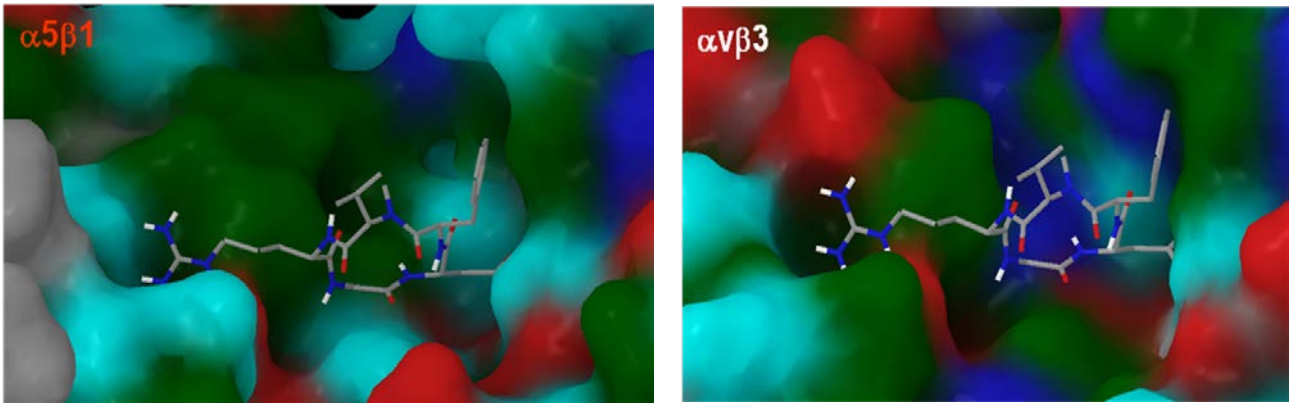


Figure 2.2.42. Cilengitide docking best poses into the $\alpha_5\beta_1$ and $\alpha_V\beta_3$ binding sites: molecular surfaces coloured by residue property (red = acidic, blue = basic, green = hydrophobic, cyan = neutral).

Notwithstanding slight differences, the X-ray structures of $\alpha_V\beta_3$ and $\alpha_5\beta_1$ complexed with RGD ligands reveal an identical atomic basis for the interaction: RGD binds to the interface of the α and β subunits, the Arg residue fitting into a cleft in the α subunit, and the Asp coordinating a

cation in the β subunit. The RGD sequence is disposed in an extended conformation in the receptor site with the carboxylic and guanidine groups acting as an electrostatic clamp, respectively, on the Ca^{2+} ions (MIDAS) of the β subunit and on the Asp residues of the α chain.

The distance between the negatively charged carboxylate group (Asp Cγ labeled C-) and the positively charged guanidinium moiety (Arg Cζ labeled C+) of the ligand is similar in the two crystal structures (Cilengitide C+/C- distance in $\alpha_V\beta_3$ = 14.15 Å, linear RGD C+/C- distance in $\alpha_5\beta_1$ = 14.22 Å) and also the distance between the corresponding receptor anchor points is almost the same (distance Ca^{2+} - CγAsp218 in $\alpha_V\beta_3$ = 12.95 Å, distance Ca^{2+}-CγAsp227 in $\alpha_5\beta_1$ = 13.26 Å). In particular, the Arg side chain of the RGD peptide ligand in $\alpha_5\beta_1$ makes hydrogen bonds with α_5 residues Gln221 and Asp227, the Asp carboxylate directly coordinates the β_1 MIDAS Ca^{2+}, making also hydrogen bonds with the backbone NH of Tyr133 (β_1) and Asn224 (β_1). The docking protocol was developed from the recently solved crystal structure of the $\alpha_5\beta_1$ integrin (PDB entry 3VI4)[17] and validated by evaluating its ability to reproduce the X-ray binding geometry of the linear hexapeptide and to properly dock Cilengitide, known to be a nanomolar ligand of this receptor.[21]

2.2.3.2 In vitro biological assays of compounds 2 - 7

Integrin receptor competitive binding assays

The cyclic RGD peptidomimetics were examined *in vitro* for their ability to inhibit biotinylated fibronectin binding to the purified $\alpha_5\beta_1$ receptor (**Table 2.2.9.**).

Structures	$\alpha_V\beta_3$ IC_{50} [nM][a]	**$\alpha_5\beta_1$ IC_{50} [nM][a]**	$\alpha_5\beta_1/\alpha_V\beta_3$ IC_{50} [nM]
cyclo[**DKP-2**-RGD]	3.2 ± 2.7	**1647 ± 871**	514.7
cyclo[**DKP-3**-RGD]	4.5 ± 1.1	**532 ± 35**	118.2
cyclo[**DKP-4**-RGD]	7.6 ± 4.3	**314 ± 1.4**	41.3
cyclo[**DKP-5**-RGD]	12.2 ± 5.0	**1217 ± 157**	99.8
cyclo[**DKP-6**-RGD]	2.1 ± 0.6	**73.3 ± 3.7**	34.9
cyclo[**DKP-7**-RGD]-A	220.2 ± 82.3	**> 10^{-5}**	///
cyclo[**DKP-7**-RGD]-B	0.2 ± 0.09	**25.7 ± 6.2**	128.5
cyclo[RGDfV]	3.2 ± 1.3	**166 ± 28**	51.9
cyclo[RGDf(N-Me)V][21] (Cilengitide)	0.6 ± 0.1	**14.4 ± 3.1**	24

Table 2.2.9. Inhibition of biotinylated vitronectin binding to $\alpha_V\beta_3$ receptor and of biotinylated fibronectin binding to $\alpha_5\beta_1$ receptor. [a] IC_{50} values were calculated as the concentration of

compound required for 50% inhibition of biotinylated vitronectin and fibronectin binding as estimated by GraphPad Prism software; all values are the arithmetic mean ± SD of triplicate determinations.

Screening assays were performed incubating the immobilized integrin receptors with various concentrations (10^{-12}–10^{-5} M) of the RGD ligands in the presence of biotinylated fibronectin (1 μg/mL), and measuring the concentration of bound fibronectin in the presence of the competitive ligands (see "**Experimental Section**"). The ability of the new compounds to inhibit the binding of fibronectin to the isolated $\alpha_5\beta_1$ receptor was compared with that of the reference compounds *cyclo*[RGDfV] and Cilengitide.

The compounds were able to inhibit biotinylated fibronectin binding to integrin $\alpha_5\beta_1$ at concentrations ranging from nanomolar (**7-B** and **6**) to sub-micromolar (**4** and **3**) and micromolar (**2** and **5**) IC_{50} values. Compound **7-A** appeared to be the worst ligand, confirming the low conformational preorganization also for binding to $\alpha_5\beta_1$ integrin. Compared with the binding results for $\alpha_V\beta_3$ receptor, all the DKP ligands and the reference peptides lost at least one order of magnitude in affinity ($\alpha_5\beta_1$ IC_{50} values 30 - 500 times higher than $\alpha_V\beta_3$ IC_{50} values). Since the X-ray structures reveal similar RGD geometrical requirements for the binding to $\alpha_5\beta_1$ and $\alpha_V\beta_3$ integrins, the observed selectivity may depend on the physicochemical features of the binding sites. In fact, the receptor region that binds to the ligand guanidinium group is less acidic in $\alpha_5\beta_1$ than in $\alpha_V\beta_3$, thus leading to a weaker interaction of the ligand Arg side chain with the α_5 subunit. Moreover, due to both the less efficient electrostatic clamp and the larger binding pocket compared to $\alpha_V\beta_3$, the ligands interact less strongly with $\alpha_5\beta_1$ and may reveal a greater mobility within the receptor site. Interestingly, compounds characterized by a type I-trans geometry (**7-B**, **6** and **4**) showed the highest affinities for the $\alpha_5\beta_1$ integrin.

Cell adhesion assays

Adhesion assay tests were performed in two different laboratories using the same MDA-MB-231 breast cancer cell line and experimental protocol (see "**Experimental section**" for details). The results obtained for compounds **2** and **3** and the reference peptide *cyclo*[RGDfV] at the cell

biology laboratory of CISI (an Interdepartmental Centre of Milan University), that first developed the adhesion assay protocol for both ECV-304 and MDA-MB-231 cells, are shown in **Table 2.2.10.**. As for ECV-304 cells, both ligands showed a micromolar activity in the inhibition of MDA-MB-231 cell adhesion to both VN and FN.

Ligands	ECV-304		MDA-MB-231	
	Vitronectin IC_{50} [μM][a]	Fibronectin IC_{50} [μM][a]	**Vitronectin IC_{50} [μM][a]**	Fibronectin IC_{50} [μM][a]
cyclo[DKP-2-RGD]	**13.8 ± 0.2**	2.2 ± 0.4	**6.8 ± 2.9**	3.0 ± 0.9
cyclo[DKP-3-RGD]	**49.5 ± 3.0**	15.0 ± 3.9	**1.6 ± 0.0**	0.7 ± 0.1
cyclo[RGDfV]	**0.22 ± 0.00**	0.36 ± 0.09	**11.3 ± 0.1**	3.4 ± 0.4

Table 2.2.10. Effect of DKP-RGD compounds on ECV-304 and MDA-MB-231 cell adhesion to vitronectin and fibronectin. [a] Each data point was performed in triplicate in two independent experiments. N.D. = Not Detectable, N.R. = Not Reproducible.

Ligands	ECV-304		MDA-MB-231	
	Vitronectin IC_{50} [μM][a]	Fibronectin IC_{50} [μM][a]	**Vitronectin IC_{50} [μM][a]**	Fibronectin IC_{50} [μM][a]
cyclo[DKP-4-RGD]	**> 70**	0.58 ± 0.21	**9.7 ± 4.4**	0.27 ± 0.02
cyclo[DKP-5-RGD]	**17.3 ± 7.6**	N.R.	**9.4 ± 2.2**	N.R.
cyclo[DKP-6-RGD]	**6.8 ± 2.0**	> 90	**4.6 ± 2.2**	N.D.
cyclo[DKP-7-RGD]-A	**N.D.**	N.D.	**> 200**	N.D.
cyclo[DKP-7-RGD]-B	**6.6 ± 0.8**	N.D.	**3.8 ± 1.1**	> 100
cyclo[RGDfV]	**2.00 ± 0.35**	0.24 ± 0.03	**0.90 ± 0.21**	0.19 ± 0.08

Table 2.2.11. Effect of DKP-RGD compounds on ECV-304 and MDA-MB-231 cell adhesion to vitronectin and fibronectin. [a] Each data point was performed in triplicate in two independent experiments. N.D. = Not Detectable, N.R. = Not Reproducible.

The results obtained for compounds **4** - **7** and the reference peptide *cyclo*[RGDfV] at the Department of Medical Biotechnology and Translational Medicine (University of Milan) are shown in **Table 2.2.11.**. Except for **7-A**, all the compounds show a micromolar activity in the inhibition of MDA-MB-231 cell adhesion to VN. Similarly to what is observed for ECV-304 cells, only ligand **4** is able to inhibit MDA-MB-231 cell adhesion to FN at micromolar concentrations. The results of FN-mediated adhesion assays do not correlate with the data obtained in the isolated receptor binding assays, especially for compound **7-B**, the highest affinity ligand of the library in the cell-free test. An explanation of such behaviour may be difficult considering only the molecular basis of the

ligand - protein recognition process. Several aspects (*e.g.* biological variability given by the cell culture microenvironment) could affect and contribute to the complexity and dynamics of a cell-ligand system.

2.2.3.3 Type I-trans bound conformation: NMR and docking results

As already discussed, the preferred conformation adopted by ligands strongly influenced their disposition in the receptor binding pocket and, as consequence, their NMR epitope. The STD results obtained with MDA-MB-231 cells at 282 K are described on the basis of ligand bound conformation in comparison with those obtained with ECV-304 cells at the same temperature.

Compound 7-B

The *tr*-NOESY experiment of compound **7-B** in the presence of MDA-MB-231 cells at 282 K showed negative cross-peaks as the NOESY spectrum. As a result, binding evidence cannot be obtained from this experiment. However, it showed a negative cross-peak between NH_{Asp} (8.57 ppm) and NH_{10} (7.73 ppm) (**Figure 2.2.43.**), consistent with the β-turn conformation at Gly-Asp found in the free state (type I-trans).

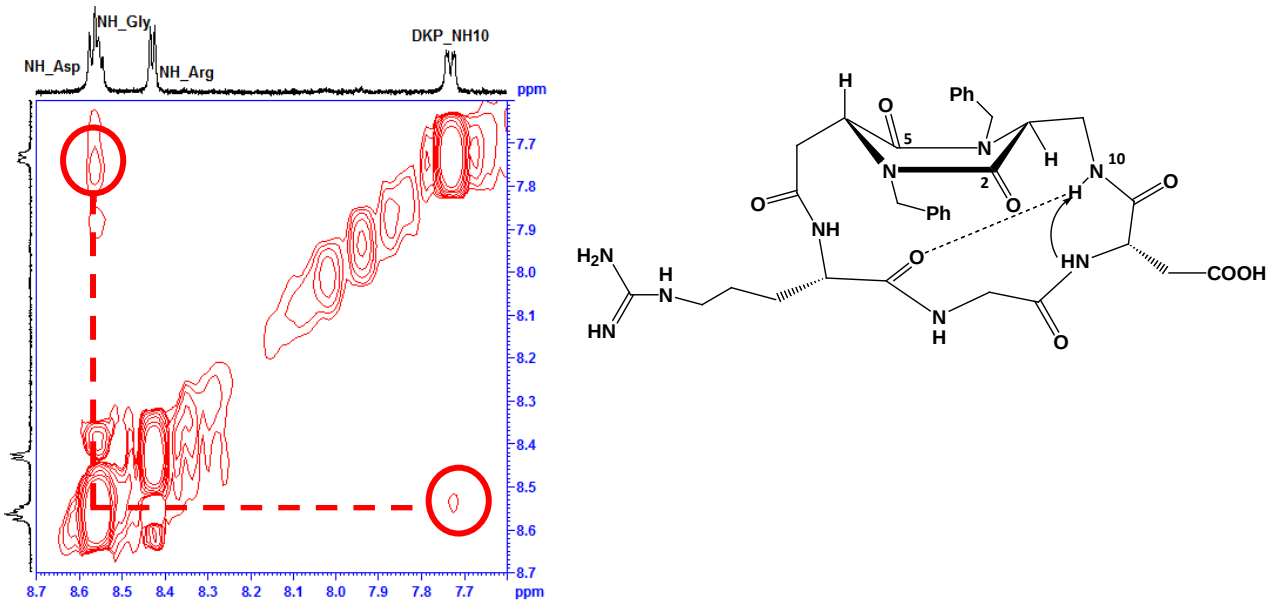


Figure 2.2.43. NOE contact between NH_{Asp} and NH_{10} indicating a Type I-trans bound conformation for compound **7-B** (β-turn at Gly-Asp, right). Hydrogen bond (dotted line) and NOE contact (arrow) are indicated.

The largest STD effects in the presence of MDA-MB-231 cells are observed for DKP-H9, DKP-H3 and $H\alpha_{Gly}$ (1.8, 1.4 and 1.1% absolute STD percentages, respectively) (**Figure 2.2.44.,** black dots and bars). Many other protons appeared in the spectrum, the guanidine NH proton and the aromatic protons (both 0.6% absolute STD percentages), the NH_{10} (0.9% absolute STD percentages), NH_{Arg} (0.6% absolute STD percentages), $H\delta_{Arg}$ and DKP-H7 (0.2 and 1.0% absolute STD percentages, respectively). Additionally, the signals of NH_{Gly} and NH_{Asp} overlap (0.2% STD absolute intensity) and it is difficult to assess their involvement in the binding process. So, the interaction of **7-B** with the receptor site involves many protons of both the RGD sequence and the DKP moiety. Compared with the STD data obtained with ECV-304 cells (**Figure 2.2.44.**, orange dots and bars), a similar pattern is observed, although higher intensities and more signals (DKP-H3) of DKP protons are detected with MDA-MB-231 cells. For the RGD signals, Arg and Asp side chains seem to interact more tightly with the receptors in the presence of ECV-304 cells.

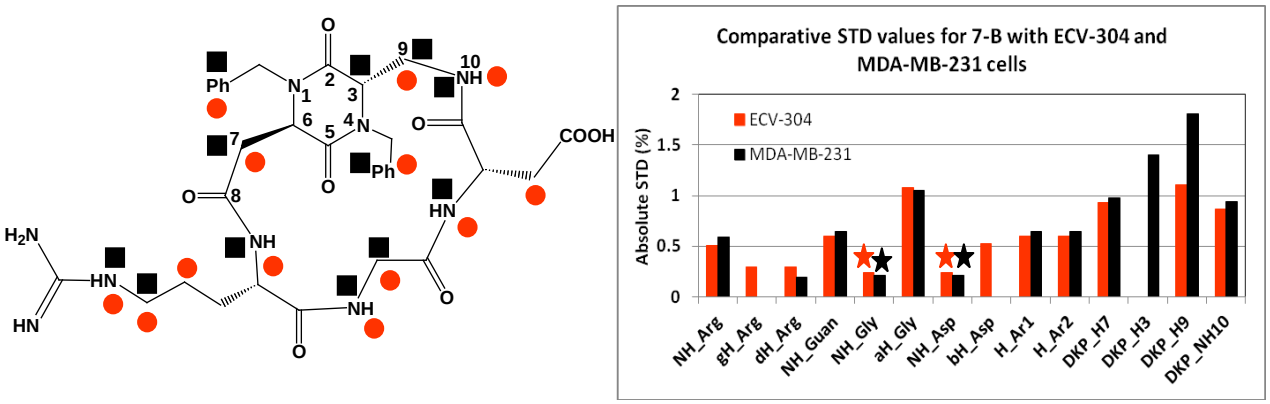


Figure 2.2.44. Binding epitope: a) map of the STD signals of ligand **7-B** in the presence of ECV-304 cells (orange dots) with respect to MDA-MB-231 cells (black squares), b) relevant absolute STD percentages for the protons of compound **7-B** obtained at T = 282 K in the presence of ECV-304 (orange bars) and MDA-MB-231 (black bars) cells. Stars indicate overlapped protons.

In general, the map of the STD signals obtained for ligand **7-B** in the presence of ECV-304 and MDA-MB-231 cells (**Figure 2.2.44.**) shows that the ligand epitope is conceivably the same.

Docking calculations of ligand **7-B** into the $\alpha_5\beta_1$ integrin binding site produced top-ranked poses conserving optimal interactions with the receptor, as shown in **Table 2.2.12.**. This result is in

agreement with the nanomolar affinity reported for the isolated $\alpha_5\beta_1$ receptor. Moreover, the ligand binding modes described by docking calculations in $\alpha_5\beta_1$ and $\alpha_V\beta_3$ models are similar (**Figure 2.2.45.**) and well interpret the NMR data.

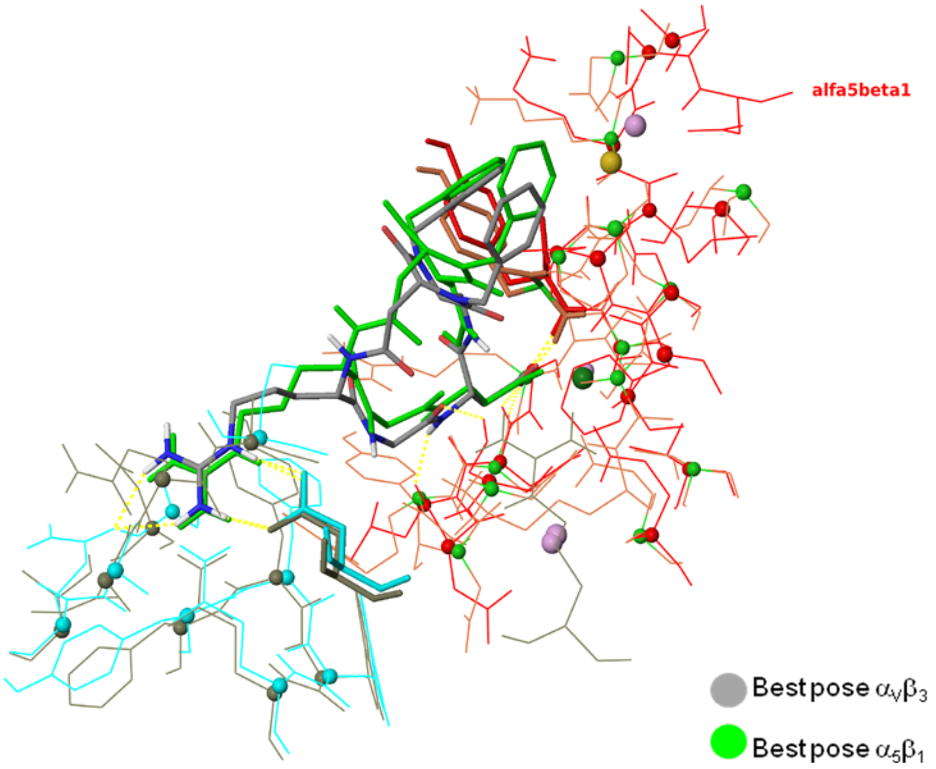


Figure 2.2.45. Schematic representation of the differences between the two docking best poses of compound **7-B** in $\alpha_V\beta_3$ and $\alpha_5\beta_1$ receptors; superimposition of the docking best pose in $\alpha_V\beta_3$ (gray), α_V chain (gray) and β_3 chain (orange) with the docking best pose in $\alpha_5\beta_1$ (green), α_5 chain (light blue) and β_1 chain (red).

Ligand protons	Pose in $\alpha_5\beta_1$	
	Residues	Mean distance
NH_Arg	βCH_2-Ser224 (α)	4.52 Å
δH_Arg	HAr-Phe187 (α)	3.21 Å
NH_Guan	O_2=C-Asp227 (α)	1.90 Å
*NH_Gly	βCH_2-Ser227 (β)	2.62 Å
αH_Gly	HAr-Phe187 (α)	2.84 Å
*NH_Asp	O=C-Leu225 (β)	2.72 Å
H_Ar	HAr-Tyr133 (β)	2.77 Å
DKP_H7	O=C-Glu320 (β)	4.72 Å
DKP_H3	βCH_2-Ser134 (β)	4.68 Å
DKP_H9	βCH_2-Ser134 (β)	2.70 Å
DKP_NH10	HO-Ser227 (β)	4.02 Å

Table 2.2.12. Average values of the distances between ligand protons (selected from STD experiment) and $\alpha_5\beta_1$ residues calculated from docking poses. Asterisks indicate overlapping protons.

For both the receptors the compound **7-B** is the most potent ligand of the series and it effectively inhibits the binding of biotinylated vitronectin and fibronectin to the isolated $\alpha_V\beta_3$ and $\alpha_5\beta_1$ integrins at sub-nanomolar and nanomolar concentrations, respectively.

Compound 6

The *tr*-NOESY experiment of compound **6** showed negative NOE contacts like in the free state and does not permit to evaluate the binding event. It showed a cross-peak between NH_{Asp} and NH_{10}, consistent with the type I-trans H-bonding pattern found in the free state and in the bound state in the presence of ECV membrane proteins (**Figure 2.2.46.**). This could indicate that, upon binding to cells, the conformational equilibrium present in the free state is shifted towards one conformation, and precisely to the most abundant in the free state.

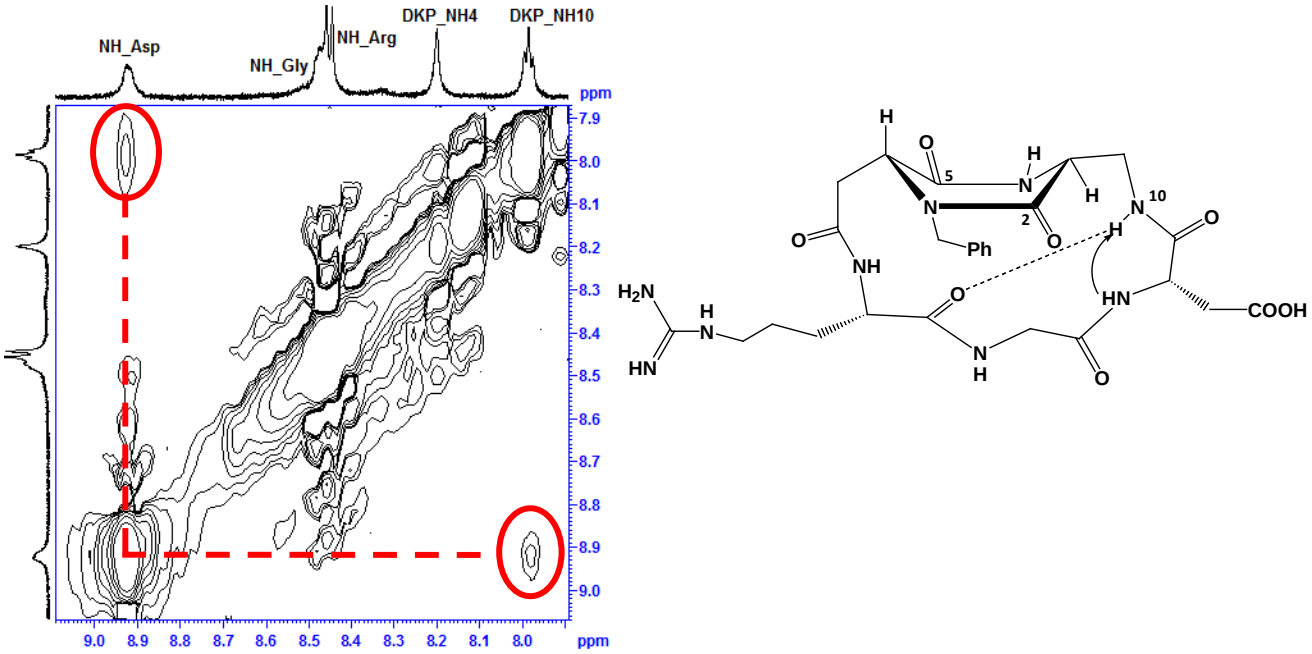


Figure 2.2.46. NOE contact between NH_{Asp} and NH_{10} indicating a Type I-trans bound conformation for compound **6** (β-turn at Gly-Asp, right). Hydrogen bond (dotted line) and NOE contact (arrow) are indicated.

The largest STD effects with MDA-MB-231 cells are observed for the aromatic protons and for the $H\alpha_{Gly}$ (1.3 and 0.8% absolute STD intensity, respectively) (**Figure 2.2.47.**). The guanidine NH proton, NH_4 and DKP-H9 showed a similar STD effect (0.7, 0.7 and 0.6% absolute STD intensity, respectively). Additionally, the signals of $H\beta_{Asp}$ and DKP-H7 overlap (0.2% STD absolute intensity) and it is difficult to assess their involvement in the binding process.

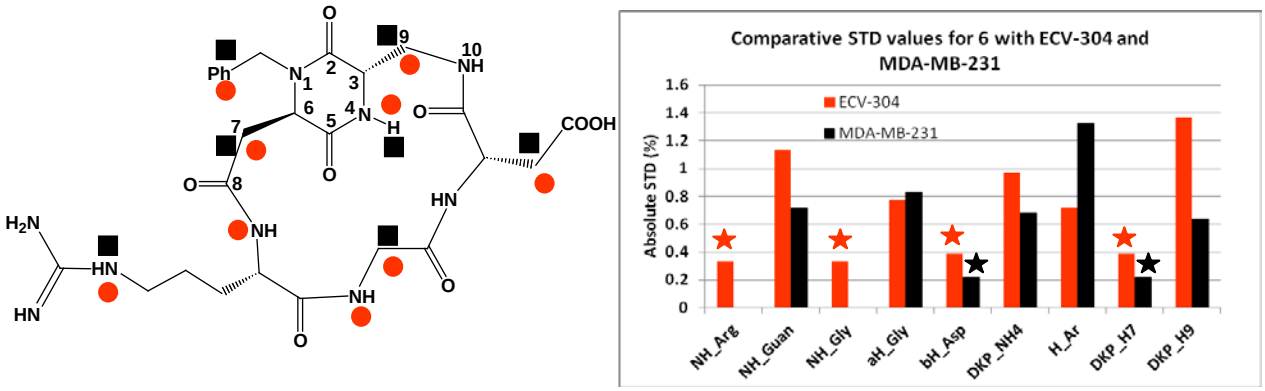


Figure 2.2.47. Binding epitope: a) map of the STD signals of ligand **6** in the presence of ECV-304 cells (orange dots) with respect to MDA-MB-231 cells (black squares), b) relevant absolute STD

percentages for the protons of cyclic DKP-RGD **6** obtained at T = 282 K in the presence of ECV-304 (orange bars) and MDA-MB-231 (black bars) cells. Stars indicate overlapped protons.

The comparison of the STD signals obtained for ligand **6** in the presence of ECV-304 and MDA-MB-231 cells (**Figure 2.2.47.**) shows that the ligand epitope is similar. Some differences in signal intensities are observed both for the DKP and RGD moieties and NH_{Gly} and NH_{Arg} signals (overlapping with ECV-304 cells) are missing in presence of MDA-MB-231 cells. As observed for compound **7-B**, it seems that the interaction of the RGD sequence with the receptors is stronger with ECV-304 than with MDA-MB-231 cells. In agreement with the nanomolar binding affinity, docking calculations of ligand **6** into the $\alpha_5\beta_1$ integrin binding site produced top-ranked poses conserving optimal interactions with the receptor (**Figure 2.2.48.**). However, the analysis of the docking poses into $\alpha_V\beta_3$ and $\alpha_5\beta_1$ models reveals different preferred binding modes of compound **6** with the two receptors. In $\alpha_V\beta_3$ ligand **6** well overlaps to the Cilengitide crystal structure (**Figure 2.2.34**) and directs the aromatic moiety toward the Tyr122 side chain of the β chain; in $\alpha_5\beta_1$ the compound orientates the scaffold and the aromatic ring toward a different region of the receptor site (**Figure 2.2.48.**). This binding mode affects the interaction of the whole ligand with the receptor and, as a consequence, could explain the lack of STD signals for NH-Arg and NH-Gly protons and the lower STD intensity observed for most signals in the presence of MDA-MB-231 with respect to ECV-304 cells.

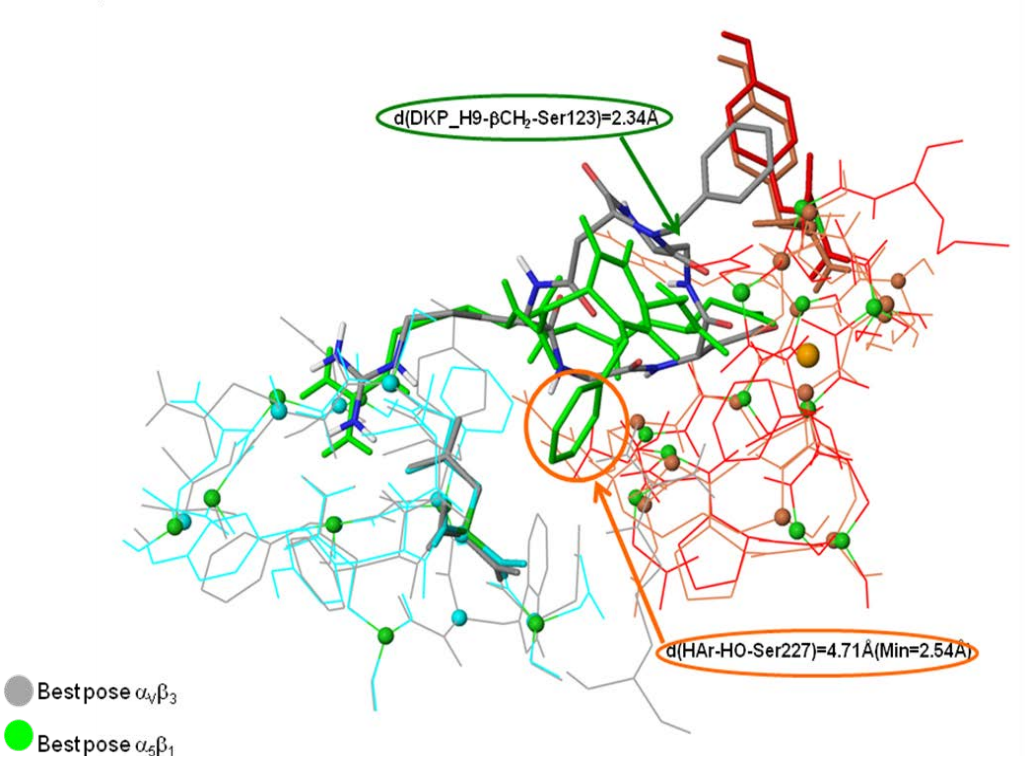


Figure 2.2.48. Schematic representation of the differences between the two docking best poses of compound **6** in $\alpha_V\beta_3$ and $\alpha_5\beta_1$ receptors; superimposition of the docking best pose in $\alpha_V\beta_3$ (gray), α_V chain (gray) and β_3 chain (orange) with the docking best pose in $\alpha_5\beta_1$ (green), α_5 chain (light blue) and β_1 chain (red).

Ligand protons	Pose in $\alpha_5\beta_1$	
	Residues	Mean distances
NH_Guan	O_2=C-Asp227 (α)	2.06 Å
αH_Gly	HAr-Phe187 (α)	2.27 Å
*βH_Asp	O=C-Asn224 (β)	3.15 Å
DKP-NH4	CH_2-Ser134 (β)	4.72 Å
H_Ar	HO-Ser227 (β)	4.71 Å Min=2.54 Å
*DKP-H7	CH_2-Ser224 (α)	4.32 Å
DKP-H9	CH_2-Ser134 (β)	2.74 Å

Table 2.2.14. Average values of the distances between ligand protons (selected from STD experiment) and $\alpha_5\beta_1$ residues calculated from docking poses. Asterisks indicate overlapping protons.

Compound 4

Even if the *tr*-NOESY and NOESY contacts are negative, the superimposition of the cross-peaks of both spectra confirms that in the bound state the ligand adopts the same conformation of the free state (β-turn at Gly-Asp, Type I-trans H-bonding pattern) (**Figure 2.2.49.**).

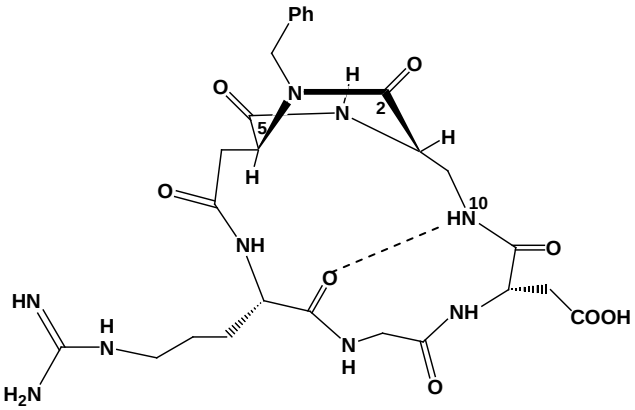
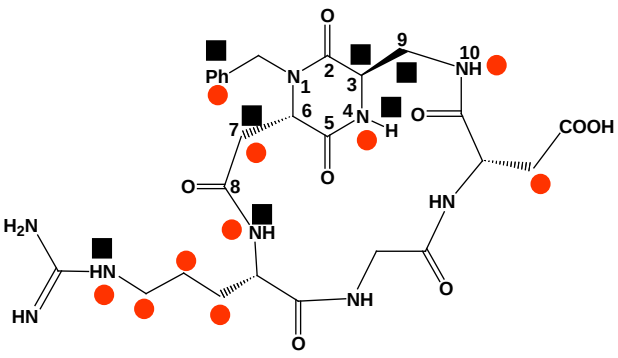


Figure 2.2.49. Bound conformation of *trans*-DKP-RGD-peptidomimetic **4**. Hydrogen bond (dotted line).

The largest STD effects with MDA-MB-231 cells are observed for the aromatic protons and for DKP-H7 (1.2 and 1.1% absolute STD percentages) (**Figure 2.2.50.**). Additionally, we note the presence of NH_{Arg}, the guanidine NH proton, NH_4, DKP-H3 and DKP-H9 (0.9, 0.6, 0.8, 0.6 and 0.5% absolute STD intensity, respectively).



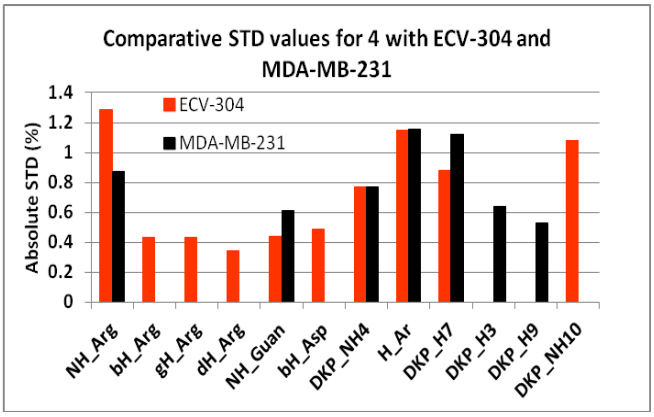


Figure 2.2.50. Binding epitopes: a) map of the STD signals of ligand **4** in the presence of ECV-304 cells (orange dots) with respect to MDA-MB-231 cells (black squares), b) relevant absolute STD

percentages for the protons of cyclic DKP-RGD **4** obtained at T = 282 K in the presence of ECV-304 (orange bars) and MDA-MB-231 (black bars) cells.

The comparison of the STD signals obtained for ligand **4** in the presence of ECV-304 and MDA-MB-231 cells (**Figure 2.2.50.**) shows that the ligand epitope is significantly different: *i)* the protons of the Arg-side chain appear with relevant STD percentages only in the interaction with ECV-304 (orange bars), *ii)* the protons DKP-H3 and DKP-H9 appear only in the interaction with MDA-MB-231 (black bars), *iii)* the protons βH_{Asp} and DKP-NH10 appear only in the interaction with ECV-304 (orange bars). This different behaviour allows us to make some considerations. As observed for compounds **7-B** and **6**, the interaction of the RGD sequence is stronger with ECV-304 cells. In particular, in the presence of MDA-MB-231 cells, compound **4** loses the interaction of the Asp side chain and prefers to interact with the DKP scaffold protons. On the basis of the different ligand epitope highlighted by STD-NMR with the two cell lines, a preferred interaction of ligand **4** with $\alpha_V\beta_3$ integrin can be hypothesized in the presence of ECV-304 cells (which express $\alpha_V\beta_3$ and $\alpha_5\beta_1$ in 1:1 ratio), whereas the interaction with $\alpha_5\beta_1$ integrin is characterized by NMR in the experiments with MDA-MB-231 cells. The higher affinity of compound **4** for $\alpha_V\beta_3$ compared to $\alpha_5\beta_1$ emerging from binding assays (see **Table 2.2.9.**) supports this hypothesis. As discussed for ligand **6**, top-ranked docking poses of compound **4** with the two receptors show different preferred binding modes (**Figure 2.2.51.**).

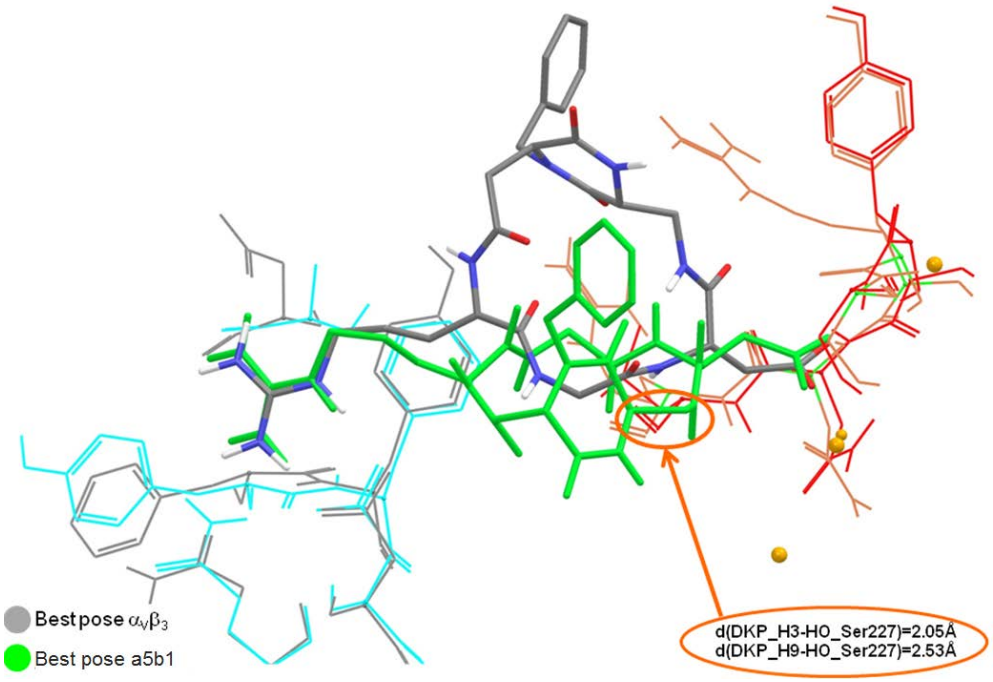


Figure 2.2.51. Schematic representation of the differences between the docking best poses of compound **4** in $\alpha_V\beta_3$ and $\alpha_5\beta_1$ receptors; superimposition of the docking best pose in $\alpha_V\beta_3$ (gray), α_V chain (gray) and β_3 chain (orange)with the docking best pose in $\alpha_5\beta_1$ (green), α_5 chain (light blue) and β_1 chain (red).

Ligand protons	Pose in $\alpha_5\beta_1$	
	Residues	Mean distance
NH_Arg	βCH_2-Ser224 (α)	3.16 Å
NH_Guan	O_2=C-Asp227 (α)	2.05 Å
DKP-NH4	HO-Ser227 (β)	2.38 Å
H-Ar	O_2=C-Glu320 (β)	2.59 Å
DKP-H7	βCH_2-Ser224 (α)	3.42 Å
DKP-H3	HO-Ser227 (β)	2.05 Å
DKP-H9	HO-Ser227 (β)	2.53 Å

Table 2.2.16. Average values of the distances between ligand protons (selected from STD experiment) and $\alpha_5\beta_1$ residues calculated from docking poses.

Moreover, due to the larger binding pocket of the integrin $\alpha_5\beta_1$, docking calculations show that compound **4** can fit into the pocket with different energetically favorable orientations, mainly differing in the position of DKP scaffold and cyclic peptide backbone. The mobility of the ligand

within the $\alpha_5\beta_1$ binding site revealed by docking poses is likely responsible of the non optimal interaction pattern observed for the RGD sequence with the MDA-MB-231 cells and of the sub-micromolar affinity for the $\alpha_5\beta_1$ receptor (IC_{50} 314 nM).

In conclusion, for the type I-trans geometry ligands we showed that small differences between epitopes were detected by STD-NMR experiments on ECV-3014 and MDA-MB-231 cells for the highest affinity $\alpha_5\beta_1$ ligands **7-B** and **6**, while relevant differences between epitopes were detected for ligand **4**. Therefore, only for this compound, the recognition of different receptors ($\alpha_V\beta_3$ or $\alpha_5\beta_1$) can be hypothesized in the interaction with ECV-304 and MDA-MB-231 cells.

2.2.3.4 Type III bound conformations: NMR and docking results

The results obtained for compounds **2**, **3** and **5** featuring a type III H-bonding pattern will be reported. These ligands are characterized by sub-micromolar or micromolar IC_{50} values in the binding assays on the isolated $\alpha_5\beta_1$ receptor.

Compound 2

The free state conformation of ligand **2** is characterized by a β-turn motif at DKP-Arg (Type III, **Figure 2.2.52.**), suggested by the relevant long-range NOE interaction between NH_{Gly} and NH_{Arg} (negative at 282 K).

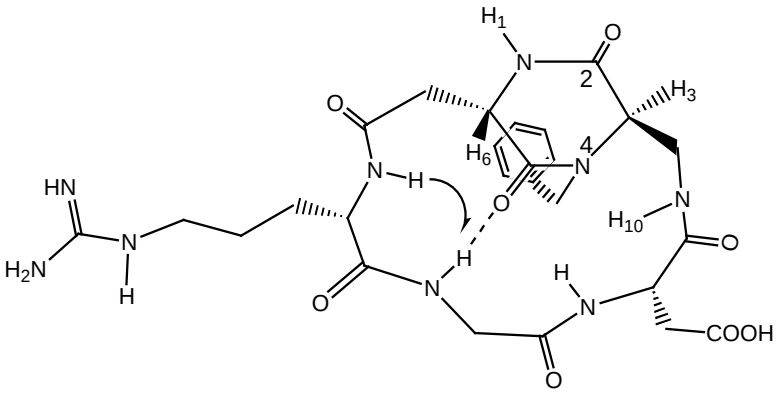


Figure 2.2.52. Free and bound conformation of *trans*-DKP-RGD-peptidomimetic **2**. Hydrogen bond (dotted line) and NOE contact (arrow) are indicated.

In the *tr*-NOESY we do not observe the NOE contact between NH_{Arg} and NH_{Gly} that characterizes ligand **2** in the free state, however the lack of differences between the other NOE contacts in the free and bound state suggests no change in the bound conformation with respect to the free state.

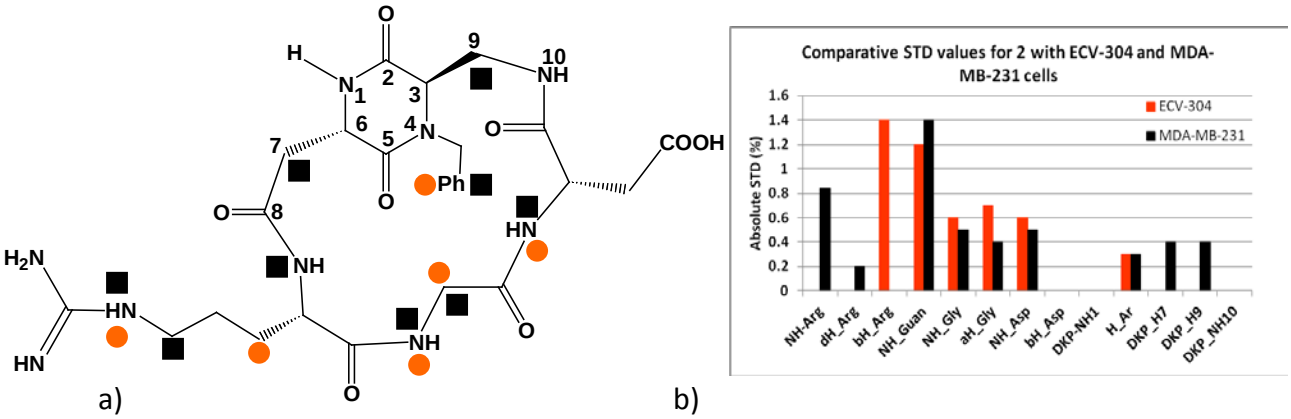


Figure 2.2.53. Binding epitopes: a) map of the STD signals of ligand **2** in the presence of ECV-304 (orange dots) and MDA-MB-231 cells (black squares), b) relevant absolute STD percentages for the protons of cyclic DKP-RGD **2** obtained at T = 282 K in the presence of ECV-304 (orange bars) and MDA-MB-231 (black bars) cells.

The comparison of the STD signals obtained for ligand **2** in the presence of ECV-304 and MDA-MB-231 cells (**Figure 2.2.53.**) shows that the epitope is different. In fact, in the interaction with MDA-MB-231 cells we can observe the involvement of some DKP scaffold protons (DKP-H7 and DKP-H9, 0.4 and 0.4% STD absolute intensity, respectively) that did not appear in the interaction with ECV-304 cells, as well as NH_{Arg} and γH_{Arg} (0.8 and 0.2% STD absolute intensity, respectively). The signals for the protons of the RGD sequence (NH_{Gly}, αH_{Gly} and NH_{Asp}, 0.5, 0.4 and 0.5% STD absolute intensity for MDA-MB-231) are similar in both the cell lines as well as the guanidine NH protons (1.4% STD absolute intensity for MDA-MB-231).

In the presence of MDA-MB-231 cells, the protons of both the DKP moiety and the RGD sequence are involved in the interaction. Docking calculations into the $\alpha_5\beta_1$ model identified two possible binding modes (A and B): in the best pose, representing the most populated binding mode A, the ligand forms all the key crystallographic interactions (**Figure 2.2.54.a**). In the second binding mode B, at less favourable docking scores, the ligand loses the interaction between the

guanidinium group and the side chain of α-Asp227 (the guanidinium group interacts with the aromatic ring of β-Phe321) and forms a stacking interaction between the ligand benzyl group and the β-Phe187 aromatic ring (**Figure 2.2.54.b**). A selection of the distances obtained by docking poses is shown in **Table 2.2.16.**.

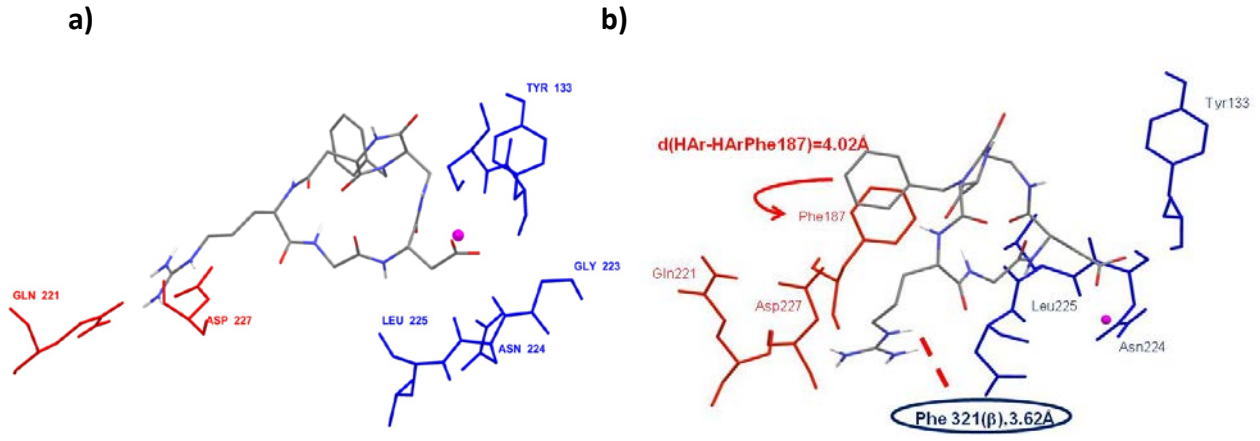


Figure 2.2.54. a) Docking best pose (binding mode A) and b) binding mode B of compound **2** (gray) in $\alpha_5\beta_1$ receptor (α_5 chain red and β_1 chain blue).

Ligand protons	Binding mode A		Binding mode B	
	Residues	Distances	Residues	Distances
NH_{Arg}	CAr-Phe187 (α)	4.90 Å (Min=3.70 Å)	βCH_2-Ser224 (α)	5.39 Å
δH_Arg	HAr-Phe187 (α)	2.74 Å	O=C-Ser224 (α)	3.19 Å
NH_Guan	O_2=C-Asp227 (α)	1.93 Å	HAr-Phe321 (β)	3.62 Å
NH_{Gly}	HAr-Phe187 (α)	4.09 Å (Min=3.54 Å)	HO-Ser227 (β)	3.21 Å
αH_Gly	βCH_2-Ser227 (β)	2.71 Å	HO-Ser227 (β)	2.09 Å
NH_{Asp}	O=C-Leu225 (β)	2.99 Å	αCH-Ser227 (β)	2.56 Å
H_Ar	HAr-Tyr133 (β)	7.42 Å (Min=5.58 Å)	HAr-Phe187 (α)	4.02 Å
DKP-H7	O_2=C-Glu320 (β)	9.28 Å (Min=6.38 Å)	βCH_2-Ser224 (α)	6.46 Å
DKP-H9	αCH-Ser134 (β)	4.69 Å (Min=2.60 Å)	HAr-Tyr133 (β)	4.37 Å

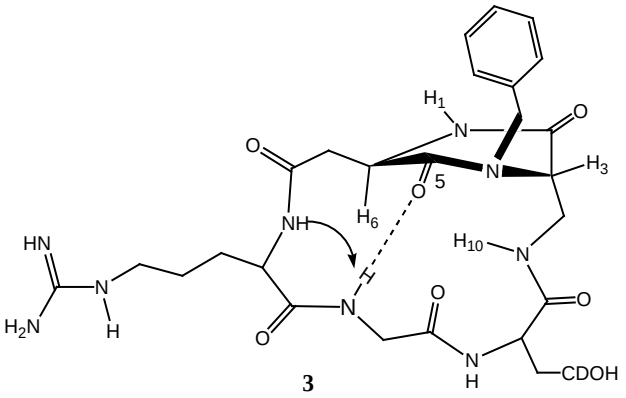
Table 2.2.16. Average values of the distances between ligand protons (selected from STD experiment) and $\alpha_5\beta_1$ residues calculated from docking poses for binding modes A and B.

The two binding modes can be used to interpret NMR results with MDA-MB-231 cells, suggesting that both the ligand arrangements in the $\alpha_5\beta_1$ binding pocket could contribute to the experimental binding epitope. The possibility of adopting different binding modes within the $\alpha_5\beta_1$ receptor site, due to the specific features of the pocket and to the weaker coordination of the

ligand to integrin charged regions compared to $\alpha_V\beta_3$, could explain the low affinity of ligand **2** for the isolated receptor and the great variability observed in the competitive binding assay ($\alpha_5\beta_1$ IC_{50} 1647 ± 871 nM).

Compound 3

Compound **3** adopts, in the free and in the bound state, mainly one conformation characterized by a β-turn at DKP-Arg (Type III, **Figure 2.2.55.**) and confirmed by the cross-peak between NH_{Arg} (8.79 ppm) and NH_{Gly} (8.12 ppm) (**Figure 2.2.56.**) in the NOE experiment. Compound **3** is the only, among those tested, that shows at 282 K positive NOE contacts in the free state and negative in the *tr*-NOESY, showing binding evidence.



3

Figure 2.2.55. Free and bound conformation of *trans*-DKP-RGD-peptidomimetic **3**. Hydrogen bond (dotted line) and NOE contact (arrow) are indicated.

Moreover, the cross-peaks between the aromatic protons and the Arg-side chain protons, which can be observed in the NOESY experiments (free ligand), are lacking in *tr*-NOESY spectrum, indicating an extended arrangement of the Arg-side chain.

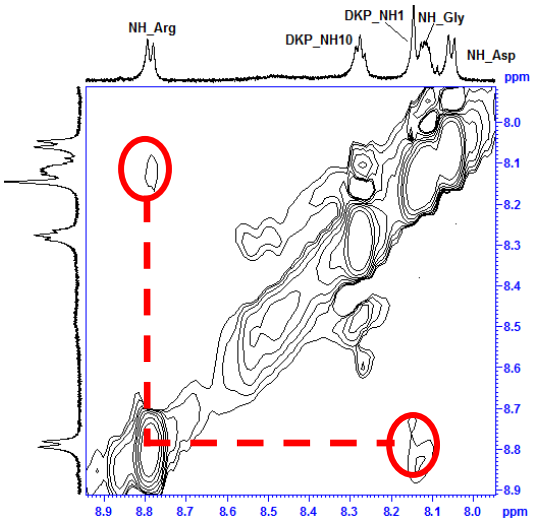


Figure 2.2.56. NOE contact between NH_{Arg} and NH_{Gly} indicating a Type III bound conformation for compound **3**.

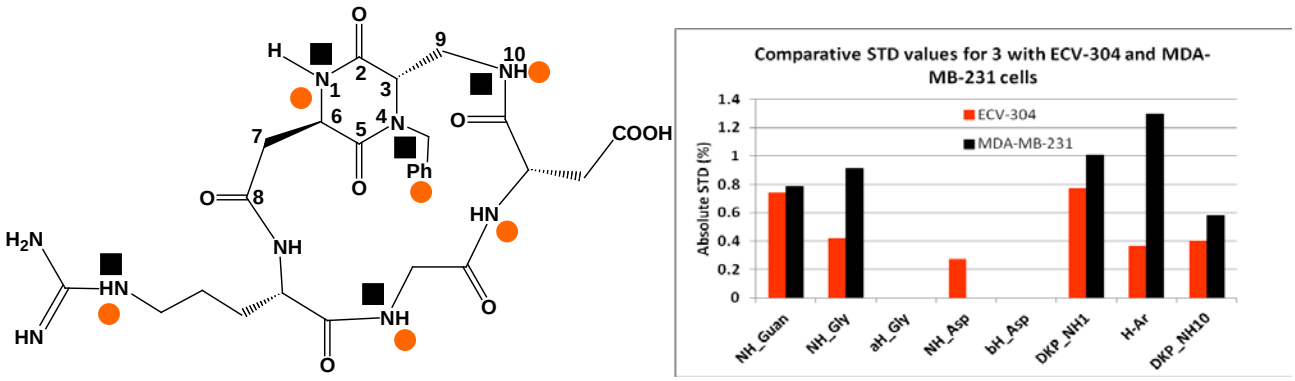


Figure 2.2.57. Binding epitope: a) map of the STD signals of ligand **3** in the presence of ECV-304 cells (orange circles) and MDA-MB-231 cells (black squares), b) relevant absolute STD percentages for the protons of cyclic DKP-RGD **3** obtained at T = 282 K in the presence of ECV-304 (orange bars) and MDA-MB-231 (black bars) cells.

The comparison of the STD signals obtained for ligand **3** in the presence of ECV-304 and MDA-MB-231 cells (**Figure 2.2.57.**) shows that the epitope is similar. The main differences concern: *i)* the presence of NH_{Asp} in the epitope with ECV-304 cells (low STD percentage) and *ii)* the higher STD percentages of the MDA-MB-231 signals compared to the ECV-304 signals, mainly for the aromatic protons. In fact, at 282 K the largest STD effects in the presence of MDA-MB-231 cells are observed for DKP-NH_1 (1.0% STD absolute intensity) and for the aromatic protons (7.26 ppm, 1.3% STD absolute intensity). The guanidine NH proton, DKP-NH_{10}, and NH_{Gly} show quite

similar STD effects (0.8, 0.6 and 0.9%, respectively) and can also be considered as lying close to the protein surface.

The analysis of the docking results highlighted two binding modes of the ligand in the receptor site of $\alpha_5\beta_1$ **(Figure 2.2.58),** both in agreement with the NMR epitope found with MDA-MB-231 cells. The comparison between NMR epitope and distance values calculated for the two families of poses is reported in **Table 2.2.18.**.

a) **b)**

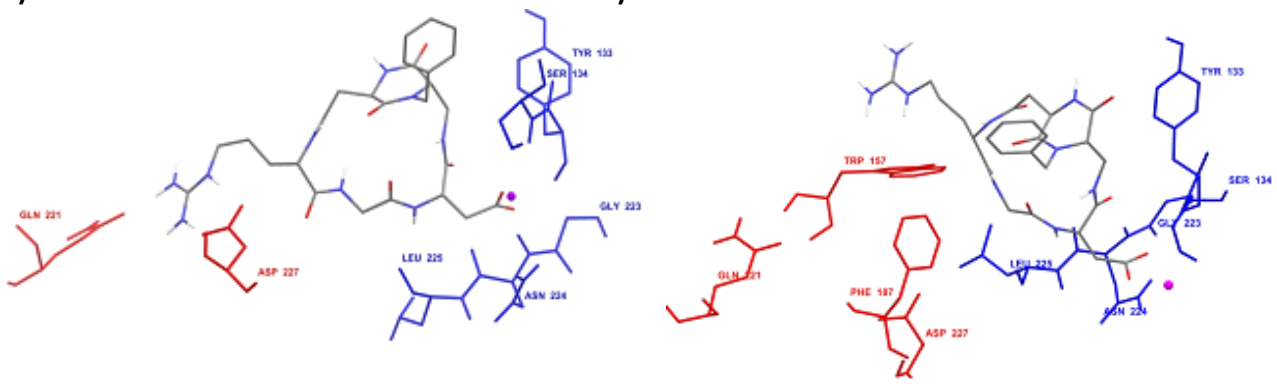


Figure 2.2.58. a) Docking best pose (binding mode A) and b) binding mode B of compound **3** (gray) in $\alpha_5\beta_1$ receptor(α_5 chain red and β_1 chain blue).

Ligand protons	Binding mode A		Binding mode B	
	Residues	Distances	Residues	Distances
NH_Guan	O_2=C_Asp227 (α)	2.02 Å	CAr-Trp157 (α)	3.46 Å
NH_{Gly}	HAr-Phe187 (α)	5.15 Å	HAr-Phe187 (α)	3.31 Å
DKP-NH1	HAr_Tyr133 (β)	6.30 Å	HAr_Tyr133 (β)	3.89 Å
H_Ar	βCH_2_Ser134 (β)	3.76 Å	HAr_Phe187 (α)	4.73 Å (Min=3.26 Å)
DKP-NH10	βCH_2_Ser134 (β)	2.98 Å	βCH2_Ser134 (β)	3.80 Å

Table 2.2.18. Average values of the distances between ligand protons (selected from STD experiment) and $\alpha_5\beta_1$ residues calculated from docking poses for binding modes A and B.

As observed for compound **2**, in the best pose (binding mode A) compound **3** forms all the key interactions with the $\alpha_5\beta_1$ integrin, but in binding mode B, characterized by a less favourable docking score, the guanidinium moiety loses the interaction with Asp227 of α chain and interact with the aromatic ring of the α-Trp157 (**Figure 2.2.58.**).

Compound 5

The *tr*-NOESY experiment of compound **5** in the presence of MDA-MB-231 cells at 282 K showed a negative cross-peak between NH_{Arg} (8.39 ppm) and NH_{Gly} (8.08 ppm); this NOE contact (**Figure 2.2.59.**) is consistent with the β-turn conformation at DKP-Arg found in the free state (type III).

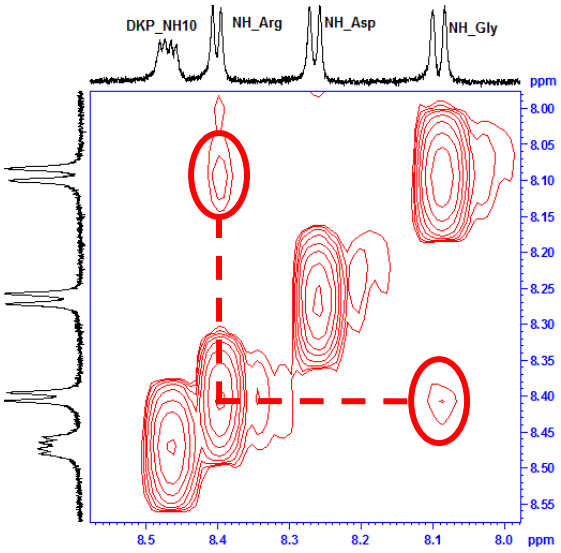


Figure 2.2.59. NOE contact between NH_{Arg} and NH_{Gly} indicating a Type III bound conformation for compound **5**.

In the presence of MDA-MB-231 cells the largest STD effect is observed for the DKP-H7 (1.1% absolute STD intensity). This signal was overlapped in the experiment with ECV-304 cells, while for MDA-MB-231 cells a slight difference in the chemical shift between the DKP-H7 and $H_{\beta Asp}$ permits to determine which of the two protons contributes to the STD effect. Additionally, we observed the presence of $H_{\delta Arg}$, the guanidine NH proton , $H_{\alpha Gly}$, the aromatic protons, and DKP-H9 (0.2, 0.8, 0.3, 0.5 and 0.5% absolute STD intensity, respectively) (**Figure 2.2.60.**). The benzyl groups are overlapped, so their contribution is halved in the graph, but it is not possible to evaluate the contribution of each ring separately.

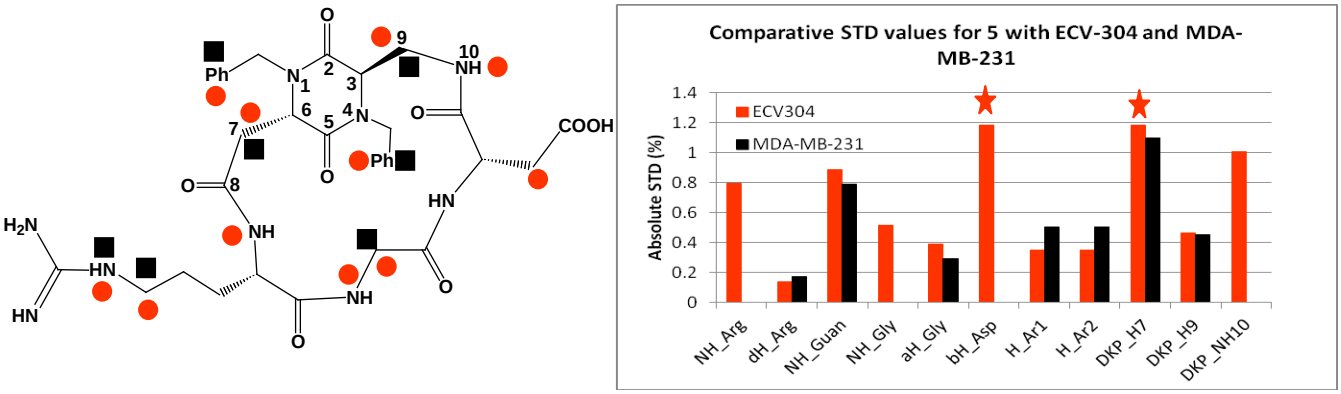


Figure 2.2.60. Binding epitopes: a) map of the STD signals of ligand **5** in the presence of ECV-304 cells (orange dots) andMDA-MB-231 cells (black squares), b) relevant absolute STD percentages for the protons of cyclic DKP-RGD **5** obtained at T = 282 K in the presence of ECV-304 (orange bars) and MDA-MB-231 (black bars) cells.

The comparison of the STD signals obtained for ligand **5** in the presence of ECV-304 and MDA-MB-231 cells (**Figure 2.2.60.**) shows that the same protons of the DKP scaffold provide significant STD percentages (except for DKP-NH10). On the contrary, some differences appear in the RGD sequence, although overlapping protons prevent to exactly define the ligand epitope with ECV-304 cells. However, it is worth noting the absence of NH_{Arg}, NH_{Gly} and βH_{Asp} in the interaction with MDA-MB-231 cells.

Docking calculations into $\alpha_5\beta_1$ receptor produced poses forming all the relevant interactions observed in the X-ray complex. Moreover, the binding modes of compound **5** in the two receptor sites of integrins $\alpha_V\beta_3$ and $\alpha_5\beta_1$, are very similar as clearly shown in the superimposition of the docking best poses (**Figure 2.2.61.**).

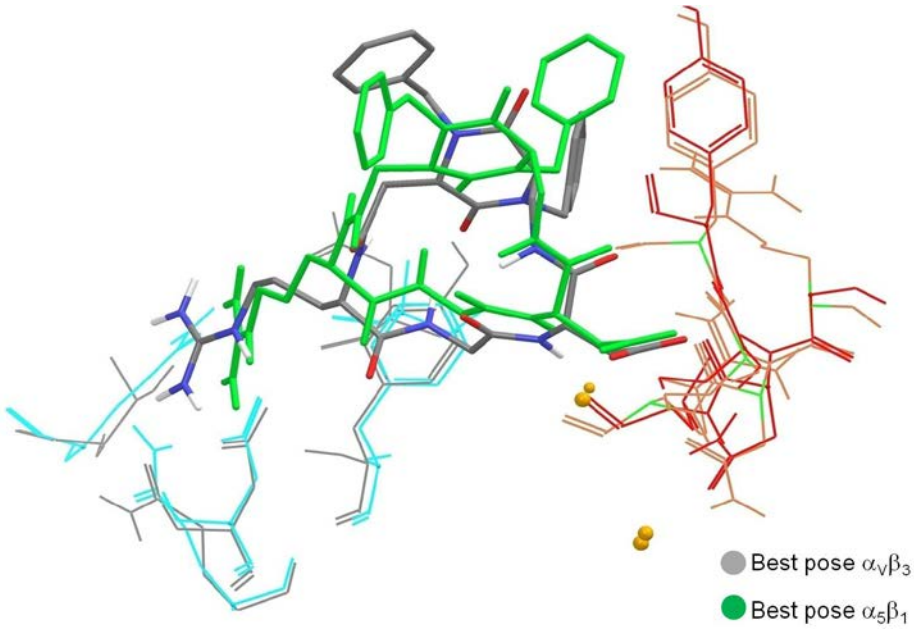


Figure 2.2.61. Superimposition of the docking best poses of compound **5** in $\alpha_V\beta_3$ and $\alpha_5\beta_1$ receptors;docking best pose in $\alpha_V\beta_3$ (gray), α_V chain (gray) and β_3 chain (orange), docking best pose in $\alpha_5\beta_1$ (green), α_5 chain (light blue) and β_1 chain (red).

Ligand protons	Pose in $\alpha_5\beta_1$	
	Residues	Mean distances
δH_Arg	HAr-Phe187 (α)	3.28 Å
NH_Guan	O_2=C-Asp227 (α)	2.01 Å
αH_Gly	βCH_2-Ser227 (β)	2.87 Å
H_Ar	HAr-Tyr133 (β)	4.93 Å (Min = 2.95 Å)
DKP_H7	HAr-Phe187 (α)	5.89 Å (Min = 4.74 Å)
DKP_H9	HAr-Tyr133 (β)	2.12 Å

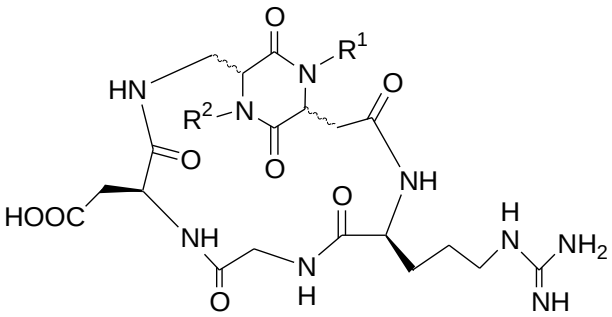
Table 2.2.20. Distances relative to the protons of the epitope, calculated as mean on all the poses.

However, the micromolar binding affinity of compound **5** for the $\alpha_5\beta_1$ isolated integrin (IC_{50} = 1217 ± 157 nM) suggests some difficulties of the docking model in properly describing ligand - receptor interactions (*e.g.* overestimation of the electrostatic contributions).

Section 3: Conclusions

In this thesis, NMR methodologies, based on saturation transfer and detection of transferred NOEs, were revealed useful methods for the investigation of ligand - protein interactions. In particular, we used these techniques to shed light on the binding epitope of peptidomimetic integrin ligands in the interaction with human platelets, known to express on their membrane integrin $\alpha_{IIb}\beta_3$, and with living cancer cells known to express integrins $\alpha_V\beta_3$ and $\alpha_5\beta_1$ (ECV-304 bladder and MDA-MB-231 breast cancer cell lines). The NMR results were supported by docking and molecular dynamics studies and were compared to the results of competitive receptor binding and cell adhesion assays.

The ligands are cyclic peptidomimetics containing the RGD sequence and chiral diketopiperazine (DKP) scaffolds, differing for the configuration at the C-3 and C-6 stereogenic centres and for the substitution at the diketopiperazinic nitrogen atoms (**Figure 2.3.1.**).[4]



(1) cyclo(DKP-1-RGD) = 3S, 6S, R^1 = H, R^2 = CH_2Ph
(2) cyclo(DKP-2-RGD) = 3R, 6S, R^1 = H, R^2 = CH_2Ph
(3) cyclo(DKP-3-RGD) = 3S, 6R, R^1 = H, R^2 = CH_2Ph
(4) cyclo(DKP-4-RGD) = 3R, 6S, R^1 = CH_2Ph, R^2 = H
(5) cyclo(DKP-5-RGD) = 3R, 6S, R^1 = CH_2Ph, R^2 = CH_2Ph
(6) cyclo(DKP-6-RGD) = 3S, 6R, R^1 = CH_2Ph, R^2 = H
(7) cyclo(DKP-7-RGD) = 3S, 6R, R^1 = CH_2Ph, R^2 = CH_2Ph

Figure 2.3.1. Cyclic RGD-peptidomimetics containing bifunctional DKP scaffolds. Compound **7** exists as two separable diastereomers, **7-A** and **7-B**.

Conformational studies of the cyclic RGD peptidomimetics by ^{1}H VT-NMR and NOESY experiments and Monte Carlo/Stochastic Dynamics (MC/SD) simulations revealed that the ligands display well-defined preferred conformations featuring intramolecular hydrogen-bonded turn motifs. For clarity, these β-turn were classified in five types of conformations: Type I-cis (β-turn at Gly-Asp), Type I-trans (β-turn at Gly-Asp), Type II (β-turn at Arg-Gly), Type III (β-turn at DKP-Arg) and Type IV (β-turn at Asp-DKP) (**Figure 2.3.2.**). In particular, Type I-cis and Type II geometries characterize compound **1**, Type III conformation characterizes compounds **2**, **3**, **5** and, in

equilibrium with Type II, compound **7-A**, Type I-trans conformation characterizes compounds **4**, **7-B** and, in equilibrium with Type IV, compound **6**. The scaffold stereochemistry and the regiochemistry of the benzyl ring strongly influenced the conformation of the ligands both in the free and in the bound states.

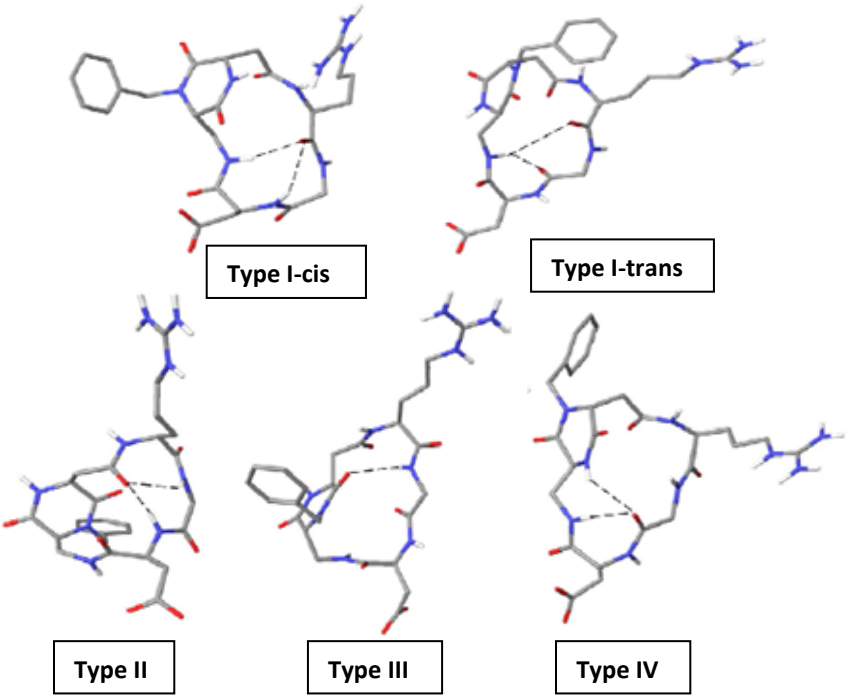


Figure 2.3.2. Representation of the five types of conformations.

Docking calculations were performed starting from the representative conformations obtained from the restrained MC/SD simulations and were analyzed using as reference models the crystal structures of the extracellular segment of the integrins $\alpha_V\beta_3$, $\alpha_{IIb}\beta_3$ and $\alpha_5\beta_1$ in complex with Cilengitide,[18] Eptifibatide[19] and a linear RGD peptide,[17] respectively.

Ligands displaying an extended arrangement of the RGD sequence [Cβ(Arg) – Cβ(Asp) average distance values ≥ 8.8 Å] produced top-ranked binding modes, conserving all the important interactions of the X-ray complex. Instead, non-extended RGD arrangements [Cβ(Arg) – Cβ(Asp) average distance values lower than 8 Å), imparted by Type I-cis and Type II geometries, seem to prevent the guanidine and carboxylic groups from achieving the required separation for binding to integrins. This feature is a necessary requirement, even if not sufficient, for a good interaction with the receptors; in fact, the acid and basic pharmacophoric groups and their orientation are

essential for the binding, because they act like an electrostatic clamp, interacting with charged regions of the receptor binding site.

The ligands, characterized by extended RGD arrangements, show the highest affinities for integrins, as demonstrated by the results of the biological assays. On the contrary, compounds **1** and **7-A**, characterized by closed geometries show the lowest receptor affinities.

First, I set up NMR methodologies by using human platelets and investigating the interaction of RGD peptidomimetic ligands **1** – **3** with the platelet-specific integrin $\alpha_{IIb}\beta_3$. The STD-NMR results were compared with docking studies. The bent conformations (Cβ(Arg) – Cβ(Asp) distances values lower than 8 Å), adopted by compound **1**, in presence of human platelets prevent the guanidine and carboxylic groups to achieve the required distance for binding to the $\alpha_{IIb}\beta_3$. In fact, for compound **1** the interaction mainly occurs through its basic moiety (guanidine NH) with the α_{IIb} subunit, for which we observed the largest STD effect. Compounds **2** and **3** are characterized by the same regiochemistry of the benzyl group and by the Type III geometry, but they differ for the stereochemistry at the two stereocenters of the DKP scaffold. Analyzing the NMR results of the experiments with platelets, the influence of a different stereochemistry on the binding epitope is evident. The compounds show the same interactions in the RGD moiety. In addition, for **3** the epitope comprises the aromatic protons and NH_{10} that are absent in the STD spectrum of **2**. The superimposition of the best poses of **2** and **3** into the $\alpha_{IIb}\beta_3$ binding site showed a similar disposition of the two compounds, perfectly overlapped in the RGD sequence, but different in the benzyl group which points into opposite direction (**Figure 2.3.3.**). According to the absence of the STD effect for the aromatic protons of compound **2**, in the docking poses no residue of the protein is in contact with the ligand benzyl group.

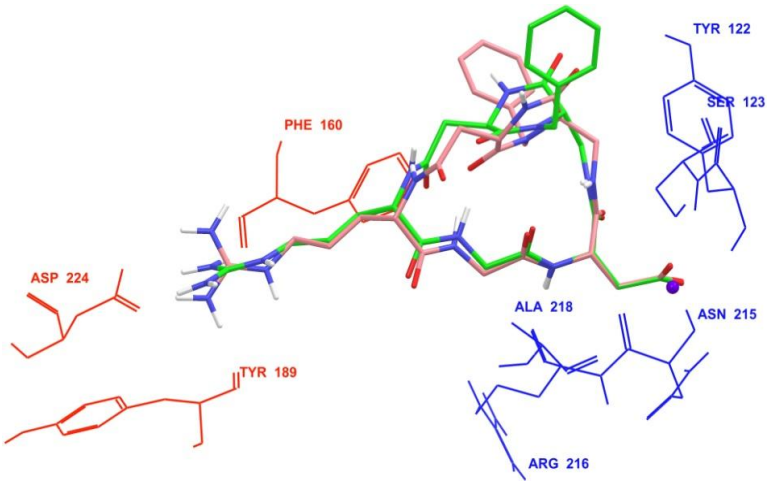


Figure 2.3.3. Docking best poses of *trans*-DKP-RGD-peptidomimetics **2** and **3** (pink and green atom colour tube representation, respectively) into the crystal structure of the integrin $\alpha_{IIb}\beta_3$ headpiece (α unit red and β unit blue wire representation).

Then, I expanded the NMR study by using ECV-304 cancer cell line and gaining insight into the interaction of the small library of ligands with the tumour-associated integrin $\alpha_V\beta_3$.

For both cell lines, human platelets and ECV-304 cells, we observed that: *i)* the primary interaction between the ligand and the receptor is performed by the guanidine residue of the Arg side-chain (see behaviour of compound **1**), *ii)* the extended conformation of the *trans* ligands **2** and **3** ensures the formation of the electrostatic clamp (STD effects on protons of the Arg and Asp residues), *iii)* the opposite configuration at C-3 and C-6 stereogenic centres of the DKP scaffold of the *trans* ligands **2** and **3** induces a different binding epitope for the non-RGD moiety of the ligands (diketopiperazine and benzyl substituents).

We completed our analysis on ECV-304 cell line with the study of peptidomimetic ligands **4** – **7** containing the two enantiomeric *trans* DKP scaffolds with different substitution at the endocyclic nitrogen atoms. Compounds **4** and **6**, monobenzylated at N-1, and the dibenzylated compound **7-B** adopt the Type I-trans conformation with respect to the Type III geometry adopted by compounds **2** and **3**, monobenzylated at N-4, and by the dibenzylated ligand **5** (**Figure 2.3.4.**).

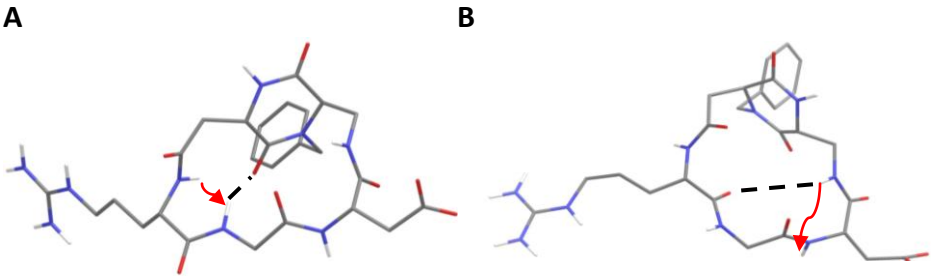


Figure 2.3.4. Representation of the two most populated extended conformations; A) Example of a Type III H-bonding pattern, B) Example of a Type I-*trans* H-bonding pattern. Red arrows indicate NOE contacts, while dashed black line the H-bonds.

The type I-trans conformation produced deep differences in the binding epitope with respect to Type III, involving in the interaction not only the RGD sequence, but also the protons of the DKP scaffold. This behaviour is evident in the comparison between the epitopes of compounds **2** and **4** (**Figure 2.3.5.**).

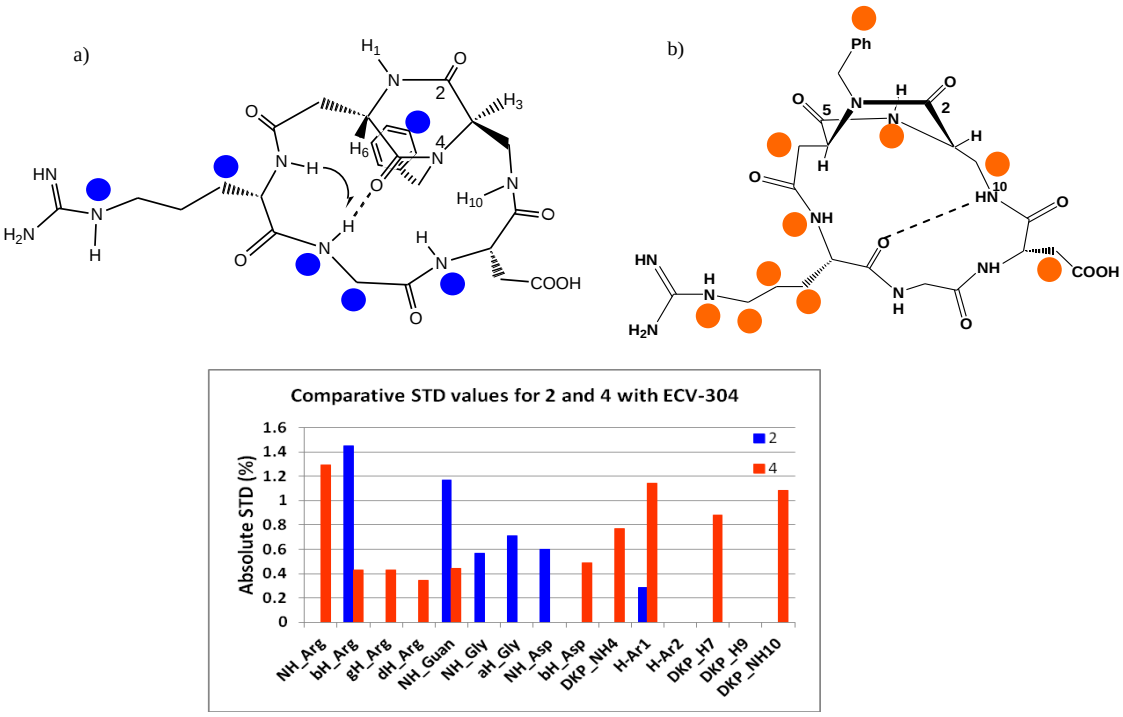


Figure 2.3.5. Comparison between the epitopes of compounds **2** and **4**. a) - b) Map of the STD signals of ligand **2** (blue dots) and **4** (orange dots). Relevant absolute STD percentages for the protons of **2** (blue bars) and **4** (orange bars) obtained at 282 K in the presence of ECV-304 cells (bottom).

While compound **2** mainly involves in the interaction the protons of the RGD sequence, compound **4** involves the protons of the Arg and Asp side chains and some protons of the DKP scaffold.

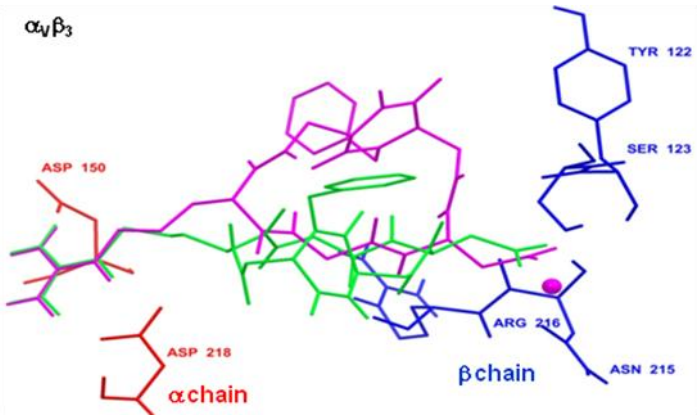


Figure 2.3.6. Docking best pose of compound **2** (purple) superimposed with the docking pose of compound **4** (green) into the crystal structure of the extracellular domain of $\alpha_v\beta_3$ integrin (α unit red and β unit blue wire representation); the metal ion at MIDAS is shown as a magenta CPK sphere.

The different epitope of the two compounds is well represented by superimposing their docking best poses into the $\alpha_V\beta_3$ receptor site, where they mainly differ in the position of DKP scaffold (**Figure 2.3.6.**).

The dibenzylated compounds, **7-A** and **7-B**, show the influence of the preferred conformation adopted by ligands on their disposition in the receptor binding pocket and, as consequence, on their NMR epitope. In fact, the most active compound of the DKP series, the diastereomer **7-B** (type I-trans), is characterized by high STD absolute percentages and lots of interactions with the receptor. On the other hand, the diastereomer **7-A**, showing the lowest affinity for the $\alpha_V\beta_3$ integrin among the trans ligands, displayed a STD signal pattern intermediate between those of the compounds **3** and **7-B**. Entropic reasons, *i.e.* the presence of both extended (type III) and non-extended geometries in the free state conformational equilibrium, can account for the lower affinity for the $\alpha_V\beta_3$ integrin of the isomer **7-A** compared to the other *trans* DKP-RGD ligands.

The ECV-304 cell line expresses different integrins on the membrane. In particular, integrins $\alpha_V\beta_3$ and $\alpha_5\beta_1$ show the same high level of expression (ratio of $\alpha_5\beta_1/\alpha_V\beta_3$ expression = 1.16). Since also integrin $\alpha_5\beta_1$ recognizes the RGD sequence, we wondered if the NMR results obtained on ECV-304 cells might be affected by the presence of both the integrins. For this reason, we performed *tr*-NOESY and STD NMR experiments using the MDA-MB-231 breast cancer cell line, known to overexpress only integrin $\alpha_5\beta_1$. MDA-MB-231 was then considered the reference cell line to investigate the interaction of the compounds with the $\alpha_5\beta_1$ integrin**.**

Compared with the binding results for $\alpha_V\beta_3$ receptor, all the DKP ligands and the reference peptides lost at least one order of magnitude in affinity ($\alpha_5\beta_1$ IC_{50} values 30 - 500 times higher than $\alpha_V\beta_3$ IC_{50} values). Since the X-ray structures reveal similar RGD geometrical requirements for the binding to $\alpha_5\beta_1$ and $\alpha_V\beta_3$ integrins, the observed selectivity may depend on the physicochemical features of the binding sites. In fact, the receptor region that binds to the ligand guanidinium group is less acidic in $\alpha_5\beta_1$ than in $\alpha_V\beta_3$, thus leading to a weaker interaction of the ligand Arg side chain with the α_5 subunit. Moreover, due to both the less efficient electrostatic clamp and the larger binding pocket compared to $\alpha_V\beta_3$, the ligands interact less strongly with $\alpha_5\beta_1$ and may reveal a greater mobility within the receptor site.

The behaviour of compound **4** is the most useful to differentiate and appreciate the interaction of the ligands of the library with the two receptors, making clear the differences that characterized the two binding pockets. The comparison of the STD signals obtained for ligand **4** in the presence of ECV-304 and MDA-MB-231 cells (**Figure 2.3.7.**) shows that the ligand epitope is significantly different: *i)* the protons of the Arg-side chain appear with relevant STD percentages only in the interaction with ECV-304 (orange bars), *ii)* the protons DKP-H3 and DKP-H9 appear only in the interaction with MDA-MB-231 (black bars), *iii)* the protons βH_{Asp} and DKP-NH10 appear only in the interaction with ECV-304 (orange bars).

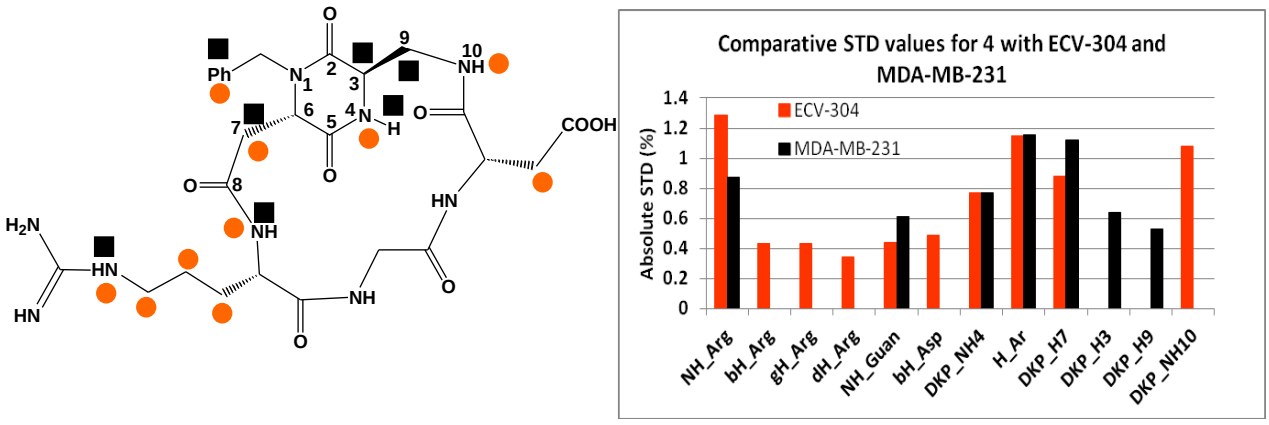


Figure 2.3.7. Binding epitopes: a) map of the STD signals of ligand **4** in the presence of ECV-304 cells (orange dots) with respect to MDA-MB-231 cells (black squares), b) relevant absolute STD percentages for the protons of cyclic DKP-RGD **4** obtained at T = 282 K in the presence of ECV-304 (orange bars) and MDA-MB-231 (black bars) cells.

On the basis of the different ligand epitopes highlighted by STD-NMR with the two cell lines, a preferred interaction of ligand **4** with $\alpha_v\beta_3$ integrin can be hypothesized in the presence of ECV-304 cells (which express $\alpha_v\beta_3$ and $\alpha_5\beta_1$ in 1:1 ratio), whereas the interaction with $\alpha_5\beta_1$ integrin is characterized by NMR in the experiments with MDA-MB-231 cells. The higher affinity of compound **4** for $\alpha_v\beta_3$ compared to $\alpha_5\beta_1$ emerging from binding assays supports this hypothesis.

Also compound **2** is characterized by a relevant different NMR binding epitope with the two cell lines. STD-NMR experiments, in fact, show for compound **2** differences in its binding epitope with the two cell lines, ECV-304 and MDA-MB-231 cells (**Figure 2.3.8.**).

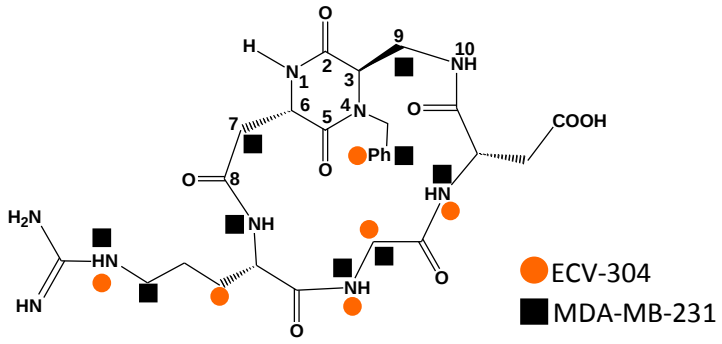


Figure 2.3.8. Binding epitopes: map of the STD signals of ligand **2** in the presence of ECV-304 cells (orange dots) and MDA-MB-231 cells (black squares).

Compound **2** with ECV-304 cell line involves in the interaction the protons of the RGD sequence, while with MDA-MB-231 it interacts also with the proton of the DKP scaffold (**Figure 2.3.8.**).

In conclusion, compounds **2** and **4** (same stereochemistry at the two stereocenters of the DKP scaffold, but different substitution pattern) are characterized by differences between epitopes in the two cell lines. Therefore, only for these compounds, the recognition of different receptors ($\alpha_V\beta_3$ or $\alpha_5\beta_1$) can be hypothesized in the interaction with ECV-304 and MDA-MB-231 cells. In the presence of ECV-304 cells, where the integrins $\alpha_V\beta_3$ and $\alpha_5\beta_1$ are expressed in the same ratio, they interact preferentially with integrin $\alpha_V\beta_3$. The other ligands, including compounds of opposite stereochemistry (*e.g.* **3** and **6**), appear to interact with the two integrins in a similar way.

Section 4: Experimental section

NMR EXPERIMENTS

Conformational analysis: NMR spectroscopy experiments were performed at 298 K and 282 K on Bruker Avance 400 and 600 MHz spectrometers. All proton and carbon chemical shifts were assigned unambiguously. The NMR experiments were carried out in a H_2O/D_2O 9:1 mixture in order to observe amide protons. Two-dimensional experiments (TOCSY, NOESY and HSQC) were carried out on samples at a concentration range of 3-6 mM. NOESY experiments were performed at 0.7 or 0.8 s. The water resonance was saturated with the excitation sculpting sequence from the Bruker library. The conformations of the cyclic peptidomimetics were analyzed with respect to hydrogen bonding of amide protons (VT-NMR spectroscopy) and NOE contacts.[4]

NMR spectroscopy studies with cells: The spectra were acquired with a Bruker Avance 600 MHz instrument at 298 and 282 K. Each sample was prepared in a 3 mm NMR tube with ligand (0.5-0.6 mg) and cells (10^6) in phosphate buffer (200 μl, pH 7.4, 20 mM additionally containing NaCl 157 mM, KCl 6 mM and $CaCl_2$ 4 mM with D_2O 10%). The interaction between our ligands and membrane proteins expressed on human platelets, ECV304 (bladder cancer cells) and MDA-MB231 (breast cancer cells) cell lines was detected using *tr*-NOESY and STD experiments.[3,44] In addition, *tr*-ROESY experiments, performed to exclude spin-diffusion artefacts, showed the same set of cross-peaks as the NOESY. In the *tr*-NOESY experiment water suppression was achieved by use of an excitation sculpting pulse sequence. In the STD experiments water suppression was achieved by use of the WATERGATE 3-9-19 pulse sequence. The on-resonance irradiation of the protein was performed at a chemical shift of -0.05 ppm and the saturation time is 3 sec. Off-resonance irradiation of the protein was applied at 200 ppm, where no protein signals were visible. STD: 60 Gaussian pulses, each requiring 50 ms. STD spectra of the ligands in the absence of cells did not show any signals. Instead, in the STD spectra of the ligands with cells, intensities of STD effects (absolute STD) were calculated by integrating the corresponding signals in ^1H NMR reference spectra. The signal intensities of the individual protons are best analyzed from the integral values in the reference and STD spectra. $(I_0 - I_{sat})/I_0$ is the fractional STD effect, expressing

the signal intensity in the STD spectrum as a fraction of the intensity of an unsaturated reference spectrum. In this equation, I_0 is the intensity of one signal in the off-resonance or reference NMR spectrum and I_{sat} is the intensity of a signal in the on-resonance NMR spectrum.

CELL CULTURES

Platelets were extracted from human blood received from Multimedica Centre of Milan. A portion of 15 mL of CPD plasma (plasma suspended in citrate-phosphate-dextrose solution) was centrifuged and the remaining pellet cautiously suspended in 2 mL of deuterated TRIS-buffered saline containing EDTA (1 mM). This procedure was repeated twice. Four additional washing-centrifuging cycles were performed using deuterated TRIS-buffered saline containing 1 mM $CaCl_2$. After the last centrifugation, the pellets were suspended in 1 mL of the last buffer and the suspension was split into two NMR tubes. To each tube, 1 mg of ligand was added.

Both the *ECV-304* and *MDA-MB-231* cell lines were expanded after freezing and were grown in RPMI culture medium (mother medium supplemented with 1% glutamine, 1% PENSTREPT (penicillin-streptomycin), 10% FBS (fetal bovine serum)). Cells were adherent to the flask wall and cover with RPMI medium. To prepare cells for NMR tube, cells were removed by trypsin and then resuspended in 10 mL of a RPMI – serum solution. After two washing-centrifuging cycles with PBS, the pellets were suspended in 200 μL of phosphate buffer.

COMPUTATIONAL STUDIES

All calculations were run with the Schrödinger suite of programs (http://www.schrodinger.com) through the Maestro graphical interface.

Conformational analysis: Conformational preferences of the RGD peptidomimetics were investigated by Monte Carlo/Stochastic Dynamics (MC/SD)[31,32,33] hybrid simulations by using NMR spectroscopy restraints derived from the experimental NOE contacts (for distance restraints used for each calculation, see the ∗**Supporting Information**∗). All the NOE restraints were set to a distance value of 2(±0.5) Å with a force constant of 100 $Kj*mol^{-1}*Å^{-2}$. MC/SD simulations were performed at 300 K within the framework of MacroModel version 9.5 by employing the

OPLS_2001 force field[33] and the implicit water GB/SA solvation model.[32] The OPLS_2001 force field, among several commonly used force fields, provided the highest correlation between calculated and experimental conformational preferences within the series of cyclic RGD peptidomimetics containing diketopiperazine scaffolds with different stereochemistry and substitution at the piperazinic nitrogen atoms. RGD side-chain dihedral angles were defined as internal coordinate degrees of freedom in the Monte Carlo part of the algorithm. A time step of 1 fs was used for the stochastic dynamics (SD) part of the algorithm for 10 ns of simulation time. Samples were taken at 2 ps intervals during each simulation, yielding 5000 conformations for analysis. The percentages of H bonds discussed here were calculated as percentages of conformations sampled during the simulation in which donor H-acceptor O distance <2.5 Å (γ-turn) or <4 Å (b-turn).

Protein setup:

$\alpha_{IIb}\beta_3$: the solved crystal structure of the integrin $\alpha_{IIb}\beta_3$ headpiece bound to the cyclic heptapeptide Eptifibatide (PDB entry code 2VDN)[19] was used for docking studies. Docking was performed only on the globular head of the integrin because the headgroup of integrin has been identified in the X-ray structure as the ligand-binding region. The protein structure was setup for docking as follows. The protein was truncated to residue sequences 18-452 for chain α and 108-352 for chain β. The coordination shell of metal cation belonging to MIDAS (metal-ion-dependent adhesion site) and ADMIDAS (adjacent to MIDAS) in the β chain was completed on the basis of the integrin $\alpha_{IIb}\beta_3$ X-ray structures by adding two and one water molecules, respectively. The Mg^{2+} ion at MIDAS and ADMIDAS in the experimental protein structure was modeled by replacing it with a Ca^{2+} ion. The resulting structure was prepared using the Protein Preparation Wizard of the graphical user interface Maestro and the OPLSAA force field.

$\alpha_V\beta_3$: the solved crystal structure of the extracellular domain of the integrin $\alpha_V\beta_3$ receptor in complex with the cyclic pentapeptide cilengitide and in the presence of the proadhesive Mn^{2+} ion (PDB ID: 1L5G)[18] was used for docking studies. Docking was performed only on the globular head of the integrin because the head group of integrin has been identified as the ligand-binding

region in the X-ray structure. The protein structure was set up for as follows: the protein was truncated to residue sequences 41-342 for α chain and 114-347 for β chain. Because of a lack of parameters, the Mn^{2+} ions in the experimental protein structure were modelled by replacing them with Ca^{2+} ions. The resulting structure was prepared by use of the Protein Preparation Wizard in the graphical user interface Maestro and the OPLSAA force field.

$\alpha_5\beta_1$: the recently solved crystal structure of the extracellular domain of the integrin $\alpha_5\beta_1$ receptor in complex with the linear esapeptide GRGDNP (PDB ID: 3VI4)[17] was used for docking studies. Docking was performed only on the globular head of the integrin because, also in this case, the closed head group of integrin has been identified as the ligand-binding region in the X-ray structure. The protein structure was set up as follows: the protein was truncated to residue sequences 40 – 351 for α chain and 121 – 358 for β chain. Because of the absence of ADMIDAS ion in the X-ray structure, a Ca^{2+} ion was added in this region based on the ADMIDAS position in $\alpha_V\beta_3$ X-ray structure. The resulting structure was prepared by use of the Protein Preparation Wizard in the graphical user interface Maestro and the OPLSA_2005 force field.

All models were validated through docking studies conducted with the respective crystallographic ligands. The results well agree with the experimental data, both for a good "docking score" and for the number of poses that reflect the arrangement of the crystal structure of the ligand in the receptor pocket.

Molecular Docking: the automated docking calculations were performed using Glide (Grid-based Ligand Docking with Energetics)[47] within the framework of Impact version 4.5 (for $\alpha_{IIb}\beta_3$ and $\alpha_V\beta_3$) or 5.7 (for $\alpha_5\beta_1$) in a standard precision mode (SP). Glide uses a hierarchical series of filters to search for possible locations of the ligand in the active site region of the receptor. The shape and properties of the receptor are represented on a grid by several different sets of fields that provide progressively more accurate scoring of the ligand poses. The grid generation step started from the extracellular fragment of X-ray structure of the complex protein : ligand, as described in the protein setup section. The center of the grid enclosing box was defined by the center of the bound ligand, as described in the original PDB entry. The enclosing box dimensions, which are automatically deduced from the ligand size, fit the entire active site. For the docking step, the size

of the bounding box for placing the ligand center was set to 12 Å. No further modifications were applied to the default settings. The GlideScore function was used to select 20 poses for each ligand. The Glide program was initially tested for its ability to reproduce the binding geometry of the crystallized ligands. The program was successful in reproducing the experimentally determined mode of these compounds, as they correspond to the best-scored pose.

The automated docking calculations were performed by starting from the representative geometries previously obtained from computational studies (free-state molecular mechanics conformational searches and MC/SD simulations) for the RGD cyclopeptides.

Molecular Dynamics simulations: MD simulations were performed on some peptidomimetic integrin ligands within the X-ray crystal structure of the $\alpha_{IIb}\beta_3$, $\alpha_V\beta_3$ and $\alpha_5\beta_1$ integrin binding sites prepared as described in the protein setup. Starting from docking best poses, MD simulations were performed at 300 K within the framework of MacroModel version 9.5 by employing the OPLS_2001 force field[33] and the implicit water GB/SA solvation model.[32] The substructure utility was used to model the receptor atoms. Residue atoms within 5 Å of the ligand were constrained to crystallographic coordinates by a harmonic potential (with a force constant k = 100 Kj mol^{-1} $Å^2$ for atoms up to 3 Å and with k = 200 Kj mol^{-1} $Å^2$ for atoms between 3 and 5 Å) allowing the polar and aromatic hydrogen atoms free movement. Residue atoms between 5 Å and 10 Å were frozen to their X-ray position. The remaining receptor atoms were ignored during the calculations. Molecular dynamic simulations gave similar results (disposition of the ligand and interactions) to those obtained with docking calculations.

IN VITRO BIOLOGICAL ASSAYS

Integrin receptor binding assays: the cyclic RGD peptidomimetics were examined in vitro for their ability to inhibit biotinylated vitronectin binding to the purified $\alpha_V\beta_3$ integrin and to inhibit biotinylated fibronectin binding to the purified $\alpha_5\beta_1$ receptor.

$\alpha_V\beta_3$: Purified $\alpha_V\beta_3$ receptor (Chemicon International, Inc., Temecula, CA, USA) were diluted to 0.5 μg mL^{-1} in coating buffer containing 20 mM Tris-HCl (pH 7.4), 150 mM NaCl, 1 mM

$MnCl_2$, 2 mM $CaCl_2$, and 1 mM $MgCl_2$. An aliquot of diluted receptors (100 μLwell^{-1}) was added to 96-well microtiter plates (NUNC MW 96F Medisorp Straight) and incubated overnight at 4 °C. The plates were then incubated with blocking solution (coating buffer plus 1 % bovine serum albumin) for additional 2 h at room temperature to block nonspecific binding, followed by 3 h incubation at room temperature with various concentrations (10^{-12} – 10^{-5} M) of test compounds in the presence of biotinylated vitronectin (1 μg mL^{-1}) using an EZ-Link Sulfo-NHS-Biotinylation kit (Pierce, Rockford, IL, USA). After washing, the plates were incubated for 1 h at room temperature with biotinylated streptavidin-peroxidase complex (Amersham Biosciences, Uppsala, Sweden) followed by 30 min incubation with 100 μL Substrate Reagent Solution (R&D Systems, Minneapolis, MN) before stopping the reaction with the addition of 50 μL 1 N H_2SO_4. Absorbance at 415 nm was read in a Synergy HT Multi-Detection Microplate Reader (BioTek Instruments, Inc.).

$\alpha_5\beta_1$: Purified $\alpha_5\beta_1$ receptor was diluted to 0.5 μg/mL in coating buffer containing 20 mmol/L Tris- HCl (pH 7.4), 150 mmol/L NaCl, 1 mmol/L $MnCl_2$, 2 mmol/L $CaCl_2$ and 1 mmol/L $MgCl_2$. An aliquot of diluted receptors (100 μL/well) was added to 96-well microtiter plates (NUNC MW 96F MAXISORP STRAIGHT) and incubated overnight at 4°C. The plates were then incubated with blocking solution (coating buffer plus 1% bovine serum albumin (Sigma) for additional 2 hours at room temperature to block non specific binding followed by 3 hours incubation at room temperature with various concentrations of test compounds in the presence of 1 μg/mL fibronectin biotinylated using EZ-Link Sulfo-NHS-Biotynilation kit (Pierce, Rockford, IL). After washing, the plates were incubated for 1 hour at room temperature with streptavidin-biotinylated peroxidase complex (Amersham Biosciences, Uppsala, Sweden) followed by 30 minutes incubation with 100 μL Substrate Reagent Solution (R&D Systems, Minneapolis, MN) before stopping the reaction by addition of 50 μL of 1 N H_2SO_4. Absorbance at 415 nm was read.

The ability of the new compounds to inhibit the binding of vitronectin and fibronectin to the isolated $\alpha_V\beta_3$ and $\alpha_5\beta_1$ receptors was compared with that of the reference compounds c(RGDfV) and ST1646. IC_{50} values were calculated as the concentration of compound required for 50 % inhibition of biotynilated vitronectin/fibronectin binding as estimated by GraphPad Prism

software by nonlinear regression analysis; all values are the arithmetic mean ± SD of triplicate determinations.

Cell adhesion assays: to assess the activity of RGD compounds as integrin antagonists on a cell model, cell adhesion experiments were performed using ECV-304 bladder cancer cells which highly express integrin $\alpha_V\beta_3$ and MDA-MB-231 breast cancer cells which overexpress integrin $\alpha_5\beta_1$. Cells were allowed to adhere to immobilized fibronectin or vitronectin in the presence of increasing concentrations of the tested compounds. Adhesion assay tests were performed in two different laboratories using the same ECV-304 and MDA-MB-231 cancer cell lines and experimental protocol. The results were obtained for compounds **1**, **2** and **3** and the reference peptide *cyclo*[RGDfV] at the cell biology laboratory of CISI (an Interdepartmental Centre of Milan University, Dr. E. Araldi), that first developed the adhesion assay protocol for both ECV-304 and MDA-MB-231 cells. The results for compounds **4** - **7** and the reference peptide *cyclo*[RGDfV] were obtained at the Department of Medical Biotechnology and Translational Medicine by Dr. C. Tringali and Dr. I. Silvestri (University of Milan)**.**

∗Bibliography∗

[1] a) Humphries, M. *J. Biochem. Soc. Trans.* **2000**, 28, 311 - 340; b) Mould, A. P.; Humphries, M. J. *Nature* **2004**, 432, 27 - 28.

[2] Aplin, A. E.; Howe, A.; Alamari, S. K. et al. *Pharmacol. Rev.* **1998**, 50, 197 - 264.

[3] Meyer, B.; Peters, T. *Angew. Chem. Int. Ed.* **2003**, 42, 8, 864 – 890.

[4] Marchini, M.; Mingozzi, M.; Colombo, R.; Guzzetti, I.; Belvisi, L.; Vasile, F.; Potenza, D.; Piarulli, U.; Arosio, D. and Gennari, C. *Chem. Eur. J.* **2012**, 18, 6195 – 6207.

[5] Potenza, D.; Belvisi, L. *Org. Biomol. Chem.* **2008**, 6, 258 – 262.

[6] Potenza, D.; Vasile, F.; Belvisi, L.; Civera, M. and Araldi, E. M. V. *ChemBioChem* **2011**, 12, 695 – 699.

[7] Guzzetti, I.; Civera, M.; Vasile, F.; Araldi, E. M.; Belvisi, L.; Gennari, C.; Potenza, D.; Fanelli, R. and Piarulli, U. *Org. Biomol. Chem.* **2013**, 11, 3886 – 3893.

[8] a) Gottschalk, K-E.; Kessler, H. *Angew. Chem. Int. Ed.* **2002**, 41, 3767 – 3774; b) Cox, D.; Brennan, M. and Moran, N. *Nature Rev. Drug Discov.* **2010**, 9, 804 – 820.

[9] a) Shattil, S. J.; Chungo, K.; Ginsberg, M.H. *Nature Rev. Drug Discov.* **2010**, 11, 288 – 300; b) Plow, E. F.; Haas, T. A.; Zhang, L.; Loftus, J. and Smith, J. W. *JBC* **2000**, 275, 21785 – 21788.

[10] Haubner, R.; Finsinger, D.; Kessler, H. *Angew. Chem. Int. Ed.* **1997**, 36, 1374 – 1997.

[11] Wiesner, S. et al. *Cell. Mol. Life Sci.* **2005**, 62, 1081 – 1099.

[12] Arnaout, M. A. et al. *Annu. Rev. Cell Dev. Biol.* **2005**, 21, 381 – 410.

[13] Shimaoka, M.; Springer, T.A. *Nature Rev. Drug Discov.* **2003**, 2, 703 - 716.

[14] Wang, J.; Springer, T. A. *Immunol. Rev.* **1998**, 163, 197 - 215.

[15] Mousa, S. A. *Curr. Opin. Chem. Biol.* **2002**, 6, 534 - 541.

[16] a) Springer, T. A.; Zhu, J.; Xiao, T. *JCB* **2008**, 182, 4, 791 – 800; b) Van Aghtoven, J. F.; Xiong, J.-P.; Alonso, J. L.; Rui, X.; Adair, B. D.; Goodman, S. L.; Arnaout, M. A. *Nature Structural & Molecular Biology* **2014**, 21, 383 – 390.

[17] a) Nagae, M.; Re, S.; Mihara, E. et al. *JCB* **2012**, 197, 1, 131 – 140; b) Zhu, J.; Zhu, J.; Negri, A.; Provasi, D.; Filizola, M.; Coller, B. S.; Springer, T. A. *Blood* **2010**, 116, 5049 – 5059; c) Zhu, J.; Luo, B. H.; Xiao, T. *et al. Mol Cell.* **2008**, 32, 849 – 861.

[18] a) Xiong, J-P. ; Stehle, T. et al. *Science* **2002**, 296, 151 – 155; b) Dong, X.; Mi, L-Z; Zhu, J. et al. *Biochemistry* **2012**, 51, 8814 – 8828.

[19] Xiao, T; Takagi, J.; Coller, B. S. et al. *Nature* **2004**, 432, 59 – 67.

[20] Humpries, M. J. et al. *J. Cell Sci.* **2006**, 119, 3901 – 3903.

[21] Dechantsretter, M. A.; Planker, E.; Mathä, B.; Lohof, E.; Hölzemann, G.; Jonczyc, A.; Goodman, S. L.; Kessler, H. *J. Med. Chem.* **1999**, 42, 3033 - 3040; b) Mas-Moruno, C.; Beck, J. G.; Doedens, L.; Frank, A. O.; Marinelli, L.; Cosconati, S.; Novellino, E.; Kessler, H. *Angew. Chem. Int. Ed.* **2011**, 50, 9496 – 9500.

[22] a) Belvisi, L.; Riccioni, T.; Marcellini, M.; Chiarucci, I.; Efrati, D.; Vesci, L.; Potenza, D.; Scolastico, C.; Manzoni, L.; Lombardo, K.; Stasi, M. A.; Nico, B.; Ribatti, D.; Presta, M.; Carminati, P. and Pisano, C. *Mol. Cancer Ther.* **2005**, 4, 1670 – 1680; b) Belvisi, L.; Bernardi, A.; Colombo, M.; Manzoni, L.; Potenza, D.; Scolastico, C. ; Giannini, G.; Marcellini, M.; Riccioni, T.; Castorina, M.; Lo Giudice, P.; Pisano, C. *Bioorg. Med. Chem.* **2006**, 14, 169 – 180.

[23] Arosio, D. et al. *ChemMedChem* **2008**, 3, 1589 – 1603.

[24] Manzoni, L. et al. *ChemMedChem* **2009**, 4, 615 – 632.

[25] de Ressureicão, A. S. M.; Vidu, A.; Civera, M.; Belvisi, L.; Potenza, D.; Manzoni, L.; Ongeri, S.; Gennari, C; Piarulli, U. *Chem. Eur. J.* **2009**, 15, 12184 – 12188.

[26] Casiraghi, G. et al. *J. Med. Chem.* **2005**, 48, 7675 – 7687.

[27] a) Zanardi, F. et al. *J. Med. Chem.* **2008**, 51, 1771 – 1782; b) "Lecture Notes of Computational Organic chemistry" Bernardi, A.; Belvisi, L.; Olivucci, M.; Sinicropi, A.; Colombo, G.; Barone, V.; Crescenzi, O. Cap. 3 "3D Models of oligopeptides by Molecular Mechanics and Conformational Analysis" **2006**.

[28] Auzzas, L.; Zanardi, F.; Battistini, L.; Burreddu, P.; Carta, P.; Russo, G.; Curto, C. and Casiraghi G. *Current Medicinal Chemistry* **2010**, 17, 1255 – 1299.

[29] Cini, N.; Trabocchi, A.; Menche, G.; Battoncetti, A. et al. *Bioorg. Med. Chem.* **2009**, 17, 1542 - 1549.

[30] Mas-Moruno, C., Rechenmacher, F. and Kessler, H. *Anti-Cancer Agents Med. Chem.* **2010**, 10, 753 – 768.

[31] Guarnieri, F.; Still, W. C. *J. Comput. Chem.* **1994**, 15, 1302 – 1310.

[32] Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, 112, 6127 – 6129.

[33] Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, 118, 11225 – 11236.

[34] a) Curnis, F; Longhi, R.; Crippa, L.; Cattaneo, A.; Dondossola, E.; Bachi, A.; Corti, A. *J. Biol. Chem.* **2006**, 281, 36466 – 36476; b) Corti, A.; Curnis, F. *J. Cell Sci.* **2011**, 124, 515 – 522.

[35] Clarke, S. *Int. J. Peptide Protein Res.* **1987**, 30, 808 – 821.

[36] a) Curnis, F.; Sacchi, A.; Gasparri, A.; Longhi, R.; Bachi, A.; Doglioni, C.; Bordignon, C.; Traversari, C.; Rizzardi, G.-P.; Corti, A. *Cancer Res.* **2008**, 68, 7073 – 7082; b) Spitaleri, A.; Mari, S.; Curnis, F.; Traversari, C.; Longhi, R.; Bordignon, C.; Corti, A.; Rizzardi, G.-P.; Musco, G. *J. Biol. Chem.* **2008**, 283, 19757 – 19768; c) Curnis, F.; Cattaneo, A.; Longhi, R.; Sacchi, A.; Gasparri, A. M.; Pastorino, F.; Di Matteo, P.; Traversari, C.; Bachi, A.; Ponzoni, M.; Rizzardi, G.-P.; Corti, A. *J. Biol. Chem.* **2010**, 285, 9114 – 9123; d) Ghitti, M.; Spitaleri, A.; Valentinis, B.; Mari, S.; Asperti, C.; Traversari, C.; Rizzardi, G.-P.; Musco, G. *Angew. Chem. Int. Ed.* **2012**, 51, 7702 – 7705.

[37] Mingozzi, M.; Dal Corso, A.; Marchini, M.; Guzzetti, I.; Civera, M.; Piarulli, U.; Arosio, D.; Belvisi, L.; Potenza, D.; Pignataro, L. and Gennari, C. *Chem. Eur. J.* **2013**, 19, 3563 – 3567.

[38] a) Frank, A. O.; Otto, E.; Mas-Moruno, C.; Schiller, H. B.; Marinelli, L.; Cosconati, S.; Bochen, A.; Vossmeyer, D.; Zahn, G.; Stragies, R.; Novellino, E.; Kessler, H. *Angew. Chem. Int. Ed.* **2010**, 49, 9278 – 9281; b) Spitaleri, A.; Ghitti, M.; Mari, S.; Alberici, L.; Traversari, C.; Rizzardi, G.-P.; Musco, G. *Angew. Chem. Int. Ed.* **2011**, 50, 1832 – 1836.

[39] Meinecke, R. and Meyer, B. *J. Med. Chem.* **2001**, 44, 3059 – 3065.

[40] Ptaff, M.; Tangemann, K.; Müller, B.; Gurrath, M.; Müller, G.; Kessler, H.; Timpl, R. and Engel, J. *J. Biol. Chem.* **1994**, 269, 20233 – 20238.

[41] Keenan, R. M.; Miller, W. H.; Kwon, C.; Ali, F. E.; Callahan, J. F.; Calvo, R. R.; Hwang, S. M.; Kopple, K. D.; Peishoff, C. E. et al. *J. Med. Chem.* **1997**, 40, 2289 – 2292.

[42] Claasen, B.; Axmann, M.; Meinecke, R. and Meyer, B. *J. Am. Chem. Soc.* **2005**, 127, 916 – 919.

[43] a) Plantefaber, L. C. et al. *Cell* **1989**, 56, 281; b) Felding-Habermann, B. et al. *J. Clin. Invest.* **1992**, 89, 2018 - 2022; c) Mizejewski, G. J. *Proc. Soc. Exp. Biol. Med.* **1999**, 222, 124.

[44] Potenza, D.; Vasile, F.; Seminars in Organic Synthesis, "A. Corbella" Summer School, 37th, Gargnano, Italy, June 18 – 22, 2012, 225 – 248.

[45] a) Müller, G.; Gurrath, M.; Kessler, H. *J. Comput.-Aided Mol. Des.* **1994**, 8, 709 – 730; b) Haubner, R.; Gratias, R.; Diefenbach, B.; Goodman, S. L.; Jonczyk, A.; Kessler, H. *J. Am. Chem. Soc.* **1996**, 118, 7461 –

7472; c) Haubner, R.; Schmitt, W.; Hölzemann, G.; Goodman, S. L.; Jonczyk, A.; Kessler, H. *J. Am. Chem. Soc.* **1996**, 118, 7881 – 7891.

[46] Tamkun, J. W. et al. *Cell* **1986**, 46, 271 – 282.

[47] Eldrige, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. *Journal of Computer-Aided Molecular Design* **1997**, 11, 425-445.

CHAPTER 3
ANALYSIS OF LIGAND – CADHERIN INTERACTIONS

Section 1: Introduction

This part of my PhD thesis is focused on recent findings obtained in the field of rational design of peptidomimetic targeting the protein - protein interactions (PPIs) of cadherins, in the framework of the FIRB project coordinated by Dr. M. Civera (Computer-aided design, synthesis and biological evaluation of peptidomimetics targeting N-cadherin as anticancer agents, MIUR-FIRB "Futuro in Ricerca" RBFR088ITV).

The role played by PPIs is central in many physiological processes, from cell-cell messaging to cellular apoptosis. Targeting the interfaces between proteins has huge therapeutic potential, but developing drug-like small molecules that modulate PPIs is a great challenge.[1] Modern molecular and cellular biology have enabled enormous progress in unveiling many aspects of this complex network. As part of an ongoing investigation into the cellular adhesion process mediated by protein – protein interactions, our group set out to rationally design and synthesize small peptidomimetic molecules aimed at providing partial or total inhibition of the cadherin – cadherin interaction.[2] Cadherins, in fact, promote cell adhesion by an homophilic interaction of their N-terminal extracellular domains. Classical cadherins, including type I and II subfamilies, consist of an extracellular portion, constituted by five domains (EC1-EC5) rigidified at the interface by calcium ions, a single pass transmembrane region and a conserved C-terminal cytoplasmic tail that interacts with the actin cytoskeleton through the binding to intracellular molecules of the catenins superfamily.[3] Dysregulation of the type I epithelial E-cadherin, neuronal N-cadherin and placental P-cadherin has been shown to contribute to different aspects of tumour progression.[3a] Despite a growing interest in the field, the rational design of small ligands targeting cadherins protein – protein interactions is still in a very early stage. Our group designed the first small library of peptidomimetics targeting the adhesive interface identified by recently resolved X-ray structure of N- and E-cadherin homophilic dimers.[4] Virtual screening results highlights three most promising peptitomimetic ligands for synthesis and biological evaluation as cadherin PPI inhibitors.[2]

Surprisingly, structural and *in silico* studies of small molecule - cadherin interactions have not been reported yet. In this context, I applied different NMR techniques to improve the

structural understanding of the binding of the peptidomimetic ligands to the extracellular domains of the cadherins. In particular, during the last part of my PhD, I focused the attention on the interaction between peptidomimetic ligands and E- and P-cadherin EC1-EC2 fragment by using different NMR techniques. NMR can give a fundamental contribution to the characterization of ligand - cadherin complexes, during my PhD I performed:

- a detailed conformational study of peptidomimetic ligands by VT-NMR and NOESY techniques.
- a study of the interactions between these ligands and the EC1-EC2 constructs of the E- and P-cadherin by *tr*-NOESY and STD-NMR experiments.[5,6] This study was performed on the isolated protein in solution. A recent paper demonstrated, in fact, that these proteins have a role in tumour progression, either if embedded into the membrane or free in solution.[7]

3.1.1 CADHERINS: function and structure

Animal tissues are bound together thanks to adhesive forces in their component cells. Two principal types of adhesive junctions exist: desmosomes and adherens junctions, and their role is the maintenance of cell shape and tissue integrity. Adherens junctions are cell-cell adhesion complexes found in a variety of cells,[8] characterized by opposed plasma membranes with an in-between space of 15 to 30 nm. Desmosomes reinforce adhesion and are mostly present in organs like heart and skin.[9]

Among the adherens junction components, cadherins are the core elements.[10] Cadherins are a superfamily of cell surface glycoproteins whose ectodomains contain multiple repeats of β-sandwich extracellular cadherin (EC) domains, in particular they are the prototypical example of calcium-dependent homophilic cell-cell adhesion.[3b]

The level of cadherin expression influences the strength of adhesion, whereas the type of cadherin expressed determines the specificity and the properties of cell interactions.

Cadherins found in adherens junctions were the first members of a broader superfamily of cadherins to be discovered, and thus are now being called classical cadherins. They are grouped in subfamilies based on selected sequence features and on the number and the arrangement of their

extracellular domain (EC) repeats (**Figure 3.1.1.**), which are common structural motifs of ca. 110 residues showing an immunoglobulin-like fold and numbered according to their distance from the membrane, being EC1 the N-terminal membrane-distal domain.

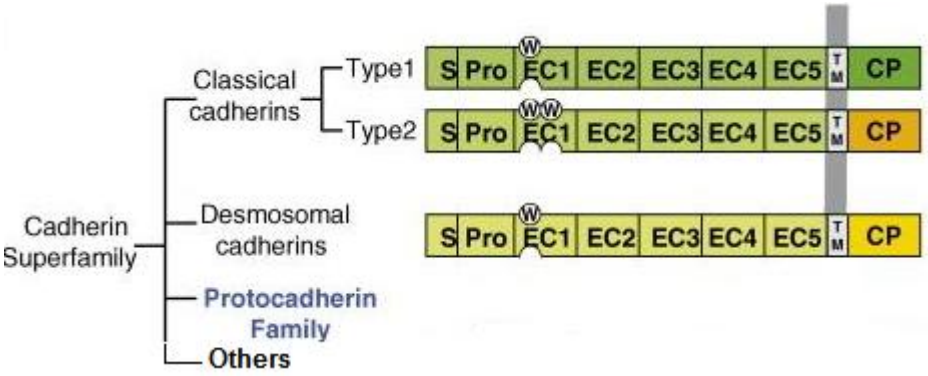


Figure 3.1.1. Schematic representation of members of the cadherin family.

Phylogenetic analyses of cadherin protein sequences reveal a number of subfamilies: type I and type II cadherins, desmosomal cadherins, protocadherins and a variety of other cadherin-related proteins. Vertebrate classical cadherins, comprising six type I and thirteen type II cadherins, are single-pass transmembrane proteins. Desmosomal cadherins, expressed in all vertebrate animals, have a domain organization similar to that of classical cadherins. However, they are attached via distinct cytoplasmic proteins to intermediate filaments forming specialized cell-cell junctions, referred to as desmosomes, in tissues exposed to high mechanical stress. All these families share a similar primary sequence and share a capacity for homophilic binding, but the molecular mechanisms used by different subfamilies to accomplish this appears to vary significantly.[3b]

Classical cadherins are structurally characterized by five extracellular domains (EC1-EC5) rigidified at the interface by calcium ions, a single pass transmembrane region and a cytoplasmic tail (**Figure 3.1.2.**). The five EC domains are preceded by a signal sequence and a prodomain that must be removed by proteolysis for the adhesive functions.

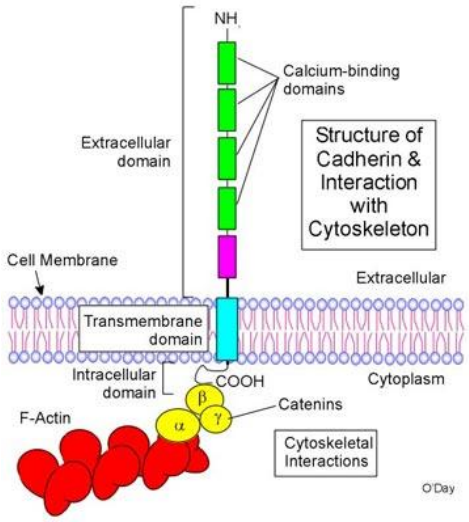


Figure 3.1.2. Overall architecture of classical cadherins.

In the adherens junctions, classical cadherins form *trans* complexes by bridging the intercellular space via their ectodomains. They project from opposing cell surfaces and form adhesive homodimers by interacting via their EC1 domains. The binding between cadherins and the actin cytoskeleton allows the structural stabilization of adherens junctions and promotes the regulation of cell morphology and motility.[11]

In addition to *trans* dimerization, the adhesion is strengthened by the lateral or *cis* association of cadherin ectodomains extending from the same cell surface. With respect to the *trans* dimer, the *cis* dimer structures appear to involve a different portion of EC1 that interacts with EC2 of a neighbouring molecule emerging from the surface of the same cell (**Figure 3.1.3.**).[4]

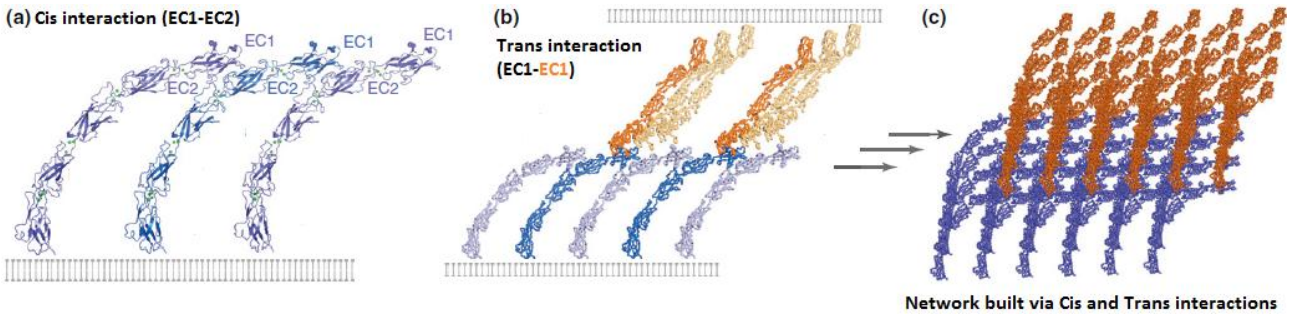


Figure 3.1.3. Proposed structural architecture of adherens junctions by X-ray dimer structures a) *cis* interaction, b) *cis* and *trans* interaction and c) the final net of *cis* and *trans* organized interactions.[4]

On the basis of the recently published crystallographic structures only of the whole type I E-(epithelial) and N (Neuronal)-cadherin ectodomain dimers,[4] a model of adherent junction architecture has been proposed (**Figure 3.1.3.**). According to the X-ray model, each cadherin monomer could simultaneously engage two molecules for lateral association and one for *trans* dimerization, resulting in an ordered two-dimensional layer that could represent the structural basis of the intercellular junction adhesion. Several experimental data[3b,4] showed that lateral binding is weaker compared to *trans* homodimerization and it appears to exert a supporting role in cadherin-mediated adhesion with respect to the primary function of the *trans* interface. The recently published structure of type I placental P-cadherin EC1-EC2 homodimers, showed a similar *trans* dimer strucuture.[12]

In all the X-ray structures of E-, N- and P-cadherin *trans* dimerization occurs through the opening of the highly conserved N-terminal six-residue portion of the EC1 (the "adhesion arm") that contains a Trp residue in position 2 (Trp2), and its swapping into an acceptor pocket of the EC1 of the interacting protein, thus resulting in the so-called "strand dimer" formation (**Figure 3.1.4.**).

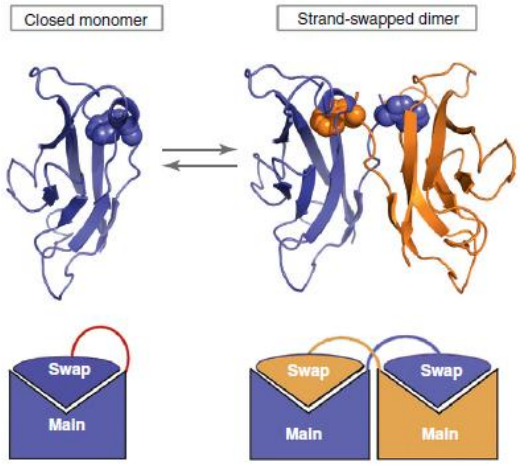


Figure 3.1.4. 3D domain swapping of EC1 domain of classical type I cadherin.[4]

Members of the same subfamily can form homo- and hetero-dimers, but different cadherins subfamilies do not adhere to each other, suggesting a high degree of specificity in the adhesion process.[13] Many doubts still remain on the path leading to the cadherin dimerized

structures.[14,15] The key feature of strand exchange dimerization is the docking of Trp2 side chain into the hydrophobic acceptor pocket of the intaracting cadherin (**Figure 3.1.5.**).[16]

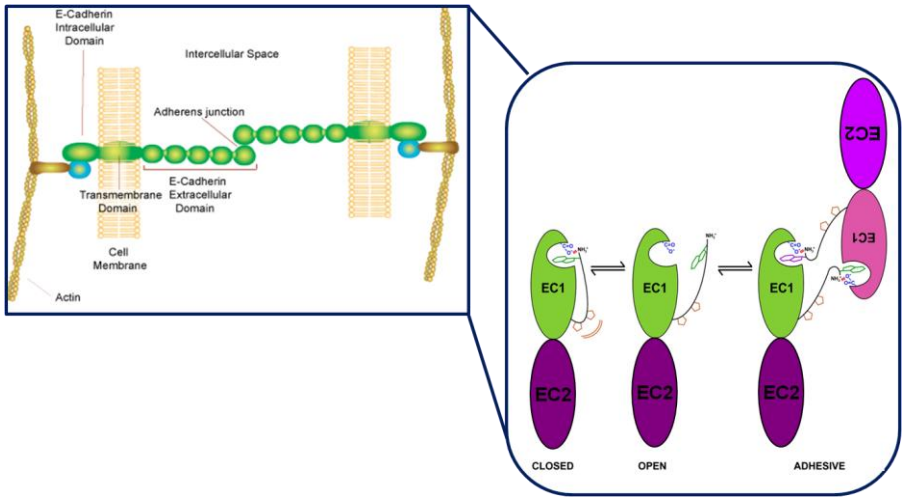


Figure 3.1.5. A model for the binding mechanism of type I cadherins.

As reported in **Figure 3.1.6.**, E-, N- and P-cadherin shared a high level of homology sequence in their EC1-EC2 domain.[4,12]

Human E-cadh **DWVIPP**ISSPENEKGPFPKNLVQIKSNKDKEGKVFYSITGQGADTPPVGVFIIERET 57
Human N-cadh **DWVIPP**INLPENSRGPFPQELVRIRSDRDKNLSLRISVTGPGADQPPTGIFILNPIS 57
Human P-cadh **DWVVAP**ISVPENGKGPFPQRLNQLKSNKDRDTKIFYSITGPGADSPPEGVFAVEKET 57

Human E-cadh GWLKVTEPLDRERIATYTLFSHAVSSNGNAVEDPMEILITVTDQNDNKPEFTQEVFK 114
Human N-cadh GQLSVTKPLDRQQNARFHLRAHAVDINGNQVETPIDIVINVIDMNDNRPEFLHQVWN 114
Human P-cadh GWLLLNKPLDREEIAKYELFGHAVSENGASVEDPMNISIIVTDQNDHKPKFTQDTFR 114

Human E-cadh GSVMEGALPGTSVMEVTATDADDDVNTYNAAIAYTILSQDPELPDKNMFTINRNTGV 171
Human N-cadh GSVPEGSKPGTYVMTVTAIDADDPNALNGMLRYRILSQAPSTPSPNMFTINNETGDI 171
Human P-cadh GSVLEGVLPGTSVMQVTATDEDDAIYTYNGVVAYSIHSQEPKDPHDLMFTIHRSTGT 171

Human E-cadh ISVVTTGLDRESFPTYTLVVQAADLQGEGLSTTATAVITVTD 213
Human N-cadh ITVAAGLDREKVQQYTLIIQATDMEGNPTYGLSNTATAVITV 213
Human P-cadh ISVISSGLDREKVPEYTLTIQATDMDGDGSTTTAVAVVEILD213

Figure 3.1.6. Sequence of E-, N- and P-cadherin EC1-EC2 fragment.

3.1.2 X-ray structures of E- and P-cadherin homodimers

The swap dimer crystal structures of the whole EC1-EC5 ectodomain of mouse E-cadherin and EC1-EC2 fragment of mouse P-cadherin were available on Protein Data Bank (3q2w.pdb and 4nqq.pdb, respectively).[4,12] For the E-cadherin also the human EC1-EC2 X-ray dimer structure was resolved (2o72.pdb)[16] and showed no significant differences compared to the mouse crystallographic homodimer.

The key crystallographic contacts of the DWVI adhesive sequence with E-cadherin binding site (3q2w.pdb, **Figure 3.1.7.**) can be summarized as follows: **1.** the formation of an intermolecular salt bridge between the charged N-terminal amino group of Asp1 and the side chain carboxylate of Glu89; **2.** the anchoring of the Trp2 side chain into a hydrophobic pocket and **3.** the formation of a hydrogen bond between the indole moiety and the carbonyl group of Asp90; **4.** the involvement of Val3-NH in a hydrogen bond with the carbonyl group of Lys25; **5.** the formation of an hydrogen bond between the backbone carbonyl group of Asp1 and the Asn27-NH group.

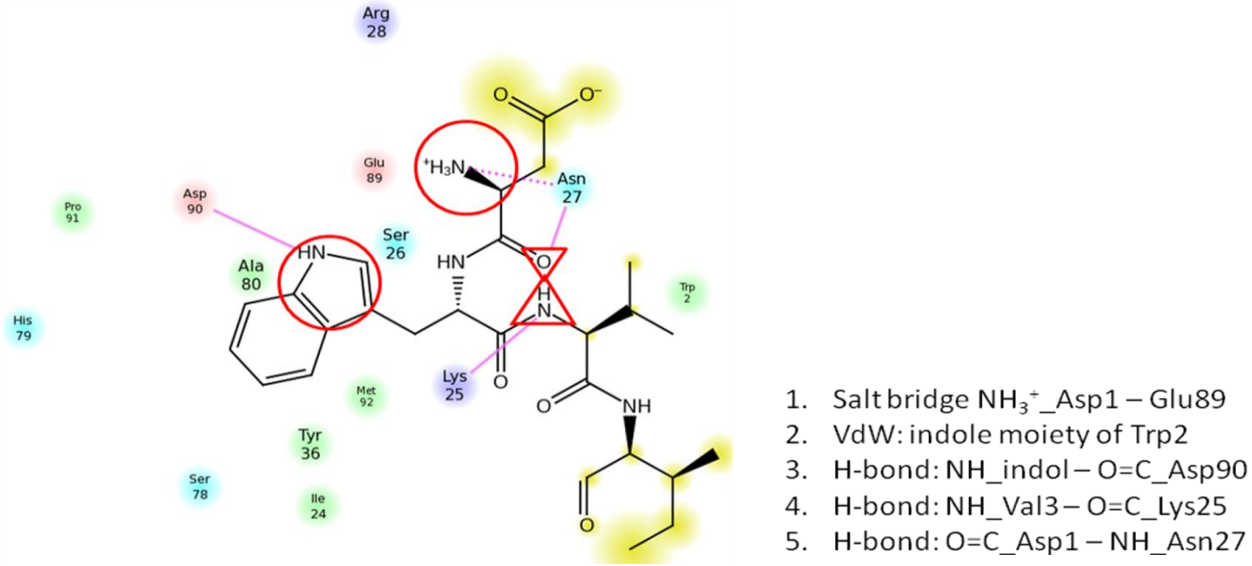


Figure 3.1.7. 2D representation of the DWVI interactions into the E-cadherin binding site (residue within 4 Å from DWVI are shown).

Mouse P-cadherin, like E-cadherin, forms strand swap dimers by exchanging the Trp2 indole moiety into the corresponding pocket. Compared to E-cadherin, the P-cadherin N-terminal tetrapeptide sequence mutates the Asp1 and Ile4 into Glu1 and Met4 amino acids, respectively. As

observed for the interactions formed by the adhesive EWVM motif into P-cadherin X-ray structure (**Figure 3.1.8.**), the mutation does not affect the binding mode of the sequence within the binding site with respect to the one described for the DWVI motif in E-cadherin pocket. In fact, the salt bridge between Glu-NH_3^+ terminal group and a conserved Glu89 side chain is formed; the Trp2 side chain docks into a hydrophobic pocket and forms an hydrogen bond with the carbonyl group of Glu90; Val3-NH is hydrogen bonded to the carbonyl group of Lys25 and also the backbone carbonyl group of Glu1 forms an hydrogen bond with the Asn27-NH group. In addition, for P-cadherin, we observed an interaction of Glu1 carboxylic group and Asn27 side chain.

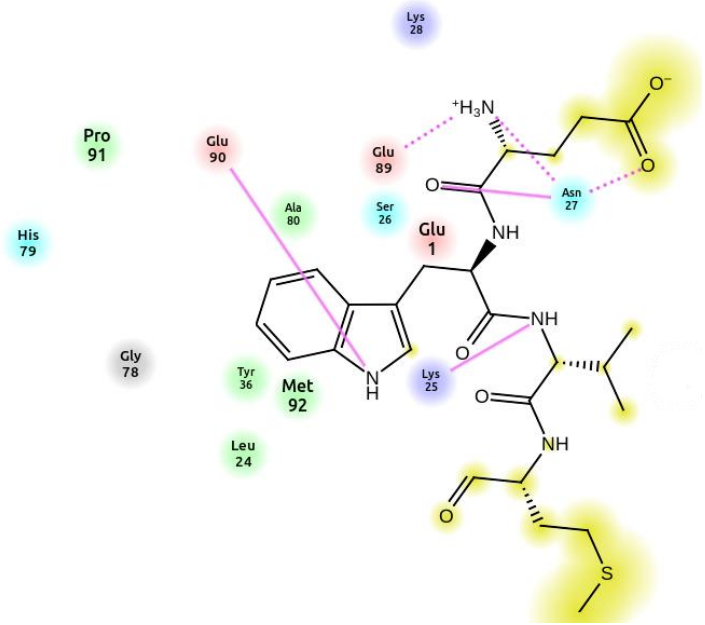


Figure 3.1.8. 2D representation of the EWVM interactions into the P-cadherin binding site (residue within 4 Å from EWVM are shown).

It is reasonable to assume that also the human P-cadherin (94% of homology sequence to mouse EC1-EC2 domains) undergoes to a domain swapping of the DWVVAP adhesive sequence and forms a swap dimers. Compared to E-cadherin DWVIPP sequence, human P-cadherin replaced the Ile4 and Pro5 with Val4 and Ala5 residues, respectively. The Pro5-Pro6 motif is supposed to mediate conformational changes necessary for strand swapping[17] by introducing a strain in the monomer adhesive arm intramolecularly docked into its own pocket in the monomer state that is

released in the swap dimer. In the case of P-cadherin, the replacing of one proline with an alanine residue should produce a more flexible adhesive arm.

3.1.3 E-cadherin in cancer

E-cadherin, expressed by epithelial cells, is regarded as the prototypical example of type I cadherin. E-cadherin expression on cells varies during the epithelial-mesenchymal transition (EMT), characterized as a phenotypic transition able to transform a quiescent epithelial cell in a highly motile and invasive cell. During cancer progression, in particular, epithelial cells undergoing EMT acquire the ability to migrate in a directional way, dissociating one from each other. In particular, during EMT E-cadherin is down-regulated and N-cadherin is *de novo* expressed at the same time, in a process called cadherin switching.[11,18,19]

Although E-cadh is considered a repressor for the majority of carcinomas, it appeared to play a role in the proliferation of epithelial ovarian cancer (EOC). E- and N-cadherins can be co-expressed in some advanced-stage EOC, leading to the conclusion that E-cadherin expression and its homophilic interaction contributes to the proliferation of EOC cells.[20]

3.1.4 P-cadherin in cancer

The mature P-cadherin glycoprotein has a molecular weight of 118 KDa, and its structure is similar to that of classical cadherins, but different from those of E-cadherin and N-cadherin in terms of immunological specificity and molecular mass.[21]

E- and P-cadherin are the first adhesion molecules that are expressed in the mouse embryo. P-cadherin expression was observed in the placenta, both in the embryonic and maternal regions. In contrast to E-cadherin, P-cadherin seemed to be frequently up-regulated in tumours, being reduced when tumour cells were more differentiated. Special attention is given to the role of P-cadherin in breast cancer, where a high volume of knowledge exists.[22] However, E-cadherin induced signaling pathways have been far studied in cancer; still, the signaling pathways activated by P-cadherin in cancer is not clear. The first limitation is due to P-cadherin dual role, as an invasion promoter or invasion suppressor depending on the cell model under study.[23]

3.1.5 The role of calcium

Cadherins are named for the dependence of their adhesive function on the presence of extracellular calcium. Before their structures were unknown, it was speculated that Ca^{2+} ions might bridge the adhesive interface. However, the role of calcium in cadherin function is far more complex. Calcium binds to cadherins at stereotyped binding sites situated between successive EC domains. Each of these sites binds three Ca^{2+} ions in a highly cooperative manner such that each five-domain classical cadherin coordinates twelve Ca^{2+} ions in total. The binding affinities of the Ca^{2+} sites vary, but all bind with a dissociation constant (K_d) lower than the Ca^{2+} concentration characteristic of the extracellular milieu, approximately 1 mM.[3b] Thus, it is expected that cadherin ectodomains will be fully Ca^{2+} - occupied under physiological conditions. Some roles are now understood for Ca^{2+} binding in classical cadherins. The first is rigidifying the ectodomain so that it adopts a characteristic crescent shape, although this structure retains considerable conformational flexibility. Thus, Ca^{2+}-mediated rigidification is critical to adhesive *trans* binding.

Another role of the calcium is to induce oligomerization of the N-terminal two domains of epithelial cadherin and this process was followed also by NMR spectroscopy in solution over a large range of protein (10 μM – 5 mM) and calcium (0 – 5 mM) concentrations.[24,25,26] Several spectrally distinct states could be distinguished that correspond to a calcium-free monomeric form, a calcium-bound monomeric form, and to calcium-bound higher oligomeric forms (**Figure 3.1.9.**).

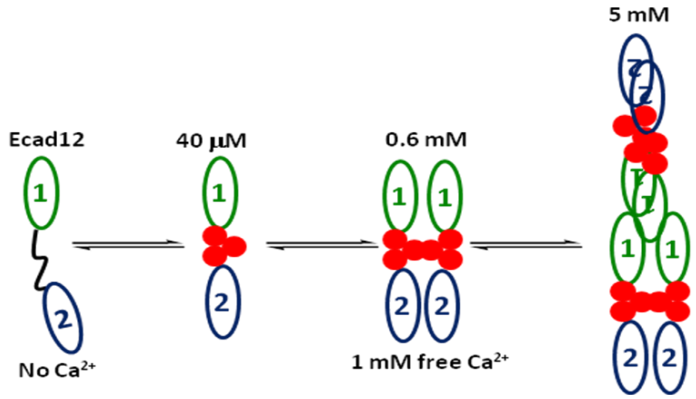


Figure 3.1.9. Schematic representation of the aggregation states.

The calcium-free monomer adopts a flexible, kinked conformation that occludes the dimer interface. In contrast, the calcium-bound monomer is already in a straight, non-flexible conformation where this interface is accessible.

3.1.6 TYPE I CLASSICAL CADHERIN: known ANTAGONISTS

Based on the first X-ray dimer of N-cadherin,[27] a first library of cyclic peptide containing the recognition sequences His79-Ala80-Val81 (HAV) and the Ile53-Asn54-Pro55 (INP) was developed.[28] Some of these peptides were shown to be able to modulate the outgrowth of neurite expressing N-cadherin either in "antagonistic" or "agonistic" modality. Among them, the antagonist peptide N-Ac-CHAVC-NH_2 (ADH-1 or Exherin^{TM}, **Figure .**) containing the HAV motif, was shown to disrupt endothelial cell adhesion, induce apoptosis and inhibit angiogenesis in millimolar (mM) concentrations with favourable results in animal models of prostate and pancreas cancer, and of melanoma.[29,30,31]

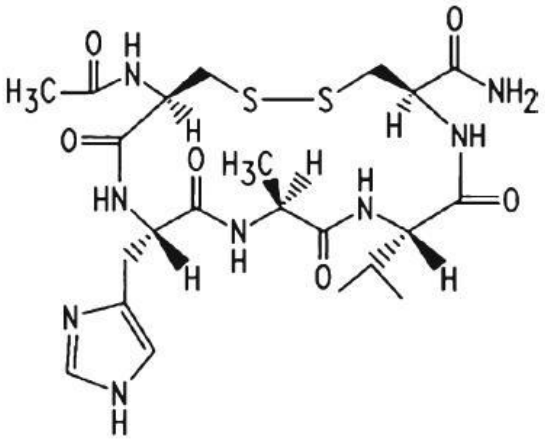


Figure 3.1.10. 2D structure of the cyclic peptide ADH-1.

Thus, it was promoted to phase I clinical investigation in patients with advanced solid tumours which express N-cadherin, showing good tolerability in high doses, and some encouraging results suggestive of a potential therapeutic value.[3e,32,33]

Libraries of small peptide or non-peptide compounds were explored by a virtual screening protocol as possible mimics of HAV and related sequences, and some compounds were shown to inhibit the N-cadherin-mediated neurite outgrowth and cell adhesion.[34,35] Phage display

technology was employed to screen libraries of 12mer peptides against chimeric proteins composed of the human N-cadh or E-cadh ectodomains fused to the Fc fragment of human immunoglobulin G1.[36] All the isolated clones contained a Trp residue in position 2, with high variability throughout the sequence, including positions 1, 3, and 4. A linear peptide (H-SWELYYPLRANL-NH_2) was reproduced by synthesis and shown to inhibit the adhesion of human breast cancer cells expressing E- and N-cadh in the mM range.

3.1.7 Rationally designed peptidomimetic ligands

This part of the investigation was performed by the research units involved in the FIRB project on N-cadherin antagonists aiming at the identification of possible inhibitors of N-cadherin-mediated adhesion in cancer cell lines.

In this project, following our interest in conformationally constrained peptidomimetic ligands taming protein - protein interactions, a small library of peptidomimetics were developed based on the tetrapeptide sequence Asp1-Trp2-Val3-Ile4 (DWVI) of the N-terminal "adhesive arm" of E- and N-cadherin (as identified by the analysis of all the relevant, available crystallographic structures).[2]

A virtual library of about 30 peptidomimetics of general formula NH_3^+-Asp-scaffold-Ile-$NHCH_3$ was designed by replacing the central dipeptide Trp2-Val3 unit of the DWVI adhesive motif with several scaffolds developed in our laboratories.[37] The peptidomimetics feature a conformationally constrained scaffold bearing a benzyl or phenyl ring to mimic the indole moiety of Trp2 of the native sequence. Depending on the configurational features of the scaffold, the aromatic ring was expected to be properly oriented into the corresponding binding pocket.

To evaluate the ability of the virtual library compounds to reproduce the DWVI key interactions observed in the dimer structures, our group built a model of EC1 fragment of E- and N-cadherin starting from the corresponding X-ray structures and set up a docking protocol using the Glide V5.7 software[38] (see "Experimental Section"). Docking results were sorted on the basis of Glide score and filtered to match the most important binding interactions observed for the DWVI motif during molecular dynamic simulations run on the X-ray EC1-EC2 dimers of E- and N-

cadherin. The interactions pattern already described for the E-cadherin X-ray dimers (**Figure 3.1.7.**), well represents also the interactions observed into N-cadherin X-ray binding site. Moreover, during the simulations, the two systems maintained the key crystallographic contacts. In fact, the same three ligands were selected by docking calculations into the E- and N-cadherin models as most promising inhibitors (**Figure 3.1.11.**).

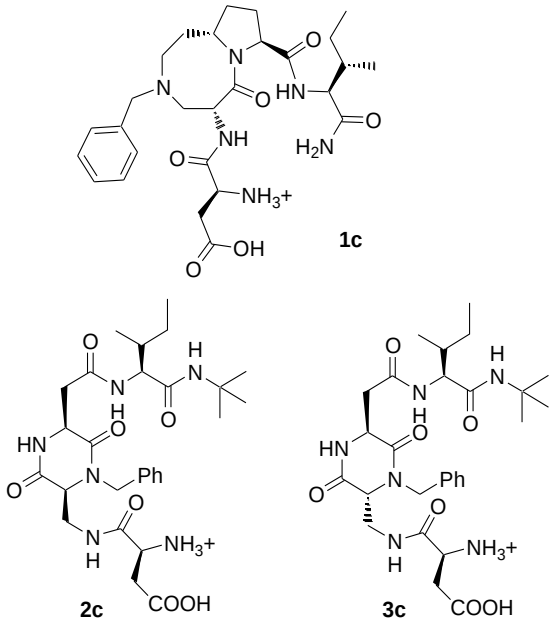


Figure 3.1.11. The three peptidomimetic ligands synthesized.

In the top-ranked poses, all the three ligands insert the scaffold benzyl ring into the corresponding hydrophobic pocket of Trp2 and form the salt bridge between the Asp1-NH_3^+ group and the Glu89 side chain. The main difference in the binding mode of the three ligands is related to their 3D alignment with respect to the DWVI sequence of the crystallographic structures: while **2c** and **3c** are able to mimic the extended conformation of the reference tetrapeptide backbone in both cadherin receptors (**Figure 3.1.12.**), **1c** orients the Ile residue back to the Asp1 amino acid and does not reproduce the experimental backbone arrangement (**Figure 3.1.13.**).

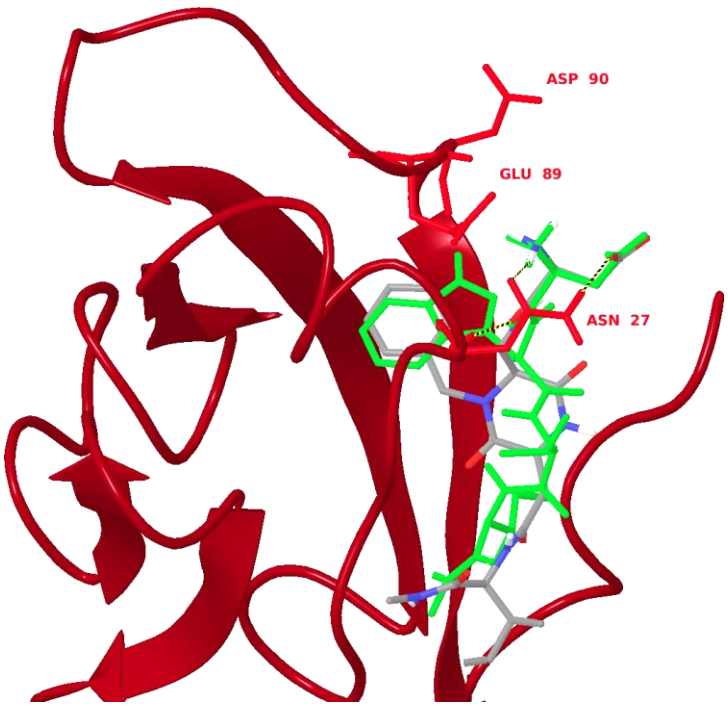


Figure 3.1.12. Best pose of **2c** (tube representation, C in grey, N in blue and O in red) into the E-cadh binding site (red cartoon representation), overlaid to the DWVI sequence (green tube representation) of the X-ray dimer. Key residues of E-cadh involved in the interaction are labelled and highlighted in tube representation.

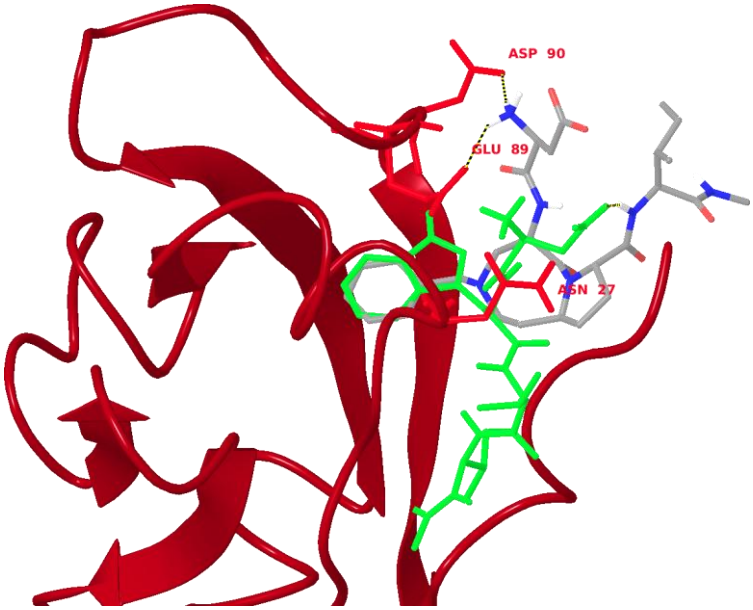


Figure 3.1.13. Best pose of **1c** (tube representation, C in grey, N in blue and O in red) into the E-cadherin (red ribbon representation) model, overlaid to the DWVI sequence (green tube representation). Key receptor residues are labeled and highlighted in tube representation.

The compounds were synthesized in the group of Prof. U. Piarulli (University of Insubria) and Dr. L. Manzoni (CNR, Consiglio Nazionale delle Ricerche). The ligands were tested by ELISA for their ability to inhibit calcium-dependent cadherin binding using the N-cadherin-expressing EOC cell line SKOV3 and N-cadherin-Fc chimeric protein. ADH-1 was also synthesized and used as reference. The biological tests were performed by the group of Dr. A. Tomassetti in the Department of Experimental Oncology and Molecular Medicine (Istituto Nazionale Tumori). Both **2c** and **3c** ligands at 2 mM concentration inhibited N-cadherin homophilic binding by 84% and 78%, respectively, and 65% and 50% at 1 mM concentration (**Figure 3.1.14.**). ADH-1 and **1c** showed about 50% inhibition of N-cadherin-Fc/cells interactions at the higher concentration, and appeared nearly ineffective at the lower concentration (**Figure 3.1.14.**). The compounds were then analyzed for their ability to inhibit the adhesion of the EOC (epithelial ovarian cancer) cell lines by observing the formation of cell monolayers in presence of each ligand at two different concentrations (2 and 1 mM). All compounds inhibited the formation of cell monolayers of N-cadherin-expressing cells at 2 mM concentration, comparably to ADH-1 (**Figure 3.1.15.**). Notably, **2c** and **3c** were also active at 1 mM concentration, and **2c** was able to inhibit cell–cell aggregation of N-cadherin-expressing cells in suspension (**Figure 3.1.14.**). When tested on the E-cadherin-expressing EOC cell line OAW42, all compounds showed lesser efficiency in inhibiting the E-cadherin homophilic interactions as well as the formation of cell monolayer with relevant effects only at the higher concentration (**Figure 3.1.14.** and **3.1.15.**). In particular, by ELISA, at the concentration of 2 mM **1c** and **2c** inhibited E-cadh-Fc binding to the cells by about 50% while ADH-1 showed 30% inhibition (**Figure 3.1.14.**), indicating a slightly better efficacy compared to ADH-1 in inhibiting also E-cadherin homophilic interactions. Compound **3c** was also able to inhibit the formation of cell monolayers at 1 mM concentration (**Figure 3.1.15.**).

Overall, compounds **2c** and **3c** resulted to be effective in inhibiting N-cadherin homophilic and, to some extent, also E-cadherin homophilic adhesion.

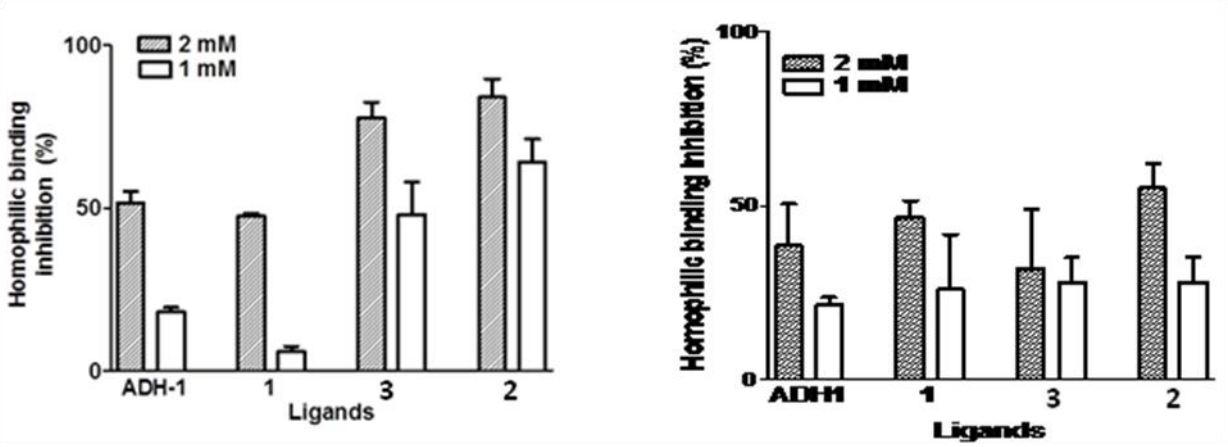


Figure 3.1.14. ELISA Inhibition assays: of N-cadherin-Fc (Left) and E-cadherin-Fc (right) binding to SKOV3 and OAW42 cells, respectively.

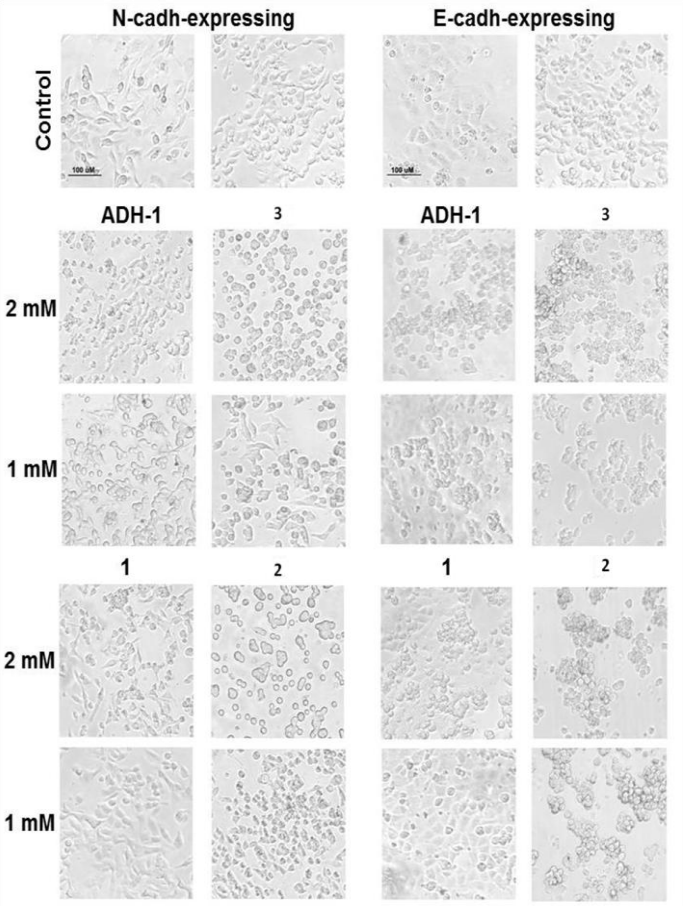


Figure 3.1.15. Adhesion assay to evaluate the inhibition of the formation of the cell monolayer by the small peptidomimetic ligands. EDTA-detached N-cadherin- (SKOV3) or E-cadherin-(OAW42) expressing cells were seeded in absence (Control) or in presence of the ligands at 2 and 1 mM. For

each cell line the controls without the ligand but with same amount of the solvent (water on the left, DMSO on the right) are reported.

Encouraged by the very promising results with N-cadherin-expressing cells, we set out to evaluate by Surface Plasmon Resonance (SPR) analysis the ability of the compounds **2c** and **3c** to specifically inhibit the N-cadherin homo-dimerization in a cell-free experiment using the N-cadherin-Fc recombinant protein. As shown in the sensogram of **Figure 3.1.16.**, at 10 μM concentrations compounds **2c** and **3c** provided 55% and 98% inhibition of N-cadh homophilic binding, respectively (**Figure 3.1.16.**), while ADH-1 showed only 28% inhibition (**Figure 3.1.16.C**). These data demonstrate the ability of both compounds **2c** and **3c** to specifically bind to N-cadherin even in the μM range.

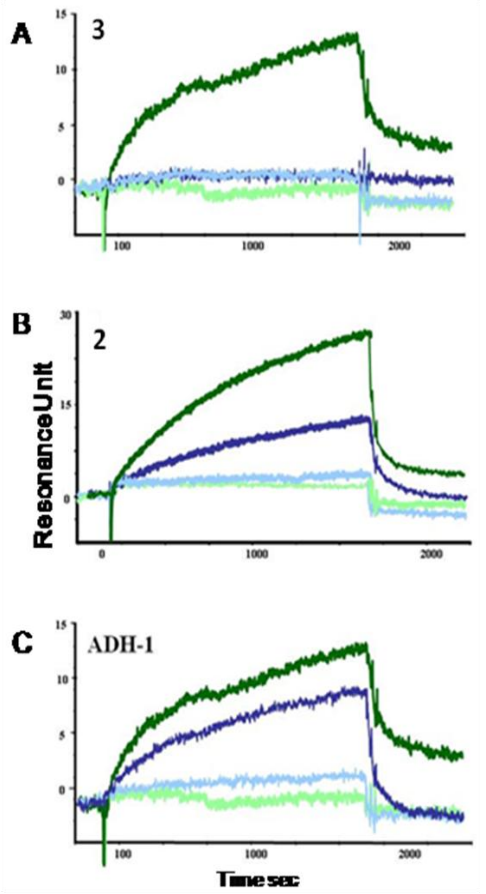


Figure 3.1.16. Biacore profiles of the homophilic binding of N-cadh in the presence or absence of inhibitor. A. Inhibition with compound **3c**. B. Inhibition with compound **2c**. C. Inhibition with ADH-1. Dark green and blue line represent the non-inhibited and inhibited N-cadh, respectively, of the

different compounds on N-cadh-Fc-bound to the sensor chip; pale green and pale blue lines represent the binding to the uncorrelated protein-bound to the sensor chip. The percentage of inhibition has been calculated considering as 100 % the resonance unit max obtained with the N-Cadh-Fc alone run immediately before the loading of each inhibitor/N-cadh-Fc complex.

However, although N- and E-cadherin share similar adhesive binding features for the DWVI sequence[4] and marked differences in the interaction mode of our peptidomimetic ligands were not observed in the two cadherin docking models, the tests on the E-cadherin-expressing EOC cell line OAW42 showed lower inhibition capability of E-cadherin homophilic binding compared with those of N-cadherin. Since the molecular basis of differences in adhesion among the cadherins are still unclear[39] and a variety of factors can contribute to the formation and dissociation of cadherin dimers on cell membranes, our results could be interpreted in light of different considerations.

NMR techniques could give a great contribution to the understanding of the molecular basis of the ligand - cadherin recognition process. In fact, by exploiting the collaboration with Dr. Parisini at Istituto Italiano di Tecnologia (IIT, expert in protein expression, purification and X-ray crystallography), in the last part of my PhD, I worked on the interaction between the peptidomimetic ligands **1c** – **3c** and the EC1-EC2 fragments of E-cadherin and P-cadherin (provided by Dr. Parisini). It was also possible the investigation of ligand binding to P-cadherin EC1-EC2 domains. This target share a common binding motif to E-cadherin, but a different a slightly different binding site. In the next future, thanks to the recently X-ray resolved P-cadherin structure, we are planning to study the peptidomimetic ligands **1c** – **3c** into a model of P-cadherin receptor.

Section 2: Results and Discussion

The complex and highly dynamical cadherin dimerization pathway has been partially elucidated by the wealth of biochemical and biophysical studies done to date.[3b,40] Ultimately, what appears to be understood is that a variety of transient yet critical key intermolecular interactions involving different residues at different stages of the full cadherin dimerization trajectory play all a critical role in guiding the system through a number of intermediate intermolecular protein - protein arrangements during the recognition process that leads from monomeric cadherin to strand-dimer formation and back. For instance, structural studies have suggested that strand-dimer formation may only represent one of several adhesive mechanisms that can all contribute to the overall cell-cell adhesion strength, the so-called X-dimer being one viable, alternative intermolecular interaction that provides adhesive properties to the system.[41] In this context, it is clear that the dynamic features of the cadherin dimerization process are likely to play against a traditional drug design approach whereby a key-key lock interaction would attempt to force the system into a stable molecular arrangement completely devoid of adhesive properties. Rather, it is conceivable that potential dimerization inhibitors may be designed by partially mimicking key cadherin structural elements known to be transiently involved in those intermediates that have been structurally characterized along the highly dynamic and reversible cadherin adhesion trajectory, thus dynamically interfering with the adhesion process by perturbing the delicate balance of interactions between different cadherin structural elements. We therefore anticipate that, given the dynamic and cooperative nature of the system, the design of a stable molecular interactor against the adhesion pocket would not just prove difficult, but it would fail to achieve complete inhibition. However, we hypothesize that peptidomimetic molecules that can transiently interfere with structurally validated intermediates in the cadherin dimerization pathway may successfully modulate cadherin-mediated adhesion, even in the absence of an unequivocal identification of the actual binding site and mode.

Due to the high flexibility of the system, NMR has proved to be a useful method to evaluate these kind of interactions. NMR studies were performed on compounds **1c** – **3c** in the

presence of E-cadherin EC1-EC2 and P-cadherin EC1-EC2 constructs. The experiments, STD and *tr*-NOESY, were conducted at three different temperatures (283 – 290 – 298 K).

3.2.1 COMPOUND 1c

3.2.1.1 NMR Characterization and conformational analysis

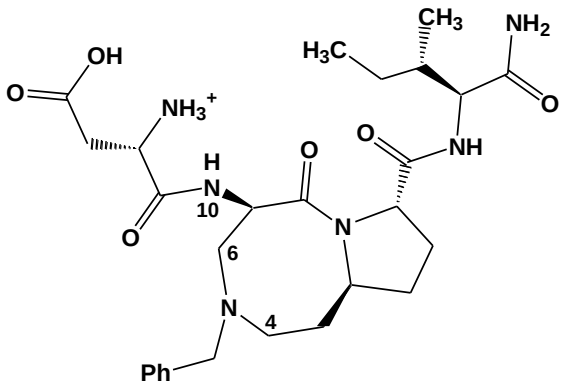


Figure 3.2.1. Structure of compound **1c**.

The complete characterization of the molecule (^{1}H and ^{13}C chemical shifts) both at 298 K and 283 K is reported in the ***Supplementary information***.

The NOESY spectra did not show any significant NOE contacts typical of a preferred conformation at both the temperatures, except for some NOE contacts between the aromatic protons and the CH_2_ 4 and 6 of the scaffold, suggesting a movement of the benzyl ring over the scaffold. In fact, the *trans* stereochemistry of the scaffold, inserted in the compound **1c**, prevents the folding of the side chains and the formation of conformational motifs stabilized by hydrogen bonds. The phase of the NOE cross peaks at 283 K is negative, while at 298 K is positive.

3.2.1.2 Interaction with E-cadherin EC1-EC2 construct

NMR experiment

We performed *tr*-NOESY[6] experiments of compound **1c** in the presence of EC1-EC2 fragment of E-cadherin. At 283 K the cross peaks are negative and the experiment does not provide information about the binding. *tr*-NOESY and NOESY did not show differences in the NOE

contacts demonstrating that the conformation of compound **1c** remains unaltered in the presence of the protein. The first STD[6] experiments at 290, 298 and 283 K showed a change in the NMR ligand epitope with temperatures (**Figure 3.2.2.**).

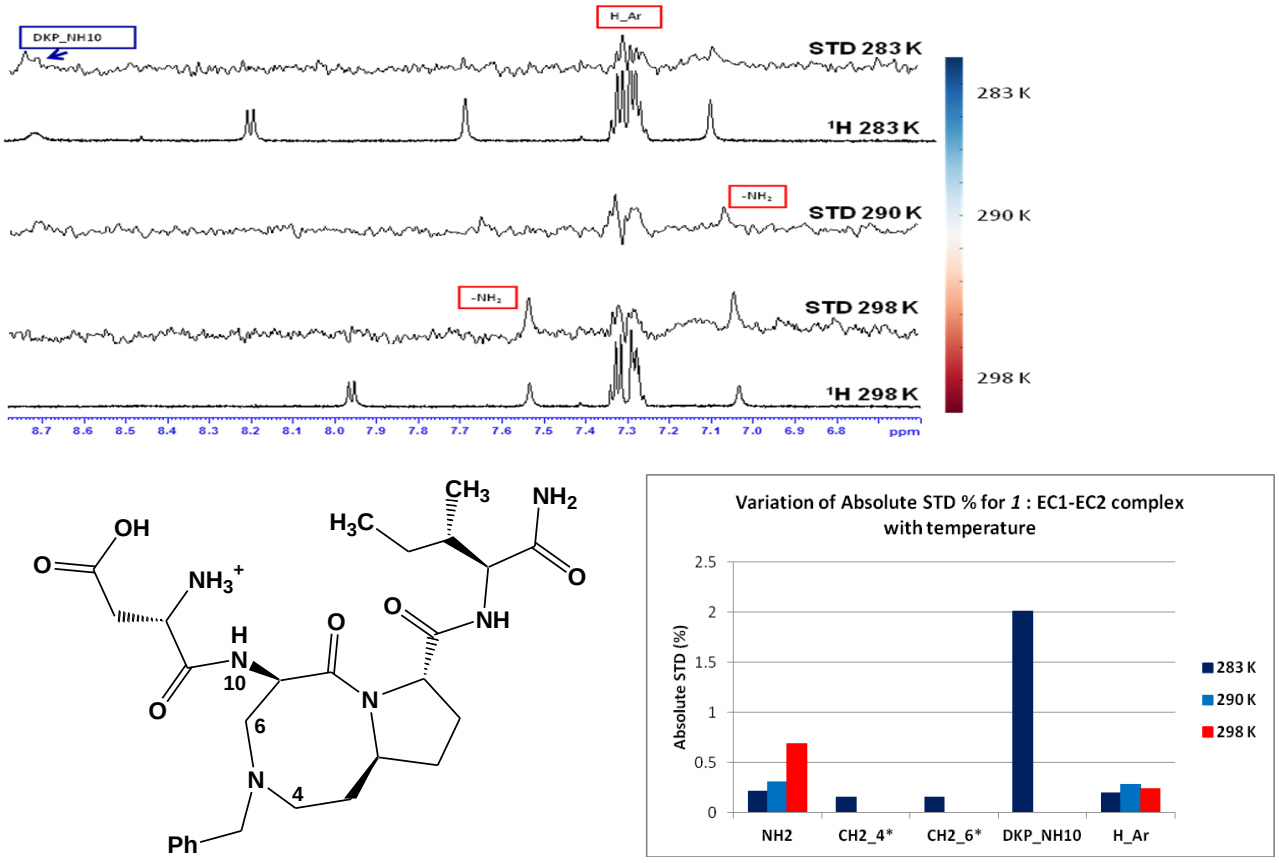


Figure 3.2.2. Low field region of the STD conducted at three different temperatures (298 K, 290 K and 283 K) with the graph of the STD absolute percentages at these three temperatures. The asterisk indicates an overlapping.

The STD signals relative to the aromatic protons (~0.3% STD absolute intensity) are present at the three temperatures with about the same intensity (**Figure 3.2.2.**), while the intensity of the STD signals of the C-terminal NH_2 protons increase with the temperature (from 0.2% at 283 K to 0.7% STD absolute intensity at 298 K), and this suggests their proximity to the protein. In contrast to what happens with the C terminal NH_2 protons, CH_2_4/CH_2_6, the protons of the scaffold, are present in low percentages (0.2% STD absolute intensity) only at 283 K. The NH10 proton shows a

decreasing STD absolute percentage raising the temperature, from a value of 2% STD absolute intensity at 283 K to its absence at 298 K. However, in the graph we have not reported the STD absolute intensity that characterized this proton at 290 K, in fact, even if it was present in the STD spectrum, its intensity was not quantifiable.

The STD epitope change with temperature suggesting three possible explanations: *i)* the conformation of the protein varies with temperature and thus the interaction with ligand change, *ii)* the ligand binding change with the temperature or *iii)* a mix of both effects. In fact, , the ligand conformation change upon binding can be excluded by the NOESY spectra. It is important to note that these preliminary experiments do not allow an exhaustive description of the ligand - cadherin interaction and other experiments have to be carried out.

Docking studies[2]

The poses of compound **1c** into E-cadherin model were analyzed in the light of the NMR epitope at different temperatures. It is worth noting that docking results do not include the effect of protein movement since during the calculation the protein is kept fixed to its X-ray structure. However, the rough estimation of ligand - protein contacts of docking poses could be useful for some preliminary considerations.

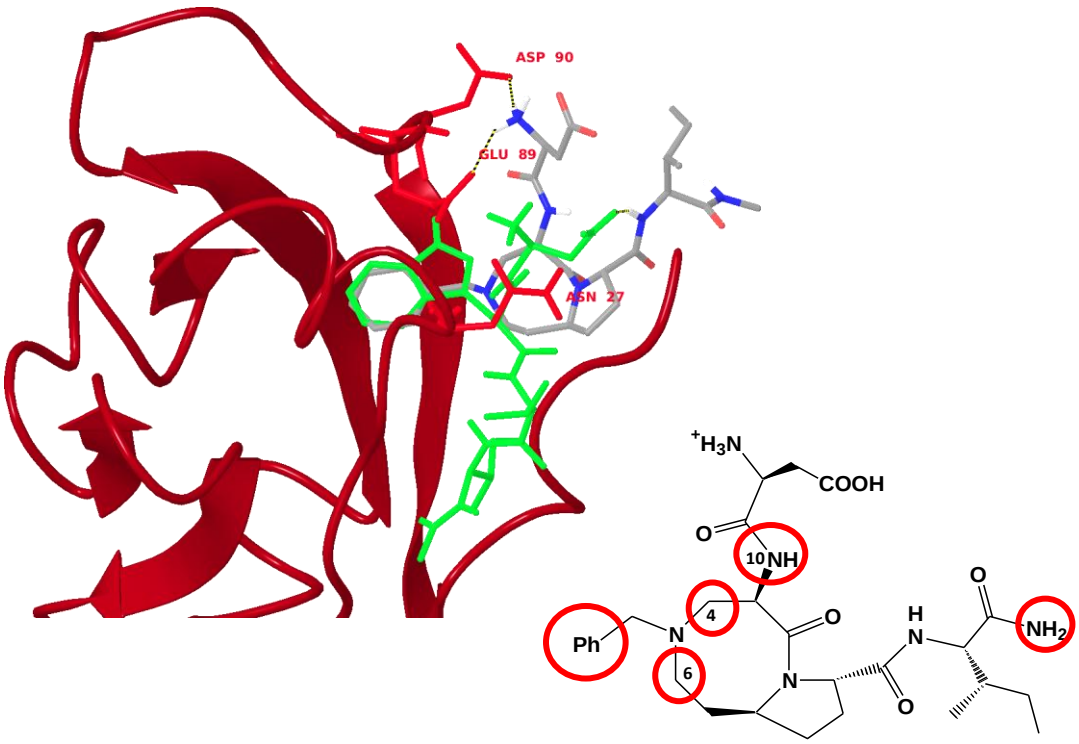


Figure 3.2.3. Left figure: best pose of compound **1c** (tube representation, C in grey, N in blue and O in red) into E-cadherin (red ribbon representation) models, overlaid to the DWVI sequence (green tube representation). Key receptor residues are labeled and highlighted in tube representation. Below is reported a 2D representation of compound **1c** with the protons involved in the NMR epitope highlighted by circles.

In all the docking poses into E-cadherin, compound **1c** places the benzyl ring into the corresponding pocket favouring also the interaction of protons CH_2_4 and CH_2_6 of the scaffold with the protein residues. These contacts are in agreement with the epitope found by NMR at 283 K for aromatic and CH4/CH6 protons. In the docking poses, for NH10 proton an average distance to Hβ protons of Asp1 side chain of 3.507 +/- 0.519 Å was calculated, while for C-terminal NH_2 protons an hydrogen bond with carboxylic group of Asp1 side chain is observed only in the best pose. In general, it seems that the best pose well represent the STD signal at 283 K. Moreover, docking results suggest a possible interaction of ligand protons NH10 and C-terminal NH_2 with the Asp1 residue, that belongs to the adhesive arm of the receptor and lies close to the binding site. In fact, in the hypothesis that the opening of the adhesive arm leaves the pocket free for the binding of our ligand, the receptor DWVI sequence should be free to move in solution and could interact

with the compounds itself. Moreover, based on the analysis of several X-ray structures, it seems that the adhesive arm does not adopt a specific secondary structure motif, *i.e.* β-sheet, helix etc., either in the dimer complex or in the monomer state. In this case, our rigid-receptor docking model introduce an important bias factor into analysis. To overcome this limit, in the next future we are planning to perform MD simulations where both ligand and protein will be free to move and their dynamical interaction should be more close to the experimental NMR conditions.

In order to investigate the effect of protein on ligand binding, we performed NMR experiments using the human P-cadherin EC1-EC2 domains, a type I cadherin strongly related to E-cahderin. In this case, the ligand binding epitope could be affected by two main factors: the binding site (slightly different form E-cadherin) and, as previously discussed, the increased flexibility of the adhesive arm (lying close to the binding site) due to the mutation of Pro5 (E-cadherin) into Ala5 (P-cadherin).

3.2.1.3 Interaction with P-cadherin EC1-EC2 construct (NMR experiments)

tr-NOESY experiments, performed in presence of P-cadherin EC1-EC2 construct at the three temperatures, did not show any variation with respect to what observed with the E-cadherin EC1-EC2. In addition, in the *tr*-NOESY experiments at 298 K, the cross-peak intensity is close to zero. This indicates a binding event. At 283 K the contacts are negative like in the free state.

STD-NMR experiments, performed at 283, 290 and 298 K, showed three different sets of signals (**Figure 3.2.4.**). A similar behaviour was observed in the interaction with the E-cadherin EC1-EC2 construct at the same three temperatures.

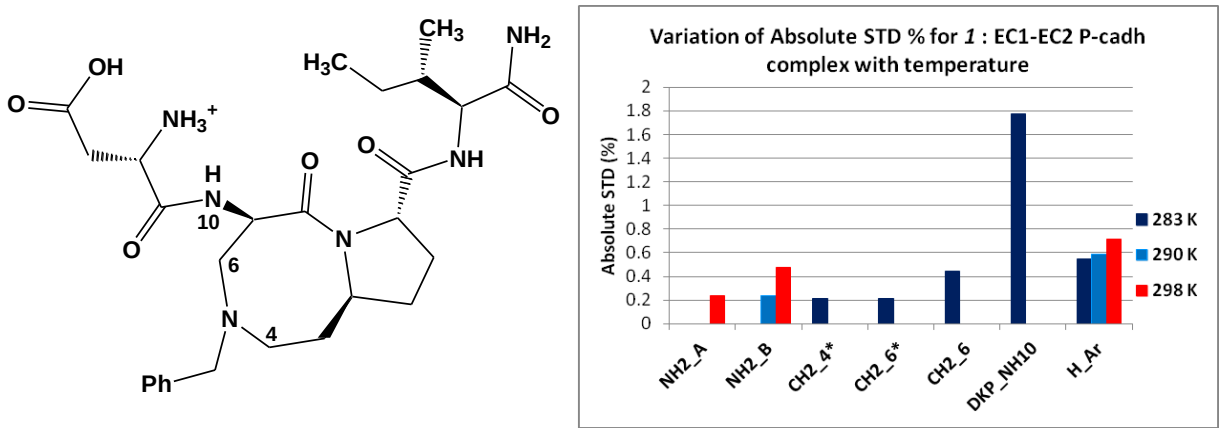


Figure 3.2.4. Low field region of the STD conducted at three different temperatures (298 K, 290 K and 283 K) with the graph of the STD absolute percentages at these three temperatures. The asterisk indicates an overlapping. CH2_6* and CH2_6 indicate the two protons insisting at C-6 of the scaffold, characterized by two different chemical shifts in the spectrum.

The aromatic protons (~0.6% STD absolute intensity) are present at the three temperatures. The protons CH_2_4 and CH_2_6*/CH2_6 of the scaffold appeared only at 283 K (0.2 and 0.4% STD absolute intensities). The proton NH10 is only present at 283 K (1.8% STD absolute intensity) as observed with the E-cadherin. The C-terminal NH_2 protons (NH2_A and NH2_B) signals appear at 290 K and 298 K, while in E-cadherin spectra were present with low STD intensity percentage also at 283 K. Since the X-ray structure of mouse P-cadherin was published only in the last month of my PhD, docking studies with this target were not performed yet. However, it seems that the differences observed in NMR epitope at different temperatures are independent of the cadherin type. In fact, the X-ray structure of mouse P-cadherin describes a binding mode of the EWVM recognition sequence very similar to the interaction pattern of DWVI sequence in E-cadherin receptor and, as consequence, also our ligand could similarly interact with the two cadherin receptor. Further NMR experiments and *in silico* investigations are needed to understand the molecular basis of this highly dynamical protein - ligand interaction.

3.2.2 COMPOUND 2c

3.2.2.1 NMR Characterization and conformational analysis

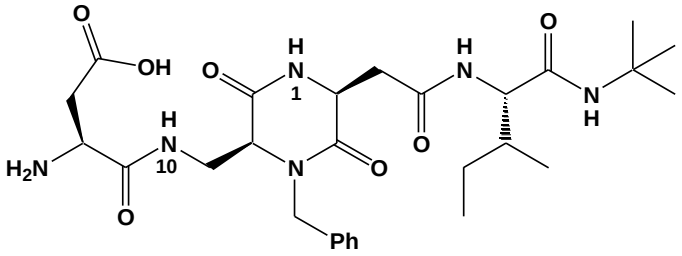


Figure 3.2.5. Structure of compound **2c**.

The complete characterization of the molecule (^{1}H and ^{13}C chemical shifts) both at 298 K and 283 K is reported in the ***Supplementary information***.

In order to investigate the conformational preference of compound **2c** in the free state we conducted variable temperature experiments. Monodimensional ^{1}H-NMR experiments were conducted to detect intramolecular hydrogen bonds, by measuring the chemical shift of the N–H protons and their temperature coefficients ($\Delta\delta/\Delta T$); in the **Table 3.2.1.**, the variation of the chemical shift on temperature ($\Delta\delta/\Delta T$) for amide protons is reported.

Residues	$\Delta\delta/\Delta T$ ($*10^{-3}$ ppb/K)	δ (ppm)
NH10	-4.56	8.50
NH1	-8.16	8.41
NH_isoLeu	-7.52	8.10
NH_tBu	-7.68	7.44

Table 3.2.1. $\Delta\delta/\Delta T$ and δ of the amide protons.

These data suggest that the compound is characterized by a dynamical conformational equilibrium (high absolute values of $\Delta\delta/\Delta T$). Only amide proton NH10 might be involved in a intramolecular hydrogen bond. These results and the NOE contacts at 298 K suggest an equilibrium between disordered conformations without specific intramolecular hydrogen bonded and a conformation with NH10 involved into an hydrogen bond (**Figure 3.2.6.a**) characterized by a

folded geometry with a 10-membered β-turn centered at the DKP scaffold. In fact, the long range NOE contacts observed at 298 K (**Figure 3.2.6.**) indicate the presence of a folded conformation (cross peaks between αH_Asp and the NH_tBu and CH_3_tBu). The NOESY performed at 283 K shows a very similar pattern of cross peaks with in addition the NOE between the aromatic protons and βH_Asp. The magnetic properties (τ_C) of ligand **2c** lead to a positive NOE at 298 K, while there is evidence of a negative Overhauser effect at 283 K.

However, NMR data (especially NOE contacts) suggest the formation of another folded geometry characterized by a 12-membered β-turn stabilized by an H-bond between N-H of the isoleucine and C=O of the aspartic residue (**Figure 3.2.6.b**).

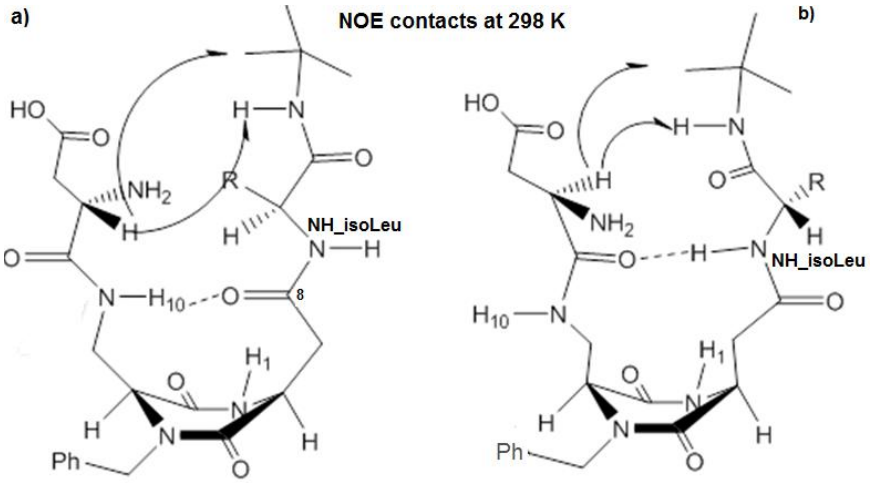


Figure 3.2.6. Folded geometries (**a** and **b**) contributing at the conformational equilibrium of compound **2c** at 298 K. NOE contacts are represented as black arrows.

A *cis* scaffold, however, is known in literature to induce a closure of the two side-chains.[37b,42] For this compound, we ran 10 ns of restrained Metropolis Monte Carlo/Stochastic Dynamics (MC/SD) simulations,[43] using the implicit water GB/SA solvation model[44] and the OPLS_2005 force field.[45] According to the NOE contact, we use a restraint on the distance between the amide protons of Leu (NH_isoLeu) and Asp Hα (2.5± 0.5 Å). Approximately 10% of the conformations sampled during the simulation is characterized by the 10-membered β-turn centered at DKP and stabilized by the hydrogen bond between the NH10 and C(8)=O and another

10% is characterized by a 12-membered β-turn stabilized by the hydrogen bond between NH_isoLeu and C=O_Asp. The unrestrained MC/SD simulation produced disordered structures without specific intramolecular hydrogen bonds.

3.2.2.2 Interaction with E-cadherin EC1-EC2 construct

NMR experiment

The interaction of compound **2c** with the E-cadherin EC1-EC2 construct was studied by *tr*-NOESY and STD[5] at 283 K, 290 K and 298 K. The *tr*-NOESY experiment of **2c** at 298 K showed positive cross-peaks as in the free state, which indicates the absence of binding. In addition, at 283 K the *tr*-NOESY data do not give information about the binding event.

As observed for compound **1c**, the binding epitope of compound **2c** with E-cadherin change with temperature.

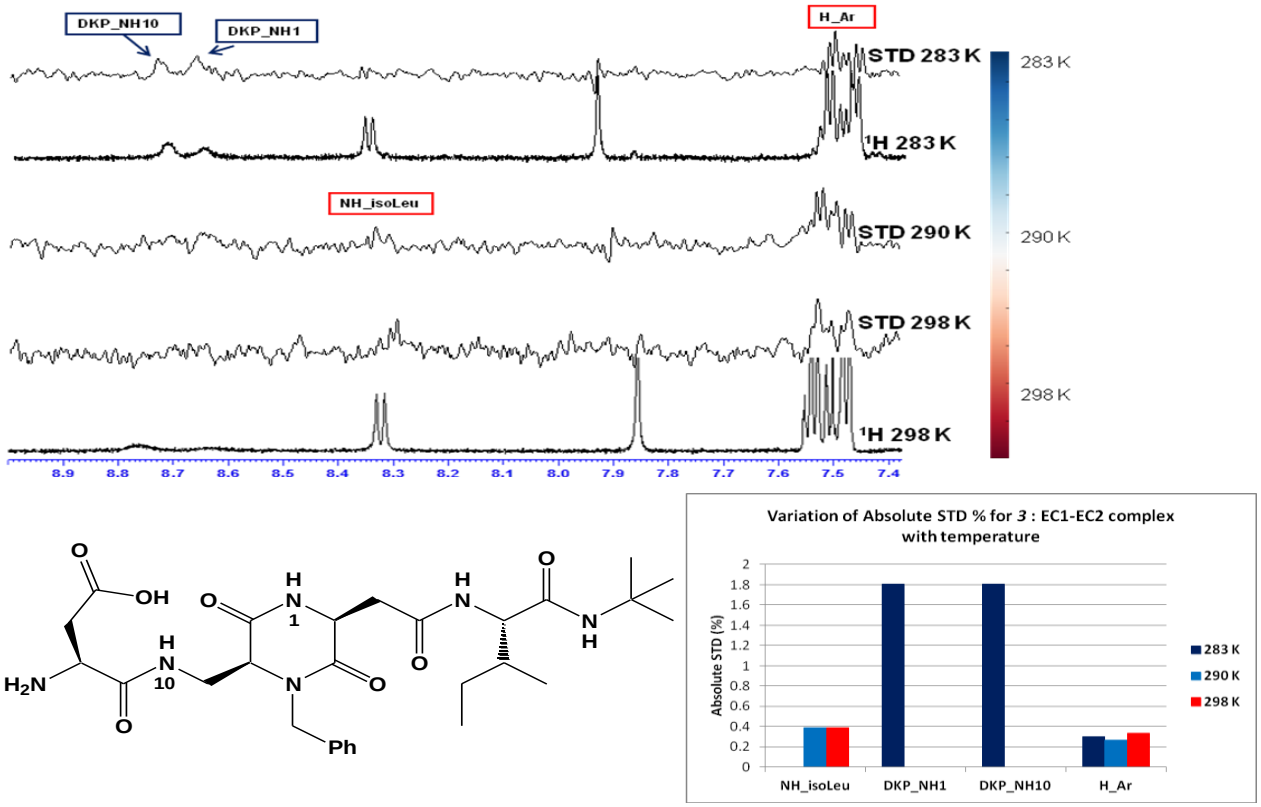


Figure 3.2.7. Low field region of the STD conducted at three different temperatures (298 K, 290 K and 283 K) with the graph of the STD absolute percentages at these three temperatures.

The signals of the aromatic protons are present in the STD spectra at the three temperatures (~0.3% STD absolute intensity) (**Figure 3.2.7.**). The NH_1 and NH_{10} disappeared increasing the temperature. However, in the graph we have not reported the STD absolute intensity that characterized the NH1 proton at 290 K; in fact, even if it was present in the STD spectrum, its intensity was not quantifiable. The proton NH_isoLeu, on the contrary, appears in the experiment at 290 and 298 K (0.4 and 0.4% STD absolute intensities, respectively).

Docking studies[2]

As already discussed, docking calculation is not the suitable tool for a comparison with experimental ligand epitope, but could provide some interesting considerations.

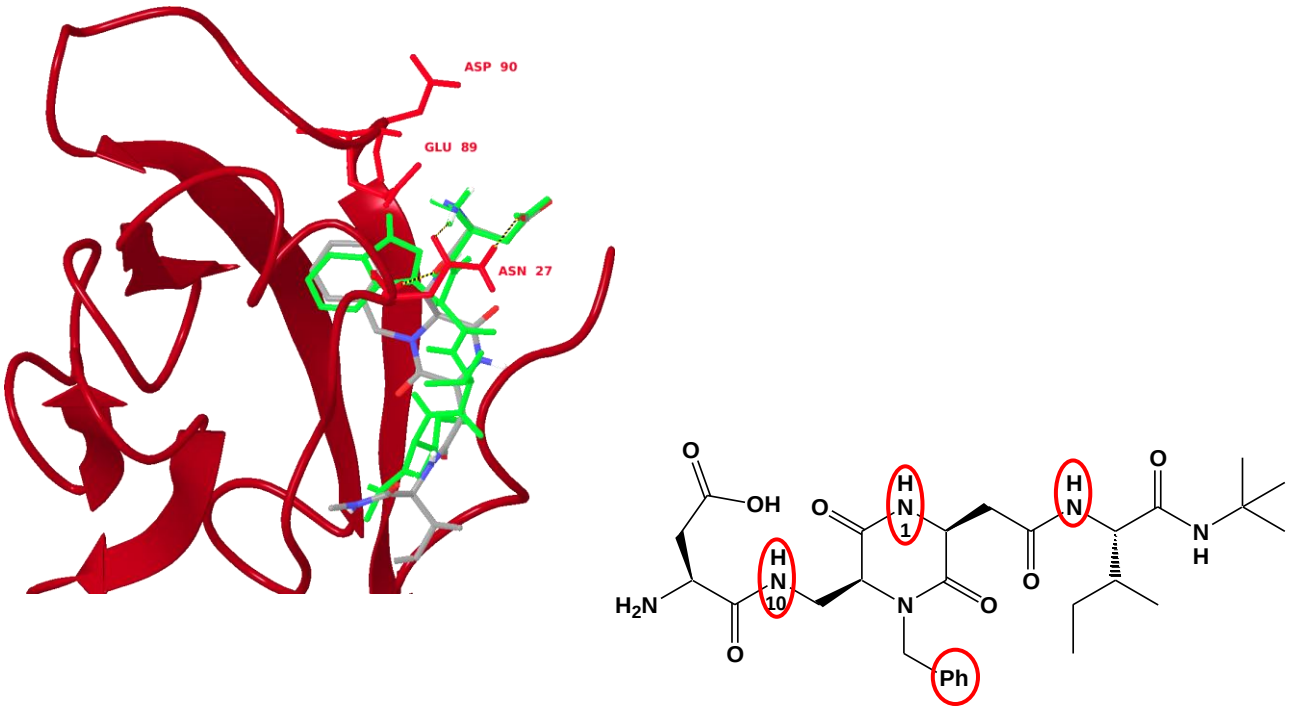


Figure 3.2.8. Left figure: best pose of compound **2c** (tube representation, C in grey, N in blue and O in red) into E-cadh (red ribbon representation) models, overlaid to the DWVI sequence (green tube representation). Key receptor residues are labeled and highlighted in tube representation. Below is reported a 2D representation of compound **2c** with the protons involved in the NMR epitope highlighted by circles.

In docking calculation the ligand is free to move and preferentially adopts an extended conformation that well overlays to the reference tetrapeptide sequence DWVI. In all the docking poses we observed a contact between NH10 and the Hγ of Glu89 side chain, while NH1 is involved in an hydrogen bond with the carbonyl backbone of Trp2 and NH_isoLeu is in contact with the Lys25 side chain aliphatic protons or moves toward the Hβ protons of Trp2 showing a greater motility.

Docking calculation starting from the two folded geometries, using specific torsional constraints to keep the intramolecular hydrogen bonds fixed, showed a different disposition of the ligand inside the binding pocket. The conformation with the 10-membered β-turn was not able to place the aromatic ring into the corresponding pocket. Interestingly, in some poses the benzyl group form stacking interaction with the indole ring of Trp2, while keeping the salt bridge with Glu89 residue, thus preserving the two key interactions for the binding to E-cadherin (**Figure 3.2.9.**, left). Moreover, in this second type of biding mode, NH_isoLeu and NH1 form an hydrogen bond with carbonyl backbone of Lys25 and Asn27 side chain, respectively, while NH1, involved in the intramolecular hydrogen bonds, is not available for further interactions with the receptors. A third type of binding mode was observed for the 12-membered β-turn geometries in the analysis of docking results. The benzyl group fit into the corresponding pocket and, in some poses, the salt bridge with Glu89 was also formed (**Figure 3.2.9.**, right). Compared to experimental epitope, in this third binding mode, the ligand NH_isoLeu intramolecularly involved in the hydrogen bond does not form interactions with the receptor, while NH1 is close to Trp2 backbone and NH10 forms an hydrogen bond with Glu89 side chain. All this three possible binding mode are compatible with STD spectra and further computational studies will be carried out to investigate the effect of binding conformation on binding.

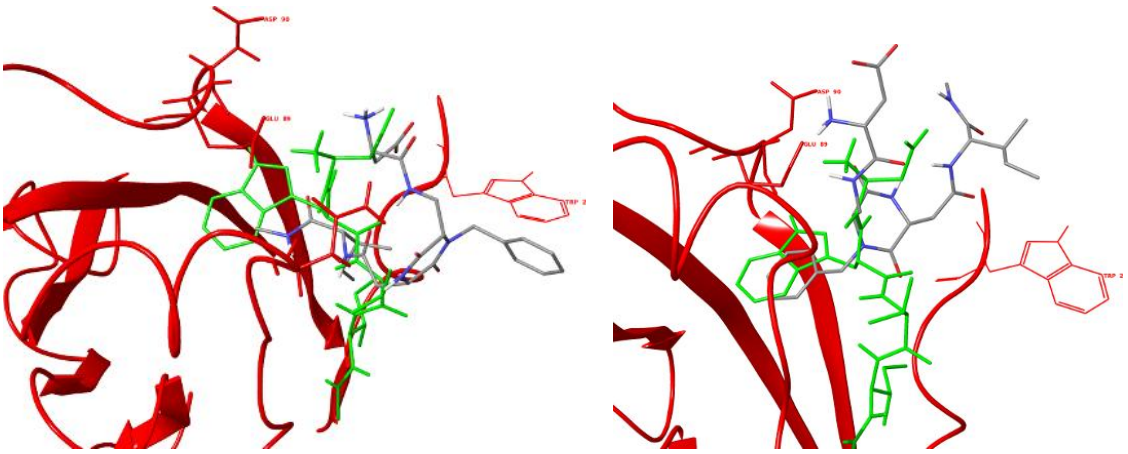


Figure 3.2.9. Docking poses of compound **2c** (tube representation, C in grey, N in blue and O in red) featuring the 10-membered β-turn geometry (left) and the 12-membered β-turn geometry (right) into E-cadh (red ribbon representation) models, overlaid to the DWVI sequence (green tube representation).

3.2.2.3 Interaction with P-cadherin EC1-EC2 construct (NMR experiments)

The comparison of the NOE contacts in the free and in the bound state do not show any significant differences, suggesting that, in presence of the protein, the conformational equilibrium remains the same. The *tr*-NOESY experiment at 298 K of compound **2c** in the presence of the P-cadherin EC1-EC2 construct suggests a binding situation, while at 282 K the NMR data do not give information about the binding event.

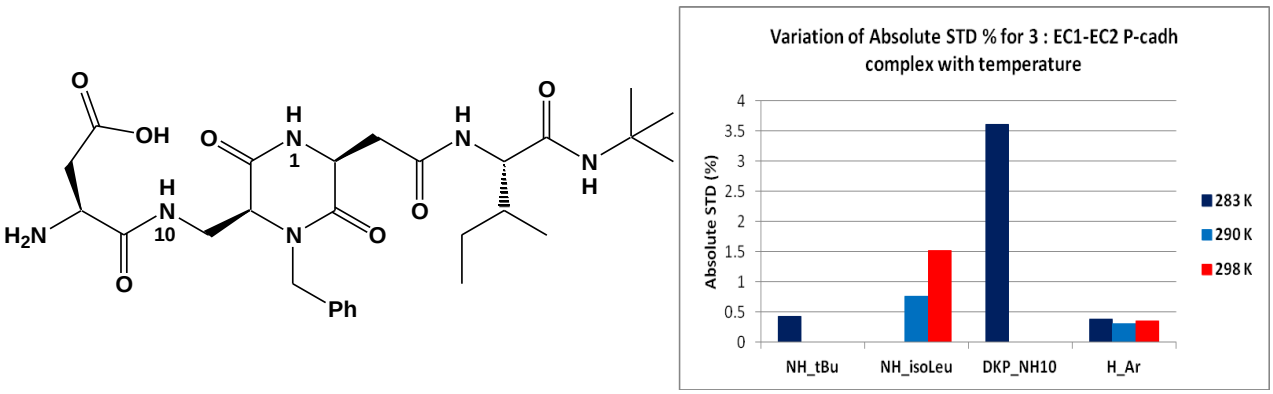


Figure 3.2.10. Low field region of the STD conducted at three different temperatures (298 K, 290 K and 283 K) with the graph of the STD absolute percentages at these three temperatures.

At 283 K the STD spectrum showed the signals of the amidic protons NH_tBu and NH10 in addition to the signals related to the aromatic protons (0.4, 3.5 and 0.3% STD absolute intensities, respectively) (**Figure 3.2.10.**). The contribution of NH_tBu proton disappeared increasing the temperature at 290 K. An opposite behaviour characterized the amidic proton NH_isoLeu, showing a 1.5% STD absolute intensity at 298 K, a decreased intensity at 290 K (0.7% STD absolute intensity) and disappeared at 283 K. The STD signals of aromatic protons are visible at all the temperatures with about the same STD intensity.

The ligand epitope varies with the temperature also for the interaction with P-cahderin, but with a slightly different pattern of interactions. Compared to E-cadherin, compound **2c** loses the signal of NH1 and adds the interaction of NH_tBu proton. This change could be due to the different binding site and further investigation using the X-ray structure of P-cadherin will be performed.

3.2.3 COMPOUND 3c

3.2.3.1 NMR Characterization and conformational analysis

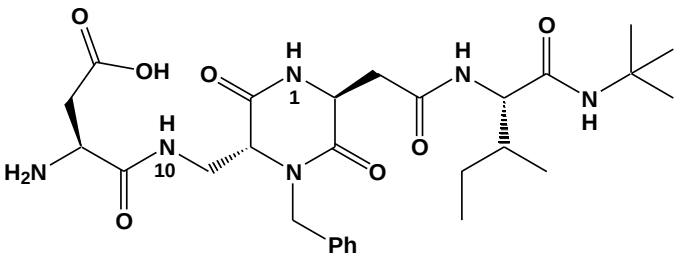


Figure 3.2.11. Structure of compound **3c**.

The complete characterization of the molecule (^{1}H and ^{13}C chemical shifts) both at 298 K and 283 K is reported in the *Supplementary information*.

In order to investigate the conformation adopted by compound **3c**, we conducted variable temperature experiments. In the **Table 3.2.2.** the data relative to the amide protons of the compound were reported. Obviously, the *trans* configuration of the diketopiperazine and the

limited length of the side chains, prevents the folding of the side chains and the formation of the conformational motifs stabilized by hydrogen bonds.

Residues	$\Delta\delta/\Delta T$ ($*10^{-3}$ ppb/K)	δ (ppm)
NH10	-6.00	8.53
NH1	-4.50	7.49
NH_isoLeu	-8.50	7.94
NH_tBu	-8.50	8.31

Table 3.2.2. $\Delta\delta/\Delta T$ and d of the amidic protons.

These data suggest that the compound is characterized by conformational mobility (high value of $\Delta\delta/\Delta T$), except for the proton NH1 that is characterized by a low value of $\Delta\delta/\Delta T$ and of the chemical shift (high field region). However, the absence of significant NOE contacts at the two temperatures (298 and 283 K) does not allow to evaluate its involvement in a intramolecular hydrogen bond that stabilized the molecule in a preferred conformation. The only NOE contacts observable are those between the aromatic protons and the methyl groups of the *ter*-buthyl group, and those between the aromatic protons and the methyl group of the isoLeu side-chain. The absence of relevant NOE contacts and the $\Delta\delta/\Delta T$ data suggest the presence of an extended conformation.

3.2.3.2 Interaction with E-cadherin EC1-EC2 construct

NMR experiments

The signal phase of the *tr*-NOESY experiments does not allow to evaluate a binding situation. In fact, in the *tr*-NOESY experiments at 298 K, the cross-peak intensity is close to zero (like in the free state) and at 283 K the contacts are negative like in the free state.

STD experiments at 298 K shows signals relative to aromatic protons and NH_isoLeu. These protons do not appear in the experiments performed at 283 K, but there are the signals related to NH1 and NH10 protons (**Figure 3.2.12.**). Unfortunately, also in the control experiment the NH1 appears, so NH1 cannot be integrated.

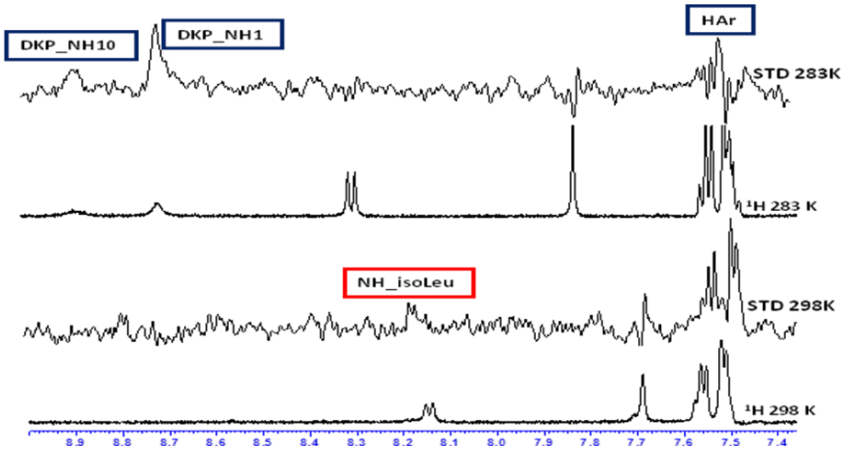


Figure 3.2.12. Low field region of the STD conducted at two different temperatures (298 and 283 K).

This precludes the quantitative analysis of the interaction of the compound with the protein, even if the STD experiments conducted at the two temperatures 298 and 283 K suggest a variation of the epitope also for this compound (**Figure 3.2.12.**, where at 298 K the STD signals of the NH_isoLeu appeared in contrast to NH10 that appeared at 283 K). It is worth noting that also the aromatic protons do not show any contribution at 283 K.

Docking studies[2]

The binding mode of compound **3c** in the E-cadherin binding site is very similar to the extended arrangement observed for compound **2c**. Both ligands adopt a similar disposition within the binding site and well reproduce the X-ray pose of the DWVI sequence.

Experiment with P-chaderin will be carried out in the next future to elucidate the behaviour of this compound with respect to the E-cadherin.

Section 3: Conclusions

Cadherins are known to play a key role in important physiological processes such as tissue morphogenesis and stability, as well as in the immune system regulation. The dysregulation of some type I cadherin have been shown to contribute to pathological processes, and specifically to tumour progression.[7,3d] In this project, we have evaluated two of the classical type I-cadherin: E- (epithelial) and P- (placental) cadherin.

Despite a growing interest in the field, the rational design of small ligands targeting cadherins protein - protein interactions is still in a very early stage. Task of the group, so, was to rationally design and synthesize small peptidomimetic molecules aimed at providing partial or total inhibition of the cadherin - cadherin interaction.

According to the results obtained from docking studies on E-cadherin and evaluated on the base of the typical interactions of the natural sequence DWVI, three ligands of general formula NH_3^+-Asp-scaffold-Ile-NHR were selected for the synthesis and subsequent biological assays (**Figure 3.3.1**).[2]

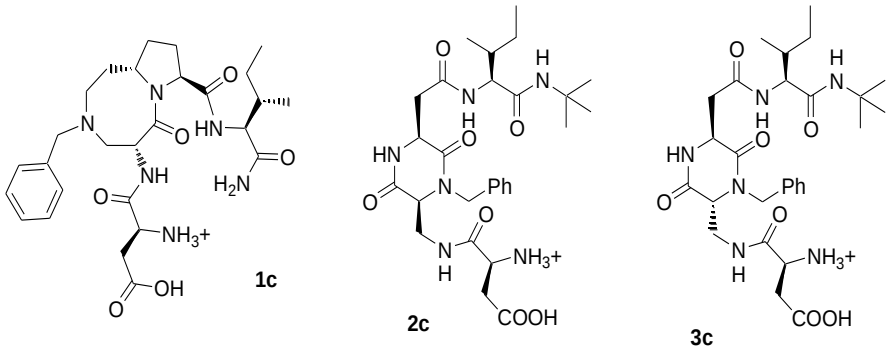


Figure 3.3.1. Peptidomimetic ligands synthesized.

A detailed investigation of the ligand conformation were conducted. The *trans* stereochemistry of the scaffold, inserted in the compounds **1c** and **3c**, prevents the folding of the side chains and the formation of conformational motifs stabilized by hydrogen bonds. In fact, these compounds adopt an extended arrangement as demonstrated by NOESY and VT-NMR experiments.

For compound **2c** (*cis*-stereochemistry), we suggest the presence of an equilibrium between open and closed conformations, being a closed geometry supported by the presence of typical NOE contacts.

In addition, I applied NMR techniques in order to detect and study the binding of compounds **1c – 3c** with the EC1–EC2 constructs of the epithelial subtype E – and placental subtype P-cadherins (expressed and purified by Dr. Parisini at IIT). Due to the difficulties to crystallize small molecule - cadherin complexes (just the dimeric forms were crystallized),[4,16] NMR has proved to be a useful method of ligand binding detection. Having evidence from biological assays that the interaction between our compounds and cadherins happened, we conducted, firstly, STD experiments at 283 K with the aim of slowing the rate of the ligand exchanging between solvent and protein environment. Then, for both these proteins, we performed STD experiments at 290 and 298 K. Surprisingly, all the three compounds change the epitope as a function of the temperature with both cadherin constructs. Moreover, the ligand binding epitope is similar between the two receptors. Several aspects could contribute to the STD binding epitope variation with temperature: *i)* the ligand binding mode change with temperature, *ii)* a conformational change of the protein binding site due to temperature or *iii)* both aspects.

To address this question further NMR and *in silico* investigations should be done. NMR studies using labeled protein constructs and molecular dynamics simulations allowing both ligand and protein free to move are in progress. It is expected that the combined use of NMR and computational approaches will significantly improve the structural understanding of the binding of peptidomimetic ligands to the extracellular domains of these cadherins.

Section 4: Experimental section

NMR EXPERIMENTS

Conformational analysis: NMR spectroscopy experiments were performed at 298 K and 283 K on Bruker Avance 400 and 600 MHz spectrometers. All proton and carbon chemical shifts were assigned unambiguously. The NMR experiments were carried out in phosphate buffer. The buffer prepared has a pH of 7.4 with 20 mM phosphate salts, 150 mM NaCl and 1 mM $CaCl_2$, with an addition of 10 % D_2O for NMR experiments.

Two-dimensional experiments (TOCSY, NOESY and HSQC) were carried out on samples at a concentration range of 3-6 mM. NOESY experiments were performed at 0.7 s. The water resonance was saturated with the excitation sculpting sequence from the Bruker library. The conformations of the linear pentapeptides were analyzed with respect to hydrogen bonding of amide protons (VT-NMR spectroscopy) and NOE contacts.

NMR spectroscopy studies with the isolated proteins: The spectra were acquired with a Bruker Avance 600 MHz instrument at 298, 290 K and 283 K. Each sample was prepared in a 3 mm NMR tube with ligand (0.5-0.6 mg) and the isolated proteins (~36 µL, ~5 mg/mL) in phosphate buffer (~200 µl).

Amount of protein used: All the experiments were conducted with the first two domain of E- and P-cadherin (Molecular Weight (EC1-EC2 construct E-cadherin)=23118.79 Da, Molecular Weight (EC1-EC2 construct P-cadherin)= 23600.00 Da) and with a ligand : protein ratio of 80 : 1. To calculate protein moles (x):

$80 : 1 = 6*10^{-7} : x$

$x = 7.5*10^{-9}$ mol

To calculate grams of protein:

$g = 7.5*10^{-9} * 23118.79 = 0.000173$ g => 0.173mg

To calculate mL of protein solution to add:

0.173 mg : 5 mg = x : 1 mL

x = (0.1733 mg*1 mL)/5 mg = 34 μL

Amount of Ca^{2+} ion: It is known that a protein solution 0.22 mM needs 1 mM of Ca^{2+} (~5-fold more), in our tube we need:

mol protein = 0.22 mM * 0.036 mL (amount of protein in the tube) = 0.00000792 mmol

M = 0.00000792 mmol / 0.2 mL (amount of buffer solution) = 0.0396 mM (concentration of protein) that corresponds to 0.198 mM Ca^{2+} (~ 5-fold more).

The interaction between our ligands and isolated membrane proteins was detected using *tr*-NOESY and STD experiments as described in the "***Experimental Section***" of the ***Chapter 2***.

COMPUTATIONAL STUDIES

Docking studies.

Models set up and validation.

The automated docking calculations were performed using Glide (Grid-based Ligand Docking with Energetics) version 5.7109.[38] The grid have been generated for E-cadh structure prepared as described in the *Protein preparation* section and selecting as receptor the EC1 domain (1-103 residues) of one monomer and as ligand the DWVIPP esapeptide sequence of the partner monomer. The center of the grid enclosing box was defined by the center of the DWVIPP sequence. The enclosing box dimensions, which are automatically deduced from the ligand size, fit the entire active site. Docking calculations were performed using the standard precision mode (SP). The receptor was considered as a rigid body while the ligand sampling was set to 'Flexible' with the option 'Penalize non planar conformation' for amides. No Epik state penalities were used in the docking score calculations. The size of the bounding box for placing the ligand center was set to 14 Å. No further modifications were applied to the default settings. The GlideScore function was used to select 10 poses for each ligand. The Glide program was initially tested for its ability to reproduce the crystallized binding mode of fragments of the N-terminal native sequence (i.e. from the tripeptide DWV up to the decapeptide). The program was successful in reproducing the experimentally determined binding mode of these peptides, as it corresponds to the best-scored pose.

BIOLOGICAL ASSAYS

The N-cadh-expressing SKOV3 cells[46] were obtained from ATCC (Rockville, MD) and maintained in RPMI 1640 medium (Sigma, Sant Louis, Missouri) supplemented with 10% fetal calf serum (FCS) (Hyclone, Logan, UT), 1% L-glutamine at 37°C in a humidified atmosphere of 5% CO_2 in air. The E-cadh-expressing OAW42 cells[47] were provided by Dr. A. Ullrich, (Max Planck Institute of Biochemistry, Martinsried, Germany) and maintained in DMEM supplemented with 10% FCS and 1% L-glutamine in a humidified atmosphere of 5% CO_2.

Prior adhesion assay or ELISA, confluent cells were detached using EDTA to protect cadherin from proteolysis and washed twice with PBS without calcium to prevent cell-cell adhesion. For adhesion assay cells were suspended in cell culture medium supplemented with 2% FCS containing varying concentrations of each of the peptidomimetic ligand or the solvent and left to form the monolayer in a humidified atmosphere of 5 % CO_2. For ELISA cells were suspended in PBS containing 1.2 mM $CaCl_2$ and varying concentrations (2 and 1 mM) of each of the peptidomimetic ligand or the solvent as above. After 1 h incubation at 37°C cells were added to cadherin-coated plates and incubated for further 2h at 37°C. Ninty-six well plates were coated overnight with 30 ng/well of E-cadh-Fc chimeric protein or human IgG1 Fc fragment (R&D Systems, Minneapolis, MN). After incubation, the wells were washed three times to remove unbound cells. The adherent cells were detected by crystal violet substrate. The intensity of the color was measured at 450 nM. Binding to human Fc fragment of IGg1 was evaluated to exclude unspecific cell binding and subtracted from those obtained with cadherin-Fc proteins. ADH-1 and compound **1** were dissolved in water, compounds **2** and **3** in 10% DMSO.

SPR analysis

SPR experiments were performed using Biacore 2000 (GE Healthcare). Standard EDC\NHS coupling was used to covalently immobilized recombinant N-cadh-Fc on CM5 (GE Healthcare) sensor chip. Briefly, CM5 chip was activated with ECD\NHS for 7 min with excess activated carboxyl groups blocked with ethanolamine for 7 min following immobilization of N-cadh-Fc, resuspended as suggested by the manufacturer, was diluted to 10 µg\ml in 10 mM sodium

acetate, pH 4.8. For all experiments with covalently immobilized N-cadh-Fc, one flow cell served as a reference surface following activation and blocking on each chip in the absence of N-cadh-Fc. One flow cell was immobilized with an uncorrelated recombinant protein (Axl-Fc, R&D) with a similar molecular weight (120 KDa). The sensor chip was equilibrated with PBS in absence of calcium ions, and N-cadh-Fc was dissociated from the chip using the same buffer and the response was recorded for at least 240 s. N-cadh-Fc (10 nM), freshly diluted in PBS plus 1 mM $CaCl_2$, was applied to the surface of the N-cadh immobilized CM5 sensor chip for 1800 s. Since the compounds to be analyzed (**2** and **3**) are only 600 Dalton and their direct binding to N-cadh is not detectable by the Biacore 2000 instrumentation, an inhibition experiment was performed. For compound **2** and **3**, 10 nM of N-cadh-Fc were pre-incubated for 30 min at 25^{o}C in the presence of each inhibitor at a concentration of 10 µM in PBS plus 1 mM $CaCl_2$ and then applied on the chip. N-cadh-Fc alone was injected after each injection of N-cadh/inhibitor to verify the functionality of the chip and the resulted RU (Resonance Units) max was considered as 100%. After each injection, the sensor chip was regenerated using 30-60 µl of 20 mM EGTA pH 7 to eliminate calcium ions from the system. A flow rate of 5µl\min was kept for all the experiments.

∗Bibliography∗

[1] Arkin, M. R.; Wells, J. *Nat. Rev. Drug Discov.* **2004**, *3*, 301–317.

[2] a) Doro, F.; Colombo, C.; Alberti, C.; Arosio, D.; Belvisi, L.; Casagrande, C.; Fanelli, R.; Manzoni, L.; Parisini, E.; Piarulli, U.; Luison, E.; Figini, M.; Tomassetti, A.; Civera, M. "Computational design of novel peptidomimetic inhibitors of cadherin homophilic interactions" *European Journal of Medicinal Chemistry* (submitted), b) Doro, F. PhD thesis, Computational modelling of biomolecular systems: applications to the study of molecular recognition processes, University of Milan, **2013**.

[3] a) Berx, G.; van Roy, F. *Cold Spring Harb. Perspect. Biol.* **2009**, *1*, a003129; b) Leckband, D.; Sivasankar, S. *Curr. Opin. Cell Biol.* **2012**, *24*, 620–627; c) Brasch, J.; Harrison, O. J.; Honig, B.; Shapiro, L. *Trends Cell Biol.* **2012**, *22*, 299–310; d) Blaschuk, O. W.; Devemy, E. *Eur. J. Pharmacol.* **2009**, *625*, 195–198; e) Blaschuk, O. W. *Cell Tissue Res.* **2012**, *348*, 309–313.

[4] Harrison, O. J.; Jin, X.; Hong, S.; Bahna, F.; Ahlsen, G.; Brasch, J.; Wu, Y.; Vendome, J.; Felsovalyi, K.; Hampton, C. M.; Troyanovsky, R. B.; Ben-Shaul, A.; Frank, J.; Troyanovsky, S. M.; Shapiro, L.; Honig, B.; Sergey, M. *Structure* **2011**, *19*, 244–256.

[5] Meyer, B.; Peters, T. *Angew. Chem. Int. Ed.* **2003**, 42, 864 – 890.

[6] Potenza, D.; Vasile, F.; Seminars in Organic Synthesis, "A. Corbella" Summer School, 37th, Gargnano, Italy, June 18 – 22, 2012, 225 – 248.

[7] Brouxhon, S. M.; Kyrkanides, S.; Teng, X.; Athar, M.; Ghazizadeh, S.; Simon, M.; O'Banion, M. K.; Ma, L. *Oncogene* **2014**, 33, 225 – 235.

[8] Adams, C. L.; Chen, Y. T.; Smith, S. J.; Nelson, W. J. *J. Cell Biol.* **1998**, *142*, 1105–1019.

[9] Garrod, D. R.; Merritt, A. J.; Nie, Z. *Curr. Opin. Cell Biol.* **2002**, *14*, 537–545; h

[10] Takeichi, M. *Science* **1991**, *251*, 1451–1455.

[11] Gumbiner, B. M. *J. Cell Biol.* **2000**, *148*, 399–404.

[12] Vendome, J.; Felsovalyi, K.; Song, H.; Yang, Z. et al. *PNAS* **2014**, 111(40), E4175-E4184, doi:10.1073/pnas.1416737111.

[13] Chen, C. P.; Posy, S.; Ben-Shaul, A.; Shapiro, L.; Honig, B. H. *Proc. Natl. Acad. Sci. USA* **2005**, 102, 8531 – 8536.

[14] Sivasankar, S.; Zhang, Y.; Nelson, W. J.; Chu, S. *Structure* **2009**, 17, 1075 – 1081.

[15] Miloushev, V. Z.; Bahna, F.; Ciatto, C.; Ahlsen, G.; Honig, B.; Shapiro, L.; Palmer, A. G. *Structure* **2008**, 16, 1195 – 1205.

[16] Parisini, E.; Higgins, J. M. G.; Liu, J.-H.; Brenner, M. B.; Wang, J.-H. *J. Mol. Biol.* **2007**, 373, 401 – 411.

[17] Vendome, J. Nat. Struct. Mol. Biol. 2011, 18(6), 693 – 700.

[18] Cavallaro, U.; Liebner, S.; Dejana, E. *Exp. Cell Res.* **2006**, *312*, 659–667.

[19] Jeanes, A.; Gottardi, C. J.; Yap, A. S. *Oncogene* **2008**, *27*, 6920–6929.

[20] De Santis, G.; Miotti, S.; Mazzi, M.; Canevari, S.; Tomassetti, A. *Oncogene* **2009**, *28*, 1206–1217.

[21] a) Nose, A.; Takeichi, M. *J. Cell Biol.* **1986**, 103, 2649 – 2658; b) Paredes, J.; Correia, A. L.; Ribeiro, A. S.; Albergaria, A.; Milanezi, F. and Schmitt, F. C. *Breast Cancer Research* **2007**, 9(5), 214 – 226.

[22] Paredes, J.; Albergaria, A.; Oliveira J. T.; Jeronimo, C.; Milanezi, F.; Schmitt, F. C. *Clin. Cancer Res.* **2005**, 11, 5869 – 5877.

[23] Paredes, J. et al. *Biochimica et Biophysica Acta* **2012**, 1826, 297 – 311.

[24] Häussinger, D.; Ahrens, T.; Sass, H.-J., Pertz, O.; Engel, J.; Grzesiek, S. *J. Mol. Biol.* **2012**, 324, 823 – 839.

[25] Boggon, T. J.; Murray, J.; Chappuis-Flament, S.; Wong, E.; Gumbiner, B. M.; Shapiro, L. *Science* **2002**, *296*, 1308–1313.

[26] Pokutta, S.; Herrenknecht, K.; Kemler, R.; Engel, J. *Eur. J. Biochem.* **1994**, *223*, 1019–1026.

[27] Shapiro, L.; Fannon, A. M.; Kwong, P. D.; Thompson, A.; Lehmann, M. S.; Grübel, G.; Legrand, J. F. et al. *Nature* **1995**, 374(6520), 327 – 337.

[28] Williams, G.; Williams, E.-J.; Doherty, P. *J. Biol. Chem.* **2002**, 277, 4361 – 4367.

[29] Li, H.; Price, D. K.; Figg, W. D. *Anticancer Drugs* **2007**, 18, 563 – 568.

[30] Shintani, Y.; Fukumoto, Y.; Chaika, N.; Grandgenett, P. M.; Hollingsworth, M. A.; Wheelock, M. J.; Johnson, K. R. *Int. J. Cancer* **2008**, 122, 71 – 77.

[31] Augustine, C. K.; Yoshimoto, Y.; Gupta, M.; Zipfel, P. A.; Selim, M. A.; Febbo, P.; Pendergast, A. M.; Peters, W. P.; Tyler, D. S. *Cancer Res.* **2008**, 68, 3777 – 3784.

[32] Perotti, A.; Sessa, C.; Mancuso, A.; Noberasco, C.; Cresta, S.; Locatelli, A.; Carangiu, M. L.; Passera, K.; Braghetti, A.; Scaramuzza, D.; Zanaboni, F.; Fasolo, A.; Capri, G.; Miani, M.; Peters, W. P.; Gianni, L. *Ann. Oncol.* **2009**, 20, 741 – 745.

[33] Yarom, N.; Stewart, D.; Malik, R.; Wells, J.; Avruch, L.; Jonker, D. J. *Curr. Clin. Pharmacol.* **2013**, 8, 81 – 88.

[34] Gour, B. J.; Blaschuk, O. W.; ali, A.; Ni, F.; Clen, Z.; Michaud, S. D.; Shoameng, W.; Hu, Z. *Peptidomimetic modulators of cell adhesion US7446120 B2*, **2008**.

[35] Burden-Gulley, S. M.; Gates, T. J.; Craig, S. E. L.; Lou, S. F.; Oblander, S.; Howell, S.; Gupta, M.; Brady-Kalnay, S. M. *Peptides* **2009**, 30, 2380 – 2387.

[36] Devemy, E.; Blaschuk, O. W. *Peptides* **2009**, 30, 1539 – 1547.

[37] a) Manzoni, L.; Arosio, D.; Belvisi, L.; Bracci, A.; Colombo, M.; Invernizzi, D.; Scolastico, C. *J. Org. Chem.* **2005**, 70, 4124 – 4132; b) Ressurreição, A. S. M.; Bordessa, A.; Civera, M.; Belvisi, L.; Gennari, C.; Piarulli, U. *J. Org. Chem.* **2008**, 73, 652 – 660; c) Arosio, D.; Belvisi, L.; Colombo, L.; Colombo, M.; Invernizzi, D.; Manzoni, L.; Potenza, D.; Serra, M.; Castorina, M.; Pisano, C.; Scolastico, C. *ChemMedChem* **2008**, 3, 1589 – 1603; d) Manzoni, L.; Belvisi, L.; Arosio, D.; Civera, M.; Pilkington-Miksa, M.; Potenza, D.; Caprini, A.; Araldi, E. M. V.; Monferini, E.; Mancino, M.; Podestà, F.; Scolastico, C. *ChemMedChem* **2009**, 4, 615 – 632; e) Marchini, M.; Mingozzi, M.; Colombo, R.; Guzzetti, I.; Belvisi, L.; Vasile, F.; Potenza, D.; Piarulli, U.; Arosio, D.; Gennari, C. *Chem. Eur. J.* **2012**, 18, 6195 – 6207.

[38] Glide, version 5.7, Schrödinger, LLC, New York, **2011**.

[39] Thiery, J. P.; Engl, W.; Viopnoff, V.; Dafour, S. *Curr. Opin. Cell Biol.* **2012**, 24, 614 – 619.

[40] Vendome, J.; Posy, S.; Jin, X.; Bahna, F.; Ahlsen, G.; Shapiro, L.; Honig, B. *Nat. Struct. Mol. Biol.* **2011**, 18, 693 – 700.

[41] Li, Y.; Altorelli, N. L.; Bahna, F.; Honig, B.; Shapiro, L.; Palmer 3rd, A. G. *Proc. Natl. Acad. Sci. USA* **2013**, 110, 16462 – 16467.

[42] Delatouche, R.; Durini, M.; Civera, M.; Belvisi, L.; Gennari, C.; Piarulli, U. *J. Org. Chem.* **2008**, 73, 652 – 660.

[43] Guarnieri, F.; Still, W. C. *J. Comput. Chem.* **1994**, 15, 1302 – 1310.

[44] Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, 112, 6127 – 6129.

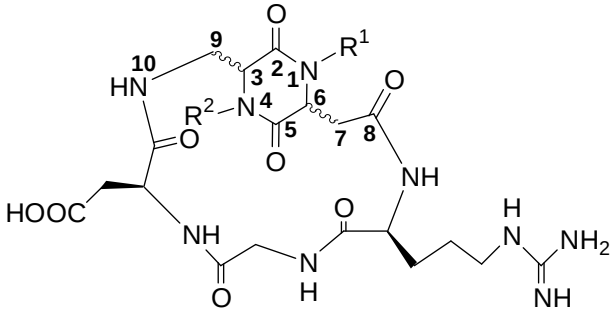
[45] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, 79, 926 – 935.

[46] Tomassetti, A.; De Santis, G.; Castellano, G.; Miotti, S.; Mazzi, M.; Tomasoni, D.; Van Roy, F.; Carcangiu, M. L.; Canevari, S. *Neoplasia* **2008**, 10 1481 – 1492.

[47] De Santis, G.; Miotti, S.; Mazzi, M.; Canevari, S.; Tomassetti, A. *Oncogene* **2009**, 28, 1206 – 1217.

∗Supplementary information for Chapter 2∗

S2.1 CHARACTERIZATION OF DKP-RGD COMPOUNDS IN THE FREE STATE AT 298 K



(1) cyclo(DKP-1-RGD) = 3S, 6S, R^1 = H, R^2 = CH_2Ph
(2) cyclo(DKP-2-RGD) = 3R, 6S, R^1 = H, R^2 = CH_2Ph
(3) cyclo(DKP-3-RGD) = 3S, 6R, R^1 = H, R^2 = CH_2Ph
(4) cyclo(DKP-4-RGD) = 3R, 6S, R^1 = CH_2Ph, R^2 = H
(5) cyclo(DKP-5-RGD) = 3R, 6S, R^1 = CH_2Ph, R^2 = CH_2Ph
(6) cyclo(DKP-6-RGD) = 3S, 6R, R^1 = CH_2Ph, R^2 = H
(7) cyclo(DKP-7-RGD) = 3S, 6R, R^1 = CH_2Ph, R^2 = CH_2Ph

Figure 2.1S Cyclic RGD-peptidomimetics containing bifunctional DKP scaffolds.

Compound **1**	^{1}H(δ,ppm) 298 K	^{13}C	n.O.e. contacts
Arg-NH	8.40, d, 3.24Hz	///	DKP_H7
Arg-Hα	4.02, sb	55.1	
Arg-Hβ	1.80, m	27.1	
Arg-Hγ	1.63, m	24.3	
Arg-Hδ	3.15, m	40.6	
Guan-NH	7.12, t, 6.01Hz	///	
Gly-NH	8.75, t, 5.95Hz	///	**NH_Asp**, γH_Arg
Gly-Hα	4.01, dd – 3.70, dd, 17.12 – 3.47Hz	43.4	NH_Asp
Asp-NH	8.10, d, 8.73Hz	///	**DKP_NH10**, **NH_Gly**, αH_Gly
Asp-Hα	4.45, sb	51.2	NH_Asp, βH_Asp
Asp-Hβ	2.77, dd – 2.57, dd, 17.35 – 8.30Hz	36.5	
DKP_NH10	7.50, d, 5.09Hz	///	**NH_Asp**, βH_Asp
DKP_NH1	8.32, s	///	
DKP-H9	3.92, ddd, 15.27/6.73/3.22Hz – 3.57, ddd, 15.27 – 4.10Hz (J_2=J_3)	39.4	DKP_NH10
DKP-H3	4.37, sb	58.9	
DKP-H7	3.19 -2.85, dd overlapped	36.8	NH_Arg
DKP-H6	4.48, sb	51.5	
CH_2-Bn	4.67, deleted by water	48.9	
H-Ar	7.30	128	

Table 2.1S ^{1}H and ^{13}C chemical shifts (δ in ppm) of compound **1** in H_2O/D_2O solution at T=298 K.

Compound **2**	^{1}H(δ,ppm) 298 K	^{13}C	n.O.e. contacts
Arg-NH	8.57, d, 7.20 Hz	///	**NH_Gly**, DKP_H7
Arg-Hα	4.21, m	-	
Arg-Hβ	1.70-1.63, m	-	
Arg-Hγ	1.61, m	-	
Arg-Hδ	3.15, m	40.5	
Guan-NH	7.12, t, 5.48 Hz	///	
Gly-NH	8.18, d, 9.6 Hz	///	**NH_Arg**
Gly-Hα	4.17-3.54, d, 18 Hz	41.7	NH_Asp
Asp-NH	8.29, d, 7.6 Hz	///	αH_Gly
Asp-Hα	4.68, influenced by water	52	
Asp-Hβ	2.82 – 2.75, dd, 17.0 – 7.0 Hz	35.3	
DKP_NH10	8.79, s br	///	αH_Asp
DKP_NH1	8.35, s	///	
DKP-H9	4.31 – 3.34, d, 16.9 Hz	38.7	
DKP-H3	4.30, m	55	
DKP-H7	2.88 – 2.39, m, 14.8 Hz	40	NH_Arg
DKP-H6	4.55, m	52	
CH_2-Bn	5.01 – 4.13, d, 16.4 Hz	47.9	
H-Ar	7.36 – 7.29, m	128.1	

Table 2.2S ^{1}H and ^{13}C chemical shifts (δ in ppm) of compound **2** in H_2O/D_2O solution at T=298 K.

Compound **3**	^{1}H(δ,ppm) 298 K	^{13}C	n.O.e. contacts
Arg-NH	8.8, d, 8.29 Hz	///	**NH_Gly**, DKP_H7, DKP_H6
Arg-Hα	4.1, m	-	
Arg-Hβ	1.89 – 1.71, m	25.7	
Arg-Hγ	1.53, m	24.6	
Arg-Hδ	3.12, q, 6.9 Hz	40.5	
Guan-NH	7.09, t, 5.76 Hz	///	
Gly-NH	8.06, dd, 7.98 – 4.14 Hz	///	**NH_Arg**, αH_Arg
Gly-Hα	4.20 – 3.55, dd, 17.6 – 8.17/5.76 Hz	42.6	
Asp-NH	7.94, d, 13.87 Hz	///	αH_Gly, DKP_H6
Asp-Hα	4.76, influenced by water	49.5	
Asp-Hβ	2.80 – 2.66, dd, 16.8 – 7.0/6.8 Hz	34.8	DKP_NH10
DKP_NH10	8.4, t, 6.12 Hz	///	αH_Asp, βH_Asp
DKP_NH1	8.1, s	///	
DKP-H9	3.89, dd, 15.3 – 7.5 Hz – 3.49, dt, 15.3 – 5.7 Hz	39.2	
DKP-H3	4.05, d, 5.52 Hz	59.8	
DKP-H7	2.79 – 2.58, dd, 14.23 Hz – 8.35/5.34 Hz	37.9	NH_Arg
DKP-H6	4.48, d	52.0	
CH_2-Bn	5.02- 4.03, d, 15.9 Hz	47.9	
H-Ar	7.4 – 7.2, m	128.0	

Table 2.3S ^{1}H and ^{13}C chemical shifts (δ in ppm) of compound **3** in H_2O/D_2O solution at T=298 K.

Compound **4**	^{1}H(δ,ppm) 298 K	^{13}C	n.O.e contacts
Arg-NH	8.30, s	-	DKP-H7
Arg-Hα	4.24, s	53.4	-
Arg-Hβ	1.77-1.67, m	27.9	-
Arg-Hγ	1.56, m	24.4	-
Arg-Hδ	3.12, q, 6.66 Hz	40.4	-
Guan-NH	7.11, t, 5.97 Hz	-	-
Gly-NH	8.27, t, 5.69Hz	-	Arg-Hα
Gly-Hα	3.94-3.61	42.6	-
Asp-NH	8.86, s	-	Gly-Hα
Asp-Hα	4.33,	52.4	-
Asp-Hβ	2.81-2.76, dd, 17.2 - 5.00 Hz	34.8	-
DKP_NH10	7.57, s (broad)	-	-
DKP_NH4	8.20, s	-	-
DKP-H9	3.64-3.31,	40.5	-
DKP-H3	3.88	52.0	-
DKP-H7	2.91-2.73,	36.3	Arg-NH
DKP-H6	4.18,	57.6	CH_2-Bn
CH_2-Bn	5.09-(influenced by water)-	-	-
H-Ar	7.33-7.23	128.4	DKP-H7

Table 2.4S ^{1}H chemical shifts (δ in ppm) of compound **5** in H_2O/D_2O solution at T=298 K.

Compound **5**	^{1}H (δ,ppm)	^{13}C	n.O.e contacts
Arg-NH	8.47, d, 6.98 Hz	-	DKP-H7, **Gly-NH**
Arg-Hα	4.09	55.5	-
Arg-Hβ	1.73-1.62, m	-	-
Arg-Hγ	1.55, m	-	-
Arg-Hδ	3.08, q, 6.55 Hz	40.3	-
Guan-NH	7.05, t(broad)	-	-
Gly-NH	8.22, d, 9.92 Hz	-	Arg-Hα, **Arg-NH**
Gly-Hα	4.21, dd, 15.87 Hz- 10.01 Hz -3.40, d, 15.87Hz	41.8	-
Asp-NH	8.41, d, 8.08 Hz	-	Gly-Hα
Asp-Hα	4.64	50.2	-
Asp-Hβ	2.84-2.67, dd 17.13 Hz- 8.03/7.23 Hz	35.2	-
DKP_NH10	8.58, t(broad)	-	Asp-Hα
DKP-H9	4.32-3.32	39.5	-
DKP-H3	4.33	-	-
DKP-H7	2.53 (overllapped)	39.0	-
DKP-H6	4.41	-	-
CH_2-Bn(1)	5.11-4.35, d, 16.56 Hz	49.6	-
CH_2-Bn(2)	4.13-5.00, d, 15.53-15.87Hz	48.3	-
H-Ar(1)	7.72-7.19	128.7	Arg-Hα, DKP-H7
H-Ar(2)	7.72-7.19	128.7	-

Table 2.5S ^{1}H chemical shifts (δ in ppm) of compound **5** in H_2O/D_2O solution at T=298 K.

Compound **6**	^{1}H (δ,ppm) 298 K	^{13}C	n.O.e contacts
Arg-NH	8.33	-	DKP-H7
Arg-Hα	4.30, broad	53.4	**Gly-NH**
Arg-Hβ	1.77-1.69, m	28.1	-
Arg-Hγ	1.57, m	24.7	-
Arg-Hδ	3.13, dd, 6.91 Hz	40.2	-
Guan-NH	7.10, t, 5.76Hz	-	-
Gly-NH	8.34	-	Asp-Hα
Gly-Hα	3.85-3.69, dd, 15.55 Hz - 5.53Hz	42.9	Asp-NH
Asp-NH	8.86, s	-	**DKP_NH10**, Gly-Hα
Asp-Hα	4.37, broad	52.0	**DKP_NH10**
Asp-Hβ	2.82-2.61, dd, 16.76 Hz - 6.23 Hz	35.7	-
DKP_NH10	7.93, t 6.18 Hz	-	Asp-Hα, **DKP_NH4**
DKP_NH4	8.01, s	-	**DKP_NH10**
DKP-H9	3.78-3.55, dd, 14.49 Hz - 4.49 Hz	-	-
DKP-H3	3.94, d, 8.29 Hz	55.1	-
DKP-H7	2.75-2.71, dd	34.8	Arg-NH
DKP-H6	4.24, broad	57.6	**H-Ar**
CH_2-Bn	5.07-4.22, d, 16.06 Hz	47.7	-
H-Ar	7.33/7.27	128.4	DKP-H6, DKP-H7, Gly-Hα

Table 2.6S ^{1}H chemical shifts (δ in ppm) of compound **6** in H_2O/D_2O solution at T=298 K.

Compound **7-A**	^{1}H (δ,ppm) 298 K	^{13}C	n.O.e contacts
Arg-NH	8.66, d, 8.03 Hz	-	**Gly-NH**, DKP-H7, DKP-H6
Arg-Hα	4.11	Under water	Gly-NH
Arg-Hβ	1.93-1.64, m	25.4	-
Arg-Hγ	1.50, m	24.4	-
Arg-Hδ	3.11, q, 6.62Hz	40.86	-
Guan-NH	7.07, t, 6.10Hz	-	-
Gly-NH	7.93, broad	-	**Arg-NH**, **Asp-NH**, Arg-Hα
Gly-Hα	4.37-3.56, dd, 17.38 Hz - 2.48 Hz	43.2	-
Asp-NH	7.76, d, 9.14 Hz	-	**Gly-NH**, Gly-Hα
Asp-Hα	4.89, influenced by water	Under water	-
Asp-Hβ	2.83-2.63, dd, 4.97 Hz - 17.38 Hz	36.94	-
DKP-NH10	8.04, t, 6.34 Hz	-	Asp-Hα, Asp-Hβ
DKP-H9	3.92-3.66, dd, 14.90 Hz - 6.62 Hz	39.7	-
DKP-H3	4.34	Under water	-
DKP-H7	2.55-2.81, dd, 11.59 Hz - 14.07Hz	37.7	Arg-NH, H-Ar
DKP-H6	4.69, influenced by water	Under water	-
CH_2-Bn(1)	5.03-4.04, d, 15.73Hz	48.0	DKP-H3, DKP-H9
CH_2-Bn(2)	4.57-4.80, influenced by water	Under water	-
H-Ar(1-2)	7.39/7.19, m	128.1	DKP-H7

Table 2.7S ^{1}H chemical shifts (δ in ppm) of compound **7-A** in H_2O/D_2O solution at T=298 K.

Compound 7-B	^{1}H (δ,ppm) 298 K	^{13}C	n.O.e contacts
Arg-NH	8.34, d, 5.85 Hz	-	DKP-H7
Arg-Hα	4.15, d, 7.79 Hz	53.59	-
Arg-Hβ	1.77, m	27.81	-
Arg-Hγ	1.61, m	-	-
Arg-Hδ	3.16, q, 6.91 Hz	40.55	-
Guan-NH	7.13-	-	-
Gly-NH	8.45, t, 5.47Hz	-	DKP-H7
Gly-Hα	3.80, d, 16.06 Hz -3.67, d, 16.06 Hz	43.45	-
Asp-NH	8.55, d, 7.55Hz	-	**DKP-NH10**, Gly-Hα
Asp-Hα	4.52	51.37	-
Asp-Hβ	2.77-2.61, dd, 7.12 Hz - 16.50 Hz	36.69	-
DKP-NH10	7.72, dd, 2.83 Hz - 8.19 Hz	-	**Asp-NH**, Asp-Hα
DKP-H9	3.74-3.33, broad	39.3	H-Ar
DKP-H3	4.10, dd, 1.94 Hz - 7.58 Hz	60.74	-
DKP-H7	2.93-2.85, dd, 6.37 - 3.0 Hz	35.63	-
DKP-H6	4.44, d	56.78	-
CH_2-Bn(1)	5.22-4.20, d, 14.63 Hz	47.51	-
CH_2-Bn(2)	5.17-4.32, d, 16.46 Hz	49.63	-
H-Ar(1-2)	7.42/7.17, m	~128.4	DKP-H6, DKP-H9, Gly-Hα

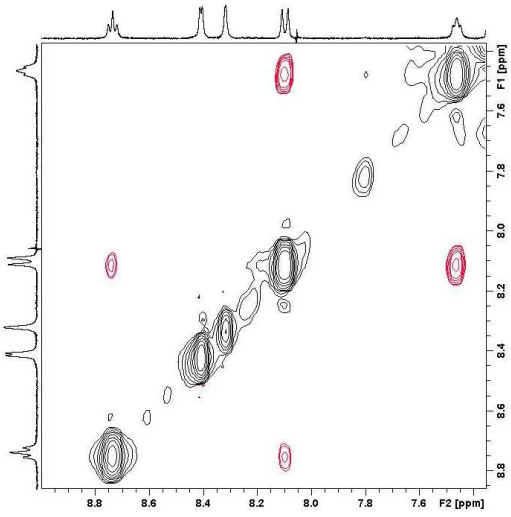
Table 2.8S ^{1}H chemical shifts (δ in ppm) of compound **7-B** in H_2O/D_2O solution at T=298 K.

The n.O.e contacts reported in the Table were used for the investigation of the ligands conformation. The NH-/NH- n.O.e. contacts are reported in bold. They are used to determine the conformation and as constraints for the MC/SD simulations.

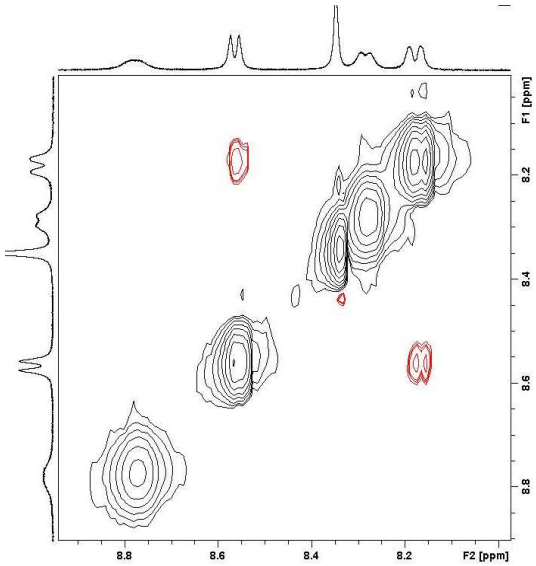
S2.2 NOESY SPECTRA FOR DKP-RGD COMPOUNDS (1 – 7)

NOESY spectra of the compounds **1** - **7** performed with mixing time = 700 ms in H_2O/D_2O (9:1) solution at 298 K. Dipolar correlation of amidic protons (**Figure 2.2S**).

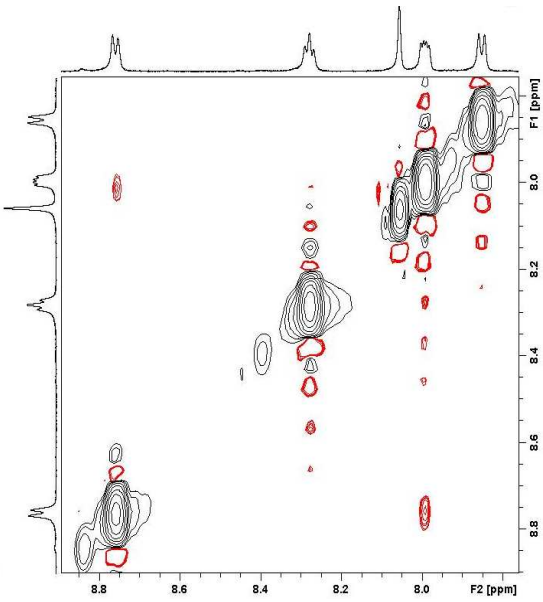
Compound 1



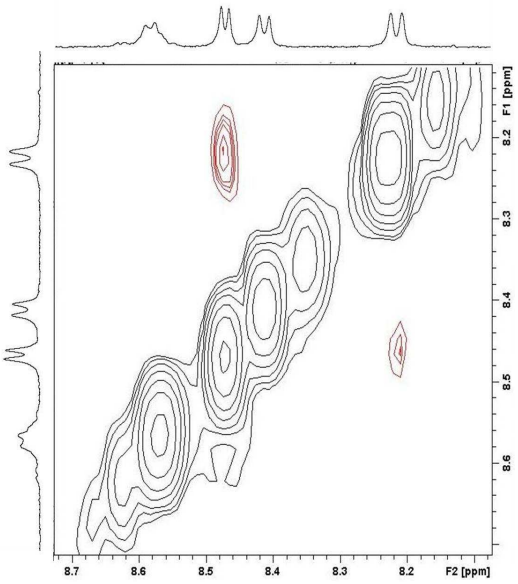
Compound 2



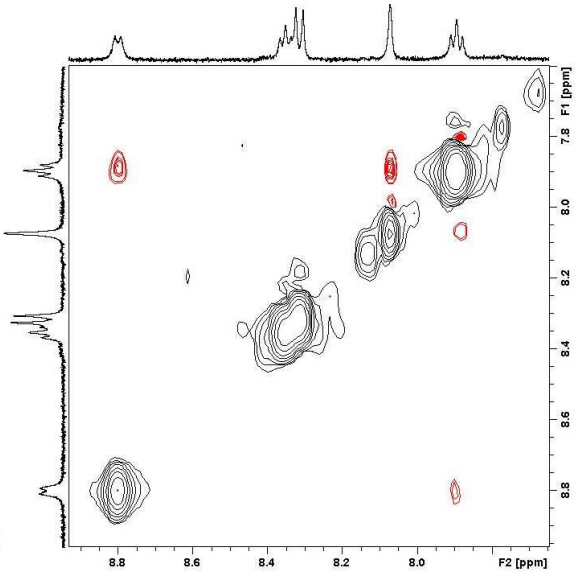
Compound 3



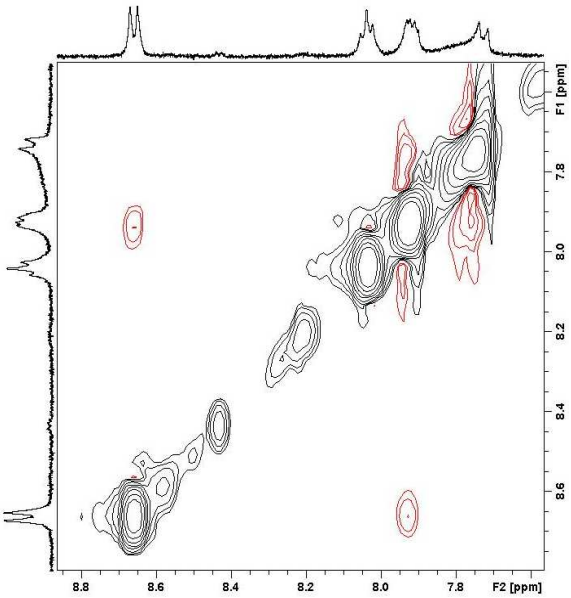
Compound 5



Compound 6



Compound 7-A



Compound 7-B

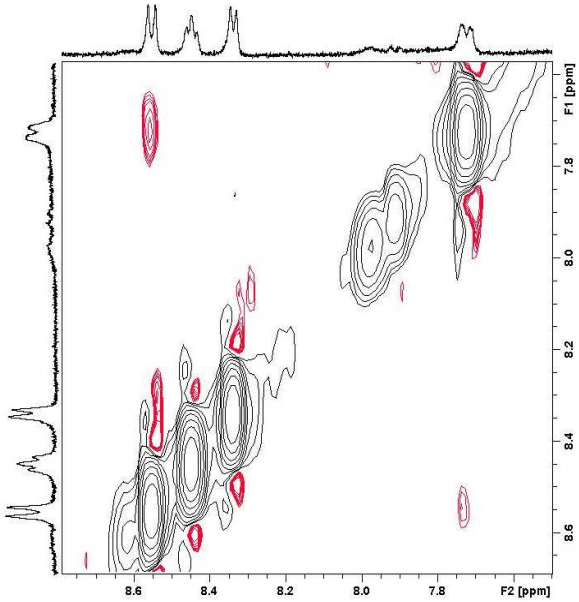


Figure 2.2S NOE contacts for DKP-RGD compounds **1 – 7**.

S2.3 WORKFLOW AND DISTANCE RESTRAINTS USED FOR THE CALCULATIONS

In a conformational equilibrium, NMR observables - such as NOEs - reflect the average properties of the individual conformations of a molecule. Thus, the area of cross peaks in a NOESY spectrum is a function of both the distance between the interacting protons and the molar fraction of each conformation at the equilibrium. For this reason, as the contribution of either factor cannot be clearly determined, all significant NOE long range interactions between amide protons have been assumed to produce the same distance restraint information. Therefore, in order to investigate the conformational features fitting a specific experimental NH-NH contact, a single restraint has been applied in each MC/SD run by imposing the same distance value of 2±0.5Å with a force constant of 100 kJ/mol·Å^2. At least two 10 ns MC/SD simulations (three on average) were conducted for each *cyclo*[DKP-RGD] peptidomimetic, using different starting conformations. When two mutually exclusive long-range NOE contacts are present (see ligands **DKP-1**, **DKP-6** and **DKP-7-A**), only one restraint at a time was applied for each 10 ns run.

Structure	Distance restraints = 2±0.5 Å (in the case of two mutually exclusive NOE contacts, only one restraint at a time was used for each 10 ns run)
cyclo[DKP-**1**-RGD]	NH_{Asp}-NH_{10} NH_{Asp}-NH_{Gly}
cyclo[DKP-**2**-RGD]	NH_{Arg}-NH_{Gly}
cyclo[DKP-**3**-RGD]	NH_{Arg}-NH_{Gly}
cyclo[DKP-**4**-RGD]	--
cyclo[DKP-**5**-RGD]	NH_{Arg}-NH_{Gly}
cyclo[DKP-**6**-RGD]	NH_{Asp}-NH_{10} NH_4-NH_{10}
cyclo[DKP-**7**-RGD]-A	NH_{Arg}-NH_{Gly} NH_{Asp}-NH_{Gly}
cyclo[DKP-**7**-RGD]-B	NH_{Asp}-NH_{10}

Table 2.9S List of distance restraints used for each *cyclo*[DKP-RGD] peptidomimetic during MC/SD simulations.

S2.4 CHARACTERIZATION OF DKP-*iso*DGR COMPOUNDS 8 AND 9 IN THE FREE STATE

Compound **8**	^{1}H (δ, ppm) 298 K	^{13}C
Asp-NH	8.25 (s overlapped)	///
Asp-Hα	4.66 (influenced by water)	50.9
Asp-Hβ	2.80, dd, 9.1 – 3 Hz	37.6
Gly-NH	7.90, t, 6.35 Hz	///
Gly-Hα	4.03-3.83, dd, 16.94 – 5.93 Hz	41.6
Arg-NH	8.27 (s overlapped)	///
Arg-Hα	4.15, d, 6.77 Hz	54.0
Arg-Hβ	1.69, m	27.1
Arg-Hγ	1.58-1.51, m	24.1
Arg-Hδ	3.13, q, 6.73 Hz	40.4
Guan-NH	7.11	///
DKP-NH10	8.44, t, 6.56 Hz	///
DKP-NH1	8.35, s	///
DKP-H9	4.01-3.46, dm	39.0
DKP-H3	4.02 (s overlapped)	59.3
DKP-H7	2.78-2.60, dd, 15.15 – 10.77 Hz	41.1
DKP-H6	4.83 (influenced by water)	51.2
CH_2-Bn	5.09-4.12, d, 15.91 Hz	47.1
H-Ar	7.38-7.25, m	128.2

Table 2.10S ^{1}H and ^{13}C chemical shifts (δ in ppm) of compound **8** in H_2O/D_20 at T=298 K.

Compound **9**	^{1}H (δ, ppm) 298 K	^{13}C
Asp-NH	8.62, d, 8.57 Hz	///
Asp-Hα	4.84 (influenced by water)	49.1
Asp-Hβ	2.95-2.63, dd, 14.82Hz - 12.12 Hz	37.0
Gly-NH	8.16, t, 5.71 Hz	////
Gly-Hα	4.17-3.71, dd, 17.90Hz - 7.29/5.30 Hz	40.8
Arg-NH	8.27, d, 6.75 Hz	///
Arg-Hα	4.13	53.4
Arg-Hβ	1.73-1.64, m	27.3
Arg-Hγ	1.55, m	23.9
Arg-Hδ	3.10	39.8
Guan-NH	7.1	///
DKP-NH10	8.45, t, 6.49 Hz	///
DKP-NH1	7.60, s	///
DKP-H9	3.88-3.59, dd 15.25 - 6.63 Hz	38.3
DKP-H3	4.04	58.8
DKP-H7	3.07-2.30, dd, 14.82 - 12.12Hz	37.5
DKP-H6	4.34, d, 12.43 Hz	50.5
CH_2-Bn	5.15-4.10, d, 16.08Hz	46.9
H-Ar	7.35-7.22, m	128.0

Table 2.11S ^{1}H and ^{13}C chemical shifts (δ in ppm) of compound **9** in H_2O/D_20 at T=298 K.

S2.5 NOESY SPECTRA FOR DKP-*iso*DGR COMPOUNDS (8 – 9)

NOESY spectra of the compounds **8** - **9** performed with mixing time = 700 ms in H_2O/D_2O (9:1) solution at 298 K. Dipolar correlation of amidic protons.

Compound 8 **Compound 9**

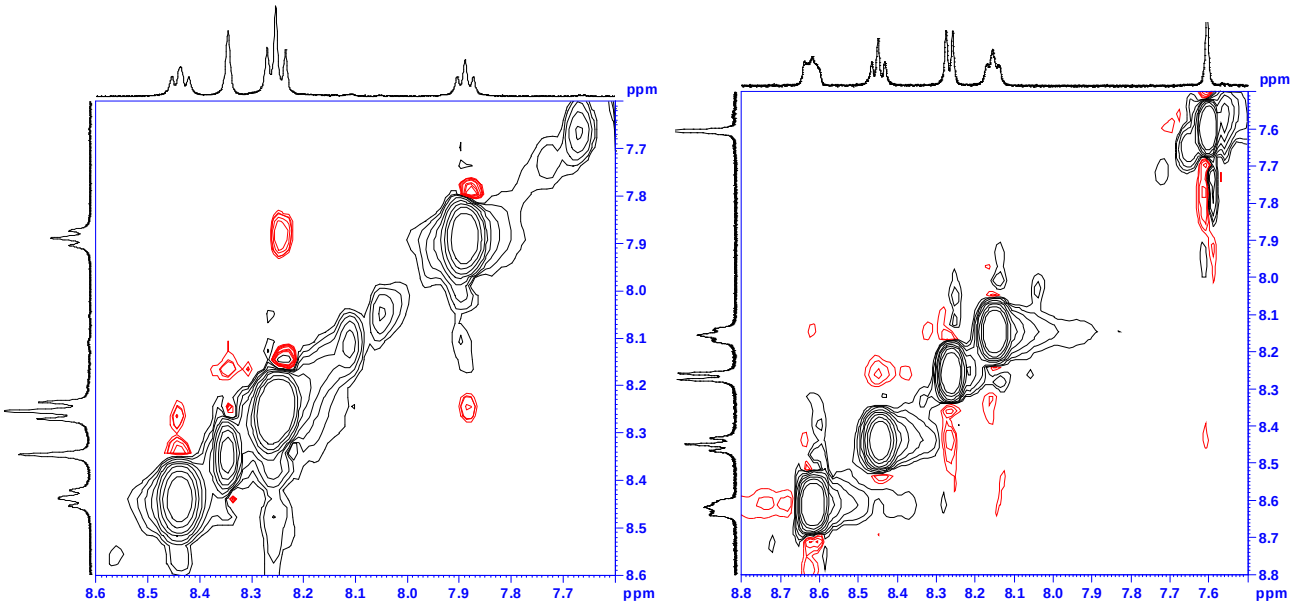


Figure 2.3S NOE contacts for DKP-*iso*RGD compounds **8 – 9**.

The n.O.e contacts reported are maintained in the three cell lines; when there is a selection between two conformations at the equilibrium, the observations about the n.O.e contacts selected are discussed in **Chapter 2** ("**Results and Discussion**").

S2.6 CHARACTERIZATION OF COMPOUNDS 1, 2 AND 3 WITH PLATELETS AT 298 K

Compound **1** / Protons	^{1}H(δ,ppm) 298 K
NH_Arg	8.40 (s overlapped)
αH-Arg	3.98 (s overlapped)
βH-Arg	1.76-1.62 (m)
γH-Arg	1.57 (m)
δH-Arg	3.14 (s overlapped)
NH_guan	7.14 (m br)
NH_Gly	8.69 (t, 6.23 Hz)
αH-Gly	4.00-3.64 (dd, 17.26-5.52 Hz)
NH_Asp	8.04 (d, 8.28 Hz)
αH-Asp	4.38 (br)
βH-Asp	2.59-2.40 (dd, 17.26-9.66 Hz)
DKP_NH1	8.38 (s overlapped)
DKP_NH10	7.49 (m br)
DKP_H9	3.91-3.47 (dt, 15.19-4.14 Hz)
DKP_H3	4.29 (br)
DKP_H7	3.14-2.82 (dd, 17.26-4.14 Hz)
DKP_H6	4.45 (br)
CH_2-Bn	///
H-Ar	7.32-7.26

Compound **2** / Protons	^{1}H(δ,ppm) 298 K
NH_Arg	8.60, d, 6.65Hz
αH-Arg	4.19, m
βH-Arg	1.78-1.69, m
γH-Arg	1.62, m
δH-Arg	3.14, q, 7.00 Hz
NH_guan	7.12, t, 5.68 Hz
NH_Gly	8.18, d br
αH-Gly	4.14-3.51, d br
NH_Asp	8.27, d, 8.29 Hz
αH-Asp	///
βH-Asp	2.60-2.54, dd, 16.2 – 7.7 Hz
DKP_NH1	8.34, s
DKP_NH10	8.64 (br)
DKP_H9	4.30-3.32, br
DKP_H3	///
DKP_H7	2.88-2.38, dd, 14.0 – 10 Hz
DKP_H6	///
CH_2-Bn	5.01-4.13, d, 16.6 Hz
H-Ar	7.38-7.28

Compound **3** / Protons	^{1}H(δ,ppm) 298 K
NH_Arg	8.72 (br)
αH-Arg	4.26, dd, 9.10 – 6.26 Hz
βH-Arg	1.89-1.72, m
γH-Arg	1.57, m
δH-Arg	3.16, td, 7.96 – 1.14 Hz
NH_guan	7.14 (br)
NH_Gly	8.03, s overlapped
αH-Gly	4.21-3.62, d
NH_Asp	8.03, s overlapped
αH-Asp	Under water
βH-Asp	2.80-2.65, dd, 14.3 – 8.53 Hz (2.55)
DKP_NH1	8.03, s overlapped
DKP_NH10	8.16, t br
DKP_H9	3.94-3.55, dd, 15.9 – 5.7 Hz
DKP_H3	4.07, s overlapped
DKP_H7	2.80-2.65, dd, 14.3 – 8.53 Hz (2.55)
DKP_H6	Influenced by water
CH_2-Bn	5.08-4.08, d, 16.5 Hz
H-Ar	7.36-7.26

Table 2.12S ^{1}H chemical shifts (δ in ppm) of compounds **1 – 2 - 3** in buffer solution in presence of human platelets suspension at T=298 K.

S2.7 CHARACTERIZATION OF DKP-RGD COMPOUNDS WITH ECV-304 AT 298 K AND 282 K

Compound **1** / Protons	^{1}H(δ,ppm) 298 K	^{1}H(δ,ppm) 282 K
NH_Arg	8.45 (s br)	8.55 (s br)
αH-Arg	4.04 (s)	4.04 (s)
βH-Arg	1.82 -1.66 (m)	1.80-1.66 (m)
γH-Arg	1.62 (m)	1.61 (m)
δH-Arg	3.18 (s)	3.18 (dd,15.4-5.8Hz)
NH_guan	7.23 (br)	7.25 (s br)
NH_Gly	8.72 (br)	8.83 (t, 6.0Hz)
αH-Gly	4.03-3.68 (d, 20 Hz)	4.04-3.68 (d, 17.80 Hz)
NH_Asp	8.08 (d, 8.1 Hz)	8.13 (d, 4.31Hz)
αH-Asp	4.40 (br)	4.39 (s br)
βH-Asp	2.59 (dd, 15-4 Hz)-2.40 (dd, 15-4 Hz)	2.59 (dd,16.0-9.9 Hz)-2.40 (dd,16.0-9.9 Hz)
DKP_NH1	8.45 (s br)	8.55 (s br)
DKP_NH10	7.54 (s br)	7.57 (s br)
DKP_H9	3.95-3.50 (dd, 16-3 Hz)	3.95 (d, 18.0 Hz)-3.50 (d, 18.0 Hz)
DKP_H3	4.32 (br)	4.33 (s)
DKP_H7	3.16-2.86 (dd, 18.7-5.3 Hz)	3.16 (d,18.5Hz)-2.89 (dd,17.5-3.9 Hz)
DKP_H6	4.51 (s br)	4.51 (s)
CH_2-Bn	- (deleted by water saturation)	- (deleted by water saturation)
H-Ar	7.38-7.30	7.36-7.31

Table 2.13S ^{1}H chemical shifts (δ in ppm) of compound **1** in buffer solution in presence of ECV-304 cancer cells suspension at T=298 K and at 282 K.

Compound **2** / Protons	^{1}H(δ,ppm) 298 K	^{1}H(δ,ppm) 282 K
NH_Arg	8.62 (s br)	8.72 (d,7.1Hz)
αH-Arg	4.19 (m)	4.18 (m)
βH-Arg	1.80-1.70 (m)	1.79-1.70 (m)
γH-Arg	1.63 (m)	1.64 (m)
δH-Arg	3.15 (t, 6.2 Hz)	3.15 (m)
NH_guan	7.12 (s br)	7.18 (s br)
NH_Gly	8.18 (s br)	8.26 (d, 10.0 Hz)
αH-Gly	4.09-3.50 (m)	4.10 (m) -3.48 (d, 18.8Hz)
NH_Asp	8.27 (s br)	8.40 (d, 8.5Hz)
αH-Asp	- (deleted by water saturation)	4.63 (m)
βH-Asp	2.58 (dd, 15.9-7.5 Hz)-2.50 (dd, 15.9-7.5 Hz)	2.60 (dd, 15.4-6.8 Hz)-2.49 (dd, 15.4-6.8 Hz)
DKP_NH1	8.60 (s br)	8.50 (s)
DKP_NH10	8.62 (s br)	8.79 (m)
DKP_H9	4.24 (br)-3.32	3.32 (m)
DKP_H3	4.30 (br)	4.33 (dd, 17.10-6.8 Hz)
DKP_H7	2.90 (dd, 15.7- 6.8 Hz)-2.38 (d, 15.7 Hz)	2.87 (dd,14.5-10.3Hz)-2.39 (d,14.5 Hz)
DKP_H6	4.60 (br)	4.65
CH_2-Bn	5.03(d, 16.56 Hz)-4.10	- (deleted by water saturation)
H-Ar	7.38-7.28	7.41-7.33

Table 2.14S ^{1}H chemical shifts (δ in ppm) of compound **2** in buffer solution in presence of ECV-304 cancer cells suspension at T=298 K and at 282 K.

Compound **3** / Protons	^{1}H(δ,ppm) 298 K	^{1}H(δ,ppm) 282 K
NH_Arg	8.74 (d, 8.4 Hz)	8.83 (d, 8.5 Hz)
αH-Arg	4.30 (br)	4.30 (q, 9.1 Hz)
βH-Arg	1.93-1.74 (m)	1.92-1.74 (m)
γH-Arg	1.60 (m)	1.61 (m)
δH-Arg	3.19 (br)	3.19 (m)
NH_guan	7.16 (br)	7.20 (t, 5.4 Hz)
NH_Gly	8.11 (s br)	8.15 (dd,4.6-7.8 Hz)
αH-Gly	4.24(dd, 15.5-7.1 Hz)-3.65 (d, 15.5 Hz)	4.24 (dd, 18.0-7.9 Hz)-3.65(dd, 18.0-3.6 Hz)
NH_Asp	8.07 (s br)	8.09 (d, 8.3 Hz)
αH-Asp	-	4.71 (s br)
βH-Asp	2.59(dd, 16.2-6.5 Hz)-2.56 (dd, 16.2-6.5 Hz)	2.59 (dd, 16.5-6.4Hz)-2.56 (dd, 16.5-6.4Hz)
DKP_NH1	8.11 (s br)	8.18 (s)
DKP_NH10	8.18 (t, 6.5 Hz)	8.30 (t, 6.8Hz)
DKP_H9	3.97(dd, 17.5-7.8 Hz)-3.57 (dd, 17.5-7.8 Hz)	3.97 (dd, 15.7-7.3 Hz)-3.58(m)
DKP_H3	4.05	4.05 (s br)
DKP_H7	2.81(dd, 14.2-8.4 Hz)-2.68 (dd, 14.2-8.4 Hz)	2.82 (dd, 13.9-8.5 Hz)-2.69 (dd, 13.9-8.5 Hz)
DKP_H6	- (deleted by water saturation)	- (deleted by water saturation)
CH_2-Bn	5.10(d, 16.2 Hz)-4.10 (d, 16.2 Hz)	5.12 (d,17.1Hz)-4.10 (d, 17.0Hz)
H-Ar	7.41-7.28	7.38-7.29

Table 2.15S ^{1}H chemical shifts (δ in ppm) of compound **3** in buffer solution in presence of ECV-304 cancer cells suspension at T=298 K and at 282 K.

Compound **4** / Protons	^{1}H(δ,ppm) 298 K	^{1}H(δ,ppm) 282 K
NH_Arg	8.35 (s br)	8.47 (s br)
αH-Arg	4.29 (s br)	4.30 (s br)
βH-Arg	1.82 (m)-1.72 (m)	1.82(m)-1.72(m)
γH-Arg	1.62 (m)	1.62 (m)
δH-Arg	3.17 (quintupletto, 6.5 Hz)	3.17 (s br)
NH_guan	7.16 (t, 5.6 Hz)	7.20 (m)
NH_Gly	8.29 (t, 5.7 Hz)	8.44 (t, 5.7 Hz)
αH-Gly	4.00 (s br)-3.65 (s br)	3.99-3.66
NH_Asp	8.77 (s br)	9.04 (s br)
αH-Asp	4.27 (s br)	4.31 (s br)
βH-Asp	2.66 (dd, 16.4-4.1 Hz)-2.56 (dd, 16.4- 10.0 Hz)	2.80 (s br)
DKP_NH4	8.22 (s)	8.35 (s)
DKP_NH10	7.56 (s br)	7.58 (s br)
DKP_H9	3.89 (s br)-3.71 (s br)	Not clearly visible
DKP_H3	Under water	Not clearly visible
DKP_H7	2.99 (d, 16.0 Hz)- 2.78 (d, 15.4 Hz)	3.00 (m)-2.80 (m)
DKP_H6	4.73 (under water)	-
CH_2-Bn	5.17 (s br)-4.18 (s br)	5.17 (s br)- 4.16 (d, 14.6 Hz)
H-Ar	7.38-7.29	7.35-7.26

Table 2.16S ^{1}H chemical shifts (δ in ppm) of compound **4** in buffer solution in presence of ECV-304 cancer cells suspension at T=298 K and at 282 K.

Compound **5** / Protons	^{1}H(δ,ppm) 298 K	^{1}H(δ,ppm) 282 K
NH_Arg	8.39 (s br)	8.39 (d,6.7 Hz)
αH-Arg	3.88	-
βH-Arg	1.54 (m)- 1.42 (m)	1.57 (s br)-1.44 (s br)
γH-Arg	1.37 (m)	1.38 (s br)
δH-Arg	2.91 (s br)	2.91 (s br)
NH_guan	6.94 (s br)	6.95 (s br)
NH_Gly	8.08 (m)	8.08
αH-Gly	4.06 (m)- 3.17 (m)	3.22 (d, 17.3 Hz) – 4.06 (dd, 16.1-9.7 Hz)
NH_Asp	8.25 (m)	8.26 (d, 8.7 Hz)
αH-Asp	4.36 (s br)	4.42 (d, 8.6 Hz)
βH-Asp	2.29 (s br)	2.38-2.27
DKP_NH10	8.45 (s br)	8.46 (dd, 9.1- 4.3 Hz)
DKP_H9	4.1 (s br)- 3.11 (s br)	4.14 (m)-3.15 (dd, 14.9-3.7 Hz)
DKP_H3	4.10 (s br)	Not clearly visible
DKP_H7	2.32 (m)	2.30 (m)
DKP_H6	Not clearly visible	4.16 (s br)
CH_2-Bn	Influence by water	4.99 (d, 17.4 Hz) -4.17 (d, 17.4 Hz)
H-Ar	7.22-7.04	7.22-7.00

Table 2.17S ^{1}H chemical shifts (δ in ppm) of compound **5** in buffer solution in presence of ECV-304 cancer cells suspension at T=298 K and at 282 K.

Compound **6** / Protons	^{1}H(δ,ppm) 298 K	^{1}H(δ,ppm) 282 K
NH_Arg	8.33 (d, 8.0 Hz)	8.46 (d, 7.9 Hz)
αH-Arg	4.32 (m)	4.27 (s br)
βH-Arg	1.79 (m)-1.70 (m)	1.80 (m)-1.71 (m)
γH-Arg	1.57 (m)	1.59 (m)
δH-Arg	3.15 (q, 6.9 Hz)	3.15 (s br)
NH_guan	7.14 (s br)	7.20 (s br)
NH_Gly	8.37 (t, 4.6 Hz)	8.48 (s br)
αH-Gly	3.95 (d, 8.6 Hz)-3.85 (s br)	3.85 (s br)- 3.70 (dd, 15.8-6.2 Hz)
NH_Asp	8.80 (d, 5.2 Hz)	8.92 (s br)
αH-Asp	4.37 (m)	4.34 (m)
βH-Asp	2.76 (s br)- 2.59 (dd, 16.2- 8.6 Hz)	2.74 (m)- 2.59 (dd, 16.2-8.8 Hz)
DKP_NH4	8.12 (s)	8.21 (s)
DKP_NH10	7.91 (t, 5.8 Hz)	7.99 (t, 5.8 Hz)
DKP_H9	3.70 (dd, 15.3-5.5 Hz)	3.88(s br)-3.50 (m)
DKP_H3	Not clearly visible	Not clearly visible
DKP_H7	2.79 (dd, 14.3-5.1 Hz)- 2.74 (m)	2.79 (dd, 13.8-4.5 Hz)- 2.73 (m)
DKP_H6	4.26 (s br)	4.25 (s br)
CH_2-Bn	5.07 (d,13.7 Hz)-4.26 (s br)	5.08 --4.26 (s br)
H-Ar	7.37- 7.28	7.37-7.28

Table 2.18S ^{1}H chemical shifts (δ in ppm) of compound **6** in buffer solution in presence of ECV-304 cancer cells suspension at T=298 K and at 282 K.

Compound **7 A** / Protons	^{1}H(δ,ppm) 282 K
NH_Arg	8.65 (d, 8.4 Hz)
αH-Arg	4.27
βH-Arg	1.96-1.64
γH-Arg	1.52
δH-Arg	3.15 (calibrato con quello a 298 K)
NH_guan	7.15 (m)
NH_Gly	8.05 (s br)
αH-Gly	3.58
NH_Asp	8.04 (s br)
αH-Asp	4.85 (under water)
βH-Asp	2.63 (dd, 16.4-7.5 Hz)-2.54 (dd, 16.4-5.7 Hz)
DKP_NH10	8.10 (t, 6.9 Hz)
DKP_H9	3.92 (m)-3.71 (m)
DKP_H3	4.35 (s br)
DKP_H7	2.90 (d, 15.2 Hz)-2.48 (dd, 14-12 Hz)
DKP_H6	Under water
CH_2-Bn	5.08 (under water)-4.06 (s br)
H-Ar	7.41-7.20

Table 2.19S ^{1}H chemical shifts (δ in ppm) of compound **7-A** in buffer solution in presence of ECV-304 cancer cells suspension at T=282 K.

Compound **7 B** / Protons	^{1}H(δ,ppm) 298 K	^{1}H(δ,ppm) 282 K
NH_Arg	8.31 (d, 6.37 Hz)	8.42 (d, 6.09 Hz)
αH-Arg	4.12 (q, 6.97 Hz)	4.11 (q, 7.39 Hz)
βH-Arg	1.72 (m)-1.59(m)	1.71(m)-1.59 (m)
γH-Arg	1.54 (m)	1.53 (m)
δH-Arg	3.12 (m br)	3.12 (m br)
NH_guan	Overlapped with H_Ar	Overlapped with H_ar
NH_Gly	8.44 (t, 5.57 Hz)	8.54 (s br)
αH-Gly	3.76 (dd, 15.74-4.41 Hz)-3.64 (dd, 15.74 Hz-6.61 Hz)	3.76 (dd, 15.9-4.61 Hz)-3.64 (15.9-6.63 Hz)
NH_Asp	8.48 (d, 7.43 Hz)	8.56 (s br)
αH-Asp	4.48 (d, 7.54 Hz)	4.47 (d, 8.13 Hz)
βH-Asp	2.65 (dd, 16.41-6.94 Hz)-2.49 (dd, 16.41-7.87 Hz)	2.65(dd, 16.7-7.5 Hz) -2.49(dd, 16.7-7.6 Hz)
DKP_NH10	7.67 (dd, 8.65-3.24 Hz)	7.72 (dd, 8.94-3.09 Hz)
DKP_H9	3.71 (ddd, 14.53-8.89-3.25 Hz)- 3.29 (ddd, 14.53-7.37-3.25 Hz)	3.27 (dd, 7.42-2.73 Hz)-3.25(dd, 7.42-3.32 Hz)
DKP_H3	4.07 (dd, 7.54-2.26 Hz)	4.04 (m)
DKP_H7	2.79 (dd, 14.55-6.14 Hz)- 2.90 (dd, 14.55-2.54 Hz)	2.90 (dd, 14.8-6.36 Hz)-2.78 (dd, 14.8-2.07 Hz)
DKP_H6	4.41 (d, 6.03 Hz)	4.41 (dd br)
CH_2-Bn (1)	5.20 (d, 15.6 Hz)-4.17 (d, 15.5 Hz)	5.20 (d, 15.51 Hz)-4.17 (d, 15.5 Hz)
CH_2-Bn (2)	5.17 (d, 17.15 Hz)-4.29 (d, 17.25 Hz)	5.16 (d, 16.17 Hz)-4.29 (d, 16.99 Hz)
H-Ar (1)	7.37-7.14	7.34
H-Ar (2)	7.37-7.14	7.25-7.14

Table 2.20S ^{1}H chemical shifts (δ in ppm) of compound **7-B** in buffer solution in presence of ECV-304 cancer cells suspension at T= 298 K and at 282 K.

S2.8 CHARACTERIZATION OF DKP-RGD COMPOUNDS WITH MDA-MB-231 AT 282 K

Compound **2** / Protons	^{1}H(δ,ppm) 282 K
NH_Arg	8.68 (d, 6.6Hz)
αH-Arg	4.18 (m br)
βH-Arg	1.76 (m)-1.67 (m)
γH-Arg	1.60 (m)
δH-Arg	3.13 (t, 5.23 Hz)
NH_guan	7.16 (s br)
NH_Gly	8.22 (d, 9.67 Hz)
αH-Gly	4.18 (q, 7.0 Hz)-3.45 (d, 15.6 Hz)
NH_Asp	8.36 (d, 8.14 Hz)
αH-Asp	4.60 (s br)
βH-Asp	2.57 (dd, 16.5-8 Hz)- 2.47 (dd, 16.5-7.4 Hz)
DKP_NH1	8.45 (s)
DKP_NH10	8.75 (dd, 9.7-4.1 Hz)
DKP_H9	4.15 (m)-3.29 (d, 16.8 Hz)
DKP_H3	-
DKP_H7	2.84 (dd, 14.2 – 9.7 Hz)-2.35 (d, 14.2 Hz)
DKP_H6	Under water
CH_2-Bn	5.05 (under water)-4.14 (m br)
H-Ar	7.36-7.27

Table 2.21S ^{1}H chemical shifts (δ in ppm) of compound **2** in buffer solution in presence of MDA-MB-231 cancer cells suspension at T= 282 K.

Compound **3** / Protons	^{1}H(δ,ppm) 282 K
NH_Arg	8.79 (d, 8.44 Hz)
αH-Arg	4.26 (q, 7.70 Hz)
βH-Arg	1.87 (m)-1.69 (m)
γH-Arg	1.57 (m)
δH-Arg	3.13 (m)
NH_guan	7.16 (t, 5.50 Hz)
NH_Gly	8.12 (dd br)
αH-Gly	4.19 (dd, 17.80-8.15 Hz) -3.60 (dd, 17.80- 4.34 Hz)
NH_Asp	8.05 (d, 8.61 Hz)
αH-Asp	4.66 (near water)
βH-Asp	2.54 (d, 6.60 Hz)
DKP_NH1	8.15 (s)
DKP_NH10	8.28 (t, 6.36 Hz)
DKP_H9	3.93 (dd, 15.48-7.23Hz) -3.53 (dt, 15.48-6.42 Hz)
DKP_H3	4.03 (s overlapped)
DKP_H7	2.76 (dd, 14.29-8.80 Hz) -2.63 (dd, 14.29-5.28 Hz)
DKP_H6	4.67 (near water)
CH_2-Bn	5.70 (d, 7.94 Hz) – 3.88 (d, 13.86 Hz)
H-Ar	7.33-7.23

Table 2.22S ^{1}H chemical shifts (δ in ppm) of compound **3** in buffer solution in presence of MDA-MB-231 cancer cells suspension at T= 282 K.

Compound **4** / Protons	^{1}H(δ,ppm) 282 K
NH_Arg	8.42 (d, 5.92 Hz)
αH-Arg	4.23 (s sovrapposto)
βH-Arg	1.78 (m)-1.68 (m)
γH-Arg	1.58 (m)
δH-Arg	3.11 (t, 5.96 Hz)
NH_guan	7.16 (br)
NH_Gly	8.35 (t, 5.92 Hz)
αH-Gly	3.96 (d, 19.59 Hz) -3.59 (d, 19.59 Hz)
NH_Asp	8.84 (s br)
αH-Asp	4.21 (s overlapped)
βH-Asp	2.60 (dd, 16.71-4.05 Hz)-2.51(dd, 16.71-10.38 Hz)
DKP_NH4	8.29 (s)
DKP_NH10	7.49 (d, 9.86 Hz)
DKP_H9	3.67 (m br)-3.23 (t, 11.08 Hz)
DKP_H3	3.81 (d br)
DKP_H7	2.99 (d, 17.04 Hz)-2.74 (17.04 Hz)
DKP_H6	4.16 (s)
CH_2-Bn	5.16 (d, 18.7 Hz)-4.10 (d, 18.7 Hz)
H-Ar (2)	7.31-7.23

Table 2.23S ^{1}H chemical shifts (δ in ppm) of compound **4** in buffer solution in presence of MDA-MB-231 cancer cells suspension at T= 282 K.

Compound **5** / Protons	^{1}H(δ,ppm) 282 K
NH_Arg	8.37 (d, 6.91 Hz)
αH-Arg	3.90 (m)
βH-Arg	1.54 (m)- 1.42 (m)
γH-Arg	1.36 (m)
δH-Arg	2.92 (m)
NH_guan	6.93 (s br)
NH_Gly	8.06 (d, 9.92Hz)
αH-Gly	4.04 (dd, 16.00-10.18 Hz) - 3.20 (d, 16.00 Hz)
NH_Asp	8.23 (d, 8.42 Hz)
αH-Asp	4.48 (near water)
βH-Asp	2.39 (dd, 15.93-7.54 Hz)- 2.25 (dd, 15.93-7.75 Hz)
DKP_NH10	8.44 (dd, 9.32- 4.51 Hz)
DKP_H9	4.14 (s overlapped)- 3.14(dd, 14.54-3.64 Hz)
DKP_H3	4.42 (not clearly visible)
DKP_H7	2.30 (m overlapped)
DKP_H6	4.18 (s overlapped)
CH_2-Bn	4.96 (d, 17.45 Hz) -4.16 (s overlapped)
CH_2-Bn	4.83 (near water) -3.99 (d, 15.27 Hz)
H-Ar (1)	7.23 -7.10
H-Ar (2)	7.18 -7.04

Table 2.24S ^{1}H chemical shifts (δ in ppm) of compound **5** in buffer solution in presence of MDA-MB-231 cancer cells suspension at T= 282 K.

Compound **6** / Protons	^{1}H(δ,ppm) 282 K
NH_Arg	8.46 (d, 7.9 Hz)
αH-Arg	4.27 (s br)
βH-Arg	1.80 (m)-1.71 (m)
γH-Arg	1.59 (m)
δH-Arg	3.15 (s br)
NH_guan	7.20 (s br)
NH_Gly	8.48 (s br)
αH-Gly	3.85 (s br)- 3.70 (dd, 15.8-6.2 Hz)
NH_Asp	8.92 (s br)
αH-Asp	4.34 (m)
βH-Asp	2.74 (m)- 2.59 (dd, 16.2-8.8 Hz)
DKP_NH4	8.21 (s)
DKP_NH10	7.99 (t, 5.8 Hz)
DKP_H9	3.88(s br)-3.50 (m)
DKP_H3	Not visible
DKP_H7	2.79 (dd, 13.8-4.5 Hz)- 2.73 (m)
DKP_H6	4.25 (s br)
CH_2-Bn	5.08 (near water)-4.26 (s br)
H-Ar	7.37-7.28

Table 2.25S ^{1}H chemical shifts (δ in ppm) of compound **6** in buffer solution in presence of MDA-MB-231 cancer cells suspension at T= 282 K.

Compound **7 B** / Protons	^{1}H(δ,ppm) 282 K
NH_Arg	8.43 (d, 6.24 Hz)
αH-Arg	4.13 (q, 6.66 Hz)
βH-Arg	1.72 (m)
γH-Arg	1.60 (m) - 1.54 (m)
δH-Arg	3.12 (m br)
NH_guan	7.16 (s overlapped)
NH_Gly	8.56 (s sovrapposto)
αH-Gly	3.75 (dd, 15.73- 4.75 Hz overlapped)-3.64 (dd, 15.73-6.89 Hz)
NH_Asp	8.57 (s overlapped)
αH-Asp	4.48 (d, 7.95 Hz)
βH-Asp	2.66 (dd, 16.54-7.33 Hz) -2.49 (dd, 16.54 - 7.80 Hz)
DKP_NH10	7.73 (dd, 8.93- 3.28 Hz)
DKP_H9	3.72 (m overlapped)- 3.26 (ddd, 14.55-8.60-2.99 Hz)
DKP_H3	4.04 (m br)
DKP_H7	2.91 (dd, 15.02-6.38 Hz) - 2.78 (dd, 15.02-1.91 Hz)
DKP_H6	4.43 (m)
CH_2-Bn (1)	5.16 (d, 17.32 Hz) - 4.29 (d, 17.32 Hz)
CH_2-Bn (2)	5.21 (d, 15.97 Hz)- 4.18 (d, 15.97 Hz)
H-Ar (1)	7.37-7.16
H-Ar (2)	7.37-7.16

Table 2.26S ^{1}H chemical shifts (δ in ppm) of compound **7-B** in buffer solution in presence of MDA-MB-231 cancer cells suspension at T= 282 K.

S2.9 STD SPECTRA OF COMPOUNDS 2 – 7 WITH MDA-MB-231 at 282 K

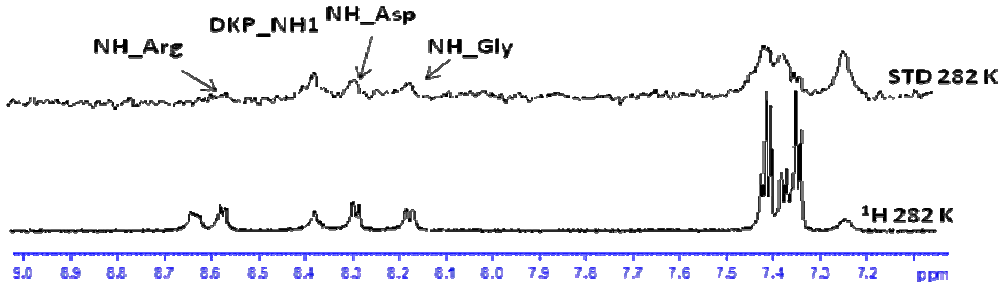


Figure 2.4S STD spectrum of compound **2** in buffer solution in presence of MDA-MB-231 cancer cells suspension at T= 282 K.

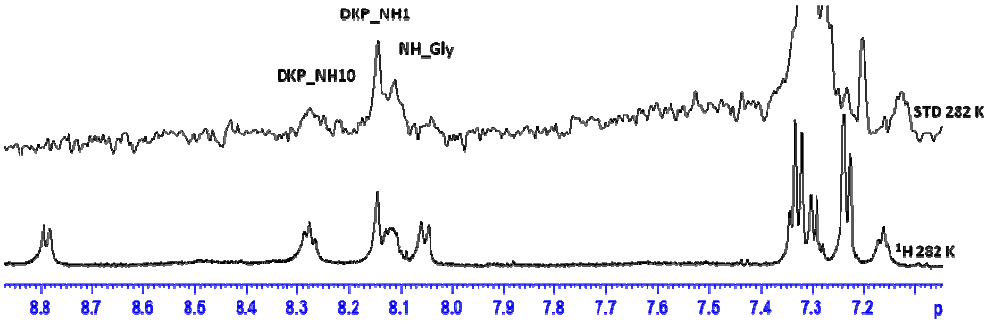


Figure 2.5S STD spectrum of compound **3** in buffer solution in presence of MDA-MB-231 cancer cells suspension at T= 282 K.

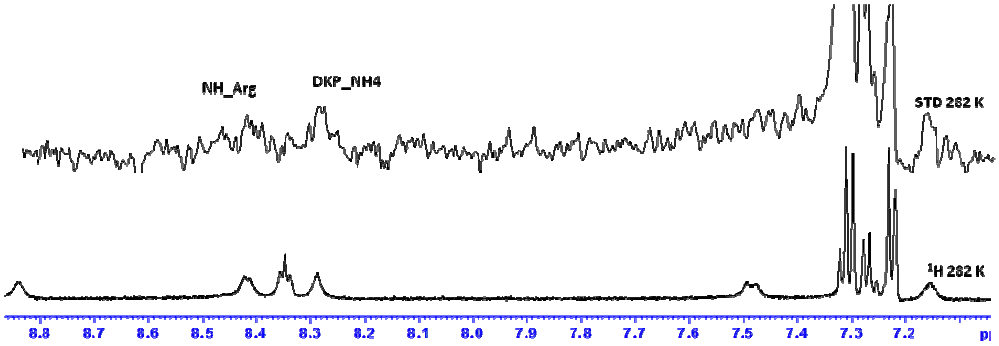


Figure 2.6S STD spectrum of compound **4** in buffer solution in presence of MDA-MB-231 cancer cells suspension at T= 282 K.

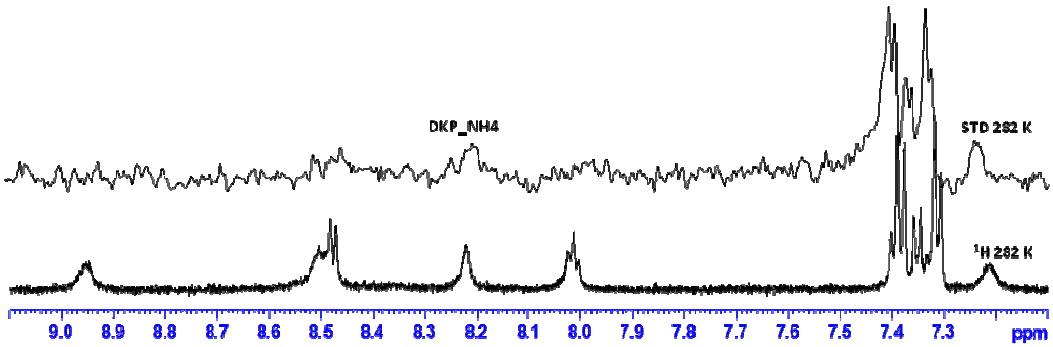


Figure 2.7S STD spectrum of compound **6** in buffer solution in presence of MDA-MB-231 cancer cells suspension at T= 282 K.

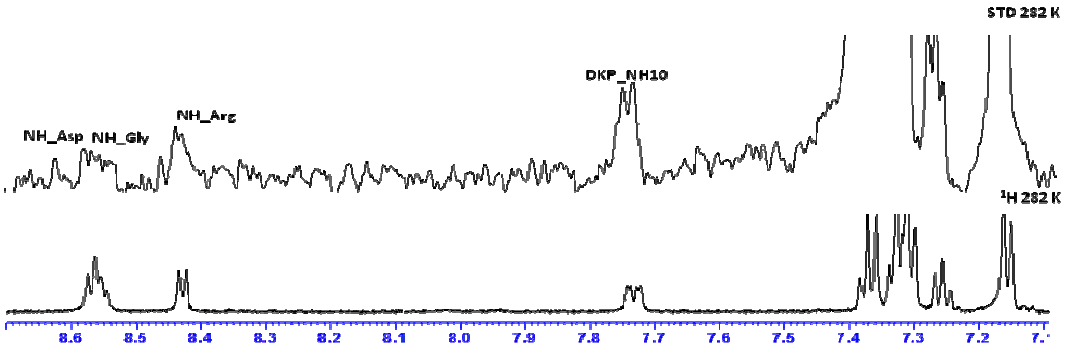
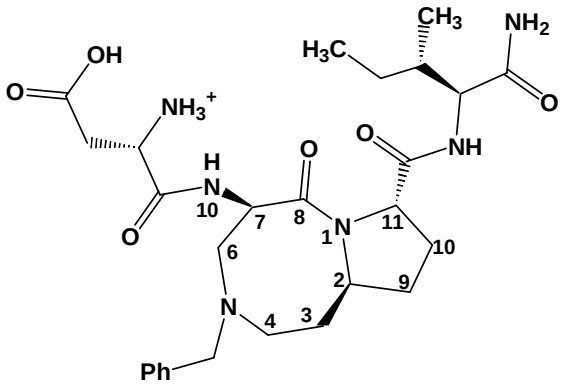


Figure 2.8S STD spectrum of compound **7-B** in buffer solution in presence of MDA-MB-231 cancer cells suspension at T= 282 K.

∗Supplementary information for Chapter 3∗

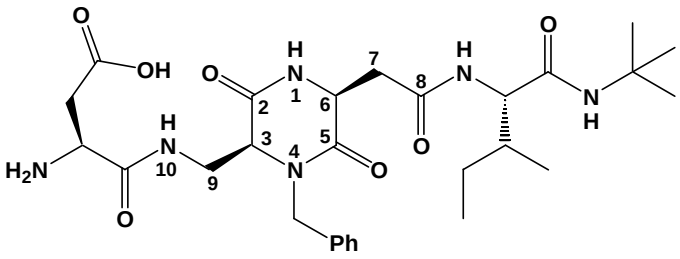
S3.1 CHARACTERIZATION OF THE COMPOUND 1c IN THE FREE-STATE AT 298 AND 283 K



Compound **1c**	^{1}H(δ,ppm) 298 K		^{1}H(δ,ppm) 283 K		^{13}C
NH_isoLeu	7.98	d, 8.21 Hz	8.16	d, 7.91 Hz	///
αH_isoLeu	4.02	t, 8.19 Hz	4.01	t, 8.19 Hz	58.2
βH_isoLeu	1.81	m	1.81	m, 8.47 Hz	35.6
γH_isoLeu	1.47 – 1.17	m	1.49	m overlapped	24.9
CH_3_isoLeu	0.89	d, 7.31 Hz	0.89	d, 7.16 Hz	14.2
CH_3_isoLeu	0.82	t, 7.69 Hz	0.80	t, 7.63 Hz	9.8
NH10	8.44	s broad	8.61	s broad	///
αH_Asp	4.18	dd, 7.82 – 5.55 Hz	4.17	dd, 8.00 – 5.84 Hz	50.8
βH_Asp	2.72	dd, 17.14 – 5.55 Hz	2.71	dd, 17.32 – 5.74 Hz	37.0
	2.66	dd, 17.14 – 8.07 Hz	2.66	dd, 17.32 – 8.00 Hz	
H7	5.05	dd, 12.45 – 5.49 Hz	5.07	influenced by water	51.0
H6	2.92	m overlapped	2.93	m	57.5
	2.82	t, 11.85 Hz	2.84	t, 11.77 Hz	
H11	4.49	m	4.48	m	59.7
H10	2.32	m	2.33	m	27.2
	1.88	dd, 14.62 – 7.60Hz	1.89	dd, 14.59 – 7.53 Hz	
H9	2.11	m	2.12	m, 6.59 Hz	30.6
	1.67	dd, 13.36 – 6.81 Hz	1.67	dd, 13.08 – 6.68 Hz	
H2	Under water		4.51	m	Under water
H3	1.55 – 1.51		1.58 – 1.54	m, 6.78 Hz	35.5
H4	Overlapped with H6		Overlapped with H6		///
CH_2_Bn	3.77	d, 13.49 Hz	3.82	d, 13.37 Hz	61.2
	3.67	d, 13.49 Hz	3.73	d, 13.37 Hz	
H_Ar	7.35 – 7.29	///	7.35 – 7.31	///	129.4
-NH_2	7.55 – 7.04	s	7.68 – 7.15	s	///

Table 3.1S ^{1}H and ^{13}C chemical shifts (δ in ppm) of compound **1c** in buffer solution at 298 K and 283 K.

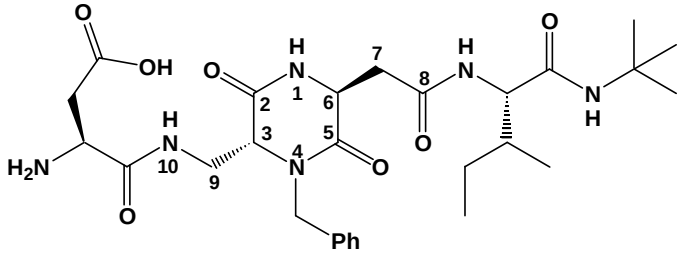
S3.2 CHARACTERIZATION OF THE COMPOUND 2c IN THE FREE-STATE AT 298 AND 283 K



Compound **2c**	^{1}H(δ,ppm) 298 K		^{1}H(δ,ppm) 283 K		^{13}C
NH_tBu	7.65	s	7.78	s	///
CH_3_tBu	1.23	s	1.23	s	28.0
NH_isoLeu	8.11	d, 8.96 Hz	8.23	d, 8.96 Hz	///
αH_isoLeu	3.94	t,8.58 Hz	3.93	t, 9.37 Hz	58.5
βH_isoLeu	1.69	q, 7.93 Hz	1.69	q, 8.84 Hz	36.1
γH_isoLeu	1.40, 1.12	m, 8.36 Hz	1.41, 1.13	m, 7.80 Hz	24.6
CH_3_isoLeu	0.84	d, 7.29 Hz	0.85	d, 7.42 Hz	14.3
CH_3_isoLeu	0.79	t, 7.93 Hz	0.79	t, 8.02 Hz	9.3
NH10	Not observable	///	8.64	s broad	///
αH_Asp	4.05	dd, 10.28 – 3.91 Hz	4.06	dd, 10.73 – 3.80 Hz	51.0
βH_Asp	2.60	dd, 17.92 – 3.94 Hz	2.61	dd, 18.11-3.58 Hz	37.5
	2.40	dd, 17.92-10.33 Hz	2.39	dd, 18.11-10.73 Hz	
DKP_H6	4.45	m	4.44	t broad	51.2
DKP_H7	3.06	dd, 17.43-4.84Hz	3.10	dd, 18.21-4.68Hz	37.7
	2.83	dd, 17.43-4.34Hz	2.82	dd, 18.21-4.16Hz	
DKP_NH1	Not observable	///	8.56	s	///
DKP_H9	3.74	d, 14.87Hz	3.73	d, 16.31Hz	40.9
	3.52	dd, 14.87-5.29Hz	3.50	d	
DKP_H3	4.29	m	4.33	s	59.1
CH_2_Bn	Under water	///	4.57	d broad	///
	Under water	///	Under water	///	
H_Ar	7.28	///	7.31	///	128.6

Table 3.2S ^{1}H and ^{13}C chemical shifts (δ in ppm) of compound **2c** in buffer solution at 298 K and 283 K.

S3.3 CHARACTERIZATION OF THE COMPOUND 3c IN THE FREE-STATE AT 298 AND 283 K



Compound **3c**	^{1}H(δ,ppm) 298 K		^{1}H(δ,ppm) 283 K		^{13}C
NH_tBu	7.49	s	7.65	s	///
CH3_tBu	1.23	s	1.23	s	27.7
NH_isoLeu	7.93	d, 8.26 Hz	8.10	d, 8.86Hz	///
αH_isoLeu	3.92	t, 8.51 Hz	3.92	t, 8.51Hz	58.8
βH_isoLeu	1.71	q, 8.99 Hz	1.70	m	35.6
γH_isoLeu	1.40, 1.08	m, 7.49 Hz	1.41-1.07	m, 7.87Hz	24.7
CH_3_isoLeu	0.78	t, 7.74 Hz	0.78	t, 7.77Hz	9.8
CH_3_isoLeu	0.83	d, 7.28 Hz	0.83	d, 7.19Hz	14.6
NH10	Not visible	///	8.65	s broad	///
αH_Asp	4.09	dd, 7.36-5.52 Hz	4.13	dd, 7.73 -5.52Hz	50.9
βH_Asp	2.67	dd, 17.50-5.79Hz	2.70	dd, 17.30-5.52Hz	37.2
	2.61	dd, 17.50-8.18Hz	2.63	dd, 17.30-8.10Hz	
DKP_H6	4.45	m	4.48	t broad	51.0
DKP_H7	3.05	dd, 17.17-3.98Hz	3.07	dd, 17.22-3.88Hz	36.7
	2.79	dd, 17.17-5.03Hz	2.78	dd, 17.22-5.01Hz	
DKP_NH1	Not visible	///	8.49	s broad	///
DKP_H9	3.82	dd, 15.00-5.14Hz	3.83	dd, 15.08-5.00Hz	39.1
	3.72	dd, 15.00-3.11Hz	3.73	d, 15.00Hz	
DKP_H3	4.01	m	4.02	s broad	59.0
CH_2_Bn	5.14	d, 16.12Hz	4.24	d, 15.82Hz	47.3
	4.20	d, 16.12Hz	5.18	d, 15.82Hz	
H_Ar	7.39-7.34	///	7.36	///	128.7

Table 3.3S ^{1}H and ^{13}C chemical shifts (δ in ppm) of compound **3c** in buffer solution at 298 K and 283 K.

S3.4 REFERENCE COMPOUNDS

NMR experimet of ADH-1 with EC1-EC2 fragment of E-cadherin

Preliminary NMR experiments were performed on the reference compound ADH-1 and E-cadherin EC1-EC2 domain. However, we were not able to find a suitable irradiation frequency for the STD experiments[5] since the STD signals in the presence of E-cadherin were the same observed for the ligand in the absence of protein (control experiment).

NMR experimet of tripeptide AWA with EC1-EC2 fragment of E-cadherin

For this reason, we used other compounds to set up the experimental conditions. Unfortunatly, the natural DWVI sequence was not soluble in the phosphate buffer used for our experiment. We then performed NMR experiment using a commercially available tripeptide containing a Trp residue in position 2, *i.e.* AWA. In docking calculation into E-cadherin model, the peptide was able to satisfy the key binding requirements placing the indole ring into the hydrophobic pocket and forming the salt bridge between N-terminal NH_3^+ group and receptor Glu89 side chain. The NOESY spectra both at 298 K and 282 K did not give evidences of preferential conformation and in the *tr*-NOESY spectra the NOE contacts were the same that characterized the compound in the free state. STD-NMR experiments at 283K showed a signal for the indole NH (**Figure 3.1S**). The other aromatic protons signals overlap to the STD spectrum performed in the absence of protein (control experiment).

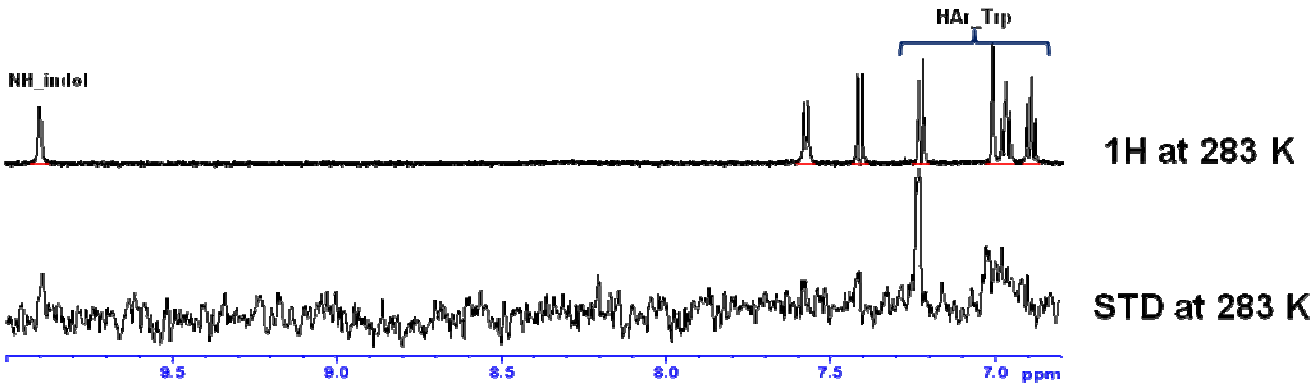
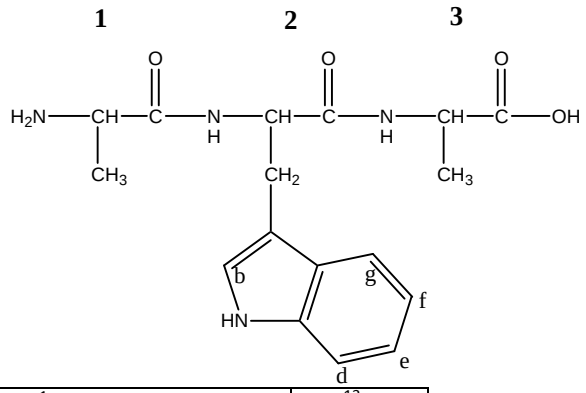


Figure 3.1S Low field region (amidic and aromatic region) of the STD and ^{1}H of AWA at 283 K.

S3.5 CHARACTERIZATION OF THE COMPOUND AWA IN THE FREE-STATE AT 283 K



Protons	^{1}H(δ,ppm) 283 K	^{13}C
NH_2_Ala1	///	///
αCH_Ala1	3.70, q, 7.28 Hz	49.1
CH_3_Ala1	1.20, d, 7.58 Hz	16.9
NH_Phe	///	/////
αCH_Phe	4.42, s broad	54.7
CH_2_Phe	3.08, dd, 14.83-5.59Hz 2.98, dd, 14.83-8.39Hz	26.8
HAr-Phe (b)	7.02, (d, 1.72Hz)	124.7
HAr-Phe (e,f)	6.98, t, 7.79Hz 6.91, t, 8.09Hz	121.7 119.33
HAr-Phe (g, d)	7.42, d, 8.49 Hz 7.24, d, 8.49 Hz	118.31 111.92
NH_Ala3	7.59, d, 6.86 Hz	////
αCH_Ala3	3.77, q, 7.28 Hz	51.0
CH_3_Ala3	1.02, d, 7.58 Hz	17.9
NH_indol	9.89, s	/////

Table 3.4S ^{1}H and ^{13}C chemical shifts (δ in ppm) of AWA in buffer solution at 283 K.