UNIVERSITÀ DEGLI STUDI DI MILANO

DIPARTIMENTO DI CHIMICA

PhD COURSE IN INDUSTRIAL CHEMISTRY- XXIX CYCLE

Computational techniques to evaluate at atomic level the mechanism of molecular binding

CHIM/06 Organic Chemistry

Cristina Paissoni
R10582

Tutor: Prof. Dr. Laura Belvisi (Università degli Studi di Milano)
Co-Tutor: Dr. Giovanna Musco (IRCCS Ospedale San Raffaele)

Coordinator: Prof. Dr. Maddalena Pizzotti

A.Y. 2015/2016

Contents

Abstract

Integrins are an important class of transmembrane receptors that relay signals bidirectionally across the plasma membrane, regulating several cell functions and playing a key role in diverse pathological processes. Specifically, integrin subtype αIIbβ3 is involved in thrombosis and stroke, while subtypes αvβ3 and α5β1 play an important role in angiogenesis and tumor progression. They therefore emerged as attractive pharmacological targets. In the past decades several peptides and peptidomimetics targeting these proteins and based on the integrin recognition motif RGD (Arg-Gly-Asp) have been developed, whereby their affinity and selectivity for a specific integrin subtype have been fine-tuned by modulation of RGD flanking residues, by cyclization or by introduction of chemical modifications. Thus far, the design and development of RGD-based cyclopeptides have been mainly based on empirical approaches, requiring expensive and time-consuming synthesis campaigns. In this field, the employment of computational tools, in the design and drug optimization process, has been limited by the inherent difficulties to predict in silico the three-dimensional structure and the inhibitory activity of cyclopeptides. However, recent improvements in both computational resources and in docking and modeling techniques are expected to open new perspectives in the development of cyclopeptides as modulators of protein-protein interactions and, particularly, as integrin inhibitors.

Within the PhD project described in this thesis, I have investigated the applicability of computational techniques in predicting and rationalizing how the environment of the recognition-motif in cyclopeptides (i.e. flanking residues and introduction of chemical modification) could influence their integrin affinity and selectivity. These

features can regulate integrin affinity both by favoring direct interactions with the receptor and/or by modulating the three-dimensional conformation properties of the recognition motif. To take into account both these aspects, I have proposed and optimized a multi stage computational protocol in which exhaustive conformational sampling, docking calculations and re-scoring techniques are combined. Specifically: i) the exhaustive sampling could be achieved by using the enhanced sampling technique Metadynamics in its Bias Exchange variant (BE-META), which represents a valuable methodology for the acceleration of rare events, allowing to cross the high free energy barriers characteristic of cyclopeptides and providing reliable estimations of the populations of the accessible conformers. ii) The docking calculations, complemented with the re-scoring technique MM-GB/SA (Molecular Mechanics Generalized Born Surface Area) and the cluster analysis of the decoy poses, allow to evaluate the ability of each peptide to engage interactions with the receptors and to rank the docking poses according to their binding ability; iii) a joint analysis of the previous outcomes results in a reliable ranking of cyclopeptides based on their binding affinity and in the rationalization of their structure-activity relationship. This computational protocol has been exploited in two different applications, illustrated within the thesis.

In the following is presented a brief description of the thesis outline.
In Chapter 1 I have introduced the investigated biological system, integrins, providing information on their structures and their mechanism of action, meanwhile evidencing why they are interesting therapeutic targets. Additionally, I have provided some background concerning the role of cyclopeptides as therapeutics, briefly reviewing their application as integrin inhibitors.
In Chapter 2 I have presented the computational techniques employed during my PhD work: Molecular Dynamics methods, especially focusing on the enhanced sampling technique Metadynamics in its Bias Exchange variant, and docking techniques, with a focus on the utilized docking software Glide and HADDOCK.
In Chapter 3 the first application of the computational protocol has been illustrated. Herein I have rationalized how the introduction of chemical modifications, specifically backbone N-methylation, impacts on the equilibrium conformation and consequently on the integrin affinity of five RGD containing cyclic hexapeptides,

which were previously generated by the group of professor Kessler to modulate their selectivity for αIIbβ3 integrin.
In Chapter 4 I have focused my attention on cyclopeptides harboring a recently discovered integrin recognition motif: isoDGR (isoAsp-Gly-Arg). Firstly, I have systematically tested the accuracy of eight Molecular Dynamic force fields in reproducing the equilibrium properties of isoDGR-based cyclopeptides. Then, I applied the computational protocol to rationalize the diverse selectivity and affinity profiles for two cancer-related integrins, αvβ3 and α5β1, displayed by three isoDGR-based cyclic hexapeptides. They differ in the residues flanking the isoDGR motif and show appealing tumor-homing properties. Herein, I have also proposed a model explaining why the conjugation of one cyclopeptide with a sulfo-SMCC-derived chemical linker results in an improved affinity and selectivity for αvβ3 integrin.

Overall, in this PhD work I have shown that the combination of different computational techniques, BE-META, docking and MM-GB/SA re-scoring, could be a reliable approach to perform structure-activity relationship studies in cyclopeptides. Specifically, the proposed protocol is able to predict the influence of the recognition motif environment (i.e. chemical modification and flanking residues) on integrin affinities. These two features regulate integrin affinity differently: the first one by conformational modulation of the recognition motif, the second by engaging direct interactions with the receptor. Of note, the approach can deal with both these mechanisms of affinity modulation. We expect that the protocol herein described could be used in future to screen novel peptides library or to complement biochemical experiments during the drug optimization stages, assisting organic chemists in the design of more effective integrin-targeting peptides.

Chapter 1

Introduction: the biological context

In the following I am going to provide a background about the system investigated in this thesis: integrins. I will focus on their structure, their mechanism of action and signaling, and I will explain why they are important therapeutic targets. Since the aim of this project is to study the interaction of these receptors with cyclopeptides, I will also briefly review the growing role of peptides as therapeutic and diagnostic agents, focusing on two classes of integrin-targeting cyclopeptides based on the Arginine-Glycine-Aspartate (RGD) and isoAspartate-Glycine-Arginine (isoDGR) sequence.

1.1 Integrins

Integrins are large, heterodimeric, transmembrane proteins that are expressed on the cell surface of multicellular animals. They are essential mediators of cell-extracellular matrix (ECM) interactions and cell-cell interactions, regulating diverse biological processes as cell attachment, proliferation, migration and differentiation.[1] The name "integrin" was introduced for the first time in 1986 by Tamkun et al.,[2] to denote the role of these proteins as an integral membrane complex that linked

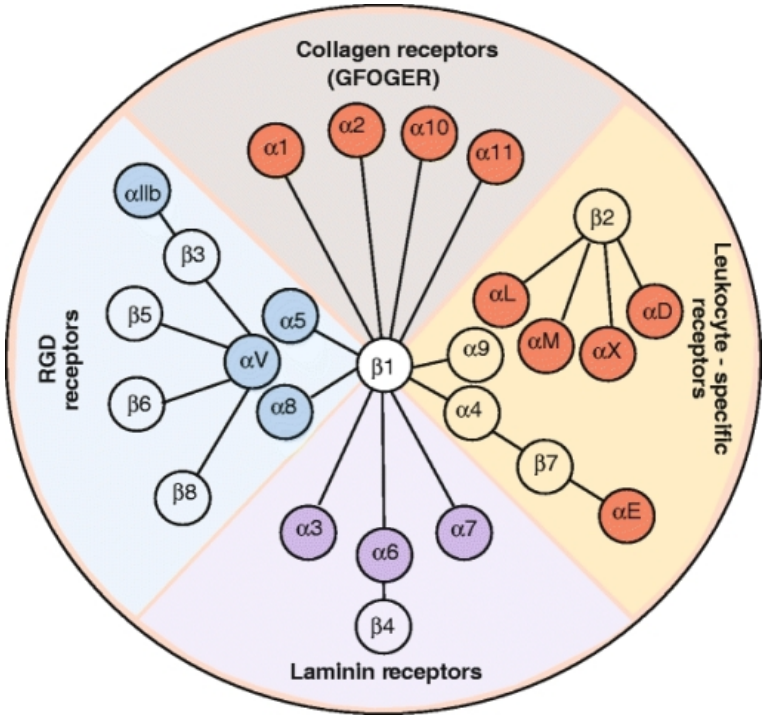


Figure 1.1: Representation of the 24 heterodimers belonging to the integrin family. Figure is reprinted from reference [6].

the extracellular matrix and the cytoskeleton. Since that first characterization, the integrin family has been widely studied and its role in different pathologies, such as thrombosis, cardiovascular disorders, inflammation, tumor invasion and metastasis, is now well established. Based on both their easy accessibility as cell surface receptors interacting with extracellular ligands and their involvement in several diseases, integrins are currently considered important therapeutic and diagnostic targets. [3–5]

All integrins are composed of two non-covalently bound domains; in vertebrates 24 heterodimers consisting of 18 α and 8 β subunits have been identified. A common feature of these receptors is their ability to bind a wide variety of ECM ligands with diverse binding affinities. [6] This characteristic, and their different tissue distribution, makes them able to exert and regulate diverse biological functions. Specifically, these 24 heterodimers can be classified, based on the nature of their molecular interactions, in four subgroups: Collagen-, Laminin-, Leukocyte-specific-

and RGD-receptors (see Figure 1.1). The group of RGD-receptors, that includes all the αv integrins, as well as αIIbβ3, α5β1 and α8β1, share the ability to bind ECM ligands at the α-β subunit interface through the recognition of the three amino acids sequence RGD (Arginine-Glycine-Aspartate). The integrins of this group are among the most promiscuous in the family, since the RGD sequence is present in several ECM proteins such as Fibronectin, Vitronectin, Fibrinogen, Laminin, Osteopontin, etc. Despite this wide variety of common ligands, the rank of their affinity varies, presumably depending on how properly the RGD-ligand conformation fits inside the specific integrin binding pocket.[7]

In the following the attention will be focused on this group of RGD-recognizing integrin, and specifically on integrins αIIbβ3, αvβ3 and α5β1.

1.1.1 Integrins structures

From a structural point of view each integrin subunit is composed by a large extracellular domain (700-1100 residues), a single transmembrane region (>20 residues) and a short cytoplasmic tail (13-70 residues).[5] The domains that comprise the extracellular region adopt a shape that resembles a large head on two legs, with the head containing the site that is responsible for the extracellular ligand recognition. The cytoplasmatic tail is fundamental for the interactions with intracellular signaling molecules and for the formation of focal adhesion. The two subunits α and β of these heterodimers do not show homology to each other, although the different α domains present similarities among themselves and also in the β domains conserved regions are found.[6]

The publication of the crystal structure of integrin αvβ3 extracellular segment in its free state[8] and, the year later, in its bound state to the artificial ligand Cilengitide, c(RGDf(NMe)V),[9] gave precious information about both the structure of the ectodomain and the positioning of the RGD motif at the α-β subunits interface. Later, the appearance of crystal structures of integrin αIIbβ3 in complex with its antagonist Eptifibatide,[10] reveal an identical atomic basis for ligand-integrin interaction.
In Figure 1.2 a schematic representation of all the subunits composing integrin

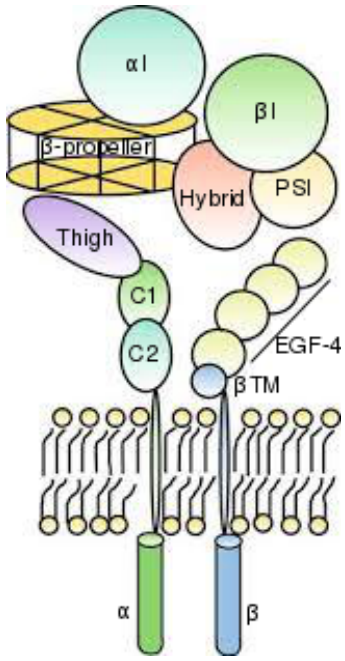


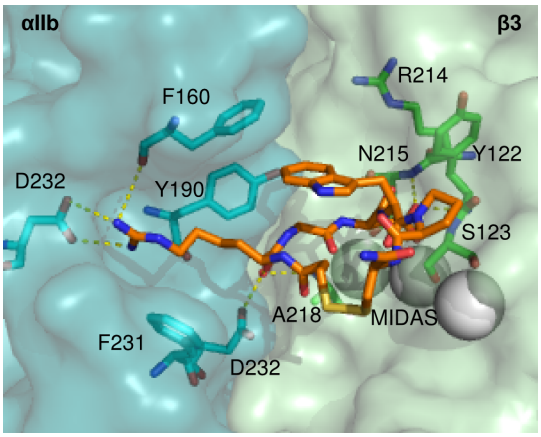
Figure 1.2: Schematic representation of the subunits arrangement in αI-containing integrins. Figure is adapted from reference [6].

domains is shown. The ectodomain of the α chain consists of a seven-bladed β-propeller, which is linked to a thigh and two calf domains, together forming the leg supporting integrin head. [11] In a subset of integrins, not including the ones belonging to the group of RGD-receptors, the α chain contains also a fifth extracellular domain called αI domain, which is found between blades 2 and 3 of the β-propeller. [12] In these αI-containing integrins, αI domain is the primary ligand binding region.

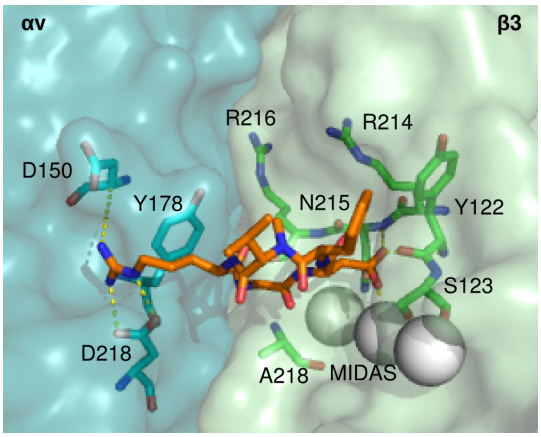
The β chain contains a βI domain, inserted in a hybrid domain, which in turn is found in a plexin-semaphorin-integrin (PSI) region; these domains are followed by four cysteine-rich epidermal growth factor (EGF) repeats and a β-tail domain. In the case of αI-lacking integrins (that include all the members of the RGD-receptors group), the site of ligand binding is found at the interface between the β-propeller of the α chain and the βI-domain of the β subunit. This βI-domain, which is homologous of the αI-domain, adopts a Rossmann-fold characterized by the presence of up to seven β-sheets surrounded by α-helices, and presents three binding sites for divalent cations (Ca^{2+}, Mg^{2+} or Mn^{2+}). It is known that the binding of

divalent cations to the integrin head can determine the integrin activation state. In particular it seems that Mn^{2+} and Mg^{2+} stimulate ligand binding, while Ca^{2+} is typically inhibitory. Out of the three sites, the metal-ion-dependent adhesion site (MIDAS) is fundamental for ligand binding, it is located at the top of the central β-strand and it is generally occupied by Mg^{2+} (or Mn^{2+}) in ligand bound integrins. The second binding site, called ligand-associated metal-ion binding site (LIMBS), stabilizes the metal ion at MIDAS and consequently acts as a positive regulator of ligand binding to integrins. Finally, the third site adjacent to MIDAS (ADMIDAS) can be occupied by Ca^{2+} or Mn^{2+} metal ions both in the bound and in the unbound state.[13]

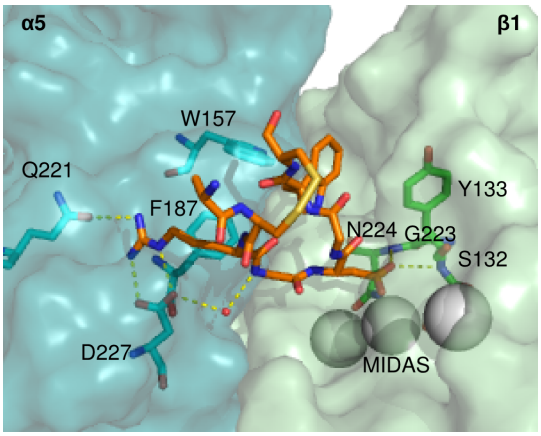
As already outlined, crystal structures of different integrins ectodomain (as $\alpha v\beta 3$, $\alpha IIb\beta 3$ and $\alpha 5\beta 1$) in complex with RGD-based peptides or peptidomimetics,[9,14–16] revealed similarity in the positioning of the RGD motif at the interface between the α and β subunits. In particular, two fundamental anchor-points in the binding pocket can be individuated: i. the Arginine sidechain of the RGD sequence engages electrostatic interactions with acidic residues belonging to the β-propeller of the α chain; ii. the RGD-Aspartate carboxylate directly coordinates the MIDAS of the β subunit and can also interact or form hydrogen bonds with other residues of this region. This canonical binding pose is often associated to an "electrostatic clamp". Figure 1.3 shows the binding poses, based on crystallographic data, of the following integrin RGD-peptide complexes: $\alpha v\beta 3$ in complex with Cilengitide (c(RGDf(*N*Me)V), Protein Data Bank code: 1l5g), $\alpha IIb\beta 3$ with Eptifibatide (MpaHarGDWPC-NH2, PDB: 2vdn), $\alpha 5\beta 1$ with a disulfide cycled peptide (ACRGDGWCG, PDB: 4wk4). It is recognized that the residues flanking the RGD sequence can play a role in fine-tuning the affinity and selectivity of these peptides. This result is achieved both by engaging direct contacts in the integrin binding pocket and by regulating the conformation adopted by the RGD sequence (specifically, a distance between the Cβ atoms of Arg and Asp of approximately 9 Å is known to be suitable for the formation of the electrostatic clamp).[17]



(a) αIIbβ3 with Eptifibatide. PDB: 2vdn



(b) αvβ3 with Cilengitide. PDB: 1l5g



(c) α5β1 with ACRGDGWCG. PDB: 4wk4

Figure 1.3: In the picture are represented the ligand-integrin binding sites of αIIbβ3, αvβ3, and α5β1, in complex with cyclic peptides. The integrin is represented in surface, with the α chain in blue and the β chain in green; the residues important for the interactions with the RGD-based cyclopeptides are highlighted in licorice. The ions in MIDAS, ADMIDAS and LIMBS position are represented with white spheres, water molecules are red small spheres, while ligands are in orange licorice. Hydrogen bonds are shown in yellow.

1.1.2 Integrin activation and signal transduction

It is well established that integrins can exist in multiple activation states, which are linked to the ligand binding stability. These different activations states, which correspond to global rearrangements of integrin tertiary structures, have been observed by electron microscopy, exposure of activation epitopes (LIMBS, ligand-induced binding sites) and have been confirmed by solution X-ray scattering.[10] Mainly three different activation and conformational states can be identified (Figure 1.4): a bent inactive state, associated with low affinity for both ECM ligands and intracellular activators, and characterized by the α and β transmembrane domains in close proximity; an extended-primed state, associated with higher affinity for ECM ligands; and a fully-activated state, in which integrin is coupled to both intracellular and ECM ligands.[18,19] The mechanistic models describing this large conformational rearrangement from the inactive to the active state are still controversial. Currently two models of integrin activation, based on NMR and crystal structures, are widely

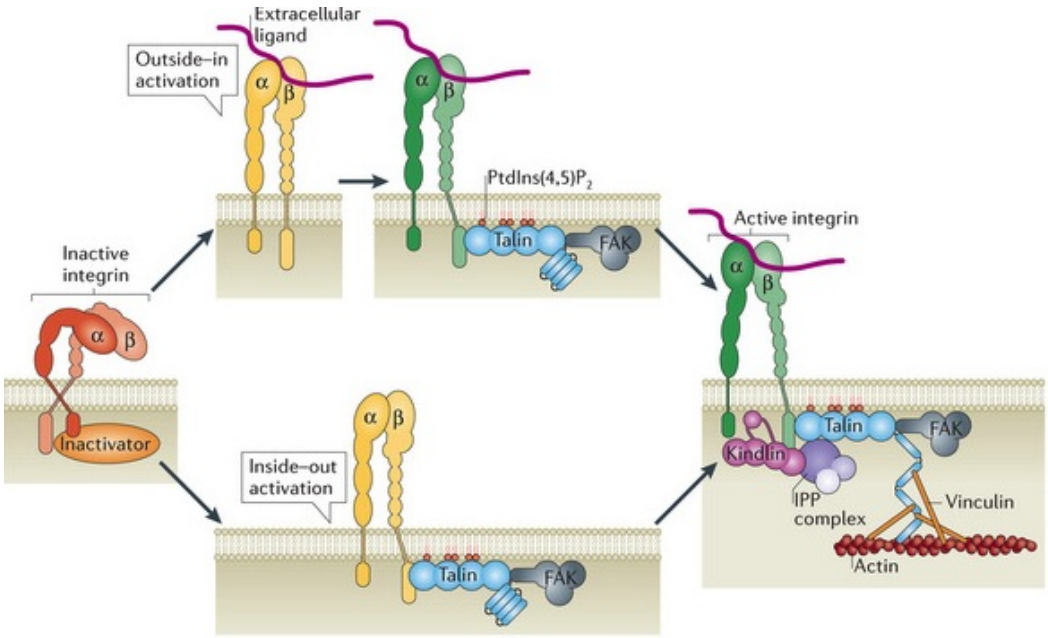
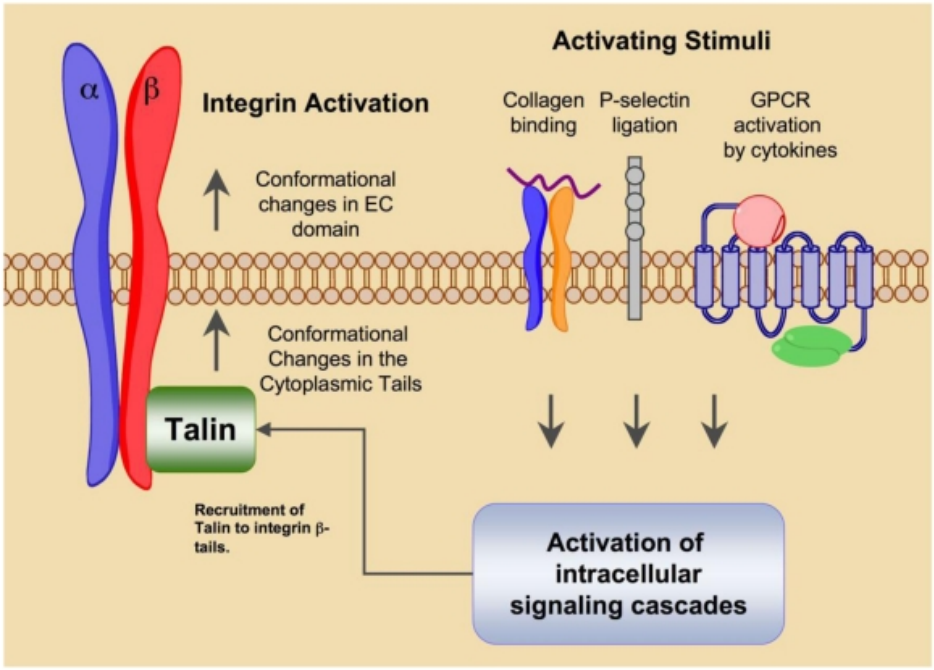


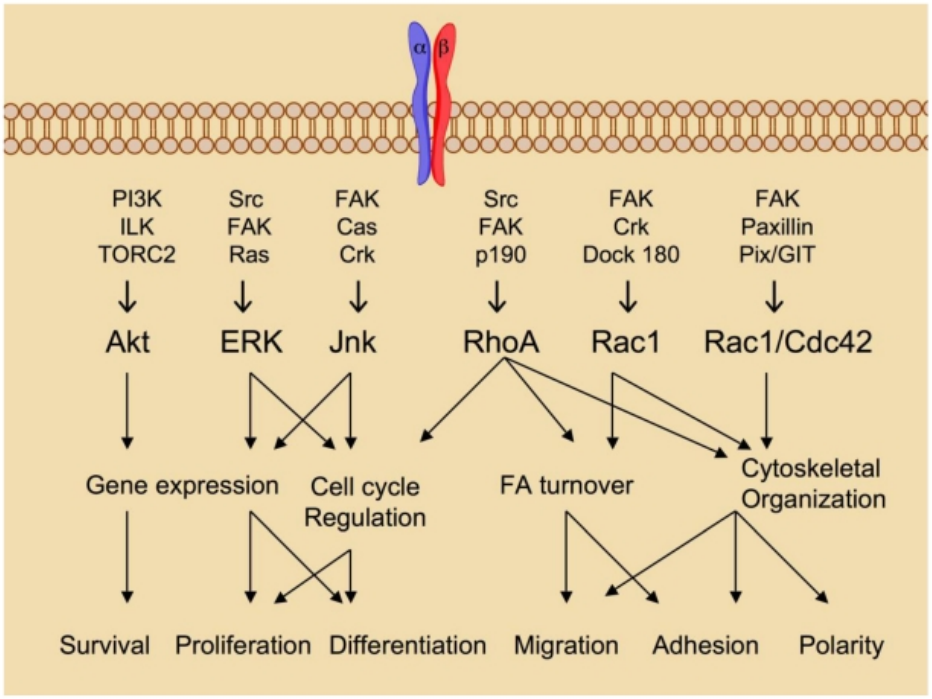
Figure 1.4: Schematic representation of the conformations adopted by integrins upon outside-in and inside-out activation. Three main conformations can be observed: inactive bent, extended-primed and fully-activated state. Figure is reprinted from reference[19].

cited: the *switchblade* model, in which activation results in a stretched integrin conformation and a loss of the dimer contacts in the transmembrane region, and the *deadbolt* model, in which smaller changes are proposed, where the activation state retains the bent conformation, with a movement in the transmembrane domain and a subsequent opening of the ligand binding site.[20–22] However, according to both the models, the proposed conformational changes lead to active structures with heightened receptor avidity.

The bidirectional nature of integrin signaling allows the activation to be regulated by both intra- and extracellular factors (referred as inside-out and outside-in signaling events).
Inside-out signaling is triggered by changes in the cytoskeleton environment that allows binding of intracellular activators to the cytoplasmatic integrin tails (Figure 1.5a). These changes in the intracellular environment are, on their side, caused by external stimuli, such as integrin-collagen binding, P-selectin ligation or G-protein coupled receptor activation by cytokines. The generated signaling cascades induce changes in integrin tail conformations that facilitate the docking of intracellular activators and the subsequent conformational rearrangements of the receptor extracellular domain. This promotes the activation of integrin and increases its binding affinity for ECM ligands.[20]
Integrin activation can be regulated also upon extracellular stimulation. The main factors are the concentration of divalent cations, ligation and mechanical stress. The binding of ECM ligand to integrin induces conformational changes that allow the recruitment of intracellular proteins, including activators, various kinases and related adaptor proteins (Figure 1.5b). The consequence is an outside-in signaling cascade that governs diverse cellular processes including survival, proliferation, differentiation, migration and adhesion. Inside-out and outside-in signaling events are not independent and can occur simultaneously, reinforcing each other. Importantly ligand binding promotes integrin clustering and formation of focal adhesion. This happens thanks to the formation of large protein complexes and the organization of the cytoskeleton: talin, which is bound to the activated integrin β-tail, engages crosslinks firstly with actin, to initiate focal adhesion growth, and then with both actin and vinculin, to provide stable adhesion.[20,23] It is known



(a) Integrin inside-out signaling



(b) Integrin outside-in signaling

Figure 1.5: Representation of the "inside-out" and "outside-in" signaling cascade. Figure is reprinted from reference [20].

that inhibition of integrin focal adhesion has disruptive effects on cell adhesion, motility and migration, finally inducing apoptotic cell death.[3]
It must be noted that, even if generally the bound-inactive state of integrin is considered its default state, passively adopted by the receptor, several evidences suggest that both active and inactive state can be actively regulated. Indeed several molecules are now established as negative regulators of integrin activity. An appropriate integrin function should be then the consequences of a proper balancing between activators and inactivators.[19]

1.1.3 Integrins αIIbβ3, αvβ3 and α5β1 and their role as therapeutic targets

According to the possible combinations of the α and β chains, 24 integrin families exist in humans, all playing diverse biological functions and being involved in different pathological processes (Table 1.1). Among these 24 integrin subtypes, 8 belong to the group of RGD receptors, which share the ability to recognize their ECM ligands through the three-amino acids RGD motif; this thesis will focus on integrins αIIbβ3, αvβ3 and α5β1, all belonging to this subgroup. In the following I am going to describe their properties and their role as therapeutic targets.

αIIbβ3

Integrin αIIbβ3 is highly expressed on the platelets surface, at levels of approximately 50000-100000 receptors per platelet, being involved in blood coagulation events. It recognizes two different consensus motifs: the canonical RGD sequence, found in Fibrinogen, Fibronectin and Von Willebrand Factor; and the QAGDV sequence, which is present in the C-terminus of the γ subunit of Fibrinogen. Several evidences suggest that, although RGD-based peptides can compete with Fibrinogen in the binding of αIIbβ3, the biologically important site for αIIbβ3 recognition within Fibrinogen resides in its γ subunit.[14,24]Integrin αIIbβ3 is activated by inside-out signaling, caused by the binding of soluble factors in the blood (e.g. thrombin) to their respective platelet receptors. Activation of αIIbβ3 increases

Table 1.1: Therapeutic indications for all the integrins subfamilies.[6,47,48] This Table has been adapted from reference[46].

Integrins	Main Ligands	Potential Implications
RGD-receptors		
αIIbβ3	Fbg, vWf	Thrombosis, stroke, myocardial ischemia
α5β1	FN	Cancer, age-related macular degeneration
α8β1	Npn, FN, VN	None
αvβ1	VN, FN	Cancer
αvβ3	VN, Opn, vWf, FN, Fbg	Cancer, osteoporosis
αvβ5	VN	Cancer
αvβ6	FN, TGF-β1,-3	Fibrosis, transplant rejection, cancer
αvβ8	FN, TGF-β1,-3	Cancer
Collagen-receptors		
α1β1	Col	Fibrosis, cancer
α2β1	Col	Fibrosis, cancer
α10β1	Col	None
α11β1	Col	None
Leukocyte-specific-receptors		
α4β1	VCAM-1, FN	Multiple sclerosis, autoimmune, Crohn's disease, inflammatory bowel disease
α4β7	MAd-CAM-1	Multiple sclerosis, autoimmune, arthritis
α9β1	VCAM-1, Opn, VEGF-C,-D	Cancer
αLβ2	ICAM-1,-2,-3	Inflammation, psoriasis, stroke, ischemia, fibrosis
αMβ2	iC3b, Fbg	Inflammation, autoimmune
αXβ2	iC3b, Fbg	Inflammation
αDβ2	ICAM-3, VCAM-1	Inflammation
αEβ7	E-cadherin	Inflammation
Laminin-receptors		
α3β1	LN-5	None
α6β1	LN-1,-2	None
α7β1	LN-1,-2	None
α3β1	LN-2,-4,-5	None

[1]Abbreviations: Col, collagens; Fbg, fibrinogen; FN, Fibronectin; LN, laminin; Npn, nephronectin; Opn, osteopontin; VN, vitronectin; vWF, von Willebrand factor; ICAM, intercellular adhesion molecule; VCAM, vascular cell adhesion molecule; MAd-CAM, mucosal vascular addressin cell adhesion molecule; VEGF, vascular endothelial growth factor; TGF, transforming growth factor.

receptor affinity and avidity to the ECM proteins (Fibrinogen, Fibronectin and Von Willebrand Factor) and promotes platelets aggregation. Mis-regulated platelet aggregation is related to the formation of thrombi and can be involved in several pathologies as multiple Acute Coronary Syndromes, unstable angina, ischemic stroke and sickle cell anemia.[20,25] Compounds targeting αIIbβ3 integrin can thus act as anti-thrombotic agents. Currently, three FDA approved drugs targeting αIIbβ3 are available: i. Abciximab,[26] a Fab fragment of monoclonal antibody, which binds integrin in a site proximal to the RGD binding pocket, inhibiting ligand binding and platelet aggregation via steric hindrance; ii. Eptifibatide,[27] a synthetic cyclo-heptapeptide based on homoArgGD sequence and cycled with a disulfide bridge; iii. Tirofiban,[28] a non-peptide small molecule RGD-mimetic. Both Eptifibatide and Tirofiban compete with ECM proteins for the binding pocket of αIIbβ3 integrin. Large and randomized clinical trails tested these antagonists, showing their ability in preventing thrombosis and mortality. The major drawback associated with the administration of these antagonists is the increasing risk of bleeding events.
From a structural point of view, crystal structures with good resolution (2.9 Å) are available for αIIbβ3 integrin, both in its free and bound state.[10,14] In this thesis I will use as reference the crystallographic structure deposited in the PDB 2vdn, where Eptifibatide has been co-crystallized with αIIbβ3 integrin. A representation of the binding pose of this complex can be found in Figure 1.3.

$\alpha\mathbf{v}\beta\mathbf{3}$

Integrin αvβ3, as well as integrin αvβ5, are highly expressed on angiogenic endothelial cells, osteoclasts and a variety of cancer cells, including glioma, melanoma, breast, ovarian and pancreatic cancer.[29] The binding to ECM proteins (as Vitronectin, Fibronectin and Fibrinogen) occurs through the canonical RGD motif and promotes cell adhesion, migration and proliferation, thus favoring the formation of new vasculatures that feed the growing tumor. Additionally cell-platelet interactions, mediated by both integrin αvβ3 and αIIbβ3, facilitate the capture of circulating tumor cells from the blood, initiating the metastatic cascade.[20] It has been shown that antagonism of αvβ3 is able to inhibit angiogenesis and tumor

growth,[30] and that it can induce endothelial cell apoptosis.[31] For this reason several $\alpha v\beta 3$ antagonists are currently in clinical trials as antiangiogenic and antitumor agents. Among these, it is worth mentioning the humanized monoclonal antibody Vitaxin,[32] and the RGD-based cyclo-pentapeptide c(RGDf(*N*Me)V), known as Cilengitide.[33] However, these antagonists did not appear to be as successful as expected in clinical trials for cancer treatment. In particular c(RGDf(*N*Me)V), despite the promising results observed both in preclinical studies and in phase 2 trials, failed in phase 3 trial for the treatment of patients with newly diagnosed glioblastoma.[34] The reasons of this failure are still debated and will be addressed in section 1.2.2. Nevertheless, Cilengitide still continues to be investigated in clinical trials for the treatment of tumors other than glioblastoma. As well, antagonism of integrin $\alpha v\beta 3$ for the inhibition of angiogenesis and treatment of malignancies remains an attractive therapeutic target.
Also for integrin $\alpha v\beta 3$, crystallographic structures are available in both its free and bound states.[8,9] The early publication of these crystallographic structures could be among the reasons that made this integrin the most studied one in the field of tumor targeting. In this thesis I will use as reference structure the PDB 1l5g, in which c(RGDf(*N*Me)V) was soaked in crystallized $\alpha v\beta 3$. In Figure 1.3 the intermolecular interactions characteristic of this complex are represented.

$\alpha\mathbf{5}\beta\mathbf{1}$

In addition to $\alpha v\beta 3$ and $\alpha v\beta 5$, also integrin $\alpha 5\beta 1$ is over-expressed on the endothelium during tumor angiogenesis (with respect to quiescent endothelium) and is upregulated in several cancer cells, playing a role in tumor migration and invasion.[35] This integrin, in the ECM environment, primarily recognizes Fibronectin through the canonical RGD sequence, and it is often referred to as the Fibronectin receptor. It has been shown that antibody, peptide, and non-peptide antagonists of this integrin are able to block angiogenesis,[36] thus $\alpha 5\beta 1$ is currently targeted for cancer therapy. Even if $\alpha 5\beta 1$ has been far less studied with respect to $\alpha v\beta 3$, and despite the first appearance of its crystallographic structure is very recent, diverse ligands targeting this integrin have been developed. These include antibodies, RGD-like molecules and non-RGD-like peptides (the most famous one of this group

is ATN-161, based on the amino acidic sequence PHSCN[37]). The potentiality of these antagonists are explored either in preclinical studies, to clarify the integrin pathological behavior, and in clinical studies as potential antiangiogenic agents.[38] Among the interesting group of RGD-like small molecules we can cite SJ749,[39] a spirocyclic isoxazolines antagonist of $\alpha5\beta1$. The group of H. Kessler designed compounds analogue of SJ749, in addition to new original compounds, in order to find highly selective $\alpha5\beta1$-targeting molecules.[40,41] Among these, one compound (K34c) was shown to affect survival of glioma cells.[42] Also the pharmaceutical groups Jerini AG and AstraZeneca developed RGD-like molecules showing selectivity for $\alpha5\beta1$ compared to $\alpha v\beta3$.[43,44]
It is believed that the relatively recent elucidation of $\alpha5\beta1$ ectodomain crystal structure,[15,16] will strongly help the development of new potent and highly selective antagonists, increasing also the efforts in relative clinical studies.[38] In this thesis I will use as reference structure for this integrin the one deposited in the PDB 4wk4, which contains $\alpha5\beta1$ in complex with a disulfide-linked cyclopeptide ACRGDGWCG that was soaked inside the receptor (see Figure 1.3).

Integrins as therapeutic targets

In summary integrins are ideal pharmacological targets both for their accessibility, being positioned on the surface of the cells, and for their role in several pathological processes. For these reasons they have been for long time a main studied by pharmaceutical industries as potential therapeutic targets.
Currently only 3 out of the 24 known human integrins, are targeted therapeutically by antibodies or small molecules or peptides.[45] These 3 integrin subtypes are all found on blood cells (leukocytes or platelets) and include integrin $\alpha IIb\beta3$, belonging to the group of RGD-receptors, and the lymphocytes integrins $\alpha4\beta1$ and $\alpha4\beta7$, which are involved in multiple sclerosis, ulcerative colitis and Crohn disease. However, according to the list ClinicalTrials.gov, in 2015, in 80 clinical trials are involved therapeutics, imaging agents or biomarkers targeting integrins.[45]
In this context, it is clear that there is a remarkable interest in the development of novel integrin antagonists, which can be selective for one integrin or that can target simultaneously different integrin subtypes (dual or multiple antagonists).

The development of integrin ligands is also attractive for their utility as diagnostic agents and for their ability to deliver chemotherapeutics, nanoparticles or other therapeutic agents selectively to tumor.[20]

1.2 Peptides targeting integrins

Integrin functions depend on several factors, including ECM ligand binding, activation through conformational rearrangements, formation of focal adhesion and interactions with proteins present in the cytoskeleton. In principle, all these events could be inhibited in order to block integrin functionalities. This can be accomplished by targeting respectively: the ligand binding pocket, the site of allosteric control or the intracellular integrin tails that are responsible of cytoplasmatic interactions. All these strategies, even if with different extents, have been explored; however, it is interesting to observe that all the integrin antagonists that are currently on the market or in late-stage clinical trials, target the integrin binding site of ECM ligands. On the contrary, all the allosteric inhibitors developed for integrin resulted to be unsuccessful as therapeutics in clinical trials, suggesting that something is still missing in order to have a comprehensive view of integrin activation mechanisms.[45]
Among the therapeutics that target the integrin binding site, competing with natural ligands, three main classes can be individuated: i. therapeutic antibodies, which are currently the most abundant integrin targeting drugs; ii. small molecules (synthetic compounds with low molecular weight); and iii. peptide based drugs. In Table 1.2 the main advantages and drawbacks for each of these classes are summarized. As this thesis will focus on integrin-targeting peptides in the following I will briefly introduce the role of peptides in medicinal chemistry.

1.2.1 Peptides as therapeutics

The development of peptides as drugs has been under-exploited for a long time due to their unfavorable pharmacokinetic properties, which include short circulation

Table 1.2: Major advantages and drawbacks for the different classes of therapeutics.[20,49]

	Therapeutic Antibodies	Small Molecules	Peptide-based drugs
Advantages	High specificity and affinity	Low cost	Good specificity and affinity
	Low toxicity	Oral administration	Low toxicity
	Easy development	Easy to synthesize	Tolerability
	Long lifetime	Good stability	Intermediate cost
Disadvantages	High production cost	Poor selectivity	Short half-life
	Host immunogenicity and infusion reactions	Off-target side effect	Not orally available
			Low stability

half-life in the body, low stability in physiological conditions, poor membrane penetration and lack of oral bioavailability.[50] Additionally, among the criteria of the famous 'rule-of-five',[51] proposed in 2000 by Lipinski and widely followed in the process of drug development, it was stated that a molecule should have a molecular weight lower than 500 Da to favor bioavailability. This rule is violated by almost all the peptides with more than four amino acids, and this was a reason for which peptides were not considered as good potential drugs for pharmaceutical applications.[52]

However in recent years this trend has changed: small-molecules have clearly showed their limitations and peptides are therefore attracting increasing interest as plausible alternative.[53] Indeed it has been reported that the market of peptide and protein-based drugs is growing faster than that of other pharmaceuticals, and according to statistics the approval rate for peptides is twice that of small-molecules.[52,54] According to data of 2015, more than 60 peptide drugs are currently approved by US Food and Drug Administration, around 140 are in clinical trials and more than 500 in preclinical development.[49] Figure 1.6 represents this trend, showing the rising of peptide patent applications observed in the last years.

The reasons of this success for peptides are multiple. These include their high affinity and specificity, as well as their safe mode of action; the development of peptide rational design, which is able to improve the characteristic weaknesses of this compounds; and the increasing challenges encountered in the development of these small-molecules.[54] Peptides can combine the advantages of both small

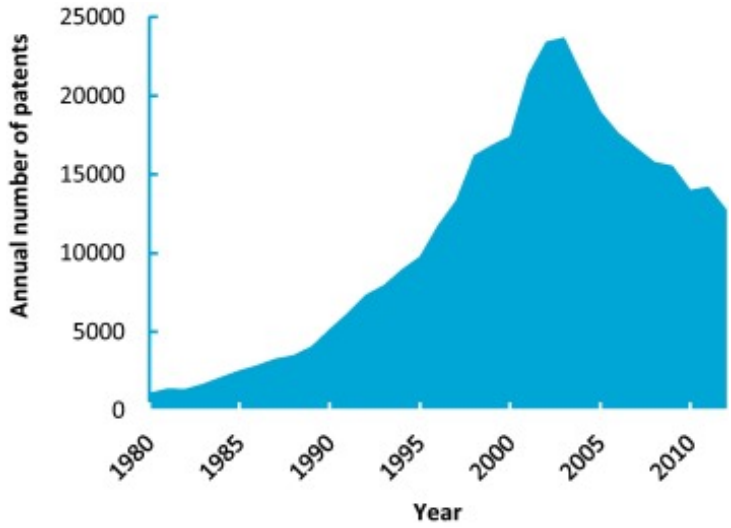


Figure 1.6: Trend in peptide patent applications from 1980 to 2012. Starting from 1996 the number of patent applications per year have always been over 10000. Figure is reprinted from reference [54].

molecules, characterized by typical molecular weights of <500 Da, and antibodies, which typically have molecular weights >5000 Da, filling the gap between these entities. They are characterized by the high target affinity and specificity of biologics, though preserving a small size; as consequence their production is cheaper and more accessible, approaching the manufacturing cost of small molecules. [49,52] An important role in this success was played by the rational design of peptide drugs, which has focused on the development of techniques able to mitigate peptides major drawbacks. A number of chemical modification strategies, that include cyclization, substitution or *N*-methylation of specific amino acids, have been used to improve their efficiency as therapeutics and to achieve oral bioavailability. [49,52] In particular cyclic backbone contributes to protection against enzymatic cleavage and buries the polar groups in the interior of the peptide; simultaneously, *N*-methylations and formation of intra-molecular hydrogen bonds can reduce the number of solvent exposed amide hydrogens, resulting in better oral bioavailability and enhanced membrane permeability. [56,58] Other strategies exploit the binding of peptides to the circulating protein albumin in order to increase their half-life in body. [49]
In this context, the seminal works of Kessler and coworkers gave advances in our understanding of the role of these structural features (i.e. cyclization and

N-methylation) on both pharmaceutical properties and biological activity.[57–59] Since peptide activity is strictly related to its three-dimensional structure, it is important to design chemical modifications that do not compromise the biologically active conformation of the peptide. From a structural point of view, it has been shown that cyclization can constrain peptides in specific stable active conformations, yielding to higher affinity compounds with respect to their linear counterparts.[57] Also *N*-methylation introduces significant changes in peptides structure, affecting the conformation of both the modified amino acid and of the preceding residue. This effect has a further long-range impact in the case of cyclopeptides and it is not easily predictable since it depends also on the specific sequence and chirality. Considering that these conformational changes can strongly influence the peptides biological activity and selectivity, cyclization and *N*-methylation can be exploited to fine tune their specificity towards particular targets and to improve their pharmacokinetic properties.[60,61]

All these data, including the unique properties of peptides in terms of potency and tolerability, together with the possibility to enhance their functionalities by introducing chemical modifications, support the idea that peptides, and in particular cyclopeptides, can represent a promising class of therapeutics to be used for the target of protein-protein interactions. Moreover, the potentialities of peptides are not limited to their role as drugs, and can be used for diagnostic and drug delivery purposes taking advantage of their possible targeting activities.[53,55]

1.2.2 RGD-based cyclopeptides

The three amino acidic sequence RGD is currently the fundamental motif of the majority of peptides targeting the integrins of the RGD-receptors subgroup. In the last decades a plethora of RGD-based peptides has been developed. Until 2002, the drug design process relied on the hypothesized pharmacophoric regions, derived only from the structures of known ligands. With the appearance of the first integrin X-ray structure alone and in complex with an RGD-based ligand,[8,9] structure-based rational design of selective inhibitors became possible.[62]

The developed peptides present different selectivity and affinity towards different

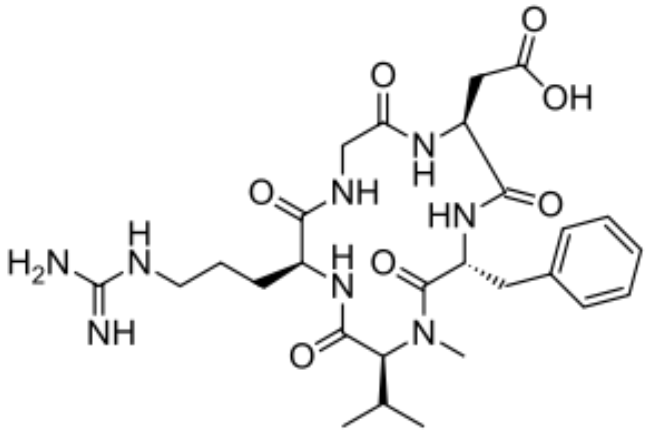


Figure 1.7: Structure of c(RGDf(*N*Me)V), known as Cilengitide.

integrins. Since the RGD motif is conserved in all the ligands, it is clear that integrin specificity is mainly dictated by the environment of the RGD sequence. The amino acids flanking the RGD motif can regulate the affinity both by direct interactions with integrin and by modulation of the conformational features of the RGD motif.[63,64] Cyclization is often utilized to constrain peptides in specific conformation: the conferred rigidity greatly improves selectivity towards specific integrin subtypes, additionally cyclopeptides are more stable being less susceptible to chemical degradation with respect to their linear counterparts. Furthermore introduction of D-amino acids and of chemical modifications have fostered the development of highly potent and selective RGD-cyclopeptides.[64,65]

Among these ligands the most famous one is c(RGDf(*N*Me)V), also known as Cilengitide, a cyclic RGD-based pentapeptide developed by the group of Kessler in the early 90s (structure in Figure 1.7).[66] This peptide displays a sub-nanomolar affinity for $\alpha v\beta 3$ integrin and has a 100-1000 fold increased activity over the linear reference peptide. The characteristic features of this molecules include: the cyclized conformation, the presence of a D-amino acid (D-Phenylalanine) and the backbone *N*-methylation of the flanking residue Valine. c(RGDf(*N*Me)V) was developed as antiangiogenic therapeutic by the pharmaceutical company Merck-Serono (Germany). It obtained excellent results in preclinical and early-stage clinical trials, showing safety profiles and no side effects in humans. It reached

phase III of clinical development for the treatment of glioblastoma and phase II for several other tumors.[67] However, in 2014, c(RGDf(*N*Me)V) failed in the treatment of glioblastoma, it did not reach the goal of increased progression-free or overall survival, and its further development as anticancer drug has been stopped.[34]
The reasons of this failure are still not fully understood. The main reason can be ascribed to the complexity of c(RGDf(*N*Me)V) mechanism of action when targeting $\alpha v\beta 3$ integrin. Indeed dose-dependent opposing effects of c(RGDf(*N*Me)V) have been observed: while at high doses it inhibits growth of blood vessels and tumor angiogenesis, at low doses this process seems to be stimulated.[68] The doses used in clinical trials might not be able to reflect the complexity of this effect. Additionally, the short half-life (2-4 hours) of c(RGDf(*N*Me)V) in body could have contributed to these negative results.[68] Finally, it must not be excluded the possibility of paradoxical effects. Indeed, the inhibition of one integrin could induce, via conformational changes, signaling in the intracellular environment; as consequence the upregulation of other related receptors, which bind the same ligands, could be promoted in order to maintain adhesion and signaling.[67,69] Overall, these data suggest that the role of integrin ligand as agonists or antagonists must be carefully considered.

1.2.3 isoDGR-based cyclopeptides

In 2006 biochemical studies showed that the NGR (Asparagine-Glycine-Arginine) motif of the ECM protein Fibronectin can spontaneously transform into the isoDGR (isoAspartate-Glycine-Arginine) sequence, resulting in a gain of function for the protein and in the creation of a new adhesion binding site for integrins.[70] The deamidation of Asparagine at the NGR site occurs via hydrolysis of the succinimide intermediate, as represented in Figure 1.8; this leads to formation of DGR and isoDGR in a 1:3 ratio.
Subsequent investigations demonstrated that isoDGR sequence can fit into the RGD-binding pocket of $\alpha v\beta 3$ integrin, recapitulating the canonical RGD/$\alpha v\beta 3$ contacts and establishing additional polar interactions.[71] This motif gained increasingly attention from 2012, when it was shown that, at variance to c(RGDf(*N*Me)V),

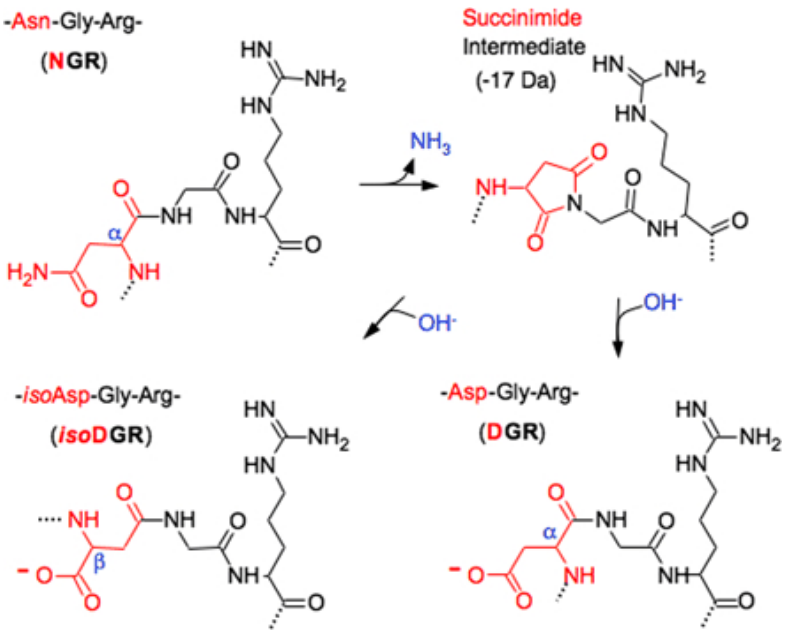


Figure 1.8: Representation of the NGR deamidation reaction, which leads to the formation of DGR and isoDGR. Figure is reprinted from reference [70].

CisoDGRC, an isoDGR-based peptide cyclized through a disulfide bridge, behaves as a pure integrin antagonist blocking $\alpha v\beta 3$ conformational changes and allosteric activation.[72] Since it cannot be excluded that the c(RGDf(NMe)V) failure in clinical trials could be related to its agonist-like activity, and the subsequent paradoxical effects, the intrinsic ability of the isoDGR motif to block receptor allosteric activation holds promise for improved real integrin inhibitors.
It was suggested that peptides containing the isoDGR motif could be exploited both as new integrin antagonists and as ligands for the targeted delivery of drugs or imaging.[73] Similarly to the case of RGD motif, structure-activity studies revealed that cyclization of isoDGR-based peptides improves ligand affinity for integrin; additionally, introduction of chemical modification (e.g. acetylation) and modification of the flanking residues are strategies that can be exploited to fine-tune affinity and selectivity towards specific integrin subtypes.[74,75]

In recent years the isoDGR-motif has been used to develop cyclopeptides and peptidomimetics. The group of Corti, which mainly contributed to the discovery the isoDGR motif as integrin binding sequence, designed a set of head-to-tail

isoDGR hexapeptides to be coupled with human serum albumin and nanoparticles via a free thiol group. Notably they found a peptide that, after conjugation with albumin, showed high affinity and selectivity for $\alpha v\beta 3$, and was able to inhibit tumor growth. All these characteristics can be favorably exploited for the delivery of nanomedicines to tumor neovasculature.[76] Kessler and coworkers investigated the role of flanking residues in cyclic pentapeptides based on the isoDGR motif, studying their affinity and selectivity for several integrins, including $\alpha v\beta 3$, $\alpha v\beta 5$, $\alpha v\beta 6$ and $\alpha 5\beta 1$.[77,78] Finally, cyclic isoDGR peptidomimetics containing a bifunctional diketopiperazine scaffolds have been developed and studied by the group of Gennari, leading to the individuation of two ligands characterized by low nanomolar activity for $\alpha v\beta 3$ receptor.[79]

Chapter 2

Methods

Computer simulations and modeling of molecular systems are widely used techniques to study the basic mechanisms of biological processes, including the motion and the assembly of proteins and molecules. The present chapter, divided in three sections, deals with the computational methods used in this thesis. In the first two sections the basic concepts of Molecular Dynamics (MD) simulations are presented, with a special focus on the problem of simulating rare events and on the use of enhanced sampling techniques to study equilibrium properties of systems involving complex free energy landscapes. The last section is dedicated to molecular docking, cluster analysis methodologies and MM-GB/SA (Molecular Mechanics-Generalized Born Surface Area) methods. All these techniques are useful to predict the binding mode and the affinity of biomolecular assemblies.

2.1 Molecular Dynamics

Molecular Dynamics (MD) simulations are currently very popular and routinely applied in the field of biochemistry to investigate the basis of protein structure, dynamic and function. Their first applications date back to 1950s and early 1960s, with pioneering works on the dynamics of liquids, but only in 1977 the first MD simulation of a biologically relevant macromolecule has been published by the

group of Martin Karplus.[80,81] While these earliest MD simulations were performed considering isolated proteins in vacuum, nowadays macromolecules can be simulated in a more realistic environment, consisting in a solvation box filled with water molecules and, optionally, ions. Simulations performed in these conditions, known as explicit solvent model, are currently the most utilized in studies on conformation and dynamics of biomolecules. In these decades the role of MD is growing faster and faster, thanks both to the increasing power of computational resources and the continuing advances in methodologies, making it possible to study bigger systems, longer time scales and larger conformational rearrangements.[82,83] An important acknowledgment to this field has been given with the 2013 Nobel Prize in Chemistry, which was awarded jointly to Martin Karplus, Michael Levitt and Arieh Warshel *for the development of multiscale models for complex chemical systems*. All these reasons are contributing in the emerging of MD as a fundamental complement of experiments in several contexts: the major applications of MD in biology include the study of conformational and allosteric changes in macromolecules, the description of protein folding, the prediction of binding free-energy and the investigation of mechanisms regulating molecular recognition.

MD simulations are based on the integration of Newton equations of motion and allow to follow the particles' position as a function of time, being able to provide information on the thermodynamic and kinetic properties of the investigated system. The link between microscopic data (i.e. distribution and motion of atoms) and macroscopic observables (e.g. pressure, heat capacity and free energies) is provided by statistical mechanics.[84] In this framework a physical quantity can be computed from MD by averaging over the microscopic states (configurations sampled in the trajectory) of the system, which are distributed according to a certain statistical ensemble. Assuming that the ergodic hypothesis holds, in the limit of infinite simulation time, the MD trajectory is able to explore all the accessible states: in this case time averages equals ensemble averages, therefore allowing the determination of thermodynamic properties.[80,85] This condition, that requires an exhaustive sampling of the accessible configuration space, is often prohibitive in the case of big systems, when dealing with complicated free-energy landscapes, or when studying biological processes that take place on long time scales. To alleviate

this problem, which is a major issue in MD, several methodologies capable to accelerate rare events, generally known as "enhanced sampling techniques", have been developed.

In the following I am providing more detailed information on the equation of motion used in MD and on the potential energy function that describes the inter- and intra-molecular interactions, also known as force-field. Afterward I will discuss the limitations of MD, specifically focusing on the problem of sampling rare events and presenting some methods developed to overcome this issue.

2.1.1 Equation of motion

In principle the motion of a molecular system is accurately described by the laws of quantum mechanics. However calculating the evolution of a system based on these laws requires a high computational cost, setting strong limitations to the size of the molecules and to the time-lengths that can be simulated. Hence classical MD, based on the numeric integration of the Newtonian equation of motion, is needed to overcome these drawbacks. This reduction implies mainly two approximations: i. the motion of electrons and nuclei can be treated separately due to their considerably different weight (as stated by the Born-Oppenheimer approximation); ii. the heavy nuclei can be approximated as classical point particles moving in a potential energy surface generated by the electron clouds (which indeed equilibrate quickly for each instantaneous configuration of the nuclei).[80] These assumptions allow to represent a molecule in terms of atoms, whose centers are represented by the nuclei, interacting through a potential energy that takes into account the effect of the electrons. In this framework, classical equation of motion can be used to follow the evolution of each atom.
Indicating the position vector of atom i at time t with $\mathbf{r}_i(t)$, its mass with m_i and the force acting on it with $\mathbf{F}_i(t)$, the Newton equation of motion can be written as:

$$\mathbf{F}_i(t) = m_i \frac{d^2\mathbf{r}_i(t)}{dt^2}. \tag{2.1}$$

Hence, the time-evolution of the system can be provided integrating equation 2.1 once it is given: i) the initial structure of the system (i.e. the set of coordinates for each of the N atoms at a reference time t_0), ii) a set of initial velocities consistent with the simulated temperature and iii) a potential energy function E_{tot}, for which $\mathbf{F}_i = -\nabla_{\mathbf{r}_i} E_{tot}(\mathbf{r}_1, \ldots, \mathbf{r}_N)$. Because of the complexity of biological systems, which generally consist of a huge number of atoms, finding an analytical solution for this equation is not possible. For this reason, numerical integrators which approximate the Taylor expansion at different orders, are used to propagate positions and velocities with a finite time step Δt. Several different integrators are available and are implemented in dedicated MD software packages (e.g. GROMACS, AMBER, NAMD and CHARMM).[86–89] The simplest integrator is the Euler one, which consists in a first-order approximation of the Taylor expansion of positions and velocities. Second-order algorithms include Verlet, velocity-Verlet and leap-frog integrators.[90–92] In Verlet, atomic position and acceleration at time t, as well as positions from the prior step $t - \Delta t$, are used to determine new positions at $t + \Delta t$. Differently, velocity-Verlet and leap-frog algorithms include explicitly the velocities $\mathbf{v}_i(t) = d\mathbf{r}(t)/dt$ in the calculations. Another family of algorithms, known as predictor-corrector methods, is available and allows to correct integration to a selected error order at the cost of extra computations and storage.[93] Within this thesis I will use the highly popular and versatile GROMACS package, which makes use of the leap-frog algorithm by default. In leap-frog integrator positions are updated at every time step Δt, while velocities are computed at intermediate times (this specificity is at the origin of the algorithm's name) according to the equations:

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \mathbf{v}_i(t + \Delta t/2)\, \Delta t, \tag{2.2a}$$

$$\mathbf{v}_i(t + \Delta t/2) = \mathbf{v}_i(t - \Delta t/2) + \frac{\mathbf{F}_i(t)}{m_i}\, \Delta t. \tag{2.2b}$$

The iteration of this procedure allows to follow the evolution of the system in time. From equations 2.2 it is clear the importance of choosing an appropriate value for the time step Δt. On one side it should be small enough to avoid significant deviation from the exact analytical trajectory, which corresponds to the limit of an

infinitesimally small time step, thus guaranteeing numerical stability and accuracy in conserving energy. On the other side it should be as big as possible in order to save computational time, allowing to sample longer trajectories and thus to simulate more relevant biological processes. Typically the time step Δt should be at least an order of magnitude smaller than the fastest time scale in the system (i.e. the oscillations of bonds involving light atoms as hydrogen). Since these bonds vibrations are generally of poor interest in the study of biomolecular system, it is possible to fix their lengths using constraining algorithms (as SHAKE or LINCS)[94,95] thus allowing to increase the time step up to a factor of four without affecting significantly the quality of the trajectory. In this condition, appropriate values for Δt in atomistic simulations are usually around 2 femtoseconds. It is worth noting that all the frequently employed integrators are time reversible, meaning that the direction of the simulation in time is arbitrary.

2.1.2 Thermodynamic ensembles

An aspect that must be considered is that Newtonian dynamics implies the conservation of energy. As consequence a dynamic which is based on the simple integration of the Newton equation would provide a set of configurations distributed according to the microcanonical ensemble NVE (corresponding to the conditions of constant number of particles N, volume V and energy E). This ensemble is generally not representative of common experimental or physiological conditions. More realistic situations can be simulated using the canonical ensemble (NVT, characterized by constant number of particles, volume and temperature) or the isothermal-isobaric ensemble (NPT, with constant number of particles, pressure and temperature). Thus, in order to mimic experimental or physiological conditions in MD, the use of a thermostat and/or a barostat along the simulation is needed.
It is worth noting that the goal of a thermostat is not to maintain constant temperature fixing the kinetic energy, but it is to keep the average simulation temperature close to the desired one. A common strategy to this aim consists in modifying or rescaling properly the atomic velocities of all the particles. Indeed the instantaneous temperature $T(t)$, which can vary due to the interconversion between kinetic and

potential terms of the total energy, is related to the kinetic energy via particles velocities as follows:

$$\frac{1}{2} N_f \, k_b \, T(t) = \frac{1}{2} \sum_i m_i \, v_i^2(t) \tag{2.3}$$

where N_f is the number of degrees of freedom, k_b is the Boltzmann's constant, m_i and v_i are the mass and velocity of atom i. Analogously, in order to maintain constant pressure, the volume of the box containing the molecule can be adjusted during the simulations, rescaling the atomic positions accordingly. Among the most famous algorithms used to control temperature there are Berendsen, Nosè-Hoover, velocity-rescaling and Langevin thermostats.[96–99]Popular barostats instead include Berendsen and Rahman-Parrinello barostat.[96,100]

2.1.3 Force field

As stated in section 2.1.1, in order to numerically solve the Newton equation of motion for a system of N atoms, it is necessary to know the potential energy $E_{tot}(\mathbf{r}_1, \ldots, \mathbf{r}_N)$, which is function of the $3\,N$ atomic positions and that determines the forces $\mathbf{F}_i = -\nabla_{\mathbf{r}_i} E_{tot}(\mathbf{r}_1, \ldots, \mathbf{r}_N)$ acting on each atom i at different time steps. Since this potential energy function must be evaluated several times during the simulation, it is advisable to maintain its form as simple as possible; for this reason anharmonic and cross-terms are often neglected and it is usually based on sets of empirical parameters. These parameters are derived and optimized by fitting to high-level Quantum Mechanics (QM) calculations and experiments (X-ray, electron diffraction, NMR spectroscopy or infrared spectroscopy), using as test-cases small molecules or fragments. The main assumption here is that these parameters are transferable and that they can be used to study larger and more complicated biomolecules of interest. The combination of the functional forms and of the associated set of parameters is called force-field. Despite their classic nature, force-fields performances are extremely good, being able to approach the high accuracy of QM calculations in a fraction of time. Several force-fields exist, differing both for the form of the energy potential and for the approaches adopted

to derive the parameters. Among these, the most popular and computationally efficient ones are additive and assign to each atom a fixed partial charge, which is not influenced by its electrostatic environment. In these force-fields the potential energy equation has the following form:

$$E_{tot} = E_{bonds} + E_{angles} + E_{dihedrals} + E_{vdw} + E_{coulomb}, \tag{2.4}$$

resulting from the summation of bonded interactions, namely the first three addends, and non-bonded interactions, i.e. the last two addends. Each term of 2.4 is calculated as:

$$E_{bonds}(r_{ij}) = \frac{1}{2}\, k^r_{ij}\, (r_{ij} - r^0_{ij})^2, \tag{2.5a}$$

$$E_{angles}(\theta_{ijk}) = \frac{1}{2}\, k^\theta_{ijk}\, (\theta_{ijk} - \theta^0_{ijk})^2, \tag{2.5b}$$

$$E_{dihedrals}(\phi_{ijkl}) = \frac{1}{2}\, k^n_\phi\, (1 + cos(n\, \phi_{ijkl} - \gamma_n)), \tag{2.5c}$$

$$E_{vdw}(r_{ij}) = \frac{A_{ij}}{r^{12}_{ij}} - \frac{B_{ij}}{r^6_{ij}}, \tag{2.5d}$$

$$E_{coulomb}(r_{ij}) = \frac{q_i\, q_j}{4\pi\epsilon_0 r_{ij}}. \tag{2.5e}$$

The terms E_{bonds} and E_{angles}, representing the bond stretching (2-body interaction) and the angle bending (3-body interaction), have the form of harmonic potentials, which maintain bonds and angles close to their equilibrium values, r^0_{ij} and θ^0_{ijk} respectively. While r^0_{ij} and θ^0_{ijk} values are derived from structural database, the force constants k^r_{ij} and k^θ_{ijk} generally result from fitting to experimental data.
The dihedral energy term $E_{dihedrals}$ approximates a Fourier series with a small number of terms; in this equation k^n_ϕ is the dihedral force constant, n is the dihedral periodicity and γ_n is a phase of the dihedral angle ϕ. This term is also used to enforce planarity of certain groups (e.g. aromatic rings) and to maintain proper chirality. Concerning the non-bonded interactions, the van der Waals (E_{vdw}) potential takes into account both the short range repulsion term and the attractive long-range term. The coefficients A_{ij} and B_{ij} depend on pairs of atom type; the distance between atoms i and j is indicated with r_{ij}. Finally, the Coulomb potential describes the electrostatic interactions involving two atoms i and j, characterized

by partial charges q_i and q_j, where $\epsilon_0 r_{ij}$ is a distance-dependent dielectric function. In the case of simulations incorporating explicit solvent representations the value ϵ_0 is usually set to unity.

The role of force-field in MD is of outstanding importance. Choosing an appropriate energy function with adequate parameters is indeed critical to have a reliable simulation. Several force-fields, sharing analogous mathematical forms but differing in the optimization of the parameters, have been developed and are routinely used in MD. Among the most popular force-field families it is worth to cite AMBER,[101] OPLS-AA,[102] CHARMM[103] and GROMOS,[104] for which several variants are available. All these force-fields use an all-atoms representation, with the exception of GROMOS that adopts an united-atom treatment, in which the parameters for non-polar hydrogens are not provided. Due to the importance of this issue, continuously updated and optimized force-field versions are published. In the last year OPLS-AA and AMBER have made available novel optimized force-field variants;[105,106] additionally new residue-specific force-fields are being developed and tested.[107,108] Even if numerous studies have been conducted to evaluate their reliability,[109–112] it is not possible to say which of these force-field variants is the best one. Indeed their performance strongly depends on the system investigated and on the information that one wants to retrieve from the simulation. In any case the guidelines originating from these studies can be helpful in the choice of a proper force-field variant.

Since most of the MD simulations are currently conducted in explicit solvent conditions, diverse models for parametrization of water molecules are also available. These include TIP3P, TIP4P, TIP5P and SPC.[113–115] The main difference among these models, which all make use of a rigid geometry, consists in the number of interaction points used to represent the water molecules. While TIP3P and SPC use three interaction points (representing the three atoms of the molecule), TIP4P and TIP5P use four and five interaction points respectively, by adding one or two dummy atoms with negative charge. The increasing of interaction sites generally improves the electrostatic distribution around the molecule, however it also increases the computational cost. As consequence three-site models, thanks to their high computational efficiency, constitute the most widely used models. To

save computational time it is also possible to use the implicit solvation model, in which the average action of water molecules is represented by a potential. This model is particularly efficient when dealing with huge systems but it can not be considered as accurate as the explicit solvent one.

Most of the force-fields herein cited have been primarily developed and optimized to deal with protein and nucleic acid systems, having only limited sets of parameters for organic molecules. However, in the field of drug discovery, where the study of protein-ligand and nucleic acid-ligand complex is fundamental, force-fields working for both biological and organic molecules are highly desirable. Several general force-fields have been developed to describe organic molecules: these include MMFF, Tripos, the CHARMM General Force Field (CGenFF) and the General AMBER Force Field (GAFF).[116–119] Specifically, the widely adopted general force-field GAFF, which has been developed in 2004 along with Antechamber,[120] a tool that allows the assignment of parameters to an arbitrary molecule in automatic fashion, has the advantage to be consistent both in form and parametrization with the traditional AMBER force-fields developed for proteins and nucleic acids. GAFF can be thus considered an extension of AMBER force field to be used for a wide range of organic molecules. This feature makes it suitable for the study of protein-ligand complexes. In any case one must consider that the accuracy of this kind of generic force-field is lower with respect to the standard ones due to the use of very general parameters.

2.1.4 Computing the interactions

It is worth noting that the computation of non-bonded interactions is the most demanding portion of an MD simulation. Indeed these terms must be calculated for each pair of atoms (separated by at least three covalent bonds): thus in a system of N atoms the number of non-bonded interactions to evaluate grows as N^2, often resulting in prohibitive calculations. To alleviate this problem, earlier MD simulations applied a cutoff truncation, where interactions beyond a chosen cutoff distance were ignored or smoothly switched to zero. While this technique introduces only small errors in the evaluation of van der Waals interactions, the electrostatic

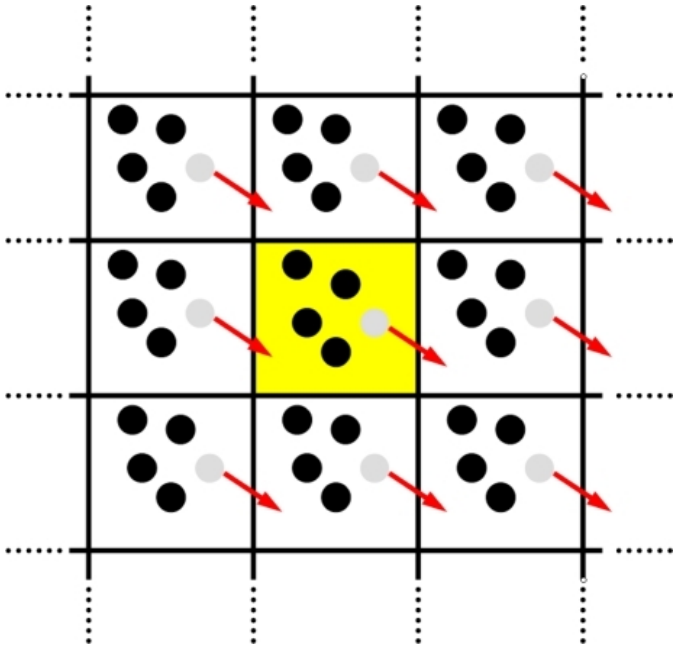


Figure 2.1: Schematic representation of periodic boundary conditions in two dimensions. If a particle moves out the simulation box an image particle comes to replace it. Figure is reprinted from reference [123].

potential is considerably affected, since it decays less rapidly with the distance. [84,121] Inclusion of longer range electrostatic interactions by simply increasing the cutoff distance is unfeasible, since the computational cost is dramatically increased. Currently, the most commonly used technique to compute electrostatic long-range interactions in MD is the Particle-Mesh Ewald method, an improvement of the traditional Ewald summation in which the long-range contribution is calculated as a Fourier transform. [122] This method allows to obtain good accuracy and reasonably fast computational time.

Particle-Mesh Ewald method is applicable in the case of infinite periodic systems and thus it is generally used in combination with the periodic boundary conditions (PBC) method, a technique able to minimize the boundary effects which inevitably affect a finite system. PBC technique consists in duplicating the box containing the system periodically in all directions. Practically only one box is simulated, while all the image particles move conjointly with the original ones (see Figure 2.1). In this framework a particle i in the box is able to interact not only with particle j in

the same box, but also with their images, virtually eliminating surface effects. The minimum image criterion, which states that among all the images of the particle j only the closer to i should be considered, guarantees that duplicate interactions are not included. In such a way the simulated system mimics the behavior of a macroscopic system, characterized by a huge number of particles (of the order of 10^{23}) and which is negligibly affected by interactions with boundaries.

2.1.5 Limitations of MD

Being aware of MD limitations is important to use simulations properly.
A first obvious limitation is related to the classical nature of the Newton equation of motion, which is the basis of MD time evolution. For this reason, all the processes involving quantum effects, as enzymatic or chemical reaction mechanisms, can not be studied by means of classical MD. To this aim hybrid quantum mechanical/molecular mechanical methods can be used, in which a restricted number of interesting coordinates is treated with QM while the residual coordinates are analyzed with MD.
A second limitation concerns the reliability of the force-fields. As already outlined, a simple mathematical form to describe the interatomic potential is advisable to save computational time; however this simplification requires the definition of a set of empirical parameters, which are critical for the fitting of the energy function to the potential acting on real atoms. The optimization of these parameters is carried out by fitting simulations' outcomes with both experimental data and QM calculations, using as data sets small molecules or fragments that are then used as building blocks. Additional refinement is usually performed comparing simulations with experimental results for well-characterized systems. Despite all the efforts dedicated to this aim and the huge improvements obtained, it is clear that having transferable empirical parameters, applicable on different systems under diverse conditions, is not an easy task. This challenge is becoming even more complicated: indeed, with the increasing power of computational resources, longer simulation times are becoming accessible, thus force-fields with improved stability and accuracy are required.

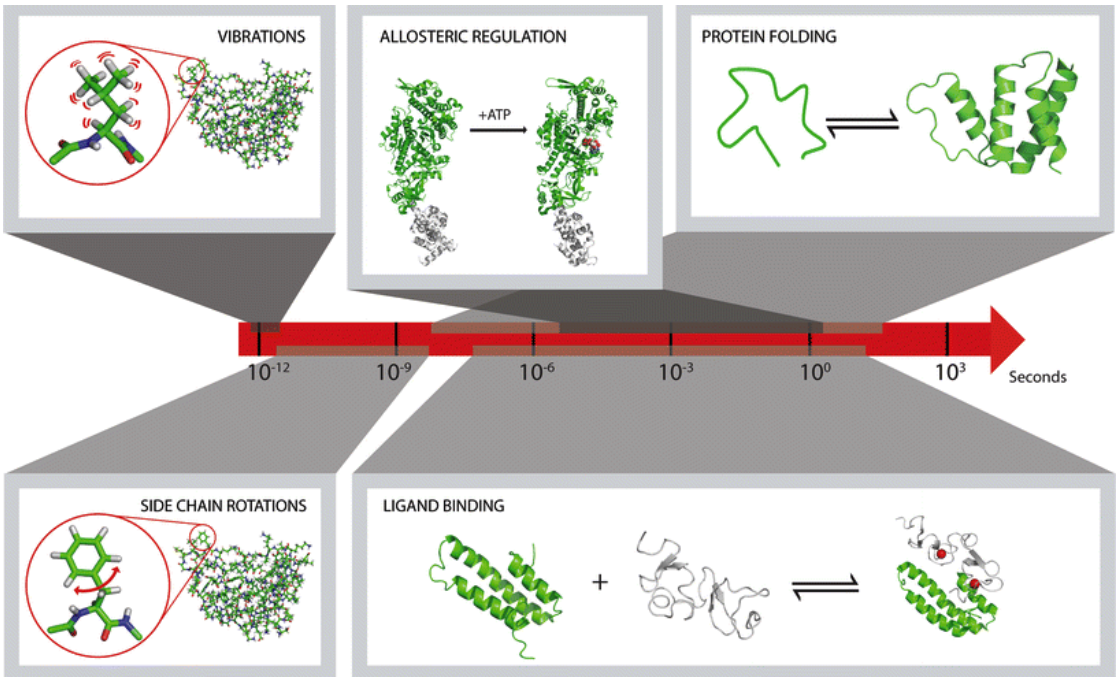


Figure 2.2: Typical time scales involved in protein dynamics. While the bonds vibrations, which set limits to the time step Δt, occur in a range of few femtoseconds, the most relevant biological processes are characterized by a time constant ranging from nanoseconds to hours. Figure is adapted from reference [124].

The most severe limitation in MD is probably related to the time-scales. As illustrated in Figure 2.2 the timescales linked to the most relevant biological processes are quite big with respect to the proper values for the time step Δt to be used in the integration of the Newton equation (usually these values are around 2 femtoseconds as explained in section 2.1.1). It follows that to simulate an event occurring on the scale of seconds, a number of time steps (and thus of integration of equation of motion and evaluation of the potential energy function) of 10^{15} is needed. Despite the enormous improvement in computational power observed in last years, this kind of simulation are commonly not affordable. In the following section I will discuss this problem, presenting few approaches, alternative to brute-force MD, that have been developed to overcome it. Specifically, I will focus on a method (Metadynamics in its Bias-Exchange variant), belonging to the group of the enhanced sampling techniques, which is widely adopted to accelerate rare events and to reconstruct the free energy of the system.

2.2 Enhancing the sampling with Bias-Exchange Metadynamics

Processes taking place on very long time scales are generally associated to complicated and rude free-energy landscapes, characterized by the presence of several local minima, in which the system remains trapped and that can be separated by high energy barriers (see Figure 2.3). In a finite MD simulation at standard ambient temperature the probability of overcoming these energy barriers can be negligible, even in several hundreds of nanoseconds. Consequently, simulating processes associated to major conformational changes or protein folding, or generally events characterized by complex free-energy landscapes, requires a huge amount of MD steps and computational time.[126] Several strategies have been explored to overcome this obstacle.

To this aim, two main classes of methodological approaches can be identified: coarse-grained methodologies and enhanced sampling techniques.[127] Coarse-grained methods rely on a simplification of the investigated system (obtained reducing a group

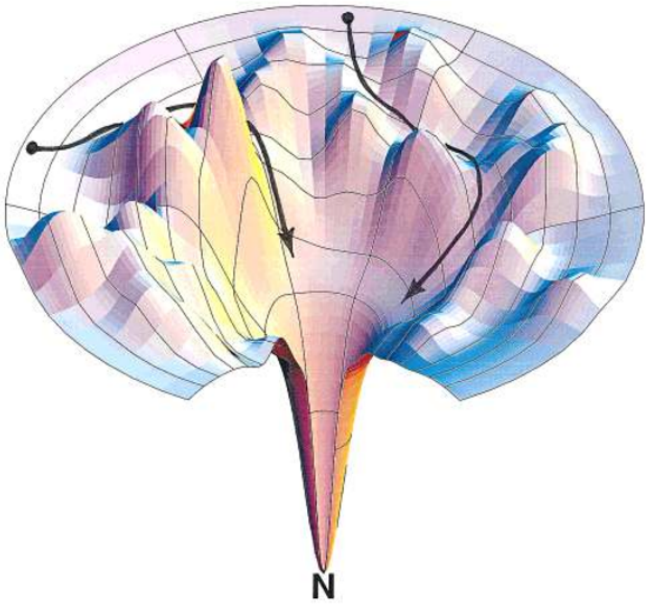


Figure 2.3: Representation of a typical rude free energy landscape characterized by the presence of several local minima. Figure is reprinted from reference[125].

of atoms to a single particle which is supposed to represent their properties) to speed up the calculation. Even if these approaches have been widely and successfully applied,[128] they lack atomic details and therefore are not suitable for the investigation of several phenomena. On the contrary enhanced sampling techniques maintain the atomic description of the system trying to enhance the sampling by accelerating the dynamic of the system. The most famous and applied methodologies include simulated annealing,[129] replica-exchange molecular-dynamics,[130] umbrella sampling[131] and Metadynamics (MetaD).[132,133] While the first two approaches are based on the idea of raising simulation temperature to overcome easier the free-energy barriers, the last two methods involve the addition of a potential bias to improve the sampling. In the following I will briefly review MetaD technique, and specifically its variant Bias-Exchange Metadynamics (BE-META),[134] which has the advantage to directly provide a good estimate of the free energy of the system as a function of few relevant degree of freedom, named Collective Variables (CVs).[85]

2.2.1 Metadynamics

MetaD is a well recognized method to accelerate rare events, which has proven its effectiveness in structural biology context. This technique allows to: i. overcome high free-energy barriers escaping the local minima, ii. reconstruct the free energy landscape of the investigated system as a function of few CVs, which should describe all the slow events relevant to the process under investigation.[132,133] MetaD is based on a history-dependent potential, defined in equation 2.6, that is added to the traditional force field potential forcing the system to visit unexplored regions. The history-dependent potential biases the dynamic of the system in a reduced dimensionality space, defined by a set of few CVs $S(x)$, which must be selected before starting the simulations. At fixed time intervals τ_G a Gaussian centered along the trajectory of the CVs is added according to:

$$V_G(S(x),t) = w \sum_{t'=\tau_G,\, 2\tau_G \ldots\, t'<t} e^{-\frac{(S(x)-s(t'))^2}{2\,\delta^2}}, \tag{2.6}$$

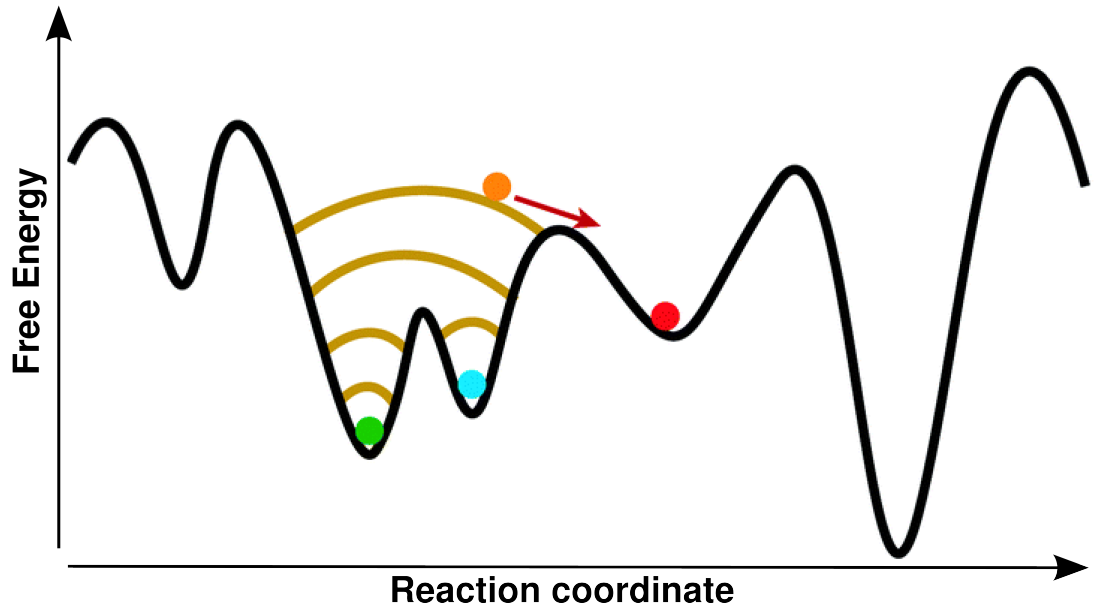


Figure 2.4: Illustration of the MetaD procedure, adapted from reference [135]. Dots represent structures, while yellow lines are representative of the biasing potential.

where $s(t) = S(x(t))$ is the value of the CV at the time t, w and δ are the height and the width of the Gaussians, respectively. The accumulation of these Gaussians along the simulation leads to the flooding of the free-energy minima, as illustrated in Figure 2.4. In such a way the system is encouraged to escape the local minima and to visit unexplored region in a reasonable simulation time. The main assumption of MetaD is that the sum of the Gaussians provides the opposite profile of the free energy, $F(s)$, once the convergence has been reached and after an initial transient time, which is necessary to fill the free energy minima:

$$\lim_{t \to +\infty} V_G(s,t) \sim -F(s). \tag{2.7}$$

Of note, the estimation of the free energy does not converge to a definite value, but fluctuates around the correct result with a mean error being proportional to the square root of w/τ_G.[136,137] Interestingly several variants of MetaD have been developed in the last decade, among which it is worth to cite the widely adopted Well-Tempered Metadynamics (WT-MetaD) and Bias-Exchange Metadynamics (BE-META). In the first variant, WT-MetaD, the height of the Gaussians w

decreases with time according to the equation $w = w_0\,\tau_G\,e^{-V(s,t)/\Delta T}$, where w_0 is the initial bias deposition rate, τ_G is the time interval at which Gaussians are deposited and ΔT is a temperature.[136] Of note, this adaptive bias allows a smooth convergence to $\lim_{t\to+\infty} V_G(s,t) = -\frac{\Delta T}{\Delta T+T}F(s)$, enabling a rigorous estimation of convergence and errors. In this framework, where the limiting cases $\Delta T = 0$ and $\Delta T \to \infty$ correspond to canonical sampling and MetaD, respectively, the parameter ΔT can be fine-tuned to increase the free-energy barriers crossing, meanwhile restricting the exploration of the CV space to physically interesting regions. The BE-META variant, instead, is a technique suitable to compensate the problems related to the proper choice of CVs.[134] It is indeed well known that the reconstruction of the free energy in MetaD is accurate only when: i. all the relevant variables, related to the slow events, are included, ii. only a limited number of CVs is used (usually not more than three since, for uncorrelated variables, the time needed to rebuild the free energy surface scales exponentially with the number of CVs).[137] This fact constitutes a major limitation when more CVs are required or when it is difficult to select a priori a limited number of variables describing the process: by combining replica exchange methods and MetaD, BE-META aims to overcome this issue. Since this technique has been widely adopted within the thesis, a more detailed description is following.

2.2.2 Bias-Exchange Metadynamics

BE-META is a combination of replica exchange techniques[130] and MetaD, whereby multiple MetaD simulations (N) are performed at the same temperature T. Each replica of the system is biased with a monodimensional potential acting on a single CV S_k, with $k = 1, ..., N$ (actually it is possible to bias more than one CV in each replica, but this option is not considered in this thesis). Exchange between pairs of replica are attempted at fixed time intervals. Considering two replica described by coordinates x_i and x_j, with bias potentials $V_G(S_a(x_i),t)$ and $V_G(S_b(x_j),t)$, an exchange is accepted with a probability P_{ab} according to:

$$P_{ab} = min(1, e^{\frac{V_G(S_a(x_i),t)+V_G(S_b(x_j),t)-V_G(S_a(x_j),t)-V_G(S_b(x_i),t)}{k_b\,T}}) \tag{2.8}$$

where k_b is the Boltzmann constant and T the simulation temperature. These exchanges enhance the ability of the system to efficiently explore a complex multi-dimensional free energy landscape, aiding the system to cross the barriers which are orthogonal to the reaction coordinates.

2.2.3 Free-energy reconstruction from BE-META trajectories

The output of the simulation consists in several mono-dimensional projections of the free energy surface along each CV. To assess the convergence, it must be verified that the bias potential grows parallel to itself: in this condition the force from the bias cancels out the average natural force (free-energy gradient) on the collective variable and the bias potential $V_G(S(x), t)$ of each replica is stable, being an unbiased estimator of the free-energy. Hence, analysis of BE-META trajectories allows to reconstruct the multidimensional free energy landscape of the system in a reduced CVs space (a subset of all the initial CVs can be used to this aim). This can provide information about its equilibrium properties.
To allow the free-energy reconstruction, the reduced CVs space is arbitrarily divided in hypercubes, called microstates. Each structure sampled during the simulation is assigned to a given microstate according to its CVs values, whereby the free energy of each microstate is estimated exploiting the mono-dimensional free energy profiles obtained from each replica by a weighted histogram analysis (WHAM) approach.[131,138] The reliability of free energy estimation strongly depends on microstate assignment, which should satisfy the following requirements: a) to densely cover the configuration space explored during the simulation, including also the barrier region; b) to be sufficiently populated to guarantee a good statistic; c) the structures belonging to the same microstate should satisfy a similarity criterion to allow for consistency.[139] To satisfy all these criteria care must be taken in the choice of the reconstructed CVs space dimensionality and of the size of the microstate.
Additionally, it is possible to cluster microstates into different metastable states, in order to individuate the most significant local free energy minima. To this aim

an approximate rate matrix among microstates is constructed and its spectrum is analyzed as described in references[140,141]. The inverse of the obtained eigenvalues provides the relaxation times (τ) of the system, which are related to the transition time between different metastable states. The number of relaxation times determines the number of metastable states composing the microstates space. In order to consider only the relevant relaxation processes of the system, it is possible to choose an appropriate number (i) of relaxation times, when a gap between τ_i and τ_{i+1} is present.[140] Finally, performing a commitor analysis, each microstate is assigned to a metastable state, which corresponds to one of the main minima of the system. All these analysis can be performed with the help of the Metagui tool.[141]

2.3 Docking techniques

Most of the processes occurring in the cells involve the interaction between two or more biological systems. The molecular characterization of these recognition processes, as well as the ability to predict possible interactions among molecules, are crucial elements for a deep comprehension of biological mechanisms, for diseases research and for the development of drugs.[142,143] These non-trivial tasks can be achieved computationally by means of molecular docking techniques, a powerful methodology able to predict the three-dimensional structure and the binding affinity of a molecular complex starting from the knowledge of the coordinates of the unbound components (which can be proteins, nucleic acids, peptides or small molecules). The initial unbound structures may come from crystallographic and NMR experiments, as well as from homology models; in the case of small-molecules or ligand, the input conformation can be also created *de novo*.[144]
During the docking procedure several possible complex models, which are commonly known as decoy poses or binding modes, are generated. Since the molecular recognition process can be associated to a free-energy landscape, all these poses represent different local minima: the aim of docking is to individuate the global free-energy minimum corresponding to the native binding mode. This procedure involves two basic steps: sampling and scoring. The sampling step consists in a search algorithm that allows the exploration and generation of several possible binding

modes, varying the conformation and the orientation of the ligand in the binding site of the receptor. The scoring procedure involves the use of some quantitative method (referred to as scoring function) to rank the generated decoy poses and to distinguish native-like poses from non-native ones. These two highly interrelated tasks, which are generally iterated until convergence to a solution of minimum energy, can be treated at different levels of accuracy in diverse docking algorithms. Since they are the basis of docking procedures a more detailed presentation will follow in this chapter.

It is worth noting that the goals of molecular docking can go beyond the prediction of a reliable binding pose and of the associated chemical intermolecular interactions. Several algorithms indeed aim also to individuate active molecules from large compounds libraries, to rank ligands in a database and to give an estimation of binding affinities towards a specific target.[145] Thanks to all these features, the docking programs have gained increasingly importance and are currently used for a wide range of applications. In the field of computer-aided drug design & development docking software are exploited both in virtual screening campaigns, to individuate novel molecules targeting proteins of therapeutic interest or nucleic acids, as well as in the phase of lead compounds optimization.[146] The description of molecular interactions underlying protein-protein or protein-ligand binding is also useful in the context of protein engineering, to identify *hot* residues which could be mutated to favor or discourage binding.[147,148] Additionally, *in silico* analysis of protein mutations and of their influence on ligand binding is helpful to provide the basis for the development of drug-resistant ligands.[144] Another possible applications of docking techniques consists in the elucidation of the molecular mechanisms of selectivity, which can be achieved by docking the same molecule into different receptor targets.[149]
This wide variety of needs is reflected by a great number of software packages devoted to perform molecular docking, which can differ for several factors including scoring function, exhaustiveness of the sampling and speed.

In the following, after an introduction about the main challenges in the docking field, which include the sampling and scoring problems, I will present two popular docking software exploited in this thesis: Glide,[150–152] a program distributed by

the Schrödinger company (Schrödinger, LLC, New York), which offers diverse ligand docking options ranging from virtual screening to more accurate protocol; and HADDOCK,[153] an information-driven docking software, developed and freely distributed by the Bonvin group, that was initially designed for the prediction of protein-protein interactions but that has been recently adapted to handle also the docking of ligands and peptides.[154,155] Finally, I will introduce how cluster analysis methods can be used to better analyze and interpret the decoy poses.

2.3.1 Sampling and Scoring

As already outlined, docking is based on two main stages: a sampling procedure, which generates a wide variety of possible binding modes, and a scoring step that aims at individuating native or near-native poses. Even if it is helpful to make a clear distinction between searching and scoring methodologies, the two tasks are strongly interrelated. Docking programs indeed make use of them iteratively and often a scoring-function is used to guide the generation of the poses.[146,156] Usually rough scoring-functions, which give priority to speed with respect to accuracy, are used to guide the search algorithms, when a huge number of poses must be considered; more accurate scoring-functions are instead applied for the ranking of the final poses. The accuracy of the docking depends on both these stages, specifically on the exhaustiveness of the sampling, which ideally should be able to generate the native binding mode, and on the reliability of the scoring-function, which should be able to rank as first the correct binding mode. Obviously, higher accuracy implies higher computational costs and thus a lower number of compounds that can be analyzed in the same time. Depending on the needs different search algorithms and scoring-functions can be used: in the following I will present a brief description of the most used ones.

Sampling

The binding process involves changes in the respective position and orientation of the binders, but also entails conformational rearrangements that can occur

in both the receptor and the ligand. Computing exhaustively all the possible binding modes, considering also molecular flexibility, is extremely expensive from a computational point of view. To handle this issue docking algorithms can adopt mainly three strategies with different degrees of exhaustiveness: i) rigid docking, in which both the receptor and the ligand are treated as rigid bodies and only the six rotational and translational degrees of freedom are explored; ii) rigid receptor and flexible ligand, in which also the conformational degrees of freedom of the ligand are sampled; iii) fully or partially flexible protein with flexible ligand, where all the degrees of freedom of the ligand are considered and few or all the relevant conformational degrees of freedom of the protein are also investigated.[145] Currently, to have good efficiency, balancing speed and accuracy, the most popular docking method treats the ligand as flexible and the receptor as a rigid-body.[157]

The sampling is performed by applying incremental modifications in the structural parameters of the ligand, which include the translational, rotational and conformational (e.g. torsional angles) degrees of freedom. Mainly two methodologies can be adopted to perform the sampling: systematic and stochastic. In the first case, by performing gradual variations in each structural parameters, the algorithm explores the free-energy landscape associated to the conformational space until convergence to a solution of minimum energy.[158] Even if this method is effective in finding free-energy minima, it is clear that its computational cost increases when handling highly flexible ligands: in this situation indeed the number of combinations of structural parameters would enormously increase, leading to a phenomenon known as combinatorial explosion. Conversely, in stochastic methodologies the structural parameters of the ligand are randomly changed at each step, allowing for the generation of a wide variety of possible solutions. Stochastic algorithms, including Monte Carlo, Genetic Algorithms, Tabu Search and Swarm Optimization, can be employed to accept or reject the proposed solutions according to a probabilistic criterion, in order to limit the computational cost of the procedure. The main drawback of this methodology is that there is no guarantee of convergence to the global minimum, thus multiple independent runs should be performed to increase the probability of reaching the optimal solution.[145]

While ligand flexibility is often included in algorithms by sampling its possible

conformations on the fly, receptor flexibility still constitutes one of the major challenges in the docking field. This aspect, despite being often overlooked, plays an essential role in the process of molecular recognition: indeed the receptor commonly experiences structural rearrangements upon ligand binding due to the induced fit effect. These changes can range from small variations, limited to the region of binding site, to more significant conformational rearrangements, involving elements of secondary and tertiary structure.[158] Additionally, it should be considered that the free receptor could exist in multiple conformational states rather than as a unique native state.[159] Even if an exhaustive treatment of this aspect is often out of the reach, diverse methods can be adopted to deal with this issue. These can be divided in five classes, which differ in the degree of exhaustiveness and accuracy, according to the classification of references[145,160]:

i) soft-docking, where the repulsive contribution of the van der Waals potential is softened to permit small atoms overlaps, allowing the ligand to accommodate more easily in the binding pocket. This method is fast but it is adequate only when small local receptor motions occur.

ii) Side-chain flexibility, where diverse side-chain conformations are sampled while backbone is kept fixed. Also this approach can be employed only for local motions of the receptor.

iii) Molecular relaxation, usually performed in post-processing steps using Monte Carlo or MD minimization. It is useful to optimize the complex structure and to evaluate its stability.

iv) Ensemble docking, in which the ligand is docked in multiple receptor conformations generated before the run. These putative structures can come both from NMR and crystallographic experiments or from computational models (e.g. MD and molecular modeling). Even if this method seems to be quite promising, a major drawback is constituted by the lack of protocols that help in *a priori* selection of an optimal subset of protein structures.[161] Additionally also this method fails when large receptor structural rearrangements occur.

v) Collective degrees of freedom approaches, that aim at considering the full flexibility of the receptor by reducing its high-dimensional conformational landscape in a representation that captures only the dominant motion modes. These results can be achieved by using techniques such as normal mode analysis or principal

component analysis. Its use is limited in several applications because of the high computational cost required.

Scoring

The main goal of a scoring-function is to individuate native or near-native poses among the huge amount of binding modes that are generated during the sampling step. Additionally, they aim at discriminating active from inactive ligands and at estimating the binding energies of diverse compounds, possibly ranking them according to their affinity.[145] The form of a scoring-function is a mathematical equation that takes into account the physical properties of the two interacting molecules. To be able to treat several compounds in a reasonable computational time, diverse assumptions and simplification are usually adopted. Some scoring-functions merely consider the properties of the non-covalent ligand receptor interface (e.g. salt bridges, hydrogen bonds, hydrophobic contacts).[146] Conversely, more complicated and time-consuming functions take into account also desolvation contributions and entropic effects.[158,162]

In this context, diverse scoring-functions, offering different accuracy vs speed balances, have been developed. Mainly three classes can be individuated: force-field-based, empirical, and statistical scoring-functions.[163]
Force-field-based approaches rely on a molecular mechanic formalism, where the estimation of the binding energy is computed as the sum of bonded (bond stretching, angle bending and torsional energy) and non-bonded terms (electrostatic and van der Waals interactions). This kind of functions can be extended by introduction of terms accounting for desolvation and entropic energies.[157] A major limitation of these methods is the high computational cost required, due to the relatively complicated functional form. Additionally, it has been observed that these scoring-functions often overestimate interactions between charged atoms.[145]
A second class of scoring-functions is represented by the empirical methods, which are based on the concept that it is possible to relate the binding free energy with a weighted sum of unrelated variables.[145] As in the case of force-field-based approaches, the scoring-function is a sum of terms describing the general features

of the complex (e.g. hydrogen-bonds, hydrophilic and hydrophobic interactions, desolvation and entropic effects). However these terms have a simplified functional form and are weighted with proper coefficients, which have been optimized to reproduce known binding affinity data of experimentally determined complexes. The parametrization of these coefficients is usually conducted with regression or machine learning approaches.[144] While these functions allow a faster scoring than the force-field-based ones, their accuracy is strictly related to the quality of the training set and it is limited by the over-simplification of some physical interactions. Finally, there are statistical (also called knowledge-based) potentials, which rely on a statistical analysis of experimentally determined receptor-ligand structures. The main assumption here is that frequent contacts, appearing in several complexes, correspond to favorable interactions.[157] These frequency distributions are thus used to construct an interaction potential that is function of the distance between atom-types. The relationship that links the frequency of a contact and the associated energy term is provided by the inverse-Boltzmann equation. These scoring-functions are fast and have the advantage that, in their interaction potential, are included also those features that can not be easily modeled explicitly.[146] However, the accuracy of this method strongly depends on the number and diversity of the training set used to create the potential.
To improve the performance of these scoring-functions, which are all affected by some limitations, it is possible to use a consensus scoring approach. In this strategy diverse docking software and functions are used and the scores obtained are combined through a consensus scheme. It has been shown that this approach provides notable improvement in terms of accuracy.[164]

Currently, most scoring-functions show good performance in the identification of native or near-native binding modes, however it is still challenging to achieve a reliable estimation of the free energy of binding, as well as a proper ranking of different compounds toward a target receptor. Indeed in the scoring functions developed to date, it has been observed a weak correlation with experimental binding affinities and a diversity in the performance depending on the target.[161] These difficulties in reproducing experimental data can be ascribed to several reasons, which include a poor quality of the input structures, a rough treatment of

long-range interactions and receptor flexibility and particularly an inappropriate handling of solvent and entropic effects.[147,165]

Even if it is recognized that conformational entropy plays a relevant role in the determination of binding free energy, its contribution is still neglected or oversimplified in most of the docking software. This is due to the enormous computational cost required to its rigorous quantification, which is not practicable especially for big systems as protein-ligand complexes.[145] Conversely, some methodologies have been developed to deal with solvent effects. In this context, a special attention is given to structural waters, which are present in approximately 65% of the crystallographic complexes mediating receptor-ligand recognition.[144] It has been shown that the inclusion of structural waters in docking calculations allows the formation of hydrogen-bonds networks between the ligand and the binding pocket, leading to a significant improvement in accuracy.[144,158] In addition, some methods that explicitly account for water molecules during the docking process have been proposed,[166–169] but they are computational expensive due to the large number of degrees of freedom associated to solvent molecules. To save computational time it is possible to treat water molecules implicitly, representing them as a continuum dielectric medium. The most widespread approaches based on this strategy are the Molecular Mechanics Poisson-Boltzmann surface area (MM-PB/SA) and the Molecular Mechanics generalized-Born/surface area (MM-GB/SA) models.[170–172] According to these techniques the solvation contribution can be computed considering separately the polar and non-polar terms. The first term, which refers to the energy needed to transfer the solute from a continuum medium with a low dielectric constant ($\epsilon = 1$) to a medium with the dielectric constant of water ($\epsilon = 80$), is computed by solving either the Poisson-Boltzmann equation or by using the simplified Generalized Born model. Specifically, in this last model the solute is represented as a set of spheres characterized by a different dielectric constant with respect to the external solvent. The non-polar contribution instead, which aims to include an approximation of hydrophobic-hydrophobic interactions, is considered proportional to the solvent accessible surface area (SASA). These approaches are usually employed in a post-processing phase to rescore the previously generated docking poses, with the aim to improve the accuracy of ranking and binding energy predictions. Of note, MM-PB/SA and MM-GB/SA have been successfully

employed to estimate the free energy of binding in diverse protein–ligand interactions.[173,174] They usually show good correlations with experimental data, being a good compromise between efficiency and efficacy.

2.3.2 Glide software

Glide (Grid-based Ligand Docking with Energetics) software is a docking program distributed commercially by the Schrödinger company. It makes use of a systematic conformational search and of a scoring-function that relies on the combination of empirical and force-field-based terms. Specifically different protocols, with different accuracies, can be used in this framework: i) Standard-Precision Glide (SP), which uses a *soft* scoring function with the aim of identifying a wide pool of possible binders minimizing false negatives and ii) Extra-Precision Glide (XP), characterized by a *harder* scoring-function, which attempts at minimizing false positives by inclusion of severe penalties.[150–152] This last approach is more demanding than the SP mode but it is useful during leads optimization or when few compounds to be synthesized must be individuated.
During the sampling step, hierarchical filters are used to approach an exhaustive systematic search of the conformation, position and orientation of the ligand within the receptor binding pocket. At a first stage, several ligand conformations are generated by exploring its torsional angle space and the ones with lower torsional energy are selected. Secondly, all the possible positions and orientations of these conformers on the active site of the receptor are screened. To allow a fast evaluation of all these placements, the energetic-like properties of the protein are represented on a grid, assigning precomputed scores (derived from a discretized version of the empirical ChemScore function[175]) to boxes of $1\,\text{Å}^3$ dimensions. This kind of score penalizes steric clashes and is able to recognize favorable hydrophobic interactions, hydrogen-bonds and metal-ligation interactions. Afterwards, starting from the best poses (typically few hundreds) herein individuated, the ligand is minimized in the field of the receptor, using a molecular mechanics scoring function (OPLS-AA) and a multi-grid strategy: in order to guarantee more accuracy the side-dimension of the boxes, which store the Coulomb/van der Waals fields of the receptor, are

gradually decreased in the area where the two molecules are in contact. Finally, the best three to six lowest-energy decoys are minimized with a Monte Carlo approach: this last phase is critical to have accurate docking poses.
In the case of Glide XP mode, the sampling method is more extensive and it is based on an *anchor and refined growth strategy*. The set of poses derived from the SP procedure are exploited to identify ligand anchor fragments, which are then used to reconstruct the whole molecule by growing one side-chain at time. Several cycles of minimization and XP scoring are run to guide the growing, in order to reduce the scoring penalties and to refine the best poses. As a result, this algorithm is able to locate an increased number of penalty-free poses.
Both in SP and XP modes the flexibility of the receptor is not considered, however scaled van der Waals radii of selected atoms can be used to allow an easier accommodation of the ligand in the binding pocket.

The Glide scoring function (GScore), which is an extension of the empirical ChemScore function, can be described as follows:[150]

$$\begin{aligned}\text{GScore} = {} & C_{lipo-lipo} \sum f(r_{lr}) + C_{hbond-nn} \sum g(\Delta r)\, h(\Delta\alpha) + \\ & + C_{hbond-nc} \sum g(\Delta r)\, h(\Delta\alpha) + C_{hbond-cc} \sum g(\Delta r)\, h(\Delta\alpha) + C_{metal-ion} \sum f(r_{lm}) + \\ & + C_{rotb} H_{rotb} + C_{polar-phob} V_{polar-phob} + C_{coul} E_{coul} + C_{vdw} E_{vdw} + solvation\ terms.\end{aligned} \tag{2.9}$$

In this equation the summations are extended over all the ligand-receptor (lr) atoms pairs that are defined as lipophilic (first term), which are involved in hydrogen-bonds (second to forth terms) or that forms ligand-metal (lm) interactions (fifth term). The f, g and h functions give a full score (1.00) to distances Δr or angles $\Delta\alpha$ that adopt values within defined cut-offs; conversely, they assign scores between 1.00 and 0.00 when Δr or $\Delta\alpha$ have values outside these limits but within a larger threshold. As an example in hydrogen-bonds, a full score is assigned for hydrogen-acceptor distances in the range 1.85 ± 0.25 Å, while the score is linearly decreased to 0.00 for distances between 2.10 Å and 2.50 Å. The coefficients C_X have been optimized using a simulated-annealing algorithm testing training sets of 1000 drug-like molecules towards sixteen receptor sites.

The first term in 2.9 rewards favorable lipophilic interactions and is derived by the hydrophobic grid potential. The hydrogen-bonds terms (second to forth) reward the formation of hydrogen-bonds, weighting their contributions differently depending on whether the donor and acceptor are both neutral (nn), one is charged and the other neutral (nc) or both are charged (cc). The fifth term concerns metal-ligand interactions involving anionic receptor atoms; the sixth term assigns a penalty for freezing rotatable bonds, while the seventh term rewards polar but non-hydrogen-bonding atoms in a hydrophobic region. A major contribution to the GScore is given by the ligand-receptor Coulomb and van der Waals interaction energies (eighth and ninth terms). Herein the reduction of net ionic charges for formally charged groups (e.g. guanidiniums and carboxylates) is employed to have a better overall prediction of the interaction energy. Finally, a contribution is given by the incorporation of solvent effects: Glide explicitly dock water molecules in the binding pocket of the best-scored poses to check and evaluate with empirical scoring terms the exposure of various groups to water. Particular attention is given to exposure to solvent of charged groups, additionally situations as water molecules trapped in hydrophobic pockets are penalized.

In Glide XP-mode this scoring-function has been further extended. Specifically the following features have been included: i) hydrophobic enclosure contribution, which accounts for groups of lipophilic ligand atoms that are surrounded on two opposite sides by lipophilic protein atoms; ii) improvements in the scoring of hydrogen bonds, by considering special neutral-neutral hydrogen bonds in hydrophobic environment and by using five different categories of charged-charged hydrogen bonds; iii) detection of pi-cation and pi-pi stacking interactions. Additionally, due to the improved sampling, higher penalties can be assigned to violations concerning solvation effects.

Overall it has been shown that Glide performs well both in pose and binding affinity predictions. It has been reported that it outperforms other docking programs, such as GOLD, FlexX, HYDE and AutoDock Vina;[176,177] however, similarly to the other software, its efficacy is lower when dealing with highly flexible ligands.[161] Moreover it has been demonstrated that the quality of the outcomes strongly depends on input structures preparation, thus it is worth, prior to performing docking, to correct

protonation and tautomeric states, missing ligand atoms, protein side chains, hydrogen atom positions, and side-chain conformational errors in crystallographic structures. The Schrödinger company offers some programs to handle this issue, specifically these are the Protein Preparation Wizard, for the optimization of the receptor structure, and LigPrep, for ligand preparation.

As already outlined binding-energy predictions can be further improved by using MM-GB/SA or MM-PB/SA methods, which can give estimation of desolvation energy by employing implicit solvent model. To this aim the Schrödinger suite of programs Prime[178,179] can be exploited. Prime MM-GB/SA takes as input the single minimized receptor-ligand poses obtained from the docking screening to refine and rescore them. For the refinement the Prime local optimization feature is utilized, while for the binding-energy prediction it is employed a combination of the OPLS-2005 molecular mechanics force-field and of the Generalized-Born/Surface Area continuum solvent model. With Prime MM-GB/SA it is also possible to partially account for receptor flexibility defining a protein region within a certain distance from the ligand, which is relaxed during the process. Different options are available for the minimization: it is possible to minimize only polar hydrogens, only side-chains or both side-chains and backbone. Obviously, the computational cost required is higher when more accurate options are selected and when a larger protein region is defined. Another possibility consists in constraining the movements of the flexible residues by applying an harmonic potential, to avoid high deviation from the receptor initial structure. Several papers have shown that binding energy predictions are substantially improved by using Prime MM-GB/SA method, leading to good correlations with experimentally determined affinities.[180,181]

2.3.3 HADDOCK

HADDOCK (High Ambiguity Driven biomolecular DOCKing) is a docking program, originally developed by the Bonvin group in 2004 to predict protein-protein binding modes by exploiting data from NMR (Nuclear Magnetic Resonance) titration or mutagenesis experiments.[153,182–184] Unlike Glide, the main goal of HADDOCK is to provide reliable binding modes for unknown complexes rather than screening

large number of compounds. The characteristic feature of HADDOCK is that, differently with respect to traditional docking protocols, for which only the initial coordinates of the unbound components are considered, it makes use of experimental or predicted information during the docking process. These data are not merely used as filters in the post-processing steps, but they are employed directly to drive the calculations, reducing the search in a fractions of the whole available interaction space (see Figure 2.5).

From 2004 HADDOCK has been extended and currently it can deal with molecules of different nature, such as proteins, peptides, small molecules, nucleic acids and oligosaccharides. Additionally, several sources of experimental information can now be used besides the mutagenesis data and the NMR-derived parameters (e.g. chemical shift perturbations, nuclear Overhauser effect distances and residual dipolar couplings). These sources include chemical cross-linking, small angle X-ray scattering (SAXS), cryo-electron microscopy (cryo-EM), collision cross section (CCS) data, as well as interface predictions obtained with bioinformatics methods. The derived information can be either translated in structural restraints or in additional scoring terms. Usually low resolution data, e.g. those derived from SAXS, cryo-EM or CCS experiments, are only considered in the scoring, by computing

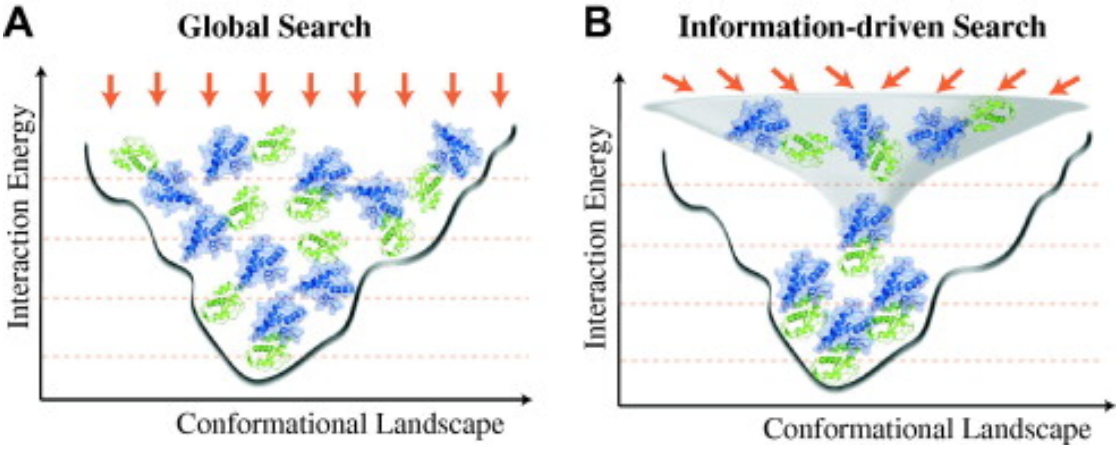


Figure 2.5: Representation of how information-driven search methods can reduce the sampling in a section of the conformational space explored by global search methods. The set of solutions generated by information-driven search are more homogeneous. Figure is reprinted from reference[185].

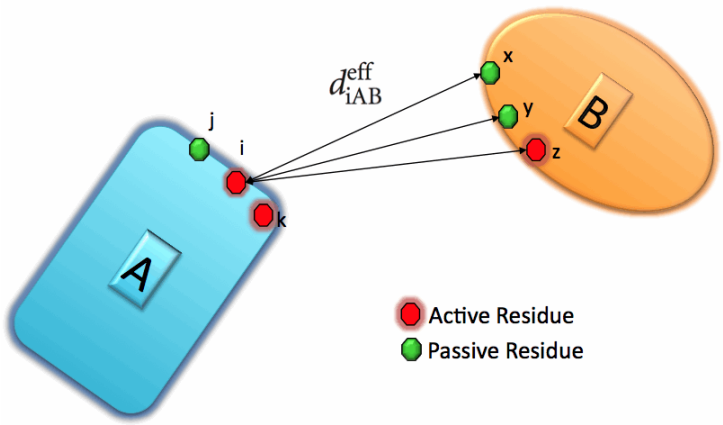


Figure 2.6: Representation of how AIRs are generated. Any atom of the active residue i of protein A is forced to interact with the atoms of all the active and passive residues (x, y and z) of protein B. Analogous restraints are applied on the other active residues k and z, but not on the passive residues j, x and y. Figure is reprinted from reference [186].

differences between the experimental values and the back-calculated ones. Conversely, information such as chemical shift perturbations, alanine-scan mutagenesis and cross-linking, are translated into distance restraints between residues. Since the exploited experimental techniques are usually able to highlight the residues engaged in interactions but are not able to specify which are the pairs of interacting residues, different levels of ambiguity (one-to-one or one-to-many) can be assigned to these distance restraints. To deal with ambiguous interaction restrains (AIRs) HADDOCK classifies residues as active or passive: the first ones are forced to interact at the interface with passive or active residues of the other molecule, otherwise penalties are applied; the second ones instead can or can not interact (see Figure 2.6). A network of AIRs is then generated, by defining an effective distance (d^{eff}_{iAB}) between each active residue i of protein A and all the active and passive residues of protein B, according to this formula: [182]

$$d^{eff}_{iAB} = \left(\sum_{m_{iA}=1}^{N_{atoms}} \sum_{k=1}^{N_{resB}} \sum_{n_{kB}=1}^{N_{atoms}} \frac{1}{d^{6}_{m_{iA} n_{kB}}} \right)^{-\frac{1}{6}} \tag{2.10}$$

where N_{atoms} are all the atoms of a given residue and N_{resB} indicates all the active and passive residues for protein B. The d^{-6} summation is used to mimic the attractive contribution of the Lennard-Jones potential; additionally it allows

to have short effective distances (and thus to satisfy AIRs) when any two atoms of the proteins to be docked are in close proximity. These effective distances are restrained by applying a flat-bottom potential, which behaves harmonically up to violations of 2 Å from the target value (usually set to 2 Å, that corresponds to real distances between 3 and 5 Å) and that is smoothly linearized after the upper bound to avoid large forces. The criteria to define active and passive residues are not universal and depend on the experimental data available and on their interpretation. Generally, active residues are the ones that are experimentally identified as interacting residues at the interface, while passive residues are defined as the neighbors of active residues with high (>40%) solvent-accessibility.

The docking procedure of HADDOCK is constituted by three steps:

- *it*0, a first docking stage by rigid-body energy minimization. The two molecules are first separated and randomly rotated around their center of mass; then their relative orientation is optimized and they are subjected to rotational and translational energy minimization. To explore as much as possible the interaction space, it is advisable to generate a large number of models ($\sim$ 1000) and to select a fraction of them ($\sim$ 200) for the subsequent refinement. The selection is performed by scoring the models as described in the following.

- *it*1, a semi-flexible refinement step in torsion angle space. Each model from *it*0 is optimized applying a three-stage simulated-annealing protocol followed by a short energy minimization to relax the poses. During this procedure the flexibility of side-chain and backbone atoms belonging to interface residues, i.e. all residues within 5 Å from the partner protein, is gradually allowed. A further selection of the models to be refined in the following stage, based on the HADDOCK score, can be performed.

- *water*, a final refinement step in a layer of explicit water. During this stage few changes in the binding modes are expected, however it should improve the reliability of the simulations and the scoring of the poses by optimizing the hydrogen-bond networks and the electrostatic interactions at the interface. All the refined docking poses are scored and saved as output.

In each of this stage and for each model, an HADDOCK score (HS) is computed. The form of this score is a sum of weighted energy-like terms according to the following equation:

$$HS = w_{vdw}\,E_{vdw} + w_{elec}\,E_{elec} + w_{desolv}\,E_{desolv} - w_{BSA}\,BSA + w_{air}\,E_{air}; \quad (2.11)$$

where E_{vdw} and E_{elec} are the van der Waals and electrostatic energy terms, E_{desolv} is an empirical desolvation contribution,[187] BSA is the buried surface-area and E_{air} is the ambiguous interaction restraint energy. The weights w_x vary at each of the three stages of the docking protocols, guaranteeing an adequate scoring-function for the goals of each step. In order to identify the best poses, expected to be the native-like ones, a cluster analysis is usually performed on all the models produced in the *water* stage: this is done following a structural similarity criterion as described in section 2.3.4. After this procedure an energetic value is assigned to each cluster, which is obtained by averaging the scores of the single binding modes (or of the N best binding modes) belonging to it. The best scored cluster is expected to be a reliable representation of the native, or near-native, complex structure.

Several variants and options to improve the efficacy of this program are available. To consider molecules flexibility, besides the flexible-refinement stage, it is possible to use ensembles of ligand and receptor structures as input. Additionally, HADDOCK is able to model multimeric molecular assemblies, handling up to six different molecules.[183] A novel methodology, called Flexible Multi-domain Docking, has been proposed to deal with significant conformational changes upon binding: in this approach the flexible molecule is represented as a collection of sub-domains connected by restrains; these separated domains are then docked to their partner simultaneously using the multi-body docking option.[185,188] Moreover, updated protocols to model protein-ligand and protein-peptide interactions have been recently published.[154,155] Finally, different clustering algorithms and methods to estimate binding affinities have been proposed.[189–191]

Of particular interest is the possibility to use a solvated docking protocol in which waters are explicitly modeled since the first stage of the docking procedure.[166,192] This

protocol includes modifications mainly in the *it*0 stage. According to this methodology the two proteins are initially solvated in explicit shell of water molecules, then a rigid-body minimization is performed to form a complex, treating each protein and its associated hydration layer as one rigid-body. After full removal of non-interfacial waters, a biased Monte Carlo approach is iteratively adopted to remove water molecules at the interface until only 25% of them remains. The Monte Carlo procedure is driven by the propensity to find water-bridged amino acid-amino acid contacts at the interface, as derived from the analysis of several protein-protein crystal structures. Afterward, a further rigid-body energy minimization is performed on the docked molecules and the remaining water. From this point on, the canonical HADDOCK procedure is employed on the resulting models, which include the remaining water molecules. It has been shown that despite docking results do not always improve by using this protocol, the scoring is improved and the water molecules at the interface are correctly positioned.

Currently HADDOCK is one of the most used and cited docking programs for the prediction of unknown protein-protein interactions.[193] It is periodically tested in both CASP (Critical Assessment of Techniques for Protein Structure Prediction) and CAPRI (Critical Assessment of PRedicted Interaction) evaluation and it has always achieved successful results.[194,195] However, HADDOCK remains mainly effective for the prediction of binding-modes and is not designed to be used in the screening and ranking of several compounds.

2.3.4 Cluster analysis

Cluster analysis is a post-processing methodology designed to gather objects sharing common features into groups (clusters). This approach is widely adopted in diverse applications, including statistical data analysis, data mining and bioinformatics. In the field of computational biology this technique is mainly applied to analyze the huge amount of data deriving from molecular dynamics simulations or to organize the several poses produced in docking calculations. It has been reported that the use of clustering algorithms generally leads to improved results both in protein structure prediction and in protein-protein docking.[190,196,197] Indeed this kind

of statistical analysis helps in identifying near-native structures better than the scoring-function alone, which is generally affected by high inaccuracies due to the simplified physics it is based on.[190]

A key point when performing cluster analysis consists in defining a similarity function that measures the distance between each pair of elements, giving rise to a distances matrix. The most commonly used distance is the root mean square deviation (RMSD) computed on the atomic coordinates of the molecules. Generally in the post-processing phase of docking calculations, the poses are aligned on the backbone atoms of the receptor and the RMSD matrix is computed on a subset of the atomic coordinates of the ligand (this kind of measurement is often referred to as ligand interface RMSD).

After this preliminary step, a wide variety of algorithms is available to group together the different poses. Within this thesis I will make use of the clustering algorithms proposed by Daura et al. in 1999,[198] which is implemented both in HADDOCK and in GROMACS. According to this algorithm the procedure to define clusters is the following: i) compute the RMSD matrix; ii) given an RMSD cutoff, compute the number of neighbors (i.e. objects with a distance lower than the cutoff) for each element; iii) define a cluster identifying the element with the highest number of neighbors and all its neighbors; iv) repeat steps *ii* and *iii* on the elements that have not yet been clustered until no more clusters, with a number of elements above a threshold, can be generated. The parameters that must be provided by the user, and which must be optimized according to the investigated system, are the RMSD cutoff and the minimum number of elements that are necessary to define a cluster. It is worth noting that this algorithm belong to the group of hard clustering approaches, in which one element can be assigned only to one group.

Chapter 3

Predicting the effect of cyclopeptide *N*-methylation on integrin affinity

Cyclopeptides are a promising class of compounds that can be used as therapeutics for the modulation of protein-protein interactions. Their favorable pharmacokinetic properties, as well as their target selectivity, can be further improved by the introduction of chemical modifications, such as *N*-methylation or acetylation. These widely adopted approaches have found successful applications in the context of integrin inhibition.

Despite all these advantages, the use of cyclopeptides has been limited due to the inherent difficulties to predict *in silico* their three-dimensional structure and their inhibitory activity. Because of these challenges, their development and optimization have been mainly based on empirical approaches, requiring expensive and time-consuming synthesis campaigns. In this context the employment of computational tools could be useful in accelerating the drug design process.

In the following chapter I am going to describe a multi-stage computational protocol that has been optimized during my thesis work to achieve reliable prediction concerning cyclopeptides conformations and their interplay with receptor affinity. This approach is based on an exhaustive conformational search of the ligands,

performed by means of an enhanced sampling technique (i.e. Bias-Exchange Metadynamics), followed by docking calculations and re-scoring via MM-GB/SA. The reliability of this strategy was tested investigating the impact of single and multiple N-methylations on the equilibrium conformation of five RGD (Arg-Gly-Asp) containing cyclic hexapeptides, which were previously generated by the group of professor Kessler to increase their selectivity for αIIbβ3 integrin. The method proposed proved its efficacy: the conformational sampling was in good agreement with available NMR data and it was possible to discriminate between integrin binders and non-binders. Additionally, I offered a structural rationale for how N-methylation modulates peptide affinities for αIIbβ3 integrin.
The approach herein described can represent a promising *in silico* screening strategy, that can be exploited to predict the effect of chemical modifications in cyclopeptides conformations before entering expensive synthesis and binding experiments.

3.1 Introduction

In recent years there has been a growing interest in the study and development of cyclic peptides, which represent a promising class of compounds for the inhibition of protein-protein interactions and other challenging therapeutic targets.[56,199] Cyclopeptides present several favorable pharmacokinetic properties, showing enhanced stability and better bioavailability with respect to their linear counterparts and being safer and more target-specific than small molecules. In particular, they provide unique opportunities to achieve enhanced binding affinity by increasing the molecular size and the contact area, meanwhile decreasing the conformational freedom of the ligand and reducing possible entropic penalties.[200] These approaches resulted to be particularly efficient in the context of integrin inhibition, where several cyclopeptides reached advanced stage in the drug development pipeline. Herein cyclization helped in constraining peptides, already in the unbound state, in specific pre-organized conformations, which could be suitable for the binding to a specific receptor.[79,201–203]
In this context, introduction of chemical modification is a frequently used strategy to modulate the conformational properties of synthetic cyclic peptides and to

influence their lipophylicity, proteolytic stability, and bioavailability.[61,204] Specifically, it has been shown that backbone *N*-methylation, commonly occurring also in natural peptides, can be used to modulate the cyclopeptides target specificity, their binding affinity and the related biological down-stream signaling.[61,205] These effects are generally ascribed to the ability of *N*-methylation in constraining the peptide backbone conformation.[206]

Despite this interest, the difficulties in predicting *de novo* the conformations adopted by cyclopeptides and consequently their binding affinity are limiting their development as drugs. In this context, the optimization of cyclopeptides remains mainly based on empirical approaches, demanding brute force synthesis of different compound's variants in order to individuate the ones with appropriate target-binding properties. Recently, improvements are being observed in computational techniques able to deal with the modeling and docking of cyclopeptides.[207,208] The main challenges to be faced by computational methods rely in the constrained geometry of macrocycles, which can populate several conformations separated by high free-energy barriers. Additionally, the introduction of chemical modifications, which can strongly affect the equilibrium between conformers, further entangles the correct prediction of their preferred conformations and, consequently, of their binding affinity.[199,202]

In this scenario, the original "lock-and-key" binding concept, in which a frozen ligand accommodates into a static receptor, had to be revisited in favor of more dynamic models of molecular recognition able to account for the conformational preferences of the ligand.[181,209,210] However, due to the constrained nature of cyclic peptides, the algorithms implemented in the common docking software are generally not suitable to this aim. Herein, the poor conformational sampling of the cyclopeptides and/or the inclusion of unrealistic high-energy ligand structures as input strongly affect the performance of molecular docking, compromising accuracy in the scoring of the decoy poses.[211] It has been suggested that an exhaustive and reliable study of cyclopeptides conformations prior to docking calculations is strongly advisable to get better computational efficiency and predictability.[207] Nevertheless, exhaustive conformational sampling of cyclopeptides is not trivial: in these molecules the transition from one conformation to another requires a concerted

motion of several dihedral angles,[207,212] complicating the investigation of their equilibrium conformation with common simulation techniques. In this framework, Metadynamics (MetaD) represents a valuable enhanced sampling methodology for the acceleration of rare events, able to provide reliable estimations of the populations of the accessible conformers, as well as relative heights of free-energy barriers (see Section 2.2).[127,132,133] The applicability of MetaD in analyzing cyclopeptides conformations had been previously tested in our laboratory, showing that it could be helpful in discriminating $\alpha v\beta 3$ integrin binders from non binders.[75] Additionally, similar approaches have been also recently adopted by other groups, confirming the reliability of this methodology.[199,208] In the work described in this chapter I have further explored the possibilities and the applicability range of this enhanced sampling technique, adopting the Bias Exchange variant of MetaD (BE-META),[134] and extending the protocol by introduction of docking calculations and rescoring via MM-GB/SA methods. The employment of BE-META in place of plain MetaD allowed to increase the number of biased collective variables, enabling a systematic bias of all the backbone dihedral angles of the investigated cyclopeptides, ultimately resulting in improved sampling and convergence. This technique also allows to easily reconstruct multidimensional free-energy landscapes of the studied system.[141]

As already outlined, cyclization is a widely used strategies in the field of integrin inhibition, to constrain peptides in specific conformations with ideal features for the binding target (i.e. proper orientation and distance of the pharmacophores). These peptides are usually based on the RGD (Arg-Gly-Asp) integrin recognition motif and their affinity and selectivity for different integrin subtypes is controlled by the RGD motif's environment, including the type and stereo-chemistry of flanking residues, the number of amino acids in the cycle and the presence of chemical modifications. All these features concur to regulate cyclopeptide affinity towards a specific integrin both by means of direct interaction with the receptor and/or by modulation of the RGD sequence conformation.[201] Thus, integrin inhibitors represent an ideal field to test the applicability of the proposed computational protocol.

Specifically, I decided to focus my attention on how cyclopeptides integrin affinity is regulated by the introduction of chemical modifications. This represents an ideal

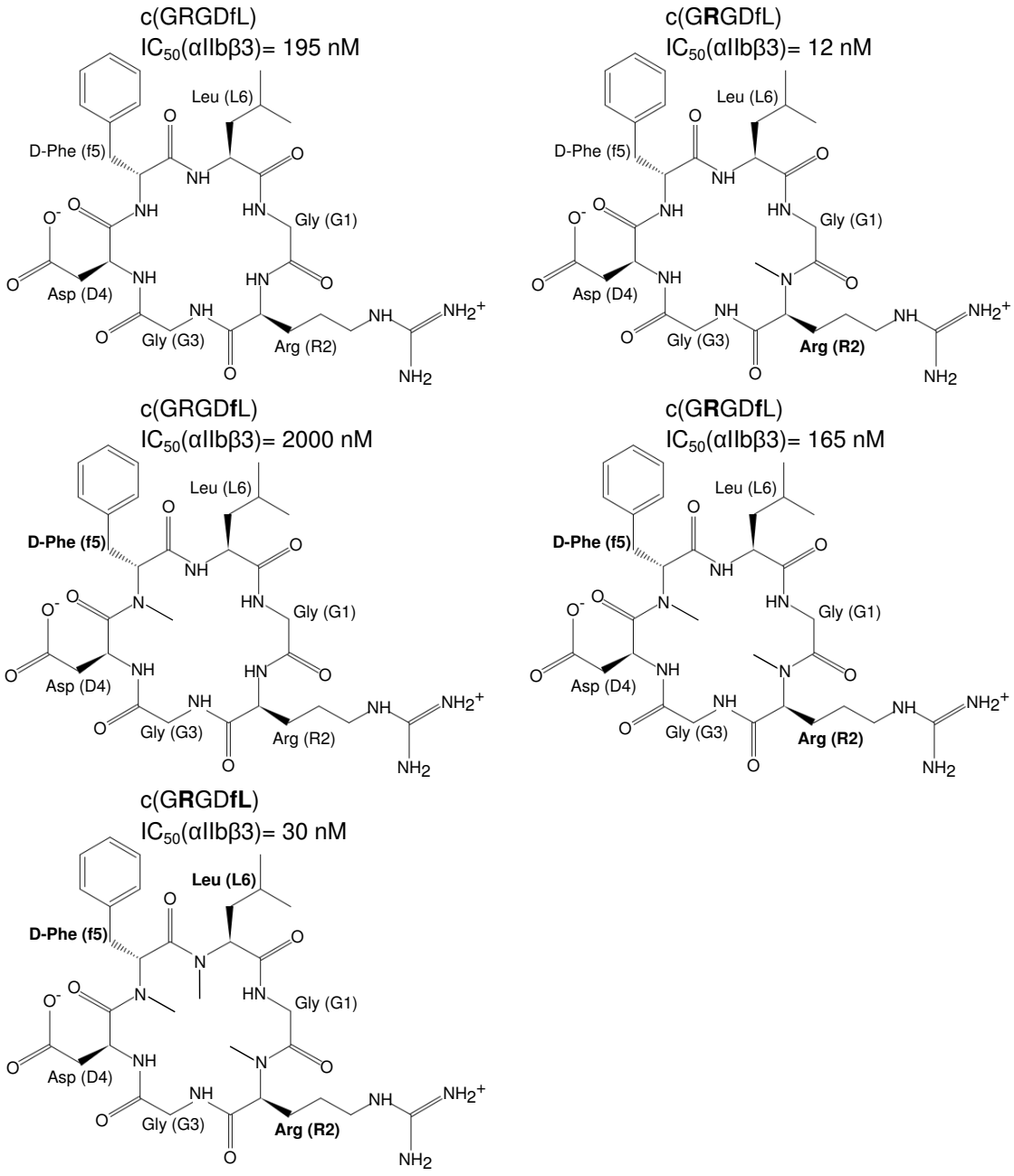


Figure 3.1: **2D-structure representation of the investigated cyclic hexapeptides.** For each molecule its IC_{50} value for integrin αIIbβ3 is reported.[205] *N*-methylated amino acids are shown in bold.

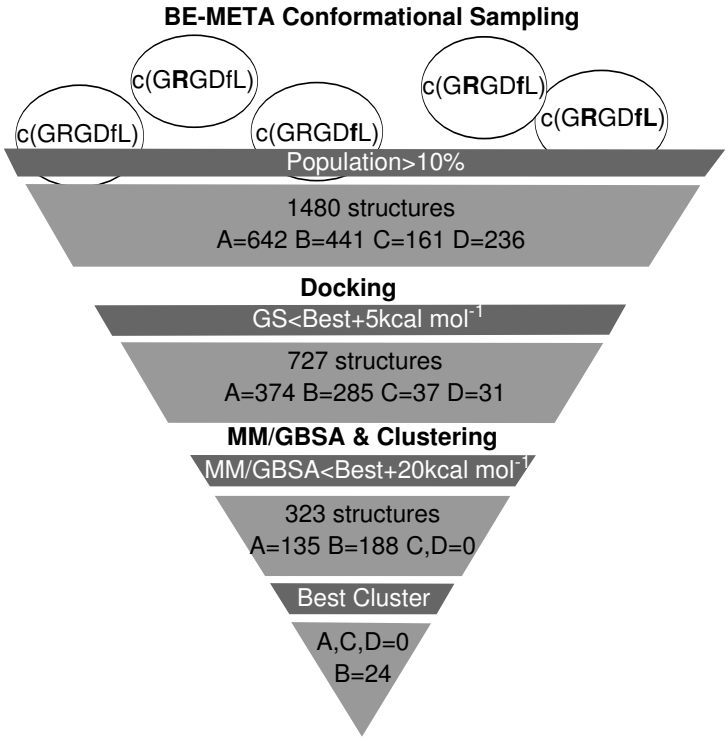


Figure 3.2: **Workflow.** The multistage approach relies on a combination of: BE-META conformational search, docking calculations, MM-GB/SA rescoring and cluster analysis. Boxes in dark gray show the filters used after each stage. Boxes in light gray report the number of structures (and how they are subdivided within the BE-META minima) satisfying the filters conditions.[213]

test-case for the proposed approach that strongly relies on a conformational search: indeed the introduction of these features is more likely to influence affinity through conformational modulation than through direct interaction with the receptor. To this aim, I reasoned to exploit a small library of head-to-tail cyclic RGD hexapeptides, previously generated by Kessler and coworkers to improve their selectivity towards αIIbβ3 integrin.[205] The five cyclopeptides of the library share the same sequence of six amino acids and differ only in position and number of backbone N-methylations. Despite these small variations, the affinities displayed by the five cyclopeptides for αIIbβ3 integrin are largely different (see Figure 3.1) ranging from 12 to 2000 nanomolar in terms of IC_{50} values. Applying the multi-stage computational protocol based on the combination of BE-META conformational sampling, docking calculations and MM-GB/SA rescoring (Figure 3.2), it has been

possible to rationalize the conformational effects of *N*-methylation on the free peptides and to highlight its interplay with receptor affinity.[213]

3.2 Results and Discussion

3.2.1 Preliminary BE-META calculations

It is known that backbone *N*-methylation can decrease the height of energy barrier between the *cis* and *trans* amide bond isomers.[214] To investigate the impact of *N*-methylation on *cis-trans* isomerization in the five studied cyclopeptides, I ran a preliminary 400 ns ten-replicas BE-META simulation on each molecule, for a total of 2 μs, using the settings described in Section 3.4. Each replica was biased on one of the following ten CVs: the six Ω dihedral angles, which describe the planarity of the peptide bond, and the ϕ dihedral angles of Arginine (R2), Aspartate (D4), D-phenylalanine (f5) and Leucine (L6). The ϕ dihedral angles of Glycines were not biased since these residues are considered the most flexible ones, as they lack the side-chain. After convergence check, the free energy surfaces were reconstructed and analyzed in a reduced six-dimensional space, defined by the Ω dihedral angles. Consistently with observations previously reported,[58] the calculations revealed that no significant isomer interconversion is present. Specifically, I found that minima containing isomers characterized by *cis* peptide bond conformations were always populated less than 10%. Accordingly, the subsequent investigation was restricted to the *trans*-isomers conformational space, excluding the Ω dihedral angles from the CVs.

3.2.2 Conformational analysis

Subsequent BE-META calculations were carried out biasing the ϕ and ψ dihedral angles of all the non-Glycine amino acids (i.e. R2, D4, f5 and L6), running for each of the five molecules an eight-replicas simulation of 320 ns, for a total length of 1.6 μs. The convergence was carefully checked monitoring the mono-dimensional

Table 3.1: **BE-META conformational analysis.**[213] For each molecule, the population of the minima (when populated more than 10%) and the experimental IC_{50} values for integrin αIIbβ3 are reported.

Peptides[1]	*N*-methylated residues	Minima population [%] A	B	C	D	IC_{50} [nM][2]
c(GRGDfL)	-	96	-	-	-	195
c(G**R**GDfL)	R	26	73	-	-	12
c(GRGD**f**L)	f	49	-	46	-	2000
c(G**R**GD**f**L)	R,f	32	37	13	18	165
c(G**R**GD**fL**)	R,f,L	18	82	-	-	30

[1] *N*-methylated residues are highlighted in bold.
[2] See reference[205].

bias potential of each replica. Specifically I compared the bias potential profiles obtained by averaging on the two halves of the simulations after an equilibration time of 1 ns. In all the cases the energy-difference between the two halves was lower than $2k_bT$, demonstrating that the potential was stable and that the simulation reached convergence (Figure A.1).

The analysis and the reconstruction of the free-energy landscape have been performed in a reduced space determined by the ϕ and ψ dihedral angles of R2, f5 and L6: these are the amino acids alternately *N*-methylated in the different cyclopeptides. These variables, which are a subset of the biased CVs, have been selected after several tests and they were shown *a posteriori* to accurately describe the free energy of the systems highlighting the differences between them.

In this six-dimensional space the main free-energy minima are identified and their population is evaluated as described in Section 3.4 (Table 3.1). Each minimum populated more than 10%, was identified according to the values adopted by its ϕ and ψ dihedral angles (see Table A.1 and Figure A.2 in Appendix A). Analysis of the conformers' positions in the six-dimensional dihedral space allowed to identify a subspace of minimal dimensions in which each minimum can be uniquely characterized. Therefore a FES projection on this bi-dimensional subspace, defined by the ϕ dihedral angles of R2 and f5, is reported in Figure 3.3 for each cyclopeptide.

It can be observed that, while the unmethylated reference molecule c(GRGDfL) significantly populates only minimum A (96%), the other cyclopeptides can adopt also

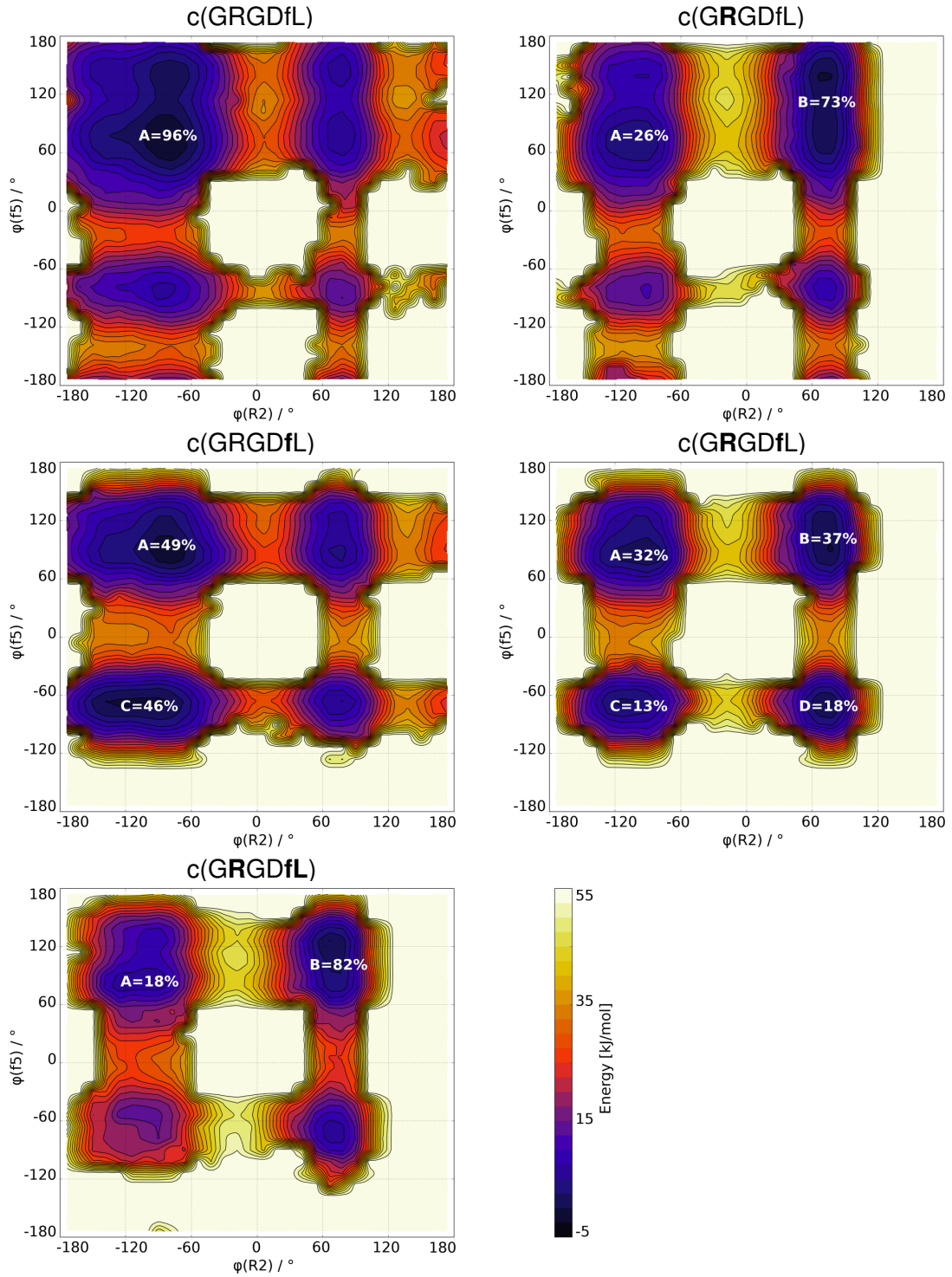


Figure 3.3: **2D-energy surfaces.** FES projections in the space defined by the ϕ dihedral angles of R2 and f5.

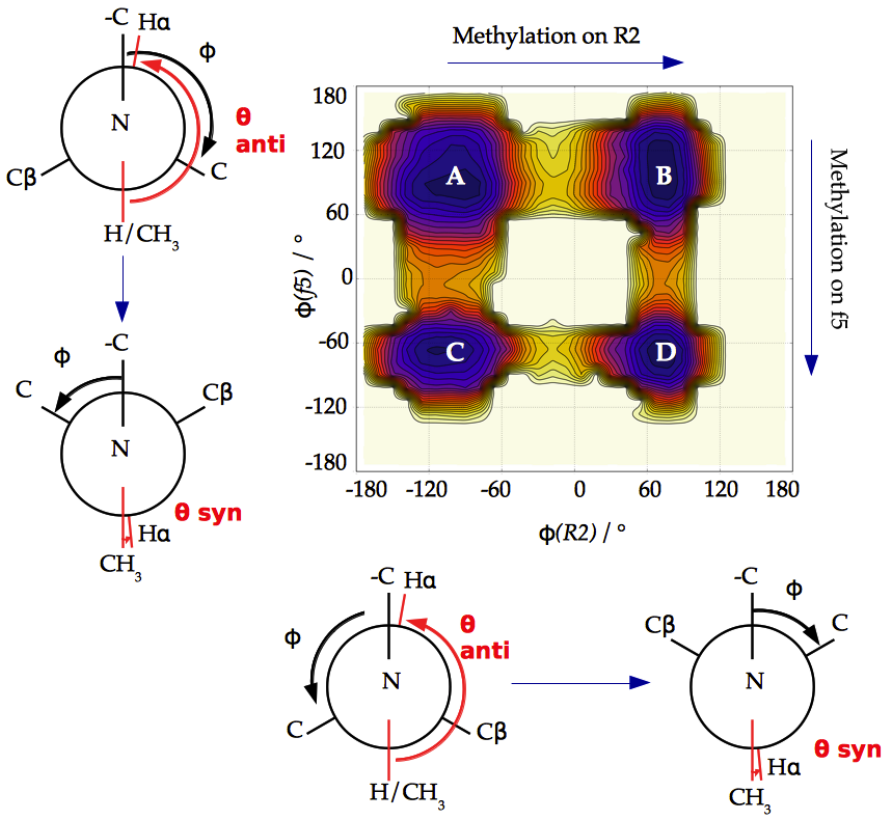


Figure 3.4: ***N*-methylation effects.** Representation of the effects triggered by N-methylation on the ϕ and θ (HN/CN-N-Cα-Hα) dihedral angles of the corresponding residue. Methylation of one amino acid induces the corresponding θ dihedral angle to adopt a *syn*-periplanar conformation.

different conformations. Specifically, the single N-methylation of R2 (c(G**R**GDfL)) and of f5 (c(GRGD**f**L)) favors the accessibility of the additional minima B (73%) and C (46%), respectively. Notably, the characteristic features that make these minima different from A are the ϕ dihedral angles of R2 (in the case of minimum B) and of f5 (for C), evidencing how the N-methylation of one residue can induce a flip in its ϕ dihedral angle (Figure 3.4). Following the same principle, simultaneous methylation of R2 and f5 (c(G**R**GD**f**L)) induces the formation of the above mentioned minima B (37%) and C (13%), and of a fourth conformational minimum D (18%), which displays different ϕ dihedral angles of both R2 and f5 with respect to A. Lastly, the triple N-methylated peptide c(G**R**GD**f****L**) shows a FES similar to the one of the molecule with single methylation on R2, resulting only in the

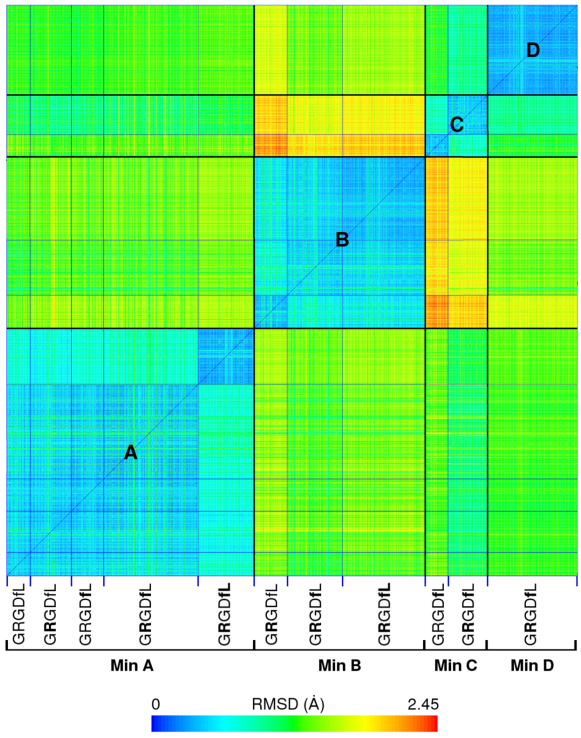


Figure 3.5: **Similarity matrix**. Matrix representing the Root Mean Square Deviation between all the conformers extracted from BE-META.[213] RMSD is calculated on the backbone, Cβ and O atoms.

formation of the additional minimum B (82%). This unconventional behavior can be explained observing that multiple methylations on two adjacent amino acids could create a steric hindrance effect, blocking the modified residues in the original conformation and eliminating the expected methylation-induced conformational changes.

I next investigated the conformations of all the cyclopeptides pooling together the structures extracted from the BE-META minima (A-D). It was found that the single minima were structurally very similar among the different molecules, as highlighted by the root mean square deviation (RMSD) matrix in Figure 3.5. Structures within the same minimum displayed an average RMSD, computed on the backbone, Cβ and O atoms, lower than 0.6 Å.
Next, the conformers within each minimum have been characterized, focusing

Table 3.2: **Structural features of the BE-META conformers.**[213] The minima have been characterized according to: the *syn* or *anti* periplanar conformation of dihedral angle θ, the presence of backbone hydrogen bonds and the average distance between the Cβ atoms of R2 and D4.

Minimum	θ [°]				H-bonds[1]	d(Cβ_{R2},Cβ_{D4}) [Å]
	R2	D4	f5	L6		
A	*anti*	*anti*	*anti*	*anti*	$HN_{G1} - O_{D4}$	7.8 ± 0.5
B	*syn*	*anti*	*anti*	*anti*	-	8.6 ± 0.6
C	*anti*	*anti*	*syn*	*anti*	$HN_{G1} - O_{D4}$	7.2 ± 0.4
D	*syn*	*anti*	*syn*	*anti*	$HN_{G1} - O_{D4}$ $O_{G1} - HN_{D4}$	5.9 ± 0.4

[1] Hydrogen bonds were individuated with the VMD[215] Hbonds plugin. A donor-acceptor distance of 3.5 Å and an angle cutoff of 35° have been considered. Only hydrogen bonds present in at least 50% of the structures are reported.

on: i. the conformations (*syn* or *anti*-periplanar) adopted by the dihedral angle θ_i of amino acid i (excluding Glycine), defined by the atoms HN/CN-N-Cα-Hα (see Figure 3.4); ii. the presence of intra-molecular backbone hydrogen bonds, if observed in at least 50% of the structures; iii. the distance between the Cβ atoms of R2 and D4 (d(Cβ_{R2},Cβ_{D4})), which is strongly related to the positioning of RGD-cyclopeptides pharmacophores, namely the charged groups of Arginine and Aspartate (Table 3.2).[14]

Concerning the θ dihedral angles, in minimum A all the residues adopt an *anti*-periplanar conformation, while in structures of minima B and C a *syn*-periplanar conformation can be observed for residues R2 and f5, respectively; finally, in conformers belonging to minimum D, both these amino acids, R2 and f5, adopt a *syn*-periplanar geometry. Overall these data suggest that *N*-methylation of a residue favors the accessibility to conformers, which otherwise would not be significantly populated, characterized by the θ dihedral angle of the corresponding residue in *syn*-periplanar conformation, i.e. with the methyl CN atom and the Hα atom pointing in the same direction (Figure 3.4).

These different geometries of the backbone dihedral angles induce a different pattern of intramolecular hydrogen bonds. On one hand, structures belonging to minima A, C and D present one or two hydrogen bonds between residues G1 and D4, showing a relatively short distance d(Cβ_{R2},Cβ_{D4}) (lower than 8 Å). Conversely, the absence

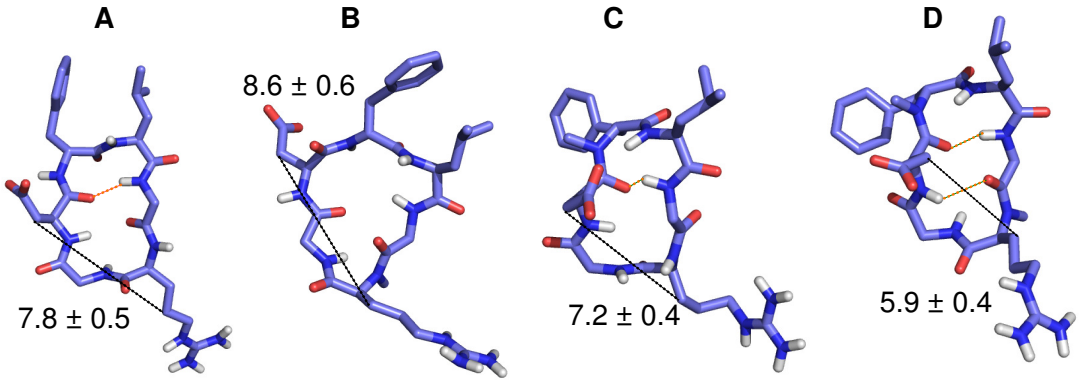


Figure 3.6: **Representative conformers.** Structures from BE-META minima are shown in sticks, with carbon, nitrogen, oxygen and hydrogen atoms colored in marine blue, blue, red and white, respectively. Hydrogen bonds and the distance $d(C\beta_{R2},C\beta_{D4})$ (for which the average is reported) are indicated with orange and black dashed lines.

of backbone hydrogen bonds in minimum B, allows more extended conformations, characterized by a distance $d(C\beta_{R2},C\beta_{D4})$ that is approximately 9 Å (see Table 3.2 and Figure 3.6).

Taken together, these data suggest that the introduction of backbone *N*-methylation in cyclopeptides modulates their structural preferences, enlarging their accessible conformational space and favoring a twist by about 180° of the peptide bond plane preceding the methylated residue (*syn*-periplanar geometry). This rearrangement displays relevant consequence in the case of R2 *N*-methylation: herein the induced θ *syn*-periplanar geometry prevents the formation of an intramolecular backbone hydrogen bond, favoring an extended conformation that satisfies the pharmacophoric requirements for the accommodation of the RGD motif in the integrin αIIbβ3 binding pocket.[14]

3.2.3 Experimental validation of the conformational sampling

To validate the hypothesis that backbone residue *N*-methylation affects the *syn-anti* conformational equilibrium of the corresponding θ dihedral angle, I quantitatively compared the experimental $^3J_{HN,H\alpha}$ scalar couplings, available only for c(G**R**GDfL)

Table 3.3: **Experimental validation of BE-META conformational sampling.** The comparison between the experimental values of $^3J_{\mathrm{HN,H}\alpha}$ scalar couplings and the back-calculated ones is shown. For N-methylated residues the inter-residue distances $d_{\mathrm{CN,H}\alpha}$, derived from the ROESY spectra, and the back-calculated distances from BE-META minima are reported. Additionally, the corresponding $\theta_{\mathrm{CN,N,C}\alpha\mathrm{,H}\alpha}$ conformation is indicated. The estimated errors for scalar couplings and distances are 0.91 Hz and 0.3 Å, respectively.

Residue	c(G**R**GDfL)					c(G**R**GD**f**L)				
	$^3J_{\mathrm{HN,H}\alpha}$ [Hz]		$d_{\mathrm{CN,H}\alpha}$ [Å]		θ^1	$^3J_{\mathrm{HN,H}\alpha}$ [Hz]		$d_{\mathrm{CN,H}\alpha}$ [Å]		θ^1
	Exp[2]	Comp	Exp[2]	Comp		Exp[2]	Comp	Exp[2]	Comp	
G1	4.4	4.5	-	-	-	3.8	5.1	-	-	-
R2	-	-	2.1	2.7	*syn*	-	-	1.9	2.6	*syn*
G3	6.0	5.8	-	-	-	5.9	5.8	-	-	-
D4	8.0	8.1	-	-	-	9.7	7.1	-	-	-
f5	7.0	8.9	-	-	-	-	-	2.9	3.4	*anti*
L6	8.0	7.4	-	-	-	-	-	2.9	3.4	*anti*

[1] the conformation adopted by the θ dihedral angle in structures extracted from BE-META minimum populated more than 70% is reported.
[2] see reference [205].

and c(G**R**GD**f**L),[205] with the back calculated ones (Table 3.3). The $^3J_{\mathrm{HN,H}\alpha}$ were computed as described in Section 3.4.3. The agreement between experimental and simulated data was quantified according to the following equation:[216]

$$\chi^2(^3J_{exp},{}^3J_{comp}) = N^{-1}\sum_{i=1}^{N}\frac{(^3J_{comp,i} - {}^3J_{exp,i})^2}{\sigma^2} \tag{3.1}$$

where $\sigma = 0.91$ Hz is the estimated systematic error and $N = 8$ is the number of J-couplings considered. The value obtained for $\chi^2(^3J_{exp},{}^3J_{comp})$ was of 1.8, indicating a satisfactory agreement.[217]

For N-methylated amino acids a direct comparison with the $^3J_{\mathrm{HN,H}\alpha}$ scalar couplings is not possible. Therefore, in these cases I calculated the average intra-residue distances $d_{\mathrm{CN,H}\alpha}$. The residues displaying θ *syn*-periplanar geometry are expected to be characterized by short $d_{\mathrm{CN,H}\alpha}$, indeed in this situation the atoms CN and Hα point in the same direction; conversely, *anti*-conformations are expected to display longer distances. The averages have been calculated using all the structures extracted from BE-META minima and weighted with their population (Section 3.4.3). A qualitative

comparison of these data with the ones obtained from the ROESY spectrum[205] was performed. In agreement with experimental data, those residues which often adopt the *syn*-periplanar conformation of the θ dihedral angle, show shorter $d_{\mathrm{CN,H}\alpha}$ distances than the ones that mainly adopt *anti* conformation (Table 3.3).
Overall, this comparison supports the finding that *N*-methylation of a residue induces its θ dihedral angle to adopt a *syn*-periplanar conformation.

3.2.4 Docking calculations

As second stage of the protocol, I investigated by means of docking calculations the relationship linking cyclopeptides' conformation and their binding affinity for αIIbβ3 integrin. To this aim I exploited the software Glide of the Schrödinger suite in its Extra Precision variant.[152] As input for the docking procedure I used the bundles of structures extracted from all the BE-META minima populated more than 10%, for a total of 1480 structures. Since an exhaustive conformational sampling had already been performed on this structure I did not apply any optimization or refinement strategy on the ligand. The structure of the protein was instead retrieved from the PDB 2vdn[14] (crystallographic integrin αIIbβ3 in complex with Eptifibatide) and optimized as described in Section 3.4. No constraints were applied during the docking procedure. The grid, necessary to drive the docking of the peptides towards the binding pocket, was defined using as reference Eptifibatide, which was subsequently removed (details are in Section 3.4).
The decoy poses obtained were scored according to the Extra-Precision Glide Score (GS) and only the ones satisfying the filter GS $<$ ($\mathrm{GS_{best}} + 5\,\mathrm{kcal\,mol^{-1}}$) were retained for the subsequent rescoring. It is worth noting that approximately the 60% of the structures belonging to minima A and B passed this first filter, while structures belonging to minima C and D were less prone to obtain a satisfactory GS: indeed, only 23% and 13% of the structures of minima C and D, respectively, were accepted (see Figure 3.2).
The remaining 727 decoy poses were then rescored using the MM-GB/SA (Molecular Mechanics-Generalized Born Surface Area) protocol implemented in the Prime software (Schrödinger company).[178] MM-GB/SA strategy indeed represents a valu-

able technique to estimate binding free-energy,[218] being a fair compromise between empirical methods and computationally intensive approaches. Even if this method generally is not able to reproduce the absolute binding free-energy values (due to the several approximations employed, including the lack of treatment of the entropy contribution), it often exhibits a good relative correlation with experiments.[218] Additionally, it has been demonstrated in several cases that the post-processing of poses with this implicit solvent technique greatly improves the prediction accuracy as compared to default Glide Extra Precision docking.[219] Accordingly, in this work, a clear improvement in the ranking of the docking poses when using MM-GB/SA with respect to the Extra Precision Glide Score was observed: while MM-GB/SA assigns low energy values to poses able to summarize canonical interactions and high energy values to the non-native poses, the Glide Score is not able to discriminate *good* poses from *bad* ones (Table A.2 in Appendix A).

Since it is known that clustering of decoy poses helps in improving structure and binding affinity prediction, I clustered the 727 decoy poses according to geometric criteria and I assigned an average MM-GB/SA score to each cluster. Specifically, all the poses were aligned on the receptor residues within the binding pocket and the RMSD was computed on all the backbone atoms and on the side-chain atoms of R2, G3, D4 and f5; clusters were identified using an RMSD threshold of 2.5 Å (see Section 3.4 for details). Herein thirteen clusters containing at least ten decoy poses have been identified. In Figure 3.7, the average MM-GB/SA score as a function of the RMSD, calculated with respect to the lowest energy structure of the best cluster (in term of MM-GB/SA score), is represented. Nicely, a funnel-like shape can be observed in this Figure, with the average MM-GB/SA score getting worser for clusters having a higher RMSD.

Details regarding the average MM-GB/SA score per cluster and the different energetic contributions, including the polar (obtained as the sum of the Coulomb and the Generalized Born electrostatic solvation energies) and non-polar (accounting for both van der Waals and lipophilic energies) terms, are reported in Table 3.4. The fact that the polar energetic contributions are positive could be disorienting, specially considering the importance of the electrostatic clamp in the integrin binding. In this situation it is worth to observe that, in all the clusters of poses, the Coulomb contribution is favorable (see Table A.3) but it is counterbalanced by

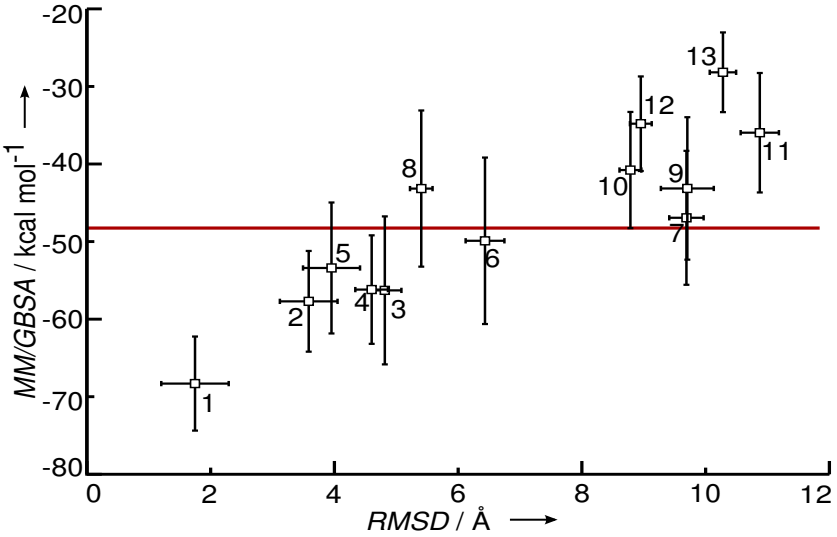


Figure 3.7: **MM-GB/SA vs RMSD**. In Figure is reported the average MM-GB/SA score energy value assigned to each cluster as a function of the cluster RMSD. The RMSD is calculated with respect to the lowest energy structure of the best cluster (in term of MM-GB/SA score).[213] Bars indicate energy and RMSD fluctuations within each cluster. The horizontal red bar represents the threshold of the filter $\langle \text{MM-GB/SA} \rangle < (\langle \text{MM-GB/SA} \rangle_{\text{best cluster}} + 20\,\text{kcal}\,\text{mol}^{-1})$.

the Generalized Born solvation energy, resulting in an overall positive polar term. Nevertheless it is easy to observe that the sum of these two contributions remains fundamental to discriminate among poses, being the best clusters characterized by a notably lower polar penalty. Lastly, it must be underlined that these kinds of calculations are reliable in estimating the variation in binding free energy but they are generally not accurate in reproducing the absolute binding free energy values. An analysis of the six best-scored clusters, i.e. the ones satisfying the filter $\langle \text{MM-GB/SA} \rangle < (\langle \text{MM-GB/SA} \rangle_{\text{best cluster}} + 20\,\text{kcal}\,\text{mol}^{-1})$, revealed that in these poses the cyclopeptide always adopts minima A or B conformations (Figure 3.8). Conversely, minima C and D resulted to be unproductive, not being able to generate decoy poses with an acceptable MM-GB/SA score.

Of note, by visual inspection of the decoy poses belonging to these six clusters, I observed that the cyclopeptides were able to stably recapitulate the characteristic electrostatic clamp, necessary for the binding, bridging the two domains of integrin αIIbβ3 and being able to coordinate the MIDAS metal ion through the

Table 3.4: **Energetic contributions per cluster.** The average MM-GB/SA values and individual contributions are reported for each cluster. The Polar term is computed as sum of Coulomb and Generalized Born electrostatic solvation energy; the Non-Polar term is the sum of van der Waals and lipophilic energy, while Additional contributions include covalent binding, hydrogen-bonding, π-π packing energy and self-contact correction.

Cluster	Energetic Contributions [kcal mol^{-1}]			
	⟨MM-GB/SA⟩	⟨Polar⟩	⟨Non Polar⟩	⟨Additional⟩
1	-68.3 ± 6.1	15.8 ± 3.3	-88.9 ± 5.9	4.7 ± 2.2
2	-57.7 ± 6.5	12.5 ± 4.1	-73.7 ± 5.7	3.5 ± 2.9
3	-56.3 ± 9.5	12.7 ± 5.7	-73.6 ± 6.3	4.6 ± 5.7
4	-56.2 ± 7.0	15.2 ± 4.4	-77.7 ± 6.7	6.3 ± 5.1
5	-53.4 ± 8.4	16.5 ± 3.9	-73.7 ± 6.3	3.8 ± 4.5
6	-49.9 ± 10.7	18.6 ± 6.0	-75.7 ± 6.1	7.2 ± 5.2
7	-46.9 ± 8.6	34.8 ± 6.5	-92.0 ± 6.7	10.2 ± 6.9
8	-43.2 ± 10.1	29.0 ± 8.8	-79.9 ± 5.1	7.7 ± 3.8
9	-43.1 ± 9.2	25.6 ± 6.6	-74.3 ± 7.6	5.6 ± 5.5
10	-40.8 ± 7.5	31.9 ± 5.7	-77.5 ± 7.1	4.8 ± 4.0
11	-36.0 ± 7.7	32.1 ± 5.9	-72.3 ± 4.6	4.2 ± 3.4
12	-34.8 ± 6.1	28.2 ± 4.4	-68.3 ± 5.6	5.2 ± 4.3
13	-28.2 ± 5.1	39.6 ± 6.1	-71.7 ± 6.0	3.9 ± 4.0

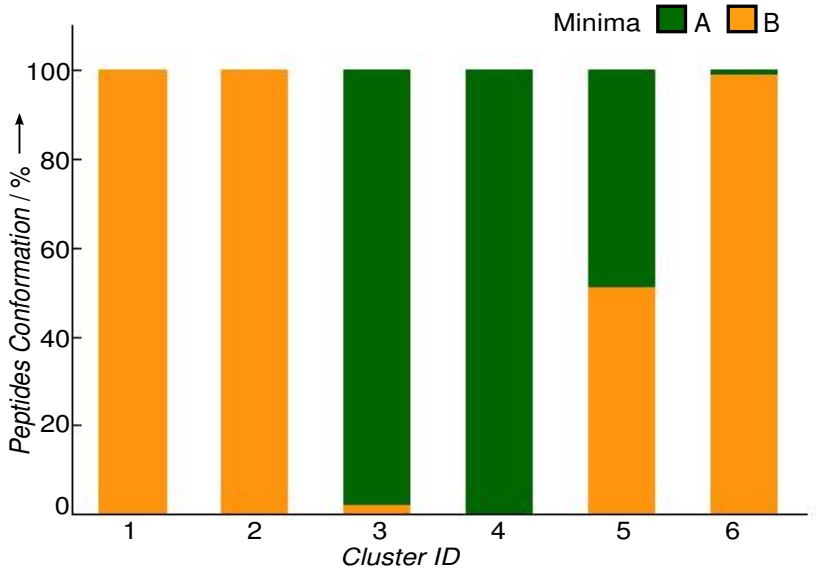


Figure 3.8: **Peptides conformation in decoy clusters.** Cyclopeptides conformational distribution, according to BE-META minima, in the six lowest energy clusters. [213]

D4 carboxylic group (Figure 3.9). Accordingly, these clusters are all characterized by a significantly lower polar contribution with respect to the others (Table 3.4). However, only in four of the six clusters the canonical interactions were perfectly summarized, with the guanidinium group of R2 engaging a salt bridge with the carboxylic group of αIIb:Asp224 and an hydrogen bond with the carbonyl of αIIb:Phe160. In the other two cases different residues of the αIIb subunit were engaged in polar interactions. Herein it is worth to observe that in the presented binding modes the MIDAS is coordinated in a bidentate manner, while it would be expected a monodentate coordination: this discrepancy could be ascribed to defects in Glide XP scoring function, specially arising in situations where a metal ion is involved in ligand binding.
Focusing on the best cluster in terms of MM-GB/SA score, further observations can be done. First of all its score is significantly better with respect to the others; herein the main contribution that differentiates it from the other clusters with acceptable (i.e. satisfying the filter) MM-GB/SA score is the non-polar term (Table 3.4). Accordingly, the decoy poses of this best cluster, besides recapitulating the canonical electrostatic clamp, establish additional stabilizing hydrophobic interactions. Specifically, L6 engages hydrophobic interactions with αIIb:Val156 and αIIb:Phe160, G3 interacts with β3:Ala218 and, finally, f5 contributes with β3:Tyr122 and β3:Tyr166 to create an aromatic cage around β3:Arg214 (see Table 3.5 and Figure 3.9). Secondly, in all the decoy poses of the best cluster, the cyclopeptide adopts minimum B conformation. Of note, all the molecules populating this minimum (i.e. c(G**R**GDfL), c(G**R**GD**f**L) and c(G**R**GD**fL**)) were able to generate decoy poses belonging to this cluster.

Overall these data suggest that: i. minimum B conformations, which are generated as consequence of R2 *N*-methylation, possess the ideal structural features to bind αIIbβ3 integrin: herein the appropriate orientation and distance between the pharmacophoric groups guarantee to summarize the canonical electrostatic clamp and to stabilize additional hydrophobic interactions; ii. minimum A conformations allow to recapitulate the canonical interaction but not to engage additional contacts; iii. minima C and D conformations do not satisfy the structural requirements to fit in the αIIbβ3 binding pocket.

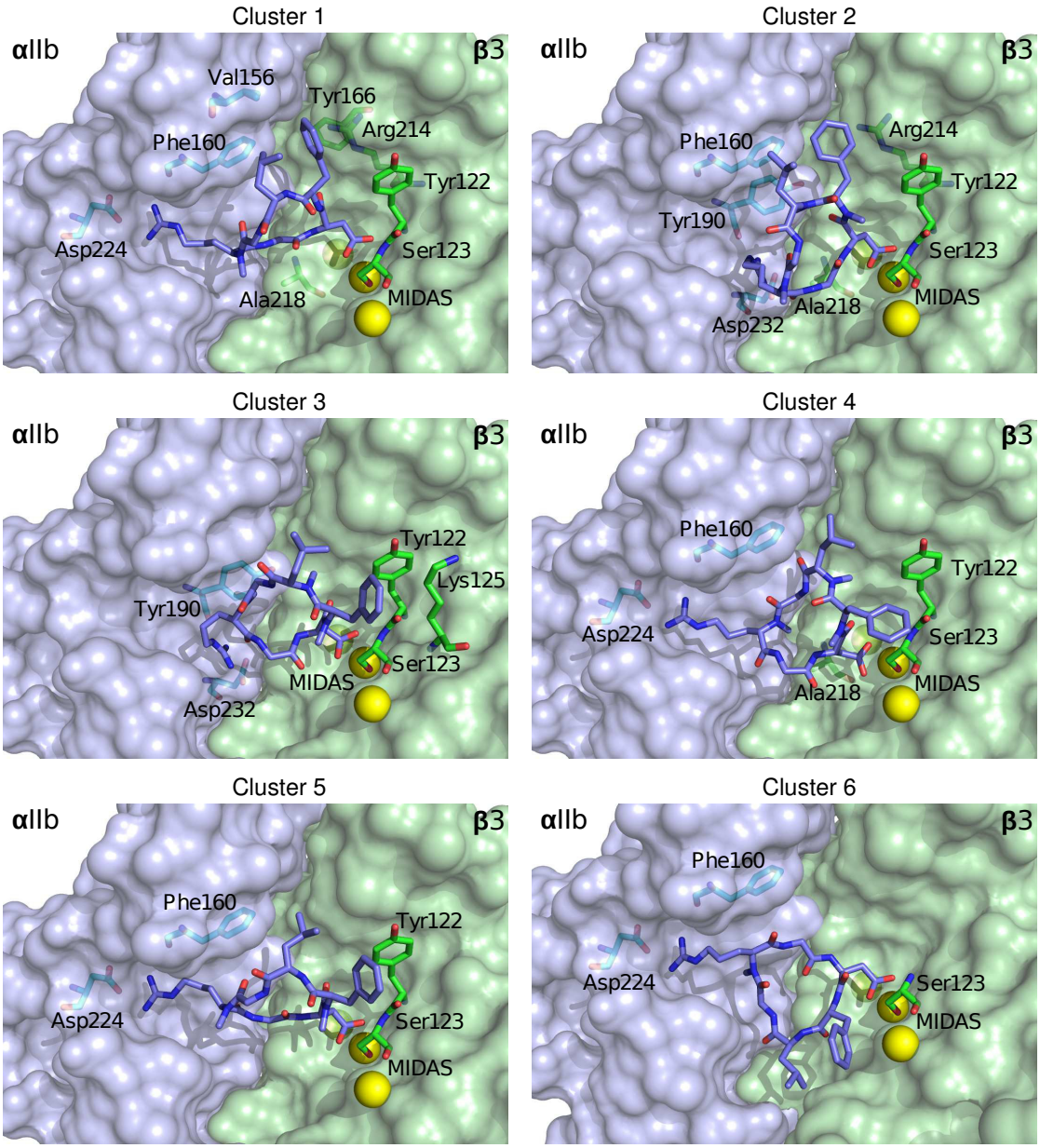


Figure 3.9: **Representative decoy poses of the six best clusters.**[213] Integrin αIIbβ3 is represented as a surface with the αIIb and β3 chains in light-blue and pale green, respectively; receptor amino acids engaging direct interactions with the ligands are shown in stick and labeled with the three-letter code. The metal binding sites, corresponding to MIDAS (Mg^{2+}), ADMIDAS (Ca^{2+}) and LIMBS (Ca^{2+}) cations, are represented as yellow spheres and the peptides are shown in marine blue sticks. Molecular graphics were produced with Pymol.[220]

Table 3.5: **Receptor-peptide interactions.** Summary of the stable interactions engaged by peptides of the best cluster with integrin αIIbβ3. Stable interactions include intermolecular distances shorter than 3 Å, occurring in more than 70% of the decoy poses. Cyclopeptide and receptor residues are labeled with one-letter and three-letter code, respectively.

Stable interactions	
Cyclopeptide residues	αIIbβ3 residues
HX_{R2}[1]	$O_{\alpha IIb:Phe160}$
	$OX_{\alpha IIb:Phe160}$[2]
$H\alpha_{G3}$	$H\beta_{\beta 3:Ala218}$
OX_{D4}[2]	Mg^{2+}
	$HN_{\beta 3:Ser123}$
Ph_{f5}[3]	$HX_{\beta 3:Arg214}$[1]
$H\delta/H\beta_{L6}$	$H\gamma_{\alpha IIb:Val156}$
	$Ph_{\alpha IIb:Phe160}$[3]

[1] HX denotes guanidinium protons.
[2] OX denotes carboxylate oxygens.
[3] Ph denotes phenyl group.

Based on these remarks and on the population distribution of the five cyclopeptides, it is possible to rationalize the interplay between conformational effects induced by *N*-methylation and peptides binding affinity for αIIbβ3. In accordance with the experimental IC_{50} values, cyclopeptides c(G**R**GDfL) and c(G**R**GD**f**L) are predicted to have the best affinity for αIIbβ3: indeed they are pre-organized, already in the free state, in the bioactive conformer B, which is populated more than 70%. In these two peptides minimum B is highly populated because only the effect of R2 methylation is appreciable; in fact the calculations revealed that the multiple methylation of the two adjacent residues f5 and L6 in c(G**R**GD**f**L) does not induce any conformational rearrangement. Conversely, c(G**R**GD**f**L), which equally populates several conformers including both the bioactive minimum B as well as the unproductive minima C and D, must pay an entropic penalty to bind the receptor, which then reflects into a lower binding strength (i.e. higher IC_{50} value, 165 nanomolar). An affinity of the same order of magnitude is displayed also by the unmethylated molecule c(GRGDfL), which only populates minimum A. Herein, the inability to adopt the ideal features characterizing minimum B results in a lower affinity with respect to the molecules that are pre-organized in this

conformation. Finally, cyclopeptide c(GRGD**f**L), in which unproductive minimum C conformations are favored, is characterized by the lowest affinity for the receptor, again obtaining a good correspondence with the available IC_{50} values.

3.3 Conclusion

Cyclopeptides, and especially their chemically modified variants, are emerging as promising therapeutics for the target of protein-protein interactions because of their several favorable properties, in terms of pharmacokinetic, bioavailability, metabolic stability and target specificity. Despite all these advantages their application and development as drugs is hindered by the difficulties to predict *in silico* their three-dimensional structure and their binding strength. Currently, empirical approaches are the most employed ones for the identification, optimization and development of cyclopeptides as potential therapeutics.

In this chapter I have shown how the combination of different computational techniques, namely BE-META, docking calculations and MM-GB/SA rescoring, could be exploited to rationalize and predict the effect of chemical modifications, and specifically backbone *N*-methylation, on receptor affinity for a small set of cyclopeptides. In particular, I have analyzed five cyclopeptides, sharing the same sequence of six amino acids and differing for the position of a single or multiple *N*-methylation, which displayed highly different affinities for αIIbβ3 integrin, ranging from 12 to 2000 nanomolar in terms of IC_{50} values. These divergences in binding strength are expected to be caused by a modulation of the cyclopeptides' preferred conformations rather than by a direct interaction of the modified atoms with the receptor. Herein it has been confirmed that pre-organization of cyclopeptides in solution strongly affects their binding strength to the receptor, demonstrating that the knowledge of the conformations preferentially adopted by the free cyclopeptide, as well as an estimation of their relative equilibrium, is fundamental to provide reliable affinity predictions. In particular it has been shown that: i. BE-META is an ideal tool to individuate the main conformers and assess their population; ii. docking calculations, complemented by MM-GB/SA rescoring and cluster analysis,

allow to classify the conformers according to their binding ability; iii. a joint analysis of the previous outcomes results in a reliable ranking of cyclopeptides based on their binding affinity and in the rationalization of their structure-activity relationship.

From a structural point of view, I observed that introduction of *N*-methylation in the investigated RGD-based cyclopeptides, does not induce a significant *cis-trans* interconversion, neither constrains the molecules in a previously existing backbone conformation, as expected. Conversely, I found that *N*-methylation regulates the structural preferences of cyclopeptides favoring additional conformations, characterized by a *syn*-periplanar geometry of the θ dihedral angle, corresponding to a 180° twist of the peptide bond plane preceding the methylated residue. In this class of cyclic hexapeptides, this effect was found to be particularly significant in the case of R2 *N*-methylation: indeed the induced rearrangement in this site results in the adoption of the bioactive conformation B, allowing a proper relative positioning of the pharmacophoric groups for the binding to αIIbβ3 integrin. The ranking of cyclopeptides according to their αIIbβ3 binding affinity, derived by the combination of conformational sampling and docking calculations, is in excellent agreement with the experimental IC_{50} values, supporting these findings.

It is expected that the computational protocol herein described could be extended and used for other applications in the future. It could be applied to investigate the impact of backbone *N*-methylation and other chemical modifications on diverse set of cyclopeptides, as well as to study the effects of residues' chemistry and stereo-chemistry on cyclopeptides structure and activity. Conceivably, the synergy between BE-META, which guarantees an exhaustive conformational sampling of the cyclopeptides, and docking, which allows the identification of unproductive poses, will allow for a successful application of this protocol, also to screen novel peptides library and before entering expensive synthesis campaign and binding assays.

3.4 Methods

3.4.1 Simulation set up

All the MD simulations have been performed using the Gromacs-4.5.3 software[87] and the Plumed 1 plugin.[221] The TIP3P[113] model was employed to simulate water molecules while the OPLS-AA force field[222] was used for cyclopeptides. This force field has been chosen among the available ones since it had been previously employed in our laboratory for the simulation of cyclic peptides[75] and since it allows a coherent treatment of the system in both the MD sampling and the subsequent docking and rescoring analysis (Prime MM-GB/SA is indeed based on this same force field). The parametrization of the *N*-methylated residues was derived from the OPLS-AA parameters for *N*-methylacetamide, postulating charges and dihedral angles coherence as done in[75]. Each cyclopeptide has been solvated in a cubic box containing 4105 water molecules, with dimension of $(5\ \text{nm})^3$. After an initial energy minimization, a 1 ns equilibration was performed in the NPT ensemble using the md integrator and a time step of 1 fs. The pressure was set to the value of 1 atm using the Berendsen barostat with a relaxation time of 1 ps, while the temperature was maintained at 300 K by using the Berendsen thermostat[96] with a coupling time constant of 100 fs (cyclopeptide and solvent molecules were coupled to independent thermostats). During the equilibration the LINCS algorithm was applied to constraint the bond lengths involving hydrogen atoms. A cut-off of 0.9 nm was used to truncate both the van der Waals and the electrostatic interactions. In the case of electrostatics, long range interactions beyond the cut-off were treated with Particle-Mesh Ewald method (Fourier grid spacing of 0.12 nm and interpolation order of 4).[122,223]
The BE-META simulations, in standard formalism, were performed in the NVT ensemble, setting the temperature at 300 K by means of the v-rescale thermostat.[98] During this production run the time step was set to 2 fs and all the bond lengths have been constrained. Every 1 ps a Gaussian hill (height: 0.24 kcal mol^{-1}, width: 0.2 rad) was added; exchanges between pairs of replica were attempted every 15 ps. Preliminary BE-META simulations have been performed to investigate the impact

of *N*-methylation on *cis-trans* peptide bond interconversion. During each of these simulations 10 CVs have been biased, namely the six Ω ($C\alpha_{i-1}$-C_{i-1}-N_i-$C\alpha_i$) dihedral angles and the four ϕ (C_{i-1}-N_i-$C\alpha_i$-C_i) dihedral angles of the non-Glycine residues (R2, D4, f5 and L6). For each peptide a total simulation time of 400 ns was carried out. Since the calculations did not reveal any significant isomer interconversion, the subsequent investigations were restricted to the *trans*-isomers conformational space. Thus, I ran for each molecule an eight-replica BE-META simulation (320 ns, 40 ns per replica). In each replica one CV, among the eight ϕ and ψ dihedral angles of R2, D4, f5 and L6, was biased.

3.4.2 Analysis of the BE-META simulations

The analysis and the reconstruction of the free energy landscape were performed, using the Metagui tool,[141] in a six-dimensional subspace defined by ϕ and ψ (N_i-$C\alpha_i$-C_i-N_{i+1}) dihedral angles of R2, f5 and L6, where it was possible to identify the main minima and to evaluate their population. The consistency of the reconstruction was assessed by varying the employed grid space (specifically, grid spaces ranging between 20° and 30° were all able to lead to consistent results). In order to consider only the relevant conformations of the system, during the analysis devoted to the identification of metastable states by clustering of BE-META microstates, I choose a number i of relaxation times, when a gap between τ_i and τ_{i+1} was present (see section 2.2.3).
The population of each minimum (X) was estimated according to:

$$\mathrm{pop_X} = \frac{\sum_{\mu \in min(X)} e^{\frac{-F_\mu}{k_b T}}}{\sum_{i=1}^{N} \sum_{\mu \in min(i)} e^{\frac{-F_\mu}{k_b T}}} 100\% \tag{3.2}$$

where k_b is the Boltzmann constant, T is the temperature, N is the total number of minima, μ indicates a generic microstate, min(i) is the ensemble of all the microstates belonging to minimum i and F_μ represents the free energy difference associated to the microstate μ and the microstate with the lowest free energy. Only minima populated more than 10% have been considered for further analysis and docking calculations.

3.4.3 Comparison with NMR data

The ${}^3J_{\mathrm{HN,H}\alpha}$ were computed using the Karplus equation:[224]

$$ {}^3J_{\mathrm{HN,H}\alpha}(\theta) = A\cos^2(\theta) + B\cos(\theta) + C, \tag{3.3} $$

where $A = 5.4\,\mathrm{Hz}$, $B = $-1.3 Hz, $C = 2.2\,\mathrm{Hz}$,[225] $\theta = \phi - 60°$ in the case of L-amino acids and $\theta = \phi + 60°$ in the case of D-amino acids. Average values for the ϕ dihedral angles ($\langle\phi\rangle$) were obtained from the ensembles of structures extracted from the BE-META main minima (minimum A and B). Then the weighted average for ${}^3J_{\mathrm{HN,H}\alpha}$ was obtained as:

$$ {}^3J_{\mathrm{HN,H}\alpha}(\theta) = \frac{J(\langle\theta\rangle_{\mathrm{A}})\,\mathrm{pop}_{\mathrm{A}} + J(\langle\theta\rangle_{\mathrm{B}})\,\mathrm{pop}_{\mathrm{B}}}{\mathrm{pop}_{\mathrm{A}} + \mathrm{pop}_{\mathrm{B}}}, \tag{3.4} $$

where $\mathrm{pop}_{\mathrm{A}}$ and $\mathrm{pop}_{\mathrm{B}}$ are the populations of minima A and B, respectively; while $\langle\theta\rangle_{\mathrm{X}}$ is the average value for θ (computed as $\langle\phi\rangle \pm 60°$) within minimum X.

Analogously the averages $d_{\mathrm{CN,H}\alpha}$ distances have been calculated according to:

$$ d_{\mathrm{CN,H}\alpha} = \frac{\langle d_{\mathrm{CN,H}\alpha}\rangle_{\mathrm{A}}\,\mathrm{pop}_{\mathrm{A}} + \langle d_{\mathrm{CN,H}\alpha}\rangle_{\mathrm{B}}\,\mathrm{pop}_{\mathrm{B}}}{\mathrm{pop}_{\mathrm{A}} + \mathrm{pop}_{\mathrm{B}}}, \tag{3.5} $$

where $\langle d_{\mathrm{CN,H}\alpha}\rangle_{\mathrm{X}}$ is the average distance within minimum X.

3.4.4 Docking Calculations

The software Glide[152] of the Schrödinger suite (version 5.8) was employed to perform docking runs. As input structures for the ligands I used the conformers extracted from the BE-META minima populated more than 10% (for a total of 1480 conformers). Since these structures had already been subjected to an exhaustive conformational search, no further energy minimization or sampling was performed. The receptor input structure (PDB: 2vdn[14], crystallographic αIIbβ3 in complex with Eptifibatide) have been prepared with the Protein Preparation Wizard tool of the Maestro Package.[226] Herein I removed all the crystallographic waters, added

missing hydrogens, optimized protonation and tautomeric states; lastly I ran a restrained minimization, employing the OPLS-AA force field,[222] with an RMSD tolerance on heavy atoms of 0.3 Å to remove steric clashes and reorient side-chain hydroxyl groups. The grid center was determined by using as reference ligand Eptifibatide, which afterwards was removed. The default cubic grid dimensions were adopted.
The decoy poses have been generated by using the Extra-Precision mode of Glide with flexible ligand (that in the case of head-to-tail cyclopeptide means flexible side-chain) and without applying any constraints. For each input structure only the best decoy pose was printed. After docking calculations the poses were ranked according to their Extra-Precision Glide Score (GS). A first filter was then applied ($GS < (GS_{best} + 5\,kcal\,mol^{-1})$) and only poses satisfying this condition were retained for MM-GB/SA rescoring and further analysis.

3.4.5 MM-GB/SA and cluster analysis

All the selected poses have been refined and rescored using the Prime MM-GB/SA tool of the Schrödinger suite. For the refinement the local optimization feature was utilized,[178] while for the rescoring it was employed a combination of the OPLS/AA molecular mechanics force-field and of the GB/SA continuum solvent model.[227] During this procedure the backbone and the Cβ atoms of the ligand were kept fixed, while the other atoms were considered flexible. Concerning the receptor, all the residues within a 5 Å cutoff from the ligand were treated as flexible.
A cluster analysis was subsequently performed using the *cluster_struc* program provided with HADDOCK.[153,198] The poses were aligned on the receptor residues within the binding pocket (i.e. residues within 10 Å from the ligand) and the RMSD was computed on all the backbone atoms and on the side-chain atoms of R2, G3, D4 and f5. For the clustering an RMSD cutoff of 2.5 Å was employed and clusters containing less than 10 poses were excluded. For each of the obtained clusters, an average MM-GB/SA value was assigned, by averaging the scores of all the members. An average RMSD was also computed, using as reference the lowest energy structure of the best cluster (in term of MM-GB/SA score). Finally a second filter was applied

($\langle$MM-GB/SA$\rangle$ < ($\langle$MM-GB/SA$\rangle_{\text{best cluster}}$ + 20 kcal mol^{-1})), in order to identify the most reliable clusters of poses. These poses have been subjected to visual inspection, investigation of energetic contributions and analysis of intermolecular stable interactions (i.e. distances shorter than 3 Å, occurring in more than 70% of the decoy poses).

Chapter 4

Structural characterization of isoDGR-based cyclopeptides

Integrins belonging to the class of the RGD (Arg-Gly-Asp) recognizing receptors constitute an ideal pharmacological target because of their role in several pathological processes, including thrombosis, stroke (αIIbβ3), angiogenesis and cancer (αvβ3, α5β1, αvβ6, αvβ8).[46] These integrins share the ability to bind extracellular matrix proteins through the three amino acidic sequence RGD, therefore several peptides and peptidomimetics mimicking this recognition motif have been developed and studied in the last decades as integrin inhibitors.[63,64]

In 2006, biochemical studies showed that the NGR (Asn-Gly-Arg) motif of the extracellular matrix protein fibronectin can spontaneously transform into the isoDGR (isoAsp-Gly-Arg) sequence, resulting in a gain of function for the protein and in the generation of a new adhesion binding site for integrins.[70] Studies on CisoDGRC, a disulphide-bond cyclic peptide mimicking the isoDGR sequence in fibronectin, demonstrated that CisoDGRC can fit into the RGD-binding pocket of αvβ3 integrin, recapitulating the canonical electrostatic clamp.[71] The same cyclopeptide was also able to inhibit endothelial cell adhesion, proliferation, and tumor growth.[70] Subsequent investigations showed also that, at variance to Cilengitide, CisoDGRC behaves as pure integrin antagonist blocking αvβ3 conformational changes and allosteric activation.[72] In recent years some isoDGR-based head-to-tail cy-

clopeptides have been developed and synthesized: for the most interesting ones structure-activity relationship studies have been conducted in order to rationalize their affinity and selectivity profiles and their ability to work as drugs-carriers has also been explored.[76–79]

Considering that this is a relatively recent research field, and since none of the developed peptides displays an affinity for $\alpha v\beta 3$ integrin comparable to the one of Cilengitide, new efforts are needed to explore the potential of isoDGR-based cyclic peptides as antiangiogenic therapeutics, drug carriers or agents for diagnostic imaging. In this context, the design and development of novel isoDGR-based integrin therapeutics could benefit from a systematic computational screening of the impact of the chemistry/stereo-chemistry of isoDGR flanking residues on ligands conformation and consequently on their integrin binding affinity. To this aim I propose to use the computational protocol previously presented in Chapter 3, relying on the combination of BE-META, docking and rescoring techniques. Nevertheless, two major issues could compromise this project: i. the force fields implemented in MD software could be not appropriate for the *in silico* structure determination of highly constrained cyclic peptides containing non-standard amino acids as isoAspartate;[228] ii. docking software could not properly evaluate differences in binding affinities upon modification of the stereo-chemistry and/or type of residues flanking the recognition motif.
These two aspects have been thoroughly investigated within this thesis. Specifically, in the next Section I have systematically tested the ability of eight different force field variants to reproduce NMR-derived experimental data for 5 isoDGR-based cyclic peptides. Herein I have shown that only two out of eight investigated force fields allowed to accurately reproduce the equilibrium properties of isoDGR-containing molecules, suggesting that transferability of parameters is a not-trivial issue. Afterwards, in Section 4.2 I have faced the problem of binding affinity predictability via docking calculation techniques. Specifically, I investigated the rationale for the affinity/selectivity profiles displayed by three cyclic head-to-tail isoDGR hexapeptides, differing for the type of residues flanking the isoDGR motif. Herein I have shown that the combination of Glide docking and MM-GB/SA rescoring allows to discriminate integrin binders (inhibition constants in the order

of tens of nanomolar) from weak binders (inhibition constants in the order of hundreds of nanomolar).
Overall, the positive results obtained holds promise for the future applications of the proposed computational protocol, which could be employed in the screening of the effects of isoDGR flanking residues on cyclopeptide conformation and integrin affinity.

4.1 Are current MD force fields reliable for the study of isoDGR-cyclopeptides?

4.1.1 Introduction: the force field problem

In the previous Chapters I have evidenced the increasing importance of cyclic peptides in the modulation of protein-protein interactions and specifically in the targeting of integrins.[53] Their constrained cyclic geometry supplies several advantages: firstly, they are characterized by improved stability and bioavailability with respect to their linear counterpart, moreover they avoid the possible side effects of small molecules. Importantly, cyclization allows the peptides to adopt, already in the unbound state, specific bioactive conformations, suitable for the selective binding to the receptor.[201]
In this context, the knowledge of the conformations preferentially adopted by cyclic peptides in their free state can help in the prediction of their binding properties.[207] Computational tools could be extremely useful to this aim. We have previously shown that enhanced sampling techniques, such as Metadynamics in its Bias Exchange variant,[134] provide a perfect framework to exhaustively sample the conformations of cyclopeptides, guaranteeing a full coverage of the relevant conformational space.[75,213] Nevertheless, the reliability of the sampling strongly depends also on the accuracy of the molecular force field employed in the MD simulation. This aspect is particularly relevant in the case of cyclic peptides, and it is even more remarkable in simulations requiring non standard amino acids.[199,228,229]

Molecular force fields are energy functions that determine the potential energy necessary to numerically solve the Newton equation of motion during the MD simulation (as described in Section 2.1.3). They consist in the sum of several energetic terms, which account for both bonded (i.e. bond stretching, angle bending and torsion potential) and non-bonded (i.e. Coulomb and van der Waals) interactions. Their general form has been reported in equations 2.4 and 2.5, where their dependencies upon several parameters is highlighted. These parameters are usually determined in order to reproduce high level Quantum Mechanics calculations for small molecules,

and are then optimized to fit experimental data (e.g. crystallography and NMR spectroscopy). This set of parameters, along with the specific functional form of the energetic function, determines the specificity of the force field.
The gain in computational power observed in the last decade, which allowed accessibility to longer simulation times, has unveiled some weaknesses and limitations in force fields accuracy. This fact has stimulated the development of new force field variants, in which improved torsional potentials for both backbone and side-chain atoms have been proposed. Additionally, an interest in testing force fields accuracy in diverse situations raised, thus allowing a better comprehension of their limitations sources.[109–112,230–232]

The problem of force field accuracy in the simulation of cyclic peptides is even more prominent. Indeed molecular force fields are generally developed and tested on protein systems or linear peptides, but they are not designed to reproduce the equilibrium properties of constrained cyclic peptides. In this context, the canonical dihedral angles distribution observed in a Ramachandran plot could not be satisfied due to the constraints imposed by the cyclic geometry. Recently this issue has gained some attention.[199,229] Specifically, the group of Wu[229] has tested the ability of some popular force fields, including their recently developed force fields RSFF1 and RSFF2 (Residue Specific Force Field 1 and 2),[107,108] in reproducing the crystallographic structures of twenty 5-12 residues cyclopeptides. The simulation of isoDGR-based cyclic peptides, where the conformation adopted by isoAspartate has relevant consequence on their integrin binding ability, could be even more challenging. Herein the presence of the non standard amino acid isoAspartate could further affect the predictability and accuracy of the simulation. Indeed, despite the high similarity shared by isoAspartate backbone and Asparagine side-chain, it is not obvious that the parameters, and particularly the torsion potentials, optimized for Asparagine side chain could be transferable to isoAspartate (see Figure 4.1).

In order to evaluate the ability of MD simulations to reproduce the equilibrium properties of isoDGR-based cyclic peptides I have performed a systematic study testing some popular force fields on 5 peptides, for which NMR experiments have been acquired. The 5 investigated molecules include two pentapeptides cycled through a

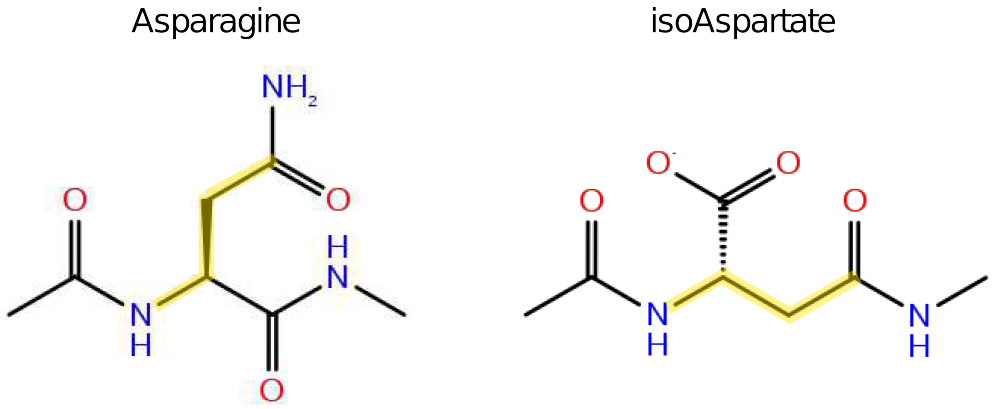


Figure 4.1: **Asparagine and isoAspartate dipeptides.** The high similarity shared by isoAspartate backbone and Asparagine side-chain is shown in a 2D-representation. Specifically the characteristic backbone dihedral angle of isoAspartate, and the corresponding dihedral in Asparagine, are highlighted in yellow.

disulfide bond, namely CisoDGRC and its acetylated variant $_{ac}$CisoDGRC,[70,75] and three head-to-tail cyclic hexapeptides, c(CGisoDGRG), c(GCisoDGRG) and c(CphgisoDGRG) (see Figure 4.2).[76] Each of the 5 cyclopeptides was simulated with 8 different force field variants in explicit water, for a total of 40 simulations. To guarantee an exhaustive coverage of the conformational space Bias Exchange Metadynamics was employed, for a total simulation time of 17 μs.

The selected force fields comprise the most recent variants among the major families of force field (AMBER, OPLS, CHARMM and GROMOS), if implemented in the Gromacs software. These include: AMBER ff99sb-ildn,[101,233] AMBER ff99sb*-ildn,[216,234] AMBER ff14sb,[106] OPLS-AA/L,[102,222] CHARMM-27[103,235] and GROMOS-54a7.[236] Additionally, two variants of the AMBER ff99sb-ildn and of the OPLS-AA/L force fields (denoted as ff99sb-ildn_ASN and OPLS-AA/L_ASN respectively) were included: herein the torsional potentials specifically optimized for Asparagine side-chain have been used in the parameterization of isoAspartate. In particular, AMBER ff99sb-ildn force field is an improvement of the f99 force field, developed in 1999 by the Kollman group.[237] Herein the energy functions associated to both the backbone ϕ/ψ dihedral terms (sb correction[101]) and to the side-chain χ dihedral angles of Isoleucine, Leucine, Aspartate and Asparagine (ildn correction[233]) have been optimized to fit high level *ab initio* quantum mechanical

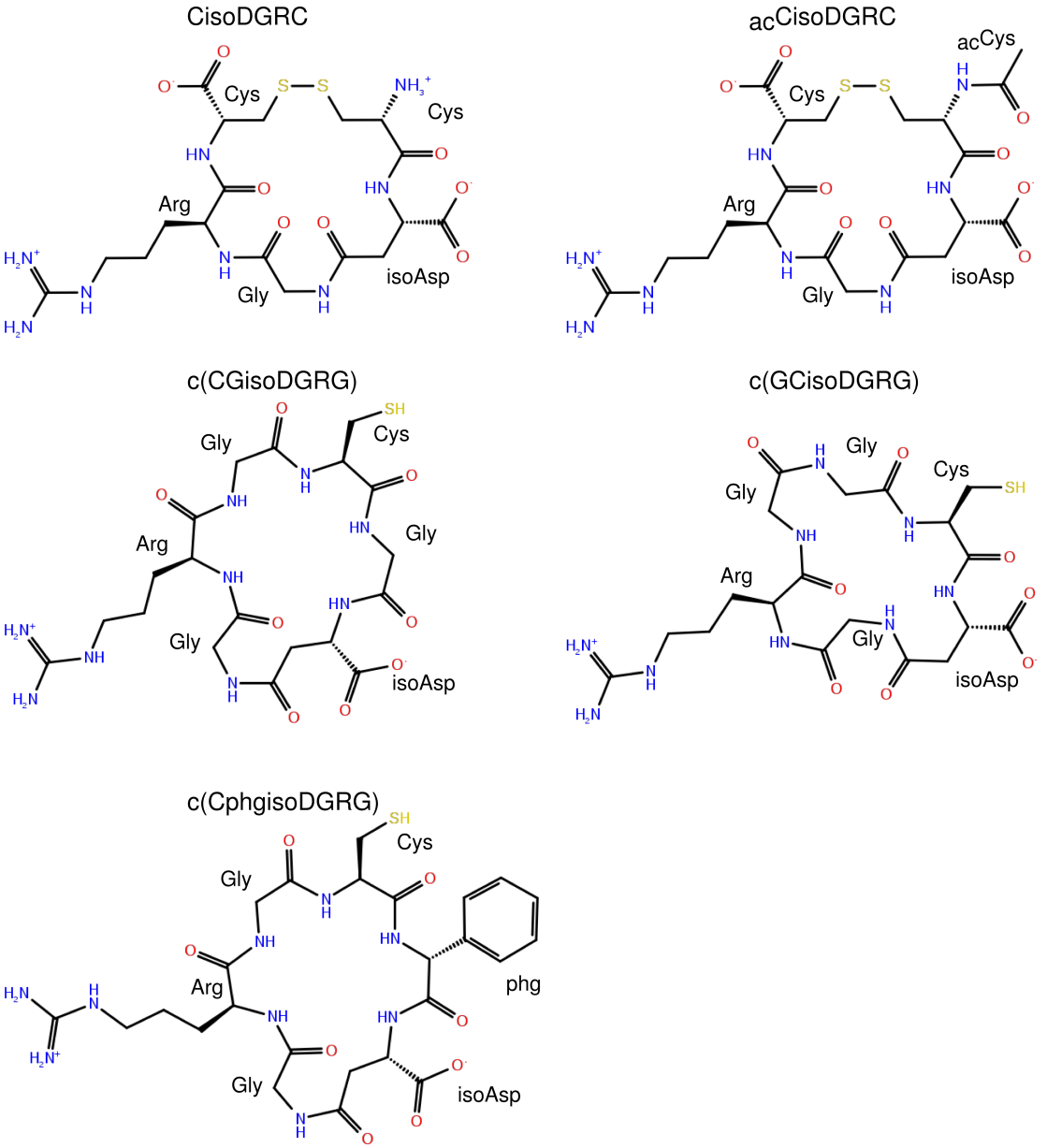


Figure 4.2: **2D-representation of the five molecules used for the force field evaluation.** 2D-structures of the isoDGR-based cyclopeptides CisoDGRC, $_{ac}$CisoDGRC, c(CGisoDGRG), c(GCisoDGRG) and c(CphgisoDGRG).

calculations and NMR-derived experimental measurements. In the force field variant denoted as ff99sb-ildn_ASN, the torsional parameters optimized for Asparagine within the ildn correction are used for isoAspartate. The force field ff99sb*-ildn differs with respect to ff99sb-ildn for simple backbone energy corrections, which were introduced to obtain a correct balance of secondary structure propensities.[216] The last force field considered for the AMBER family is the recently optimized ff14sb:[106] herein, starting from the ff99 force field with the sb correction, QM-based side-chain dihedral parameters and empirical adjustment to the backbone torsional energies have been introduced. Within the OPLS family of force field I have considered OPLS-AA/L:[222] it is an improvement of the original OPLS[102] force field where quantum mechanical data are used for refitting of side-chain and torsional parameters. Unfortunately, I have not been able to test the recently optimized OPLS-AA/M force field,[105] since a version compatible with Gromacs is currently not available. Concerning the CHARMM group I selected the CHARMM-27 variant, in which the CMAP correction, a grid-based energy correction map for backbone dihedral angles was introduced.[103,235] Finally, for the GROMOS family of force fields, which are characterized by the use of an united-atom treatment, I selected the relatively recent GROMOS-54a7,[236] where several aspects have been improved with respect to the preceding versions, including the secondary structure propensities, the N-H, C=O repulsion, the free energy of hydration of Na^+ Cl^- ions, as well as new improper dihedral angles. The new interesting Residue Specific Force Fields (RSFF1 and RSFF2)[107,108] have not been included in the evaluation since they could be not appropriate to simulate molecules containing non traditional amino acids.

The accuracy of each force field was evaluated by comparison with NMR data. Indeed NMR, thanks to its intrinsic ability to quantify the dynamic of a system, is an ideal technique to study flexible molecules and peptides, which can show a high internal mobility populating several conformational minima. NMR-derived observables, including J scalar couplings, residual dipolar couplings, nuclear Overhauser enhancement and chemical shifts, are averages of the values expected for each of the multiple possible conformations, if these conformations are in fast exchange with respect to the NMR time-scale. This condition is satisfied by cyclic peptides so that

only averages can be measured. In this situation, the experimental values should quantitatively correspond to the ensemble averages of the corresponding values computed with simulations. This is true if the following assumptions are satisfied: i. the sampling of configuration is exhaustive, ii. the relation used to compute the observable from the 3d configuration is correct, iii. the empirical potential function (or force field) that governs the statistical-mechanical distribution of configurations used to obtain a proper Boltzmann-weighted ensemble is reliable.[238]
For all these reasons NMR-based data constitute a commonly used benchmark to quantitatively assess the accuracy of force fields.[110,231,238] In this project I have used chemical shifts and J scalar couplings data. The latter are particularly relevant in the study of cyclic peptides since they are related to dihedral angles and bond distances. The agreement between experimental and back-calculated data has been quantify by means of a χ^2 metrics. These comparisons suggest that, while most of the investigated force fields can properly reproduce the equilibrium J couplings of cyclic peptides, only two (i.e. ff99sb-ildn and ff99sb*-ildn) are able to recover the NMR observables characteristic of isoAspartate with an accuracy close to the systematic uncertainty. Overall these data demonstrate that the parameterization of isoAspartate is not trivial and that care must be taken when introducing non canonical amino acids in the simulation.

4.1.2 Results and Discussion

In order to evaluate the ability of MD simulations to reproduce the equilibrium properties of isoDGR-based cyclic peptides I tested 8 force field variants on 5 peptides, for a total of 40 BE-META simulations. Each simulation has been performed biasing all the dihedral angles related to the cycles of the five molecules, with the exception of the Ω dihedral angles describing peptide bonds planarity, running 30 or 40 ns per replica (see 4.1.4.2 for details). After an accurate check of the convergence, the simulated structures have been clustered in microstates in a space defined by a subset of the biased CVs (see section 4.1.4.3), and the microstates free-energies have been estimated through a WHAM procedure.[138] For each of the structures clustered in microstates, the Karplus equation has been

used to back-calculate the J coupling constants, while the Sparta+ program[239] was employed to estimate the chemical shifts. Then, the average equilibrium values of these observables have been computed according to equation 4.1.[138]
In order to evaluate the ability of the selected force fields to reproduce experimental quantities the χ^2 metric reported in equation 4.2 was used. Herein, for each force field, the collected experimental data, including both scalar couplings and chemical shifts, are quantitatively compared with the ensemble average values of the same observables retrieved from BE-META simulations. The deviations between experimental and computed values are weighted with the uncertainty associated to these observables. These errors are the ones arising from the parameterization of either the Karplus equation or the Sparta+ chemical shift model (see Table 4.2 and 4.3).

4.1.2.1 Force field performances

Chemical Shifts

I firstly investigated the ability of the different force fields to reproduce chemical shift values. To this aim a set of 38 chemical shifts was considered, involving the C, Cα, Cβ, HN and Hα atoms of natural amino acid (see Section 4.1.4.1). As shown in Figure 4.3, for all the force fields a similar deviation between predicted and experimental chemical shifts was observed, with a χ^2 ranging from 0.5 to 0.75. Also the analysis of chemical shift predictions, subdivided per type (Figure B.1), does not evidence significant differences between force fields.
The good agreement (χ^2 always lower than 1) observed in all the cases is quite surprising for mainly two reasons: i. different behaviors were expected depending on the force field; ii. several limitations can affect the chemical shift prediction performed with the software Sparta+. Indeed this tool is based on empirical approaches and has been trained on proteins, limiting its applicability to small peptides with cyclic topology and containing non standard residues (e.g. Phenylglycine and isoAspartate).[240]
This unexpected result could be explained observing that the deviation between the experimental chemical shifts and the corresponding random coil values (as

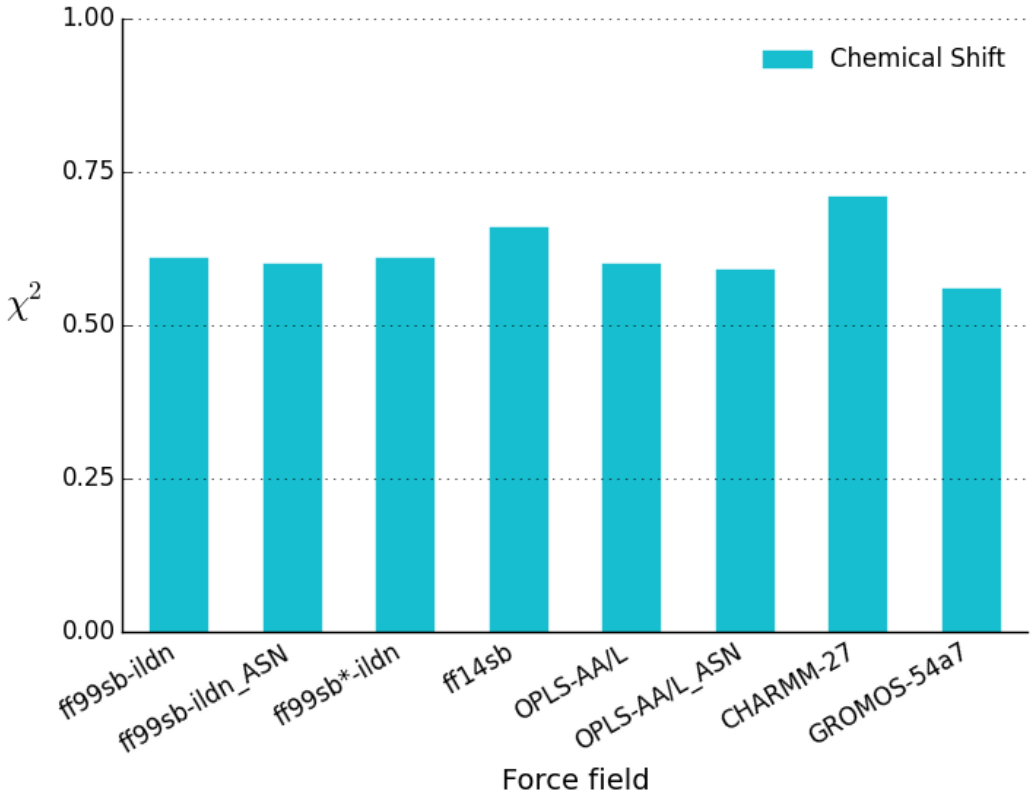


Figure 4.3: **Chemical shift χ^2.** For each force field, the ability to reproduce experimental chemical shifts is quantified through the χ^2 function.

provided by Sparta+) was negligible, indicating that the investigated peptides are highly flexible. Since I believe that this kind of comparison for this specific system is not significant to evaluate force field performance, the following of this chapter will be focused on the analysis of J scalar couplings.

J scalar couplings

Since it has been reported that coupling constants calculations are sensitive to the implementation of Karplus parameters,[106,230] I evaluated the agreement between predicted and experimental $^3J_{\mathrm{HN,H}\alpha}$ scalar couplings using two different sets of Karplus parameters (ORIG[241] and DFT[242], see section 4.1.4.3). In Figure B.2

in Appendix B the χ^2 values, estimating the accordance between experimental and computed $^3J_{\mathrm{HN,H}\alpha}$, obtained with these two different parameterization sets are compared. Even if some differences can be observed, mainly for the force fields displaying poor agreement with experiments (i.e. with high χ^2 values), the overall trend is maintained. These data suggest that this evaluation of force fields, according to their ability to reproduce the equilibrium properties of the systems, is robust with respect to variations of the Karplus parameters. In the following, the results obtained with the ORIG set of parameters will be present. For sake of completeness, the same data obtained using the DFT set, will be reported in Appendix B. It is easy to see that both the sets lead to consistent conclusions.

To evaluate the agreement between experimental and calculated J scalar couplings I analyzed separately $^3J_{\mathrm{HN,H}\alpha}$ and $^3J^{\mathrm{isoD}}_{\mathrm{H}\alpha,\mathrm{H}\beta}$. The first of these observables can provide general information about force field accuracy in simulating cyclic peptides, since it is related to the ϕ dihedral angle common to all the amino acids. Conversely, the latter observable is more focused on the appropriateness of isoAspartate parameterization: indeed $^3J^{\mathrm{isoD}}_{\mathrm{H}\alpha,\mathrm{H}\beta}$ couplings are related to the characteristic dihedral angle ξ (N-Cα-Cβ-C), which is present in isoAspartate backbone.
The χ^2 values, estimating the ability of the different force fields to reproduce $^3J_{\mathrm{HN,H}\alpha}$, $^3J^{\mathrm{isoD}}_{\mathrm{H}\alpha,\mathrm{H}\beta}$ and both these variables together, are reported in Figure 4.4.
Concerning the $^3J_{\mathrm{HN,H}\alpha}$ scalar couplings (orange bars), the χ^2 spans from a minimum value of 1.14, in the case of AMBER ff99sb-ildn_ASN force field, to a maximum of 5.53, for the OPLS-AA/L force field. Herein, the AMBER force field ff99sb-ildn, its two variants ff99sb-ildn_ASN and ff99sb*-ildn, as well as the CHARMM-27 force field, perform exceptionally well, displaying a χ^2 lower than 1.3. Good results are also achieved by the recently optimized AMBER ff14sb (χ^2=1.93), while a slightly worse accordance with experiments is observed for GROMOS-54a7. Finally, the OPLS-AA/L force field and its variants OPLS-AA/L_ASN show a poor/moderate agreement with experiments. This is not surprising since it is well recognized that a major weakness of OPLS-AA/L force field consists in its poor ability to reproduce quantities relying on torsional energetics.[105] Additionally, it must be considered that force fields have been optimized to reproduce protein dynamics and could be not perfectly suited for the simulation of small constrained

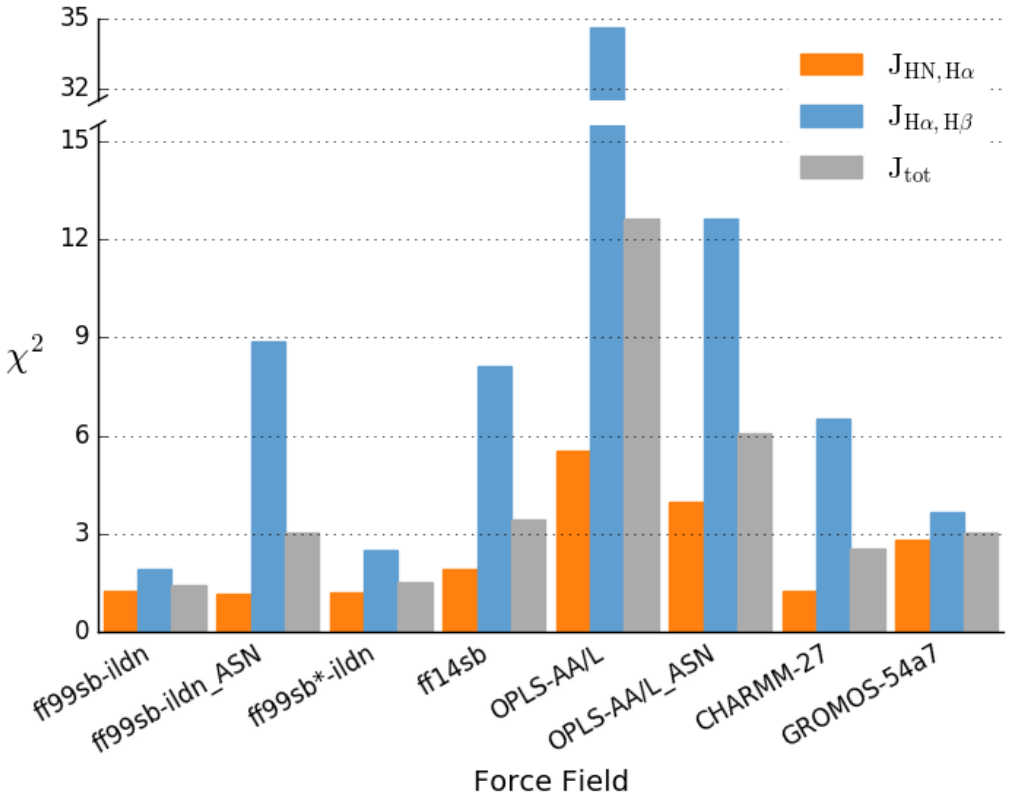


Figure 4.4: **Comparison between experimental and computed J scalar couplings.** For each force field, the ability to reproduce experimental J scalar couplings is quantified through the χ^2 function. The χ^2 values obtained, considering alternatively ${}^3J_{\mathrm{HN,H}\alpha}$, ${}^3J^{\mathrm{isoD}}_{\mathrm{H}\alpha,\mathrm{H}\beta}$, or both these variables together, is shown with orange, blue and gray bars, respectively. In order to avoid flattening of the data a gap in the y axis has been introduced.

cyclic peptides containing non standard residues.

A considerably worse general agreement is observed for ${}^3J^{\mathrm{isoD}}_{\mathrm{H}\alpha,\mathrm{H}\beta}$ scalar couplings (blue bars). Herein, only two force fields, the ff99sb-ildn and its variant ff99sb*-ildn, obtained a satisfactory agreement with experiments (χ^2=1.90 and 2.51, respectively). All the other force fields revealed their inability to properly sample isoAspartate conformations, displaying χ^2 values bigger than 3, and in most of the cases bigger than 6. These results suggest that the dynamic of isoAspartate backbone within head-to-tail cyclic peptides can not be properly approximated by the one of As-

paragine side-chain. Conceivably, the environment in which Asparagine parameters, and particularly torsional angle energetics, have been optimized is too diverse with respect to the context in which isoAspartate is found (at least when it is included in cyclic peptides). This remark is corroborated by the fact that the variant ff99sb-ildn_ASN, in which torsional potentials specifically optimized for Asparagine have been used in isoAspartate parameterization, does not improve the performance of ff99sb-ildn force field, but conversely worsens its agreement with experimental ${}^{3}J^{\mathrm{isoD}}_{\mathrm{H}\alpha,\mathrm{H}\beta}$.
Of note, the performances of ff99sb-ildn and ff99sb*-ildn are similar in both the comparisons with ${}^{3}J_{\mathrm{HN,H}\alpha}$ and ${}^{3}J^{\mathrm{isoD}}_{\mathrm{H}\alpha,\mathrm{H}\beta}$ scalar couplings. These two force fields indeed merely differ for a simple backbone energy correction,[216] introduced to achieve a better balance of proteins' secondary structure propensities. Reasonably, this modification is not relevant in the simulation of cyclic peptides. Additionally, it must be observed that the residues optimized within the ildn correction (i.e. Isoleucine, Leucine, Aspartate and Asparagine) are not present in the investigated cyclopeptides. Therfore it is expected that, in this test, analogous performances could be achieved using the force fields variants ff99sb and ff99sb*.

It is worth noting that ${}^{3}J_{\mathrm{HN,H}\alpha}$ couplings are related to the dihedral angle H-N-Cα-Hα, which is common to all the amino acids, including isoAspartate. I thus investigated whether the inclusion of these isoAspartate-related observables could affect the accuracy of force fields in reproducing ${}^{3}J_{\mathrm{HN,H}\alpha}$ scalar couplings. In Figure 4.5 for each force field the χ^2 values, estimating agreement with the experimental ${}^{3}J_{\mathrm{HN,H}\alpha}$ couplings, computed including or excluding the coupling constants associated to isoAspartic residue, are reported. Since no relevant differences between the two situations can be detected, it can be concluded that these isoAspartate-related observables can be reproduced by force fields with an accuracy comparable to the one of the other amino acids.

Overall the reported data suggest that all the AMBER force fields considered (specially ff99sb-ildn and ff99sb*-ildn) and the CHARMM-27 force field are adequate to reproduce with good accuracy cyclopeptides ${}^{3}J_{\mathrm{HN,H}\alpha}$ scalar couplings. A worse agreement with these observables was found for GROMOS-54a7 and for OPLS-AA/L force field, which has confirmed its weakness in predicting equilibrium

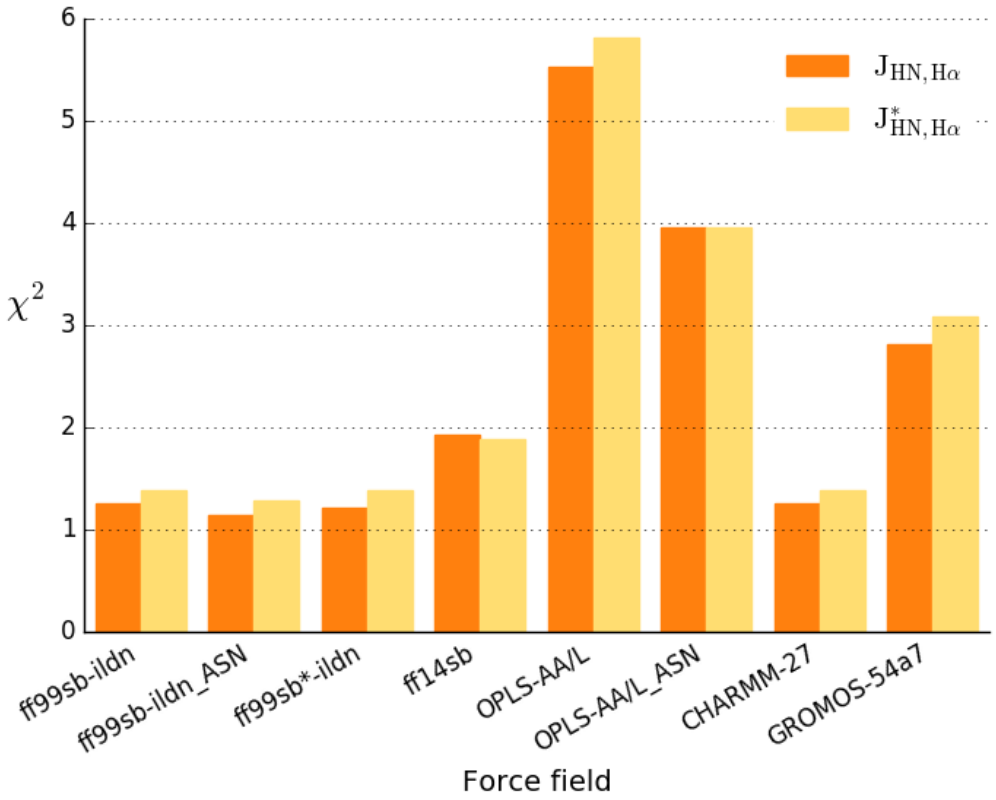


Figure 4.5: **Influence of isoAspartate on $^3J_{\mathbf{HN,H\alpha}}$ reproducibility.** For each force field, the ability to reproduce experimental $^3J_{\mathrm{HN,H\alpha}}$ scalar couplings is quantified through the χ^2 function. The χ^2 values, computed including or excluding the coupling constants associated to isoAspartate H-N-Cα-Hα dihedral, are reported as orange and yellow bars, respectively.

properties associated to torsional dihedral angles. Concerning the isoAspartate parameterization, it can be observed that most of the force fields do not provide an accurate description of its equilibrium conformations. Simulations of peptides containing isoAspartate (or more generally β amino acids) could thus benefit by an *ab initio* parameterization of these residues. Herein it has been shown that only the AMBER force fields ff99sb-ildn and ff99sb*-ildn allow to obtain a satisfactory reproducibility of $^3J^{\mathrm{isoD}}_{\mathrm{H\alpha,H\beta}}$, providing a good framework for the simulation of isoDGR-based peptides.

4.1.2.2 Influence of isoAsp parameterization on force field performance

I finally investigated whether the parameterization of isoAspartate ξ (N-Cα-Cβ-C) dihedral angle could affect the overall force field performance, influencing the conformations, and thus the back-calculated scalar couplings, of the adjacent atoms and residues. To this aim I focused the attention on two force fields: AMBER ff99sb-ildn and OPLS-AA/L, for which alternative torsion potentials for isoAspartate dihedral angles, based on the parameters optimized for Asparagine side-chains,[222,233] have been considered (AMBER ff99sb-ildn_ASN and OPLS-AA/L_ASN, respectively).

In Figure 4.6 the comparison between the $^3J_{\mathrm{HN,H}\alpha}$ scalar couplings computed with either ff99sb-ildn and OPLS-AA/L force field or with their ASN variants is reported. The $^3J_{\mathrm{HN,H}\alpha}$ scalar couplings predicted with the two AMBER ff99sb-ildn force field variants are highly correlated (R^2=0.94), suggesting that the different torsion potentials employed for isoAspartate dihedral angles have only a local effect and do not influence the conformation of the surrounding residues. Consistently, the χ^2 values, concerning $^3J_{\mathrm{HN,H}\alpha}$, displayed by these two force field variants are highly similar. Conversely, quite different $^3J_{\mathrm{HN,H}\alpha}$ coupling constants are predicted when using diverse isoAspartate parameterization in OPLS-AA/L (R^2=0.58). Conceivably, in this framework, the conformations of residues and atoms adjacent to isoAspartate are affected by the choice of the torsion potentials used to describe this amino acid. To investigate which residues are mostly affected by the different OPLS parameterizations, I re-performed the same comparison excluding alternatively:i. the $^3J_{\mathrm{HN,H}\alpha}$ couplings of isoAspartate residue (Figure 4.7 B); ii. the $^3J_{\mathrm{HN,H}\alpha}$ couplings of the residue preceding isoAspartate (Figure 4.7 C); iii. the $^3J_{\mathrm{HN,H}\alpha}$ couplings of the two residues preceding isoAspartate (Figure 4.7 D); iv. the $^3J_{\mathrm{HN,H}\alpha}$ couplings of the residue following isoAspartate(Figure 4.7 E); v. the $^3J_{\mathrm{HN,H}\alpha}$ couplings of the two residues following isoAspartate (Figure 4.7 F). It can be observed that the exclusion of the scalar couplings associated to isoAspartate and to its preceding residues does not help in improving the correlation (Figure 4.7 B, C and D). Conversely, the exclusion of the two residues following isoAspartate allowed to obtain a high correlation between the $^3J_{\mathrm{HN,H}\alpha}$ computed with the two OPLS-AA/L variants

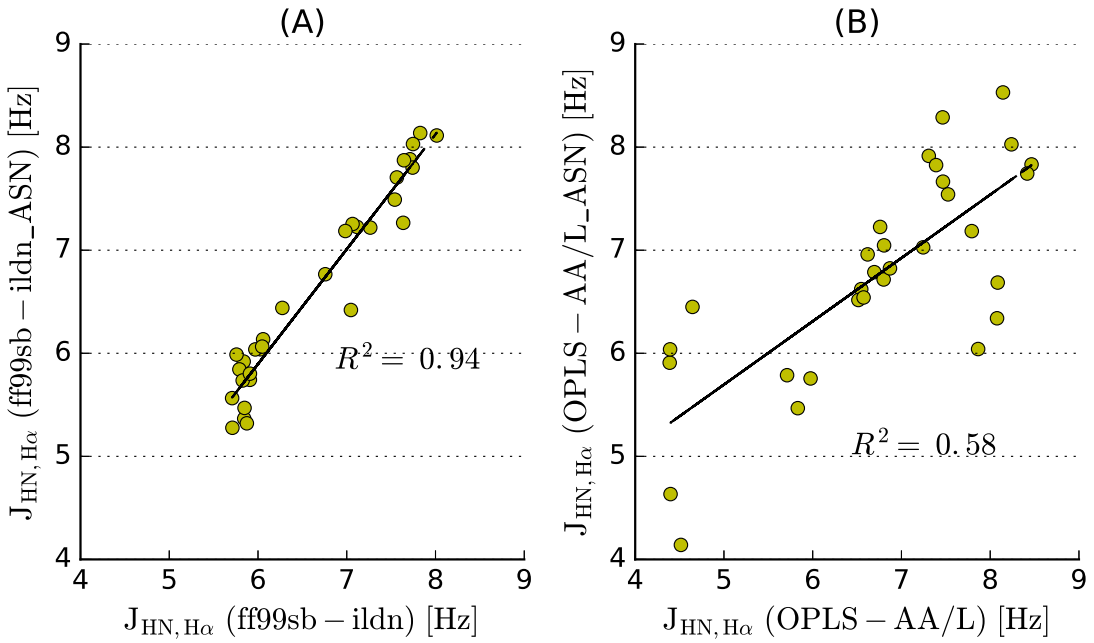


Figure 4.6: **Correlation of $^3J_{\mathbf{HN,H\alpha}}$ computed with different force field variants.** Panels show the correlation between $^3J_{HN,H\alpha}$ computed with (A) AMBER ff99sb-ildn force field or its variant ff99sb-ildn_ASN; (B) OPLS-AA/L force field or its variant OPLS-AA/L_ASN.

(R^2=0.94, Figure 4.7 E and F). This suggests that in OPLS the torsion potentials used for isoAspartate can strongly influence the conformations adopted by the two residues following isoAspartate (i.e. the Glycine and Arginine of the integrin recognition isoDGR motif). Overall, it seems that, while AMBER ff99sb-ildn force field is robust with respect to variations of local parameters, the OPLS-AA/L force field is more sensitive to these variations and needs a more careful analysis upon introduction of non standard amino acids.

I finally investigated whether the poor performance of OPLS-AA/L force fields in reproducing experimental $^3J_{HN,H\alpha}$ scalar couplings (χ^2>3, Figure 4.4) could be due to the influence of isoAspartate parameterization on the two following residues. We thus excluded the $^3J_{HN,H\alpha}$ couplings of the one/two residues following isoAspartate from the computations of χ^2 (see Figure 4.8). Since no significant variations in the overall χ^2 are observed I conclude that the poor agreement with experimental data is not dependent upon the presence of isoAspartate.

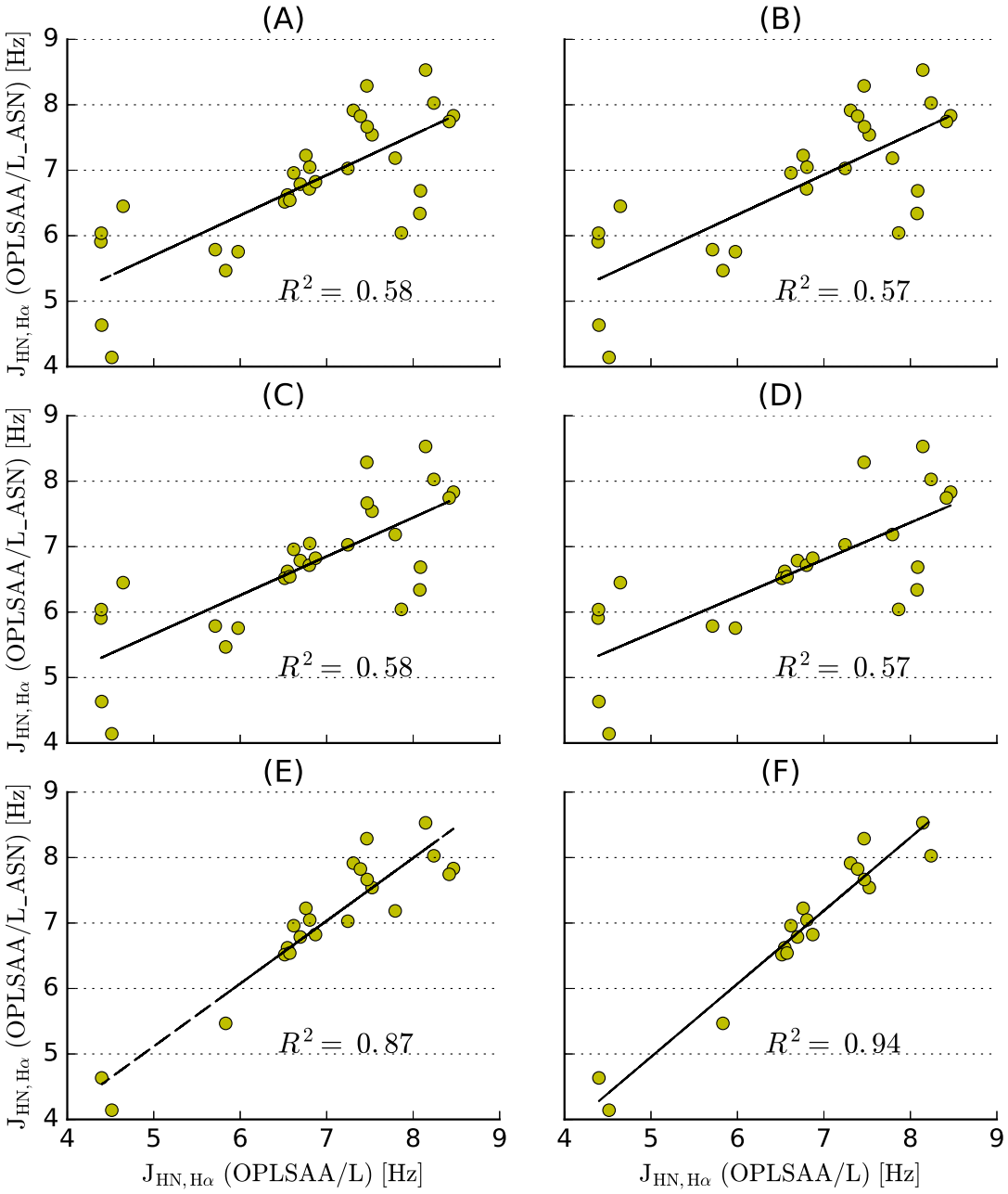


Figure 4.7: **Correlation of $^3J_{\mathbf{HN,H\alpha}}$ computed with OPLS-AA/L and OPLS-AA/L_ASN force field variants.** Panels show the correlation between: (A) all $^3J_{HN,H\alpha}$, (B) all $^3J_{HN,H\alpha}$ except isoAspartate, (C) all $^3J_{HN,H\alpha}$ except the residue preceding isoAspartate, (D) all $^3J_{HN,H\alpha}$ except the two residues preceding isoAspartate, (E) all $^3J_{HN,H\alpha}$ except the residue following isoAspartate, (F) all $^3J_{HN,H\alpha}$ except the two residues following isoAspartate, computed with OPLS-AA/L force field or its variant OPLS-AA/L_ASN.

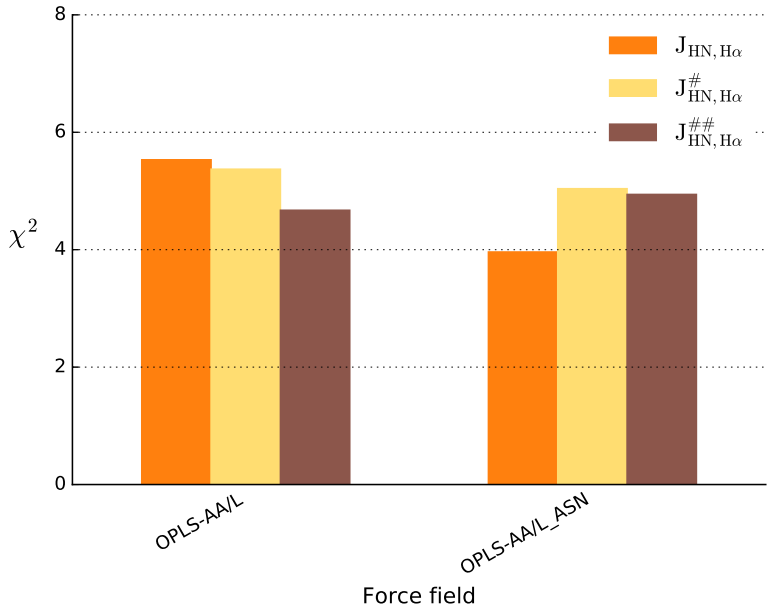


Figure 4.8: **Influence of isoAspartate parameterization in OPLS on the $^3J_{\mathbf{HN,H\alpha}}$ χ^2.** For OPLS-AA/L and OPLS-AA/L_ASN force fields, are reported the χ^2 values, estimating the ability to reproduce: all the experimental $^3J_{\mathrm{HN,H\alpha}}$ scalar couplings (orange bars), all the $^3J_{\mathrm{HN,H\alpha}}$ couplings except the ones of the residue following isoAspartate ($^3J^{\#}_{\mathrm{HN,H\alpha}}$, yellow bars), all the $^3J_{\mathrm{HN,H\alpha}}$ couplings except the ones of the two residues following isoAspartate ($^3J^{\#\#}_{\mathrm{HN,H\alpha}}$, brown bars).

4.1.3 Conclusion

The *in silico* determination of the three-dimensional conformation of cyclopeptides is challenging for mainly two reasons: i. their constrained geometry results in the presence of high free-energy barriers between local minima, entangling the achievement of an exhaustive sampling; ii. the force fields implemented in MD software could be not appropriate for the simulation of highly constrained cyclic peptides. We and others have shown that the first issue can be solved by using enhanced sampling MD techniques, such as Metadynamics.[75,199,213] The second issue, which is particularly relevant in the case of isoDGR-based cyclopeptides because of the presence of the non-standard amino acid isoAspartate, has been faced in this Chapter. Here I have presented a systematic investigation, in which the ability of eight popular force fields (AMBER ff99sb-ildn, AMBER ff99sb*-

ildn, AMBER ff14sb, OPLS-AA/L, CHARMM-27, GROMOS-54a7 plus the two variants ff99sb-ildn_ASN and OPLS-AA/L_ASN) to reproduce NMR-derived observables for five isoDGR-containing cyclopeptides has been tested. I performed 40 simulations, for a total of 17 μs, using BE-META to guarantee the exhaustiveness of the sampling. The accuracy of each force field has been quantified by means of a χ^2 metrics, where experimental J scalar couplings and chemical shifts have been compared with the corresponding back-calculated observables.

The comparison between experimental and back-calculated chemical shifts resulted to be not particularly significant since a similar agreement was found for all the force fields considered. Herein the high flexibility of the investigated peptides and the poor influence of the chemical environment resulted in chemical shifts comparable to the ones of a random coil, as confirmed both by NMR and simulations.
Therefore I focused my attention on the reproducibility of J scalar couplings, specifically highlighting the differences between the $^3J_{\mathrm{HN,H}\alpha}$, associated to the ϕ dihedral angles common to all the amino acids, and the $^3J^{\mathrm{isoD}}_{\mathrm{H}\alpha,\mathrm{H}\beta}$, related to the ξ dihedral angle characteristic of isoAspartate backbone. We found that most of the considered force fields can well reproduce the experimental $^3J_{\mathrm{HN,H}\alpha}$ scalar couplings, suggesting that in general the dynamic of cyclopeptides can be accurately predicted with exhaustive explicit-solvent MD simulations. These results are in line with the study of Wu and coworkers,[229] which reported that a good accuracy can be achieved in predicting the crystal structures of cyclopeptides. The only exceptions comprise the GROMOS-54a7 force field, which shows a poor agreement with experimental $^3J_{\mathrm{HN,H}\alpha}$, and the two variants of OPLS-AA/L force field, which confirmed their limitations and weaknesses in accurately sampling dihedral angles conformations. A direct comparison with data presented in literature, concerning the reproducibility of NMR scalar couplings in the case of linear peptides, is not straightforward: several choices can indeed influence the estimation of the error (χ^2) and in many studies diverse metric to evaluate this discrepancy has been chosen. Of note, in reference[230] the ability of different force fields in reproducing the J couplings of a polyalanine peptide was tested using the same metric and sets of Karplus parameters as in this thesis, allowing a more quantitative comparison. Herein it was shown that the force fields ff99sb, OPLS-AA/L, CHARMM-27 and a precedent

version of GROMOS (53a6) performed quite well obtaining χ^2 values ranging from a minimum of 1.8 to a maximum of 2.3. These values are perfectly in line with the ones reported in this thesis in the cases of AMBER and CHARMM force fields, suggesting that cyclopeptides can be simulated with an accuracy comparable to the one of linear peptides in these frameworks. Conversely, quite different results were obtained when using the other force fields (OPLS and GROMOS), which display a significantly worse agreement with experiments in the case of cyclic peptides. On the other side, more then one study[109,110] reported that OPLS-AA/L performs worse than AMBER ff99sb and CHARMM-27 in reproducing J scalar couplings: it is therefore difficult to conclude whether the poor performance of OPLS is due to cyclization or to intrinsic defects of the force field.

Concerning the reproducibility of the ${}^3J^{\mathrm{isoD}}_{\mathrm{H}\alpha,\mathrm{H}\beta}$ scalar couplings, characteristic of isoAspartate, I found that only two out of the eight force fields considered (i.e. AMBER ff99sb-ildn and AMBER ff99sb*-ildn) are able to achieve an accuracy comparable to the systematic uncertainty. Overall these data suggest that the presence of isoAspartate in cyclopeptides backbone can strongly affect the reliability of the simulations and that transferability of force fields parameters to non standard amino acids is not straightforward.

I finally investigated how the choice of the parameters to describe isoAspartate torsional angle energetics could influence the reproducibility of the other residues' ${}^3J_{\mathrm{HN},\mathrm{H}\alpha}$. To this aim I considered the AMBER ff99sb-ildn force field and the OPLS-AA/L force field, for which two possible parameters sets for isoAspartate torsional potentials have been used. I found that, while AMBER ff99sb-ildn is robust to variation of these parameters, in OPLS-AA/L, the torsional potential used for isoAspartate could influence the dynamic of the surrounding atoms, and specifically of the two residues following isoAspartate.

Of note all the conclusions here reported hold also when considering a different set of Karplus parameters for the back-calculation of the J couplings.

4.1.4 Methods

4.1.4.1 Experimental data used for the comparison

As benchmark for the force field evaluation, I used a set of 79 NMR measurements, experimentally obtained from NMR experiments performed on five isoDGR-containing cyclopeptides. These isoDGR-peptides include two molecules cycled through a disulfide bridge (CisoDGRC and its acetylated variant $_{\text{ac}}$CisoDGRC[75]) and three head-to-tail cyclic hexapeptides (c(CGisoDGRG), c(GCisoDGRG) and c(CphgisoDGRG)[76]). The NMR measurements consist in 38 chemical shifts, involving the C, Cα, C$_\beta$, HN and Hα atoms of natural amino acids, and 41 J scalar couplings. Specifically the J couplings can be subdivided in 31 $^3J_{\text{HN,H}\alpha}$, linked to the ϕ backbone dihedral angles, and 10 $^3J^{\text{isoD}}_{\text{H}\alpha,\text{H}\beta}$, which describe the backbone dihedral angle N-C$_\alpha$-C$_\beta$-C (ξ) characteristic of isoAspartate. In Table 4.1 all the NMR data employed, subdivided by type and by molecule, are listed.

The acquisition of the NMR data was performed on a Bruker Avance-600 spectrometer (Bruker BioSpin) equipped with a triple-resonance TCI cryo-probe with an x, y, z shielded pulsed-field gradient coil. For each molecule, ^{1}H-1D, TOCSY (TOtal Correlation SpectroscopY, tmix=60 ms), ROESY (Rotating frame nuclear Overhauser Effect SpectroscopY, tmix=100-400 ms), HSQC (Heteronuclear Single Quantum Coherence) and HMBC (Heteronuclear Multiple Bond Correlation) NMR spectra of 1-10 mM sample (90% H_2O, 10% D_2O) at 280-285 K and pH 6 have been

Table 4.1: **Summary of the NMR data.** Quantity of the experimental observables used as benchmark, subdivided by type (J coupling or Chemical Shift) and by molecule.

Peptide	J couplings		Chemical Shifts					Total
	$^3J_{HN,H\alpha}$	$^3J^{isoD}_{H\alpha,H\beta}$	CS$_{\text{C}}$	CS$_{\text{C}_\alpha}$	CS$_{\text{C}_\beta}$	CS$_{\text{H}_\alpha}$	CS$_{\text{HN}}$	
CisoDGRC	5	2	-	-	-	-	-	7
$_{\text{ac}}$CisoDGRC	5	2	-	-	-	-	-	7
c(CGisoDGRG)	5	2	3	3	2	4	3	22
c(GCisoDGRG)	9	2	3	3	1	5	3	26
c(CphgisoDGRG)	7	2	-	2	1	3	2	17
Total	31	10	6	8	4	12	8	79

recorded. Water proton signal was suppressed with excitation sculpting sequence. For c(CphgisoDGRG) a ^{1}H-1D NMR spectrum in 100% D_2O was also acquired. Proton resonances were assigned by conventional 2D experiments, including TOCSY and ROESY. Free induction decays were acquired (24-64 scans) over 5000-8000 Hz, into 2k data block for 256-512 incremental values of the evolution time (States-TPPI). Data were processed with TOPSPIN 3.2 (Bruker BioSpin GmbH, Rheinstetten, Germany) by apodization with 90° shifted sine-bell squared (qsine) window, and zero-filling in the indirect dimension to 1k points. Peptide assignments and peak lists were generated using CcpNmr.[243] The $^{3}J_{\mathrm{HN,H}\alpha}$ and $^{3}J^{isoD}_{\mathrm{H}\alpha,\mathrm{H}\beta}$ coupling constants were obtained directly from the resolved amide, α or β protons resonances of well digitized mono-dimensional spectra (40k points). The methylene stereo-specific assignments were derived by comparison of the experimental J coupling values and relative ROEs.

4.1.4.2 Simulations details

Input conformations of the peptides

An initial conformation for the cyclic head-to-tail hexapeptide c(isoDGRGGG) was generated using the Maestro 2D sketcher tool and energy minimized. The CNS[244] (Crystallography & NMR System) program was then employed to replace Glycines with proper amino acids to generate c(CGisoDGRG), c(GCisoDGRG) and c(CphgisoDGRG). A second minimization run was finally performed using Gromacs. The peptides cycled with a disulfide bond were instead directly generated with CNS and subsequently minimized with Gromacs. The two Cysteines of CisoDGRC were capped with NH_3^+ and COO^- groups, respectively, consistently with a pH value of 7. In $_{\mathrm{ac}}$CisoDGRC the first Cysteine was instead acetylated, resulting in a molecule with a total negative charge of -1.

Force fields and parameterization

Each of the cyclopeptides has been simulated with the following six force fields: AMBER ff99sb-ildn,[233] AMBER ff99sb*-ildn,[216] AMBER ff14sb,[106] OPLS-

AA/L,[222] CHARMM-27[235] and GROMOS-54a7.[236] Additionally, I performed the same simulations with two variants of the AMBER ff99sb-ildn and of the OPLS-AA/L force fields (denoted as ff99sb-ildn_ASN and OPLS-AA/L_ASN respectively), in which the dihedral angles parameters specifically optimized for Asparagine side-chain have been employed in the description of isoAspartate. Thus I simulated a total of 40 systems, consisting in 5 molecules with 8 different force field variants.

Parameterization of isoAspartic residue was performed assigning the atom-types of Asparagine side-chain to the backbone atoms of isoAspartate and the atom-types of C-terminal carboxylate backbone to the side-chain atoms of isoAspartate. Parameters for Phenylglycine were derived from the ones of Phenylalanine, applying appropriate dihedral angles. In the case of AMBER force fields the atomic partial charges of both Phenylglycine and isoAspartate atoms were derived using the R.E.D. III (RESP ESP charge Derive) package.[245] Specifically, two different conformations for both Phenylglycine and isoAspartate dipeptides (i.e. the amino acid capped with the acetyl and N-methyl groups) have been generated with Maestro,[226] then GAMESS[246] package was used for geometry optimization in the gas phase and for the computation of the molecular electrostatic potentials (MEP). For the MEP computation, the HF/6-31G* level of theory and the Connolly surface algorithm were employed; additionally four different molecular orientations for each optimized geometry have been considered using the rigid-body reorientation algorithm implemented in the R.E.D. tool. Finally, the two-stages RESP method was used for the fitting of the atomic charges following the procedure originally published by Kollman et al.[247] Intra-molecular constraints were imposed to set the charges of the acetyl and N-methyl capping groups to zero; additionally charges equivalences were imposed to hydrogens of methyl and methylene groups, to oxygens of carboxylate ions and to symmetric atoms in the phenyl group. The derived RESP partial charges for Phenylglycine and isoAspartate atoms are reported in Table B.1.

For the water molecules topology, I used the TIP3P[113] model in all the simulations with the exception of the ones performed with the GROMOS-54a7 force field. In this situation the SPC water model has been preferred: SPC is indeed strongly recommended by the GROMOS developers since this was the water model used during force field parameterization and optimization.[115]

Simulation set up

All the 40 MD simulations have been performed using the Gromacs-5.0.4 software[86] and the Plumed 2.1.3 plugin.[248] Each system has been solvated in a cubic box and, in the cases of CisoDGRC and $_{ac}$CisoDGRC, has been neutralized by adding Na^+ and Cl^- ions to a concentration of 100 mM. The dimensions of the box were chosen so that the minimum distance between any peptide's atom and the box edges was bigger than 1.2 nm. The initial energy minimization has been followed by a three-stages equilibration procedure consisting in: i. 2 ns equilibration in the NVT ensemble, where the v-rescale thermostat was used to maintain temperature at 280 K; ii. 2 ns in the NPT ensemble, where Berendsen barostat and v-rescale thermostat were employed to control pressure and temperature (1 bar and 280 K); iii. 4 ns in the NPT ensemble where the Berendsen barostat was replaced by the Parrinello-Rahman one (with the exception of simulations performed with GROMOS-54a7 in which Berendsen barostat was chosen). Positional restraints on the peptides' heavy atoms have been employed in the first two steps of equilibration and have been then released in the last stage. The relaxation time for the barostat was of 1 ps while the thermostat coupling time constant was of 100 fs (cyclopeptides and solvent molecules were coupled to independent thermostats). During the equilibration the LINCS algorithm was applied to constraint all the bond lengths. In all the cases the md integrator was used with a time step of 2 fs. A cut-off of 1.0 nm was used to truncate both the van der Waals and the electrostatic interactions; long range electrostatic interactions beyond the cut-off were treated with Particle-Mesh Ewald method (Fourier grid spacing of 0.12 nm and interpolation order of 4).[122,223]

Bias Exchange Metadynamics Simulations

The BE-META simulations, in standard formalism, were performed in the NVT ensemble, setting the temperature at 280 K by means of the v-rescale thermostat (coupling time constant of 100 fs) and using the same settings described for equilibration.[98] During this production run all the bond lengths involving hydrogen atoms have been constrained. Gaussian hills were added every 1 ps (height: 0.24 kcal mol^{-1}, width: 0.2 rad) and exchanges between pairs of replica were attempted

every 15 ps. Herein an acceptance rate between 15% and 25% was observed.
In the five molecules all the dihedral angles describing the cycle (with the exception of the Ω dihedral angles associated to the planarity of the peptide bonds) have been used as collective variables, resulting in the bias of:

i. 13 CVs for the cyclic head-to-tail hexapeptides. These comprise 5 ϕ and 5 ψ dihedral angles plus the 3 dihedral angles describing isoAspartate, namely ϕ' (C_{i-1}-N_i-$C\alpha_i$-$C\beta_i$), ξ (N_i-$C\alpha_i$-$C\beta_i$-C_i) and ψ' ($C\alpha_i$-$C\beta_i$-C_i-N_{i+1});

ii. 14 CVs for CisoDGRC, consisting in: the ϕ dihedral angles of Arginine, Glycine and Cysteine-5; the ψ dihedral angles of Cysteine-1, Arginine and Glycine; the 3 dihedral angles describing isoAspartate (ϕ', ξ and ψ'); plus the 5 dihedral angles describing the disulfide bridge. Specifically, these are the χ_1 (N-Cα-Cβ-S and N'-Cα'-Cβ'-S') and χ_2 (Cα-Cβ-S-S' and Cα'-Cβ'-S'-S) dihedral angles of the two Cysteines, plus the dihedral χ_3 (Cβ'-S'-S-Cβ). Herein the primed atoms are the ones of Cysteine-5 and the non-primed ones of Cysteine-1 (or vice versa).

iii. 15 CVs for $_{ac}$CisoDGRC, comprising the 14 CVs used for CisoDGRC plus the ϕ dihedral angle of Cysteine-1.

During the BE-META simulations, one of the listed CVs was biased in each replica. The length of each replica was of 30 ns, except for simulations that did not reach convergence and that were then prolonged to 40 ns per replica. The convergence was checked comparing the mono-dimensional free-energy profiles derived by averaging on the two halves of the simulation as described in section 3.2.2. In all the cases the convergence was reached and the difference between the two profiles was found to be lower than $2k_bT$.

4.1.4.3 Comparison between computed and NMR data

Analysis of BE-META simulations

In order to recover the equilibrium properties of the simulated systems from BE-META trajectories I used the method proposed by Laio and coworkers.[138]

Herein the structures visited during the simulations are grouped in microstates according to the values adopted by a reduced number of CVs, then the equilibrium population of each microstate is computed exploiting a weighted histogram analysis method (WHAM) based on the Metadynamics mono-dimensional potentials (see section 2.2.3).
The clustering of the structures and the evaluation of microstates free-energy was performed with the help of the Metagui tool.[141] A subset of the biased CVs was used to this aim. Specifically, in the case of head-to-tail cyclic hexapeptides 7 CVs have been used, consisting in the 6 ψ dihedral angles and in the ξ dihedral characteristic of isoAspartate. Conversely, for the two peptides cycled with a disulfide bond the CVs employed for microstates identification are the ψ dihedral angles of the first four amino acids, the χ_1 of Cysteine-5 and the ξ dihedral characteristic of isoAspartate. The variables identifying this subspaces and the optimal grid spacings (ranging between the 45° and 52°) have been individuated after several trials where different combinations of CVs have been tested. The main criteria used to select these CVs were: i. the microstates had to be characterized by low internal RMSD between conformations; ii. all the low-energy microstates (i.e. the ones with energy lower than $3k_bT$) had to be connected. Of note, despite the relatively large bin width used, necessitated by the high dimensionality of the investigated subspace, the grid employed is able to follow the energy landscape pretty finely (see Figure B.8).
In this context, it has been possible to compute the average equilibrium value of an observable X using the following equation:[138]

$$\langle X \rangle = \frac{\sum_\mu X_\mu e^{-F_\mu/k_bT}}{\sum_\mu e^{-F_\mu/k_bT}}, \tag{4.1}$$

where F_μ is the free-energy of microstate μ, estimated according to the WHAM procedure, k_b is the Boltzmann constant, T is the temperature and X_μ is the arithmetic average of X in microstate μ.

Back-calculation of J couplings and chemical shifts

For each of the structures clustered in a microstate, the Karplus relation (see equation 3.3),[224] was used to back-calculate the J coupling constants. In the case of ${}^3J_{\mathrm{HN,H}\alpha}$ I used $\theta = \phi{-}60°$ for L-amino acids, $\theta = \phi{+}60°$ for D-amino acids and both for Glycines. In the case of ${}^3J^{isoD}_{\mathrm{H}\alpha,\mathrm{H}\beta}$, θ was both $\theta = \xi$ and $\theta = \xi{-}120°$, where ξ is the dihedral angle N-Cα-Cβ-C characteristic of isoAspartate backbone. Following the examples of[230,249], and considering that Karplus relation is sensitive to the coefficients used, I employed two sets of Karplus parameters for the back-calculation of ${}^3J_{\mathrm{HN,H}\alpha}$. The first set, denoted as ORIG, comprises parameters from the original paper,[241] while the second set, called DFT, was derived from DFT calculations on Alanine dipeptide.[242] The two sets of Karplus parameters used for the back-calculation of ${}^3J_{\mathrm{HN,H}\alpha}$ and the ones used for ${}^3J^{isoD}_{\mathrm{H}\alpha,\mathrm{H}\beta}$ [225], along with the estimated systematic errors, are reported in Table 4.2. These errors, arising in the estimation of the scalar couplings using the Karplus relation, could have different sources. In fact several factors are neglected in its parameterization: these include the effect of substituents' nature, of their relative orientation, of the valence angles and of the bond lengths.[250] For ${}^3J_{\mathrm{HN,H}\alpha}$ a systematic error of 0.70 Hz was assumed, as done in[230]. In the case of ${}^3J^{isoD}_{\mathrm{H}\alpha,\mathrm{H}\beta}$ couplings a slightly higher error (0.90 Hz) was used, both because the substituents effect could be more relevant and because these parameters are less studied and developed.[251]

The chemical shifts of C, Cα, Cβ, Hα and HN atoms for the head-to tail hexapeptides, were back-calculated using the Sparta+ program on each of the structures clustered in microstates.[239] The prediction of chemical shifts has been limited to

Table 4.2: **Karplus parameters.** In Table are reported the Karplus parameters used for the back-calculation of both ${}^3J_{\mathrm{HN,H}\alpha}$ and ${}^3J^{isoD}_{\mathrm{H}\alpha,\mathrm{H}\beta}$. The associated estimated error σ is also reported.

Coupling	Set	A [Hz]	B [Hz]	C [Hz]	σ [Hz]
${}^3J_{\mathrm{HN,H}\alpha}$	ORIG[241]	7.09	-1.42	1.55	0.70
	DFT[242]	9.44	-1.53	-0.07	0.70
${}^3J^{isoD}_{\mathrm{H}\alpha,\mathrm{H}\beta}$[225]		10.2	-1.8	1.9	0.90

Table 4.3: **Sparta+ errors.** Estimated errors for the chemical shift predictions, as reported by the Sparta+ developer.[239]

	Atoms				
	C	Cα	Cβ	Hα	HN
Error [ppm]	1.09	0.94	1.14	0.25	0.49

those residues which are not neighbors of non standard amino acids (i.e. Phenylglycine and isoAspartate) because: i. Sparta+ has been parameterized empirically and trained on a database of protein composed by natural amino acids and ii. Sparta+ uses information from both residues i+1 and i-1 to predict the chemical shifts of atoms in residue i. The systematic errors associated to these chemical shift predictions are the ones determined in the original paper (see Table 4.3).[239] Average J couplings and chemical shifts (CSs) have been then computed using equation 4.1.

To evaluate the performances of the different force fields I quantified the agreement between back-calculated and experimental data using the following equation:[110,230,249]

$$\chi^2(X_{exp}, X_{comp}) = \frac{1}{N}\sum_{i=1}^{N}\frac{(\langle X_i\rangle_{\text{comp}} - X_{i,exp})^2}{\sigma_i^2}, \tag{4.2}$$

where X can be both scalar couplings or chemical shifts (deriving from the simulation average and from the NMR experiments), N is the total number of experimental data used and σ_i is the error associated to the value X_i (see Tables 4.2 and 4.3).

4.2 Rationalizing affinity and selectivity of isoDGR-based hexapeptides

4.2.1 Introduction: isoDGR-based peptides as drug delivery agents

After the discovery that the isoDGR sequence is an integrin recognition motif,[70] diverse attempts to study isoDGR-based peptides and peptidomimetics that could work as integrin inhibitors or as drugs delivery agents have been performed.[75–79] The isoDGR motif could indeed represent a new opportunity for the development of a novel generation of integrin-targeting ligands, displaying different characteristics, including affinity, selectivity and signaling effects, with respect to the widely studied RGD-based peptides.
Among the most interesting studies in this field, the group of Corti[76] has published a work in which a set of isoDGR-based cyclic hexapeptides, displaying different affinity and selectivity profiles for the two integrins $\alpha v\beta 3$ and $\alpha 5\beta 1$, both related to cancer, was presented. In particular, they have shown that one cyclic hexapeptide, c(CGisoDGRG), after conjugation to human serum albumin (HSA), a serum protein that works as a transport system for small molecules in blood, selectively recognizes integrin $\alpha v\beta 3$, binds to tumor vessels and inhibits tumor growth. Chemical conjugation of c(CGisoDGRG) with HSA was achieved by using sulfoSuccinimidyl-4-(N-Maleimidomethyl)Cyclohexane-1-Carboxylate (sulfo-SMCC), a bi-functional cross-linker often employed to link thiol-groups (which are available in the cyclopeptide thanks to the presence of a Cysteine in its sequence) and amino-groups. It has been shown that this chemical linker plays an important role in regulating cyclopeptide affinity: indeed conjugation of c(CGisoDGRG) with the linker (in which the reactive sulfo-succinimide group was blocked with ammonium chloride during the reaction as described in reference[76], see Figure C.1 and Figure 4.9) resulted in an improved affinity and selectivity of the peptide toward $\alpha v\beta 3$ integrin.

Of note this cyclopeptide, conjugated with HSA, can be used to prepare isoDGR-

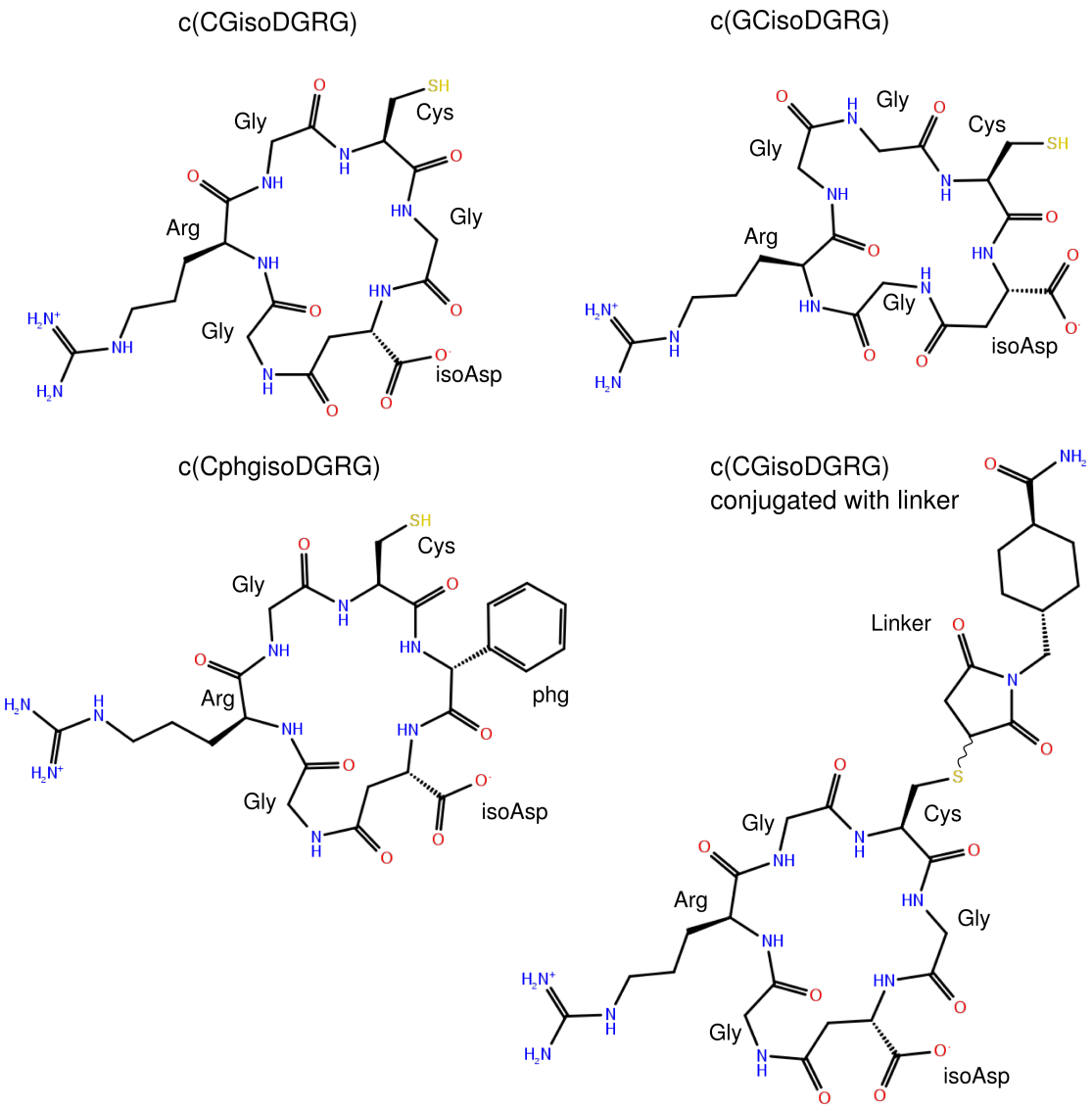


Figure 4.9: **2D-representation of the investigated molecules.** 2D-structures of the isoDGR-based cyclopeptides c(CGisoDGRG), c(GCisoDGRG) and c(CphgisoDGRG), and of the c(CGisoDGRG)-linker conjugate.

Table 4.4: **Affinity and selectivity of the investigated molecules.** The inhibition constant K_i, as measured by competitive binding assays,[76] of the three cyclopeptides c(CGisoDGRG), c(GCisoDGRG), c(CphgisoDGRG) and of the c(CGisoDGRG)-linker conjugate to integrins $\alpha v\beta 3$ and $\alpha 5\beta 1$, is reported.

Cyclopeptide	K_i [nM]		$K_i(\alpha 5\beta 1)/K_i(\alpha v\beta 3)$
	$\alpha v\beta 3$	$\alpha 5\beta 1$	
c(CGisoDGRG)	103 ± 18	320 ± 49	3
c(GCisoDGRG)	43 ± 9.6	21 ± 5.1	0.5
c(CphgisoDGRG)	1493	15	0.01
c(CGisoDGRG) + linker	35 ± 6	302	8.6

tagged gold nanoparticles, working as a nanosystem for the delivery of drugs, and specifically of cytokines, to tumor vessels. Indeed several clinical studies have shown that conjugation of cytokines, as Tumor Necrosis Factor, with peptides targeting tumor vasculatures is a useful strategy to improve their therapeutic efficacy.[252] In this context the conjugated peptide selectively recognizes tumor vasculature by targeting specific integrin subtypes and allows the administration of extremely low doses of drugs, avoiding unwanted side effects, possible toxic reactions and preventing the activation of counter-regulatory mechanisms that could contrast the therapeutic effects of cytokines.

Since the molecular scaffold of the isoDGR-based peptides is responsible of the integrin biding affinity and selectivity, it could be of interest rationalizing the structural basis of integrin recognition by the three cyclopeptides included in the study:[76] c(CGisoDGRG), c(GCisoDGRG), and c(CphgisoDGRG), the latter highly selective for $\alpha 5\beta 1$ integrin (see Figure 4.9 and Table 4.4). These three cyclopeptides display a high similarity in their sequences and differ only in the two amino acids preceding isoAspartate. Of note, these flanking residues appear to strongly influence peptide affinity to integrin. Unveiling the conformational determinants underlying these different affinities and selectivities profiles could be useful to provide the basis for a rational design of highly active and selective isoDGR-based compounds. Herein understanding why the conjugation of c(CGisoDGRG) with the chemical linker results in an improved affinity and selectivity of the peptide toward $\alpha v\beta 3$ integrin would strongly contribute in the development of improved integrin inhibitors.

From a methodological point of view, this system constitutes an ideal test case to evaluate the ability of the proposed computational protocol in predicting the effects of the type of residues flanking the isoDGR motif on cyclopeptides conformation and on integrin affinity and selectivity. Unlike the introduction of backbone *N*-methylation, this feature is expected to regulate integrin affinity by favoring additional stabilizing interactions with the receptor, and not only by conformational modulation of the recognition motif. This test would therefore represent a natural complement to the results described in Chapter 3.

Applying the multi-stage computational protocol I have been able to score the three cyclopeptides c(CGisoDGRG), c(GCisoDGRG) and c(CphgisoDGRG) according to their affinity towards integrins αvβ3 or α5β1; additionally, the visual inspection of the decoy poses herein generated provided an explanation for the high selectivity of c(CphgisoDGRG) for α5β1. Finally, thanks to a combination of NMR experiments performed in the laboratory and computational techniques, I proposed possible models that rationalize the improved affinity of c(CGisoDGRG)-linker conjugate for αvβ3 with respect to c(CGisoDGRG).

4.2.2 Results and Discussion

4.2.2.1 Conformational analysis

In order to reconstruct the free-energy conformational landscape of c(CGisoDGRG), c(GCisoDGRG) and c(CphgisoDGRG), I have performed BE-META simulations using the AMBER ff99sb-ildn force field and biasing all the backbone dihedral angles of the cyclopeptides except the ones related to the planarity of the peptide bonds, which are assumed to be always in *trans* conformation. Therefore for each peptide a 13-replicas simulation of 390 ns was run. After a careful check of the convergence (Figure C.2), the free energy has been reconstructed in a space of reduced dimensionality, defined by a subset of the biased CVs. These CVs comprise the 6 ψ dihedral angles of all the residues and the ξ dihedral (N-Cα-Cβ-C) characteristic of isoAspartate. All the details are in Section 4.2.4.1. Within this space, minima populated more than 10% have been identified and

Table 4.5: **Structural characterization of the BE-META minima.** For each cyclopeptide minima populated more than 10% have been characterized according to the presence of tight turns (α, β or γ represented in red, orange and green, respectively). The turns have been individuated detecting the presence of intramolecular backbone hydrogen bonds[1] and the turn types have been assigned comparing the average values of the dihedral angles involved in the turn with the expected reference values reported in Table C.1.

Cyclopeptide	Population	Turns	
		Type	Representation
c(CGisoDGRG)	27%	βI	c(**C**GisoDGR**G**)
	27%	βI'	c(CGisoDG**RG**)
	17%	βI	c(CGisoD**GR**G)
	14%	II-α_{RU}	c(**C**GisoDG**RG**)
c(GCisoDGRG)	39%	II-α_{RS} and γ	c(**GC**isoD**G**R**G**)
	30%	I-α_{LS}	c(**G**CisoDG**RG**)
	15%	βI'	c(GCisoDG**RG**)
c(CphgisoDGRG)	32%	βII'	c(**C**phgisoDGR**G**)
	28%	βI' and βI	c(**Cphg**isoD**GR**G)
	21%	βI	c(CphgisoD**GR**G)

[1] Hydrogen bonds were individuated with the VMD [215] Hbonds plugin. A donor-acceptor distance of 3.5 Å and an angle cutoff of 35° have been considered. Only hydrogen bonds present in at least 40% of the structures were considered.

characterized. The equilibrium minima population has been derived analyzing the BE-META trajectories following a procedure analogue to the described in Section 3.4.2 and using equation 3.2. The accuracy of these simulations is supported by the good reproducibility of the experimental J couplings, as shown in the previous Section.

The conformational analysis revealed that these cyclopeptides are quite flexible, being able to populate three or four different minima, which are structurally characterized by diverse patterns of hydrogen bonds and turns, as reported in Table 4.5 and in Figure C.3. This kind of characterization is relevant in cyclopeptides since it is well recognized that cyclic hexa-peptides generally stabilize their conformation via a double β-turns fold.[201] The definition of the turns, according to the presence of hydrogen bonds and to the values adopted by the ϕ and ψ dihedral angles of the amino acids involved, is given in Table C.1.

The structural analysis of the BE-META conformers revealed that none of the cyclopeptides, with the exception of c(CphgisoDGRG), is able to adopt the stable double β-turns fold, characteristic of hexapeptides. Locally, the regions of the cyclopeptides not including isoAspartate can organize themselves in turn structures, which can adopt either the β conformation or the α conformation, the latter occurring less frequently in proteins.[253] However, the presence of the additional isoAspartate Cβ atom in the backbone, which is absent in the other amino acids, prevents the formation of the second β-turn around isoAspartate and, consequently, the adoption of the stable double β-turns fold. The high flexibility displayed by these molecules could therefore be ascribed to the presence of this additional Cβ atom.
This situation, which prevents the pre-organization of the cyclopeptides in a stable preferential conformation, hampers a clear evaluation of the interplay between molecules' conformation and binding affinity. To account for the system conformational flexibility, I used the conformers extracted from BE-META minima as input for docking calculations. This choice is in accordance with the well-accepted assumption that an exhaustive conformational sampling of the cyclopeptides could be useful to improve the accuracy of docking results.[207]

4.2.2.2 Docking calculations

In order to rationalize the different affinity and selectivity profiles displayed by the three molecules, I docked for each cyclopeptide the conformers extracted from BE-META minima into the crystal structures of the two integrins αvβ3 and α5β1 (PDB codes 1L5G[9] and 4WK4, respectively). To this aim two different docking programs have been used: the Glide software in combination with Prime MM-GB/SA rescoring (Schrödinger suite)[152,178] and HADDOCK,[153] an information-driven docking software. Details of the docking protocols adopted are described in Section 4.2.4.2 and 4.2.4.3. For each of the twelve systems (three cyclopeptides docked in two integrins using two different docking software), a cluster analysis, according to geometric criteria, of the decoy poses was performed and an average score was assigned to each cluster. Then, the best scored cluster, in terms of MM-GB/SA score or HADDOCK score, resembling the canonical electrostatic

clamp with integrin was selected.
The correlation between the inhibition constants K_i, experimentally measured for each integrin-cyclopeptide complex as described in [76], and the corresponding average cluster scores, obtained with either the Glide software (upper panel) or with HADDOCK (lower panel), is reported in Figure 4.10. It can be observed that, while HADDOCK failed to correctly predict the best binders towards both the receptors, the combination of Glide and MM-GB/SA rescoring perfectly discriminates binders (i.e. cyclopeptides displaying a K_i in the range of tens of nanomolar for the receptor) from weak binders (i.e. K_i values in the range of hundreds of nanomolar). The poor performance of HADDOCK in ranking cyclopeptides according to their binding affinity is not unexpected, since it is well known that this program has not been developed to this aim. Accordingly, also the visual inspection of the docking poses obtained with HADDOCK does not allow to discriminate and rationalize the different affinity and selectivity profiles of c(CGisoDGRG), c(GCisoDGRG) and c(CphgisoDGRG) for integrins $\alpha v\beta 3$ and $\alpha 5\beta 1$. Therefore, in the following I will focus on the results obtained using the Glide software. A representative binding mode for each of the six systems investigated and the associated MM-GB/SA contributions are reported in Figure 4.11, Figure 4.12 and in Table 4.6, respectively. The analysis of the MM-GB/SA terms revealed that, in $\alpha v\beta 3$ integrin, the better affinity displayed by c(CGisoDGRG) and c(GCisoDGRG) with respect to c(CphgisoDGRG) is mainly due to a more favorable polar contribution; conversely, in $\alpha 5\beta 1$, the better affinity of c(CphgisoDGRG) and c(GCisoDGRG) with respect to c(CGisoDGRG) can be ascribed to a more favorable non polar term. The analysis of the decoy poses, complemented by the knowledge of these contributions, allowed to rationalize the diverse affinity of the three cyclopeptides.

Docking results on integrin $\alpha v\beta 3$ (Figure 4.11) show that all the three cyclopeptides are able to bind the receptor reproducing the canonical electrostatic clamp: on one side the guanidinium group of Arginine engages a salt bridge with the carboxylate group of either αv:Asp150 and/or αv:Asp218; on the other side the isoAspartate carboxylate group contributes to the coordination of the metal cation at the MIDAS site and establishes hydrogen bonds with $\beta 3$:Asn215. Moreover, in the three complexes, the Arginine side chain engages similar hydrophobic interactions

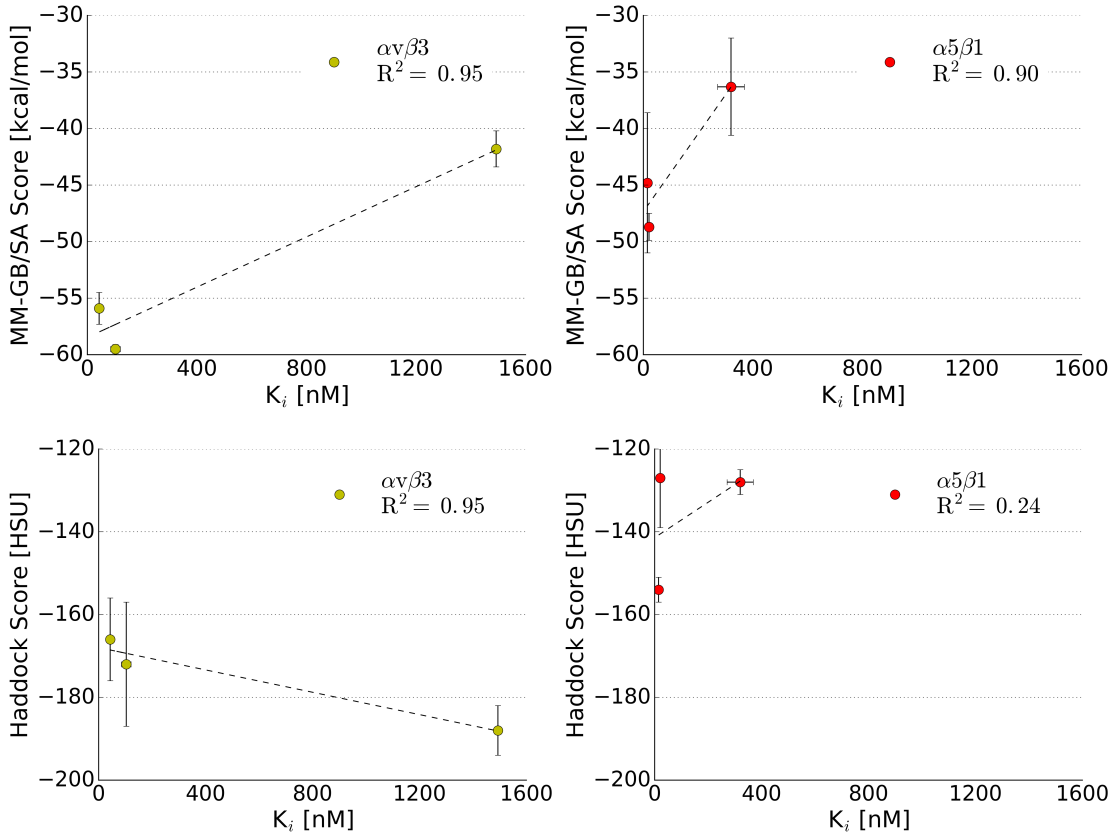


Figure 4.10: **Predicted vs Experimental affinity.** For each cyclopeptide, the correlation between K_i inhibition constants and predicted binding affinities, with either the Prime MM-GB/SA tool (upper panel) or with HADDOCK software (bottom panel), for $\alpha v\beta 3$ (left panel) or $\alpha 5\beta 1$ (right panel) integrins, is represented.

with αv:Tyr178. However, significantly different binding modes within the $\alpha v\beta 3$ binding pocket are displayed by c(CGisoDGRG) and c(GCisoDGRG) with respect to c(CphgisoDGRG). The first two molecules (Figure 4.11, top and middle panel), indeed, are able to establish a pattern of hydrogen bonds that is not observed in the $\alpha v\beta 3$-c(CphgisoDGRG) complex. Specifically, in these two molecules, the Glycine of the isoDGR motif forms an hydrogen bond with the carbonyl group of β3:Arg216; additionally the residues flanking the isoDGR sequence engage two hydrogens bonds with β3:Arg214 and αv:Tyr178. Further, this pattern of polar interactions is stabilized by additional hydrophobic contacts: the Cysteine of c(CGisoDGRG) establish non polar interactions with the side chain of β3:Met180 while the Cysteine

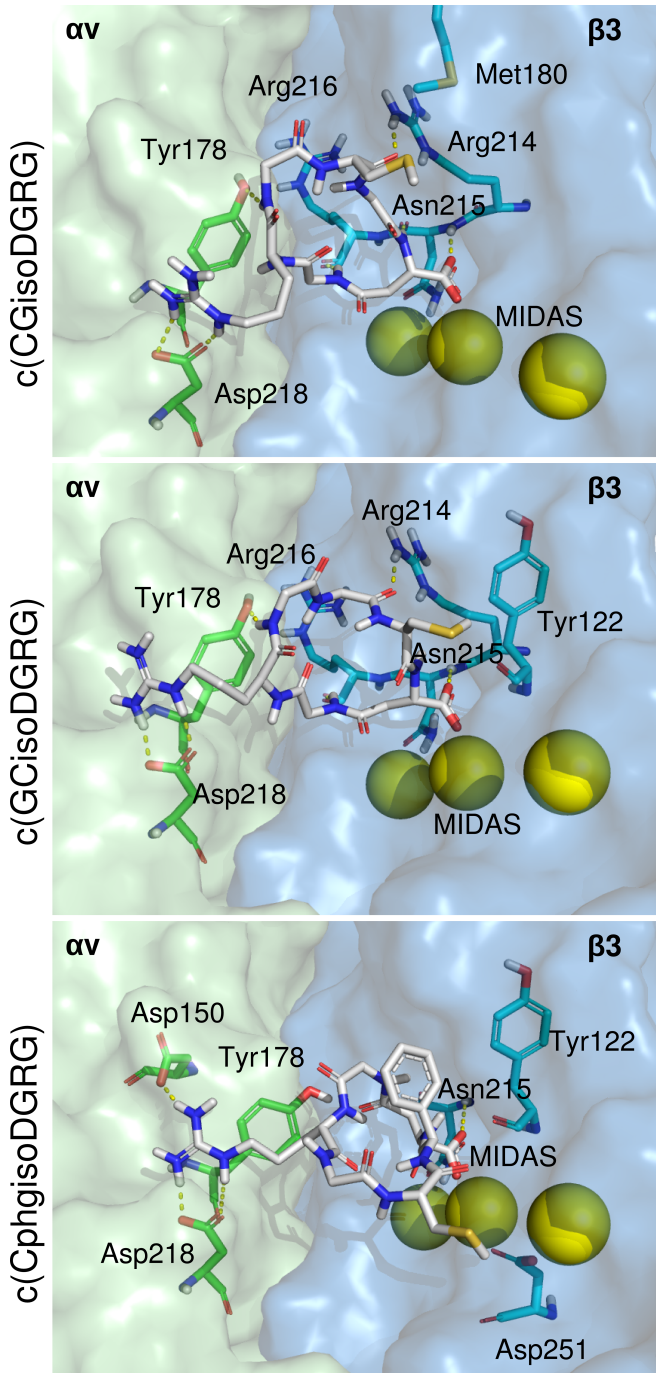


Figure 4.11: **Representative binding modes of c(CGisoDGRG), c(GCisoDGRG) and c(CphgisoDGRG) in complex with integrin $\alpha v\beta 3$, according to Glide software**. αv and β3 subunits are represented in green and blue surface, respectively. Receptor amino acids engaging direct interactions with the ligands are shown in stick and labeled with the three-letter code. Metal ions are represented as yellow spheres and peptides are shown in white sticks.

Table 4.6: **Energetic contributions.** The average MM-GB/SA values and individual contributions are reported for the selected cluster of each of the six systems investigated. All the values reported are obtained by averaging on the contributions of the best four poses of the cluster. The Polar term is computed as sum of Coulomb and Generalized Born electrostatic solvation energy; the Non-Polar term is the sum of van der Waals and lipophilic energy, while Other contributions include covalent binding, hydrogen-bonding, π-π packing energy and self-contact correction. Additionally, for each system, the inhibition constant K_i is reported.

	Cyclopeptide	K_i	Energetic Contributions [kcal mol^{-1}]			
			MM-GB/SA	Polar	Non Polar	Other
αvβ3	c(CGisoDGRG)	103±18	-59.5±0.4	-9.1±1.1	-54.9±0.7	4.5±0.5
	c(GCisoDGRG)	43±9.6	-55.9±1.4	-2.6±4.4	-57.5±5.0	4.2±3.5
	c(CphgisoDGRG)	1493	-41.8±1.6	8.2±1.7	-54.2±1.5	4.2± 0.9
α5β1	c(CGisoDGRG)	320±49	-36.3±4.3	11.3±1.4	-45.7±1.8	-1.9±2.4
	c(GCisoDGRG)	21±5.1	-48.7±1.2	12.8±3.3	-55.6±1.2	-5.9±1.6
	c(CphgisoDGRG)	15	-44.8±6.2	10.7±7.0	-54.4±3.6	-1.1±2.9

of c(GCisoDGRG) contacts β3:Tyr122. Differently, c(CphgisoDGRG) can not engage the numerous polar interactions described above (Figure 4.11, bottom panel). In the αvβ3-c(CphgisoDGRG) complex, the electrostatic clamp is stabilized only by hydrophobic interactions with β3:Tyr122 and one polar interaction, consisting in the thiol group of Cysteine that contacts the carboxylate group of β3:Asp251. Consistently with these observations, the MM-GB/SA polar contributions of αvβ3-c(CphgisoDGRG) complex is significantly lower with respect to the one obtained by the other two cyclopeptides (Table 4.6). Considering that the non polar MM-GB/SA contributions are similar in the three complexes, these observations well explain the significantly better affinity towards αvβ3 displayed by c(CGisoDGRG) and c(GCisoDGRG) with respect to c(CphgisoDGRG). Conceivably, the bulky hydrophobic residue Phenylglycine is not well tolerated by the buried binding pocket of αvβ3, characterized by the presence of several polar residues as β3:Arg214, β3:Arg216 and αv:Tyr178.

Focusing on the complexes with integrin α5β1 (Figure 4.12), docking results suggest that both c(CGisoDGRG) and c(CphgisoDGRG) perfectly recapitulate the canonical electrostatic clamp: herein the isoAspartate carboxylate group coordinates

the metal cation at the MIDAS site and the Arginine guanidinium group establishes a bidentate salt bridge with α5:Asp227, engaging also an hydrogen bond with α5:Gln221. A slightly different binding mode of the isoDGR motif is observed in the $\alpha5\beta1$-c(GCisoDGRG) complex: the Arginine guanidinium group forms hydrogen bonds with α5:Asp228 and α5:Asp227, while isoAspartate, in addition to coordinating the metal cation of the MIDAS site, establishes hydrogen bonds with β1:Ser134 and β1:Asn224. In all the three complexes hydrophobic interactions between the isoDGR motif and the α5:Phe187 stabilize the electrostatic clamp.
In addition to the electrostatic clamp, quite different patterns of interactions are observed in the three complexes. c(CGisoDGRG)-$\alpha5\beta1$ complex (Figure 4.12, top) is stabilized by an hydrogen bond between the Glycine of isoDGR motif and β1:Leu225, but the residues flanking the isoDGR motif do not contribute in engaging further interactions. Conversely, c(GCisoDGRG) (Figure 4.12, middle), in addition to the hydrogen bond with β1:Leu225, engages further stabilizing interactions with $\alpha5\beta1$, including contacts with α5:Trp157, β1:Asn224 and β1:Tyr133. Of note, the last two contacts are engaged by the Cysteine preceding isoAspartate, which in c(CGisoDGRG) is replaced by a Glycine. In agreement with this remark, the MM-GB/SA non-polar contribution in the $\alpha5\beta1$-c(CGisoDGRG) complex in significantly lower with respect to the one of $\alpha5\beta1$-c(GCisoDGRG). Finally, in $\alpha5\beta1$-c(CphgisoDGRG) (Figure 4.12, bottom), the cyclopeptide presents a π-stacking interaction between the phenyl group of D-Phenylglycine and β1:Tyr133; additionally, the thiol group of Cysteine forms an hydrogen bond with the carbonyl group of β1:Asp259, establishing further contacts with β1:Glu320 and β1:Phe321. This interactions pattern is consistent with the low nanomolar K_i displayed by this peptide; accordingly, both polar and non polar MM-GB/SA contributions are comparable with the ones of $\alpha5\beta1$-c(GCisoDGRG). These results well explain the better affinity for $\alpha5\beta1$ shown by c(GCisoDGRG) and c(CphgisoDGRG) with respect to c(CGisoDGRG), which can be mainly ascribed to non-polar contributions. Overall these data suggest that the presence of a Glycine as residue preceding isoAspartate lead to moderate affinity for $\alpha5\beta1$: indeed this amino acid, due to the lack of side chain, is not able to engage additional stabilizing interactions with the wide binding pocket of the receptor.

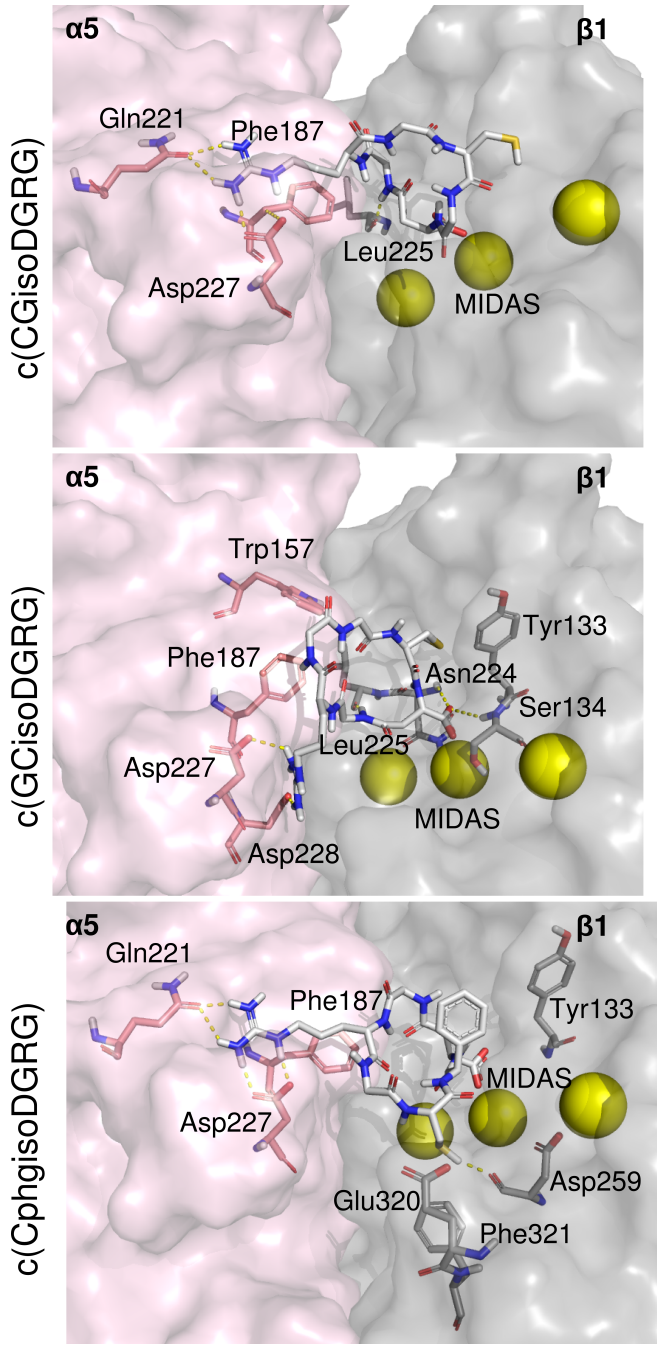


Figure 4.12: **Representative binding modes of c(CGisoDGRG), c(GCisoDGRG) and c(CphgisoDGRG) in complex with integrin $\alpha 5\beta 1$, according to Glide software**. $\alpha 5$ and $\beta 1$ subunits are represented in pink and gray surface, respectively. Receptor amino acids engaging direct interactions with the ligands are shown in stick and labeled with the three-letter code. Metal ions are represented as yellow spheres and peptides are shown in white sticks.

Table 4.7: **Comparison of Buried Surface Area.** The inter-facial Buried Surface Area (BSA) of c(CphgisoDGRG) in complex with $\alpha v\beta 3$ and $\alpha 5\beta 1$ is reported. Specifically, the involvement of polar or non polar atoms is highlighted. The data concerning BSA have been computed with the Naccess software v2.1.1 with water represented by a sphere of radius 1.4 Å.[254]

	BSA [Å^2]		
	Total	Polar	Non-Polar
$\alpha v\beta 3$	250	120	130
$\alpha 5\beta 1$	252	96	156

The availability of the crystallographic structures of $\alpha v\beta 3$ and $\alpha 5\beta 1$ integrins in complex with RGD ligands, as well as the knowledge of the key differences in the binding pockets of these two proteins, which have been exhaustively investigated in the past,[15,40,78] allowed to rationalize the highly different selectivity displayed by c(CphgisoDGRG) for $\alpha 5\beta 1$ integrin.

At a first glance, the binding modes adopted by c(CphgisoDGRG) in $\alpha v\beta 3$ or $\alpha 5\beta 1$ seem to be highly similar, with the only exception consisting in the different positioning of the cyclopeptide Cysteine. This difference in binding mode could be ascribed to the replacement of $\beta 1$:Gly261, in $\alpha 5\beta 1$, with $\beta 3$:Lys253, in $\alpha v\beta 3$. This substitution results in a narrower cleft along the dimer interface in $\alpha v\beta 3$, which prevents the accommodation of cyclopeptide Cysteine in that position. Of note, these results are in line with previous studies, where the binding mode of c(phgisoDGRk) cyclopeptide in the two integrins $\alpha 5\beta 1$ and $\alpha v\beta 3$ has been investigated.[40,78] The different selectivity of c(CphgisoDGRG), which displays a significantly higher affinity for $\alpha 5\beta 1$ with respect to $\alpha v\beta 3$, can be explained by analyzing the inter-facial buried surface area (BSA) of the two complexes, computed as the difference in receptor accessible surface area between the situations of free and bound receptor. This analysis suggests that the cyclopeptide occupies a more polar region in $\alpha v\beta 3$ with respect to $\alpha 5\beta 1$ (see Table 4.7). Conceivably, this is related to the higher acidity of the D3-A3 loop in αv as compared to $\alpha 5$, where the three Aspartates αv:Asp146,Asp148,Asp150 are replaced by $\alpha 5$:Phe155,Trp157,Ala159, respectively.[40] In this context, the desolvation energy required to accommodate the cyclopeptide within the $\alpha v\beta 3$ pocket could unfavorably contribute to its binding

affinity. Of note, this unfavorable contribution is not counterbalanced by additional polar interactions in the $\alpha v\beta 3$-c(CphgisoDGRG) complex, justifying the low affinity of c(CphgisoDGRG) for $\alpha v\beta 3$ and its higher selectivity for $\alpha 5\beta 1$.

4.2.2.3 Rationalizing the effect of c(CGisoDGRG)-linker conjugation on $\alpha v\beta 3$ integrin affinity

Biochemical studies[76] have shown that conjugation of c(CGisoDGRG) with the chemical linker (see Figure 4.9, bottom right) resulted in an improved affinity and selectivity of the peptide toward $\alpha v\beta 3$ integrin (Table 4.4). Two possible hypothesis could explain this improvement in binding affinity: i. the chemical linker influences the conformations of the cyclopeptide, either modulating the equilibrium between conformers or favoring the adoption of a novel bioactive conformation; ii. the chemical linker engages additional stabilizing interactions with the receptor. To solve this issue and to provide a model of the interacting complex we have applied a combination of NMR and MD strategies.

I started investigating the conformations adopted by the cyclopeptide c(CGisoDGRG) after conjugation with the chemical linker. Comparison of 2D-ROESY (Rotating frame nuclear Overhauser Effect SpectroscopY) spectra, acquired in our laboratory, of c(CGisoDGRG) and of c(CGisoDGRG)-linker conjugate revealed that the conformation of the cyclopeptide backbone is not influenced by the presence of the linker. Indeed, even if some chemical shift displacements can be observed due to the change of the chemical environment, the pattern of peaks related to c(CGisoDGRG) is highly similar in the two spectra (see Figure 4.13 and Table C.2). Therefore, in order to investigate the conformations adopted by the peptide conjugated with the chemical linker, I have performed BE-META simulations on eight different input structures, prepared starting from the four backbone conformations individuated in the BE-META simulation of the cyclopeptide alone (see Table 4.5 and Figure C.3) and attaching the chemical linker in both R and S configurations. Details on the linker parameterization are described in Section 4.2.4.5. Since, according to NMR data, the conformations of the peptide should not be influenced by the presence of the chemical linker, cyclopeptide backbone has been maintained in its starting

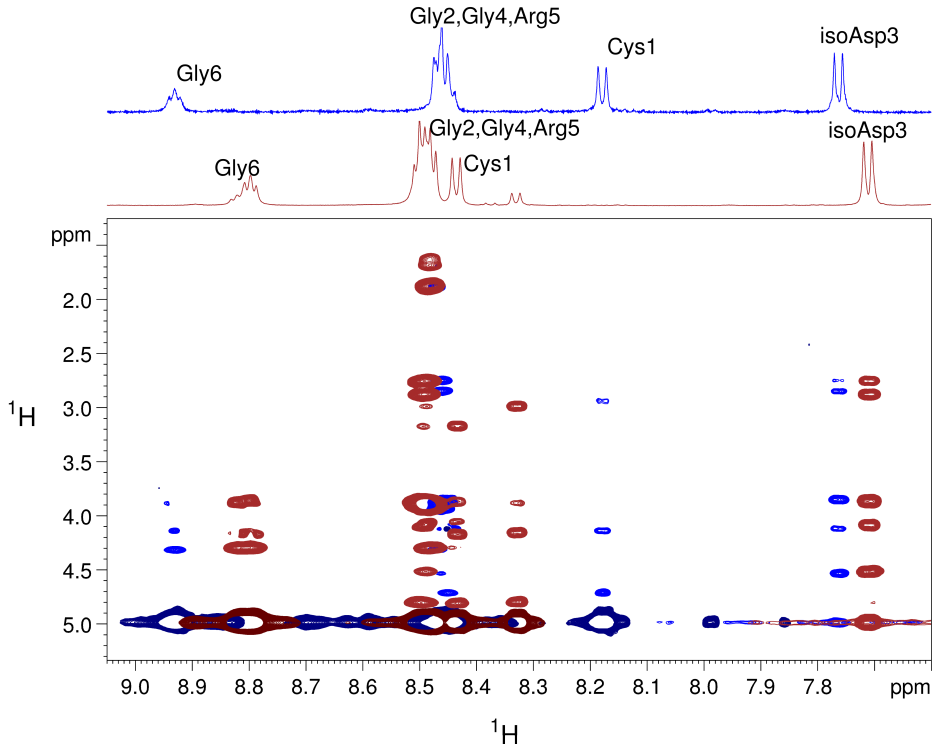


Figure 4.13: **ROESY spectra**. Superposition of bidimensional ^{1}H-^{1}H ROESY spectra of c(CGisoDGRG) (blue) and of c(CGisoDGRG)-linker conjugate (red). The region of the spectra showing the correlation between HN and Hα protons is reported. The spectra of 1-5 mM sample in 20 mM of phosphate buffer (90% H_2O, 10% D_2O) have been acquired using a mixing time of 300 ms, at temperature of 280 K and pH 6.5.

conformation through dihedral restraints, while the remaining part of the molecule was free to explore the conformational space. In particular the variables related to the degrees of freedom of the Cysteine side chain and of the linker have been biased (see Section 4.2.4.6 for details). The simulations have been analyzed as described in Section 4.2.4.6; then for each of the eight starting conformers, a pool of 100 conformations have been extracted from the main conformational minima that have been used as input for subsequent docking calculations.

During the docking calculations, performed with the Glide software,[152] positional constraints have been applied to favor the formation of the canonical interaction clamp between c(CGisoDGRG) and integrin $\alpha v\beta 3$ (see Section 4.2.4.7). Totally 381

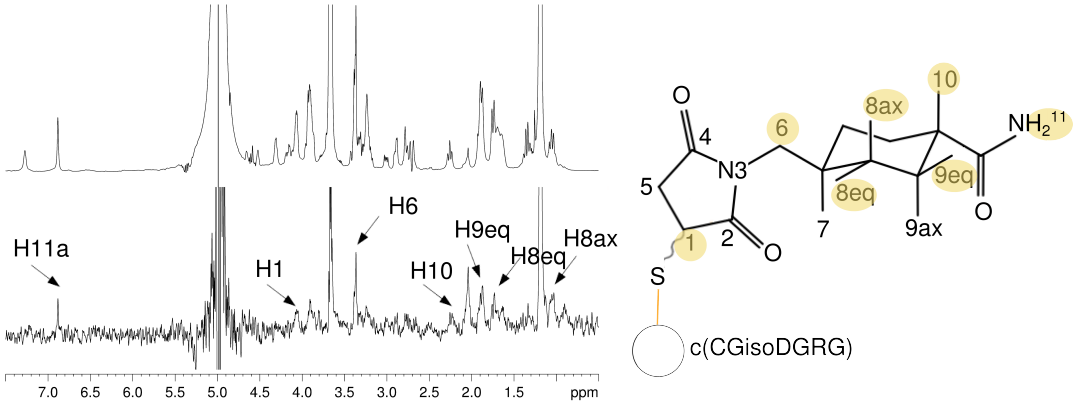


Figure 4.14: **STD experiments**. In the left panel, the off-resonance spectrum (top) and the STD spectrum (bottom) are shown. Herein the linker-related emerging peaks are assigned. The experiments have been acquired for a sample containing 5 μM of αvβ3 recombinant extracellular domain integrin and 0.3 mM of ligand (stoichiometric ratio = 1:60) in 20 mM of phosphate buffer pH 6.5, 100 mM NaCl, 0.1 mM $CaCl_2$, 5 mM $MgCl_2$ (90% H_2O, 10% D_2O). The temperature was set to 280 K. ^{1}H saturation on-resonance was set at -2 ppm and off-resonance at 100 ppm, using 3 s of total irradiation time. In the right panel it is shown a 2d-representation of the linker, in which hydrogens that are in close proximity to αvβ3 protons, according to STD experiments, are highlighted.

decoy poses (155 in the case of peptide-linker in R configuration, 226 in the case of S configuration) satisfying the positional constraints have been generated. In order to select, among these wide variety of decoy poses, the structural models that most reliably reproduce the αvβ3 c(CGisoDGRG)-linker conjugate complex, NMR data from Saturation Transfer Difference (STD) experiments have been exploited (see Figure 4.14). These data revealed which hydrogens of the cyclopeptide are in close proximity to protons of the receptor. Therefore, only decoy poses in which the hydrogens individuated from STD experiment were found at a distance lower than 4 Å from receptor protons, have been selected and afterward clustered according to geometric criteria (see Section 4.2.4.7 for details). Representative decoy poses of the clusters individuated, containing more than 4 structures, are shown in Figure 4.15. Of note, in these clusters, the cyclopeptide-linker conjugate was found in both R and S configurations, suggesting that there are no differences in the binding modes of the two diastereomers.

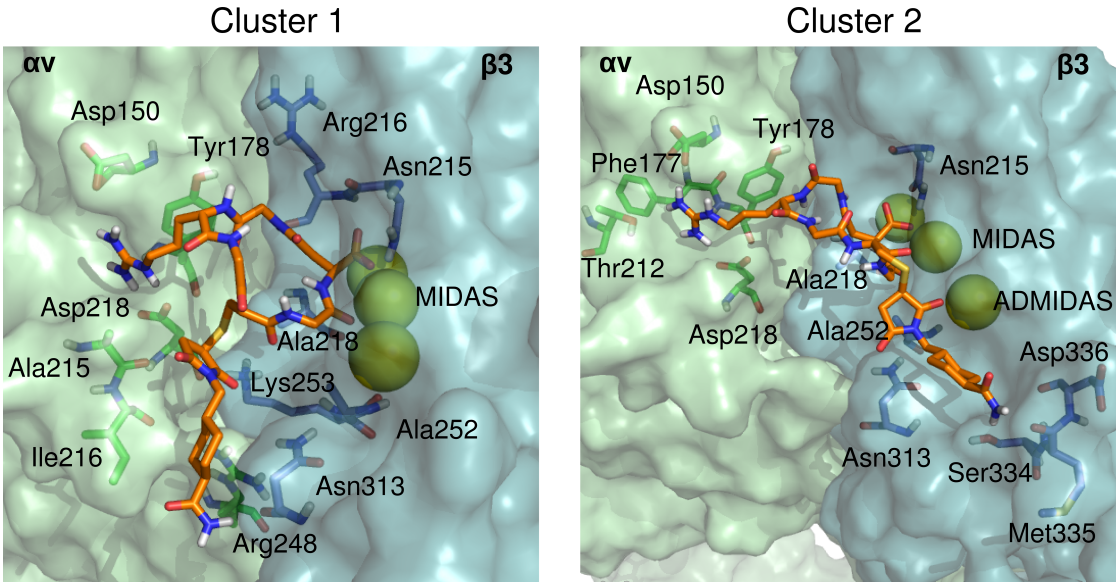


Figure 4.15: **Proposed docking models**. Two possible binding modes of c(CGisoDGRG)-linker conjugate in complex with integrin $\alpha v\beta 3$. Integrin is represented as surface with the αv and $\beta 3$ subunits in green and blue, respectively. Receptor amino acids engaging direct interactions with the ligands are shown in stick and labeled with the three-letter code. Metal ions are represented as yellow-green spheres and ligands are shown in orange sticks.

In both the proposed binding modes the electrostatic clamp is well summarized: on one side the Arginine of the isoDGR motif interacts with the αv subunit, engaging a salt bridge with Asp218 and polar contacts with Asp150 and/or Thr212; on the other side, isoAspartate coordinates the metal ion in MIDAS position and engages polar interactions with $\beta 3$:Asn215. In both the poses I observed additional hydrophobic and polar interactions stabilizing the complex.

Specifically, in the poses of cluster 1 (Figure 4.15, left panel) the cyclopeptide contributes in forming: hydrophobic interactions with αv:Tyr178, $\beta 3$:Ala218 and $\beta 3$:Ala252, polar contacts with $\beta 3$:Lys253, engaged by the Cysteine, and an hydrogen bond with $\beta 3$:Arg216, engaged by the Glycine of the isoDGR motif. Of note, also the linker actively contributes in forming additional stabilizing contacts that well justify the improved affinity of c(CGisoDGRG) after conjugation. Specifically the oxygens of the succinimide can form an hydrogen bond with the side chains of either αv:Arg248 or $\beta 3$:Lys253; while the linker terminal amide group

engages a stable hydrogen bond with the carbonyl group of β3:Asn313. Additional linker-receptor hydrophobic interactions are also observed, involving αv:Ala215 and αv:Ile216.
Concerning the poses of cluster 2 (Figure 4.15, right panel) the cyclopeptide can stabilize hydrophobic interactions with αv:Phe177, αv:Tyr178, β3:Ala218 and β3:Ala252 but not additional polar interactions. However the linker engages several polar contacts with the β3 integrin subunits: specifically with the ion in ADMIDAS position, with β3:Asn313 and with the loop β3:Ser334-Met335-Asp336. In particular the terminal amide group of the linker engages alternate hydrogen bonds with the side chain of β3:Ser334 and the carbonyl group of β3:Met335.
Importantly, in both the models, the analysis of the poses, as well as the differences in sequence and structure of the two integrins, allowed to provide possible explanation of why c(CGisoDGRG)-linker conjugate displays an improved affinity for αvβ3 but not for α5β1. With regards to cluster 1, among the αvβ3 residues directly involved in the interactions with the linker, the polar residues αv:Arg248 and β3:Lys253 are not present in α5β1, being replaced by the hydrophobic α5:Leu257 and β1:Gly261, respectively. Concerning cluster 2, the β1 loop corresponding to β3:Ser334-Met335-Asp336 (i.e. β1:Ser341-Ala342-Asn343), is significantly shifted with respect to β3 and it is involved in the coordination of the ADMIDAS, therefore not being available for interactions with the linker.

Overall, the combination of NMR calculations and modeling techniques allowed to understand that the conjugation of c(CGisoDGRG) with the chemical linker induces an improvement in αvβ3 integrin affinity and selectivity because of the engagement of novel direct interactions with the receptor by the linker. Herein, two possible models of this interacting complex have been proposed: both well reproduce the experimental data of STD experiments and justify the affinity improvement observed only in αvβ3.

In this framework further analysis are needed, therefore in the next feature I plan to: i. rescore by means of MM-GB/SA technique the selected decoy poses of the complex αvβ3 c(CGisoDGRG)-linker in order to quantify the energetic contribution of the linker to the binding affinity; ii. confirm the proposed binding modes by using the information driven docking software HADDOCK that allows to incorporate

experimental data to instruct docking calculations. Herewith, the data derived from STD experiments could be exploited as input to drive the calculations and thus to generate reliable models of the interacting complex.

4.2.3 Conclusion

Peptides based on the isoDGR sequence can be exploited as drug delivery agents to target tumor vessels, thus allowing the administration of extremely low doses of drugs and consequently avoiding unwanted side effects. In this context, the selective recognition of specific integrin subtypes by the isoDGR-based peptide is fundamental to deliver the conjugated nano-drugs or cytokines to the desired target. In this Section I have combined BE-META conformational sampling, Glide and MM-GB/SA rescoring to investigate three integrin-targeting cyclopeptides, namely c(CGisoDGRG), c(GCisoDGRG) and c(CphgisoDGRG), that were shown to be potential drug-delivery agents.

The exhaustive conformational sampling of the molecules revealed that they are quite flexible, not being able to adopt the stable double beta-turns conformations characteristic of cyclic hexapeptides. I proposed that this high flexibility could be ascribed to the presence of the additional isoAspartate Cβ atom within the cycle. The subsequent docking calculations allowed to identify low-nanomolar integrin binders and to rationalize the diverse affinity and selectivity profiles of the three molecules for the two integrins αvβ3 and α5β1, both related to cancer. Specifically, visual inspection of the binding modes, supported by analysis of the MM-GB/SA contributions, unveiled that bulky hydrophobic residues flanking isoAspartate in the isoDGR motif could be exploited to modulate the cyclopeptide selectivity towards α5β1, as confirmed also by previous findings.[78]; conversely, small residues as Glycine in the same position could be useful to fine-tune molecules selectivity towards αvβ3.
Finally, a combination of NMR experiments and computational techniques has been used to explain the improved αvβ3 affinity displayed by c(CGisoDGRG) after conjugation to a sulfo-SMCC-derived chemical linker. Since NMR data revealed that conjugation does not influence cyclopeptide conformation, it has been

concluded that the improved affinity could be related to the presence of additional interactions engaged by the linker. Therefore, I exploited docking calculations and STD experiments to propose two models of the interacting complex: herein the polar interactions observed between the linker and the receptor well justify the improved affinity.

The satisfactory results obtained in this work suggest that the proposed computational strategy, combining BE-META, docking and MM-GB/SA rescoring, can be successfully applied also on isoDGR-based peptides and with diverse receptor integrins. These data hold promises for the future applications of this computational method, especially for its employment in screening the impact of the chemistry/stereo-chemistry of flanking residues of isoDGR-based cyclopeptides on integrin affinity.

4.2.4 Methods

4.2.4.1 BE-META simulations

BE-META simulations of the three cyclopeptides c(CGisoDGRG), c(GCisoDGRG) and c(CphgisoDGRG) have been performed and analyzed as described in Section 4.1.4.2 and 4.1.4.3. Herein the AMBER ff99sb-ildn force field[101,233] was used in combination with the TIP3P[113] water model. After reconstruction of the free-energy landscape in the reduced CVs space (Section 4.1.4.3), the microstates have been clustered into different metastable states in order to individuate the most significant local free energy minima;[138,141] finally the population of each minimum has been computed according to equation 3.2. Only minima populated more than 10% have been considered for further analysis and docking calculations.

4.2.4.2 Docking calculations and MM-GB/SA rescoring with Glide

The software Glide[152] of the Schrödinger suite was employed to perform a first set of docking runs. As input ligand structures, the conformers extracted from the BE-META minima populated more than 10% were used (specifically, a bundle of

30 structures was randomly selected for each minimum of each peptide). Since these structures had already been subjected to an exhaustive conformational search, no further energy minimization or sampling was performed. The receptors input structures (PDB codes: 1L5G[9] and 4WK4[16] for $\alpha v\beta 3$ and $\alpha 5\beta 1$ in complex with Cilengitide and the disulfide cyclized peptide ACRGDGWC, respectively), have been prepared with the Protein Preparation Wizard tool of Maestro (version 10.7). All the crystallographic water molecules were removed, missing side-chains, hydrogen atoms and loops were added; the orientation of the hydroxyl groups of Serine, Threonine and Tyrosine, the side chains of Asparagine and Glutamine residues, and the protonation state of Histidine residues were optimized. Finally, structures were minimized employing an OPLS-AA force field with a root mean square deviation (RMSD) tolerance on heavy atom of 0.3 Å to remove steric clashes and reorient side-chain hydroxyl groups. Cilengitide and ACRGDGWC were used to determine the grid center on $\alpha v\beta 3$ and $\alpha 5\beta 1$, respectively, and subsequently removed. Default grid dimensions, as defined in Glide, for ligands similar in size to the input ones were used. No constrains were applied. All the ligand–receptor complexes were generated by using the Extra-Precision mode of Glide considering a flexible ligand (i.e. flexible side-chains), performing post-docking minimization and saving at most three possible poses for each ligand input structure. After docking calculations the poses were ranked according to their Extra-Precision Glide Score (GS). A first filter was then applied ($GS < (GS_{best} + 5\,kcal\,mol^{-1})$) and only poses satisfying this condition were retained for MM-GB/SA rescoring and further analysis.

The selected decoy poses have been then rescored with the MM-GB/SA method using Prime. After minimization with the local optimization features,[178] energy contributions have been calculated with the all atom OPLS force field and with the GB/SA continuum model.[227] The protein flexibility was taken into account for residues that were within a distance of 5 Å from the ligand, without setting any constraints. After the MM-GB/SA calculations, the geometry of the residues coordinating the metal ions at the binding site was carefully checked and it resulted to be perfectly recapitulated in all the cases. Concerning the ligand, all the side-chain atoms were considered flexible during the calculations.

4.2.4.3 Docking calculations with HADDOCK

A second set of docking calculations has been performed using HADDOCK, version 2.2.[184] As input ligand structures, I used the conformers extracted from the BE-META minima populated more than 10% (in this case, a bundle of 10 structures was randomly selected for each minimum of each peptide). The receptors input structures were the ones optimized for the docking calculations with Glide. The same docking procedure has been repeated for each minimum of each peptide with either αvβ3 or α5β1 integrin.

The solvated docking protocol of HADDOCK, which includes explicit treatment of water molecules, was employed.[166,192] Ambiguous interaction restraints were derived from the known interactions of the RGD motif with integrins αvβ3 or α5β1, as observed in the respective PDB crystal structures (codes: 1L5G and 4WK4). Specifically, I choose as active residues for αvβ3: αv:D150, αv:D218, β3:MIDAS; for α5β1: α5:Q221, α5:D227, β1:MIDAS; for the cyclopeptides: isoAsp and Arg residues of the isoDGR motif. Instead, as passive residues for αvβ3/α5β1 I selected all the residues with relative solvent accessibility >10% (as computed by NACCESS[254]) and within 7 Å from the RGD-based peptide in 1L5G/4WK4 PDB complex structures. During the rigid body docking step 2000 structures were calculated, allowing the ligand to explore solutions rotated by 180°. The best 500 solutions in terms of HADDOCK score were selected for a semi-flexible simulated annealing; finally the best 200 structures after this step have been subjected to a water-refinement procedure and then printed as output. The semi-flexible interface includes all receptors' residues that make intermolecular contacts within a 5 Å cutoff and the ligand. OPLS-AA and TIP3P force field were used for protein and water molecules, respectively, as default in HADDOCK. Default weights have been used for the different contributes of the HADDOCK score (HS) at each step; in particular, for the last step the HS has been computed as: $HS = 1.0\,E_{vdw} + 0.2\,E_{elec} + 1.0\,E_{desolv} + 0.1\,E_{air}$ (see Section 2.3.3).

For each of the six peptide-receptor complex (three peptides vs two receptors), the cluster analysis of the decoy poses has been performed.

4.2.4.4 Cluster analysis

The analysis have been performed clustering the docking poses for each of the six peptide-receptor complexes. The same procedure has been repeated for the decoy poses obtained with either Glide or HADDOCK docking software.
To this aim I used the clustering algorithm described by Daura et al.[153,198] with the help of the tool cluster_struc, a C++ program contained in HADDOCK. The structures have been aligned on the protein backbone residues. The RMSD matrix between the peptides has been calculated on all the backbone atoms and on the non-hydrogen side-chain atoms of the isoDGR motif. The RMSD cut-off for the cluster analysis was set between 2.0 and 2.5 Å and only clusters containing more than four structures have been considered. To each cluster an average MM-GB/SA or an average HADDOCK score (in the case of Glide or HADDOCK docking calculations, respectively) has been assigned by averaging on the score of the best four poses of the cluster.
For each peptide-receptor complex the best scored cluster of decoy poses, among the ones in which the canonical electrostatic clamp with integrin was resembled, was selected. The representation of the decoy poses and the average scores reported in this Chapter always refers to these clusters.

4.2.4.5 Parameterization of c(CGisoDGRG)-linker conjugate

In order to perform MD simulations of c(CGisoDGRG) conjugated with the linker using the AMBER ff99sb-ildn force field, it has been necessary to: i. re-compute the atomic partial charges of the Cysteine residue conjugated with the linker (denoted as CYL); ii. compute the atomic partial charges of the linker (denoted as LIN); iii. assign proper atom types and force-field parameters to the molecule.
Charges parameters for LIN and CYL were derived using the R.E.D. III (RESP ESP charge Derive) package.[245] In order to compute the partial atomic charges, four different conformations have been generated with Maestro,[226] both for the LIN capped with a methanethio group (in either S- and R-configuration) and for the CYL dipeptide capped with an N-methylsuccinimide group (in either S- and R-configuration). A 2D-representation of the structures used for the partial charges

computation are reported in Figure C.4. Afterwards the GAMESS[246] package was used for geometry optimization in the gas phase and for the computation of the molecular electrostatic potentials (MEP). Herein, the HF/6-31G* level of theory and the Connolly surface algorithm were employed. Additionally four different molecular orientations for each optimized geometry have been considered using the rigid-body reorientation algorithm implemented in the R.E.D. tool. Finally the two-stages RESP method was used for the fitting of the atomic charges, following the procedure originally published by Kollman et al.[247] During this step, intra-molecular constraints were imposed to set the charges of the capping groups (including the acetyl, N-methyl, methanethio and N-methylsuccinimide groups) to zero; additionally, charges equivalences were imposed to hydrogens of methyl, methylene and amine groups, as well as to symmetric atoms in the cyclohexane. Partial atomic charges of the atoms N, HN, C and O of CYL were set to the same partial charges of the equivalent atoms in Cysteine, according to the AMBER ff99sb-ildn force field. Since similar results were obtained for partial charges computation with capping group in S- and R-configuration (differences in partial charges of equivalent atoms lower than 0.1), only the data coming from the first set of calculations were used. The derived RESP partial charges for LIN and CYL atoms are reported in Table C.3.
Finally, the necessary topology files to perform MD simulations with the AMBER ff99sb-ildn force field of c(CGisoDGRG) conjugated with the linker were generated using the Antechamber tool.[120] Amber atom types were used, when some parameters were missing the analogous GAFF parameters have been employed.[119]

4.2.4.6 BE-META simulations of c(CGisoDGRG)-linker conjugate

Input conformations for c(CGisoDGRG) conjugated with the linker were generated with Maestro; in particular four different conformers of c(CGisoDGRG), representative of the four free-energy minima identified within the BE-META calculation of the cyclopeptide described in Section 4.2.2.1, were considered. At each of these conformers, the linker has been attached in both R and S configurations, resulting in the generation of 8 input structures.
For each of the input structures, energy minimization, equilibration and BE-META

calculations have been performed restraining the dihedral angles of the cyclopeptide previously used to identify the microstates (i.e. the 6 ψ dihedral angles and the ξ dihedral characteristic of isoAspartate). This was done in order to maintain the cyclopeptide in its initial conformational minimum and to investigate only the degrees of freedom of the linker. These dihedral angles were restrained with an harmonic-like potential to the values characteristic of their free-energy minimum (i.e. the average values computed on all the structures belonging to that minimum) with a force constant of 100 kJ/mol/rad^2 if they were far from the target value more than 30°. During both equilibration and production run, the same simulation settings described in 4.1.4.2 were used. In BE-META simulations five Collective Variables (CVs) were biased, describing the degree of freedom of the Cysteine side chain and of the linker. Specifically, these are: i. N-Cα-Cβ-S; ii. Cα-Cβ-S-C1; iii. Cβ-S-C1-C5; iv. C4-N3-C6-C7; v. N3-C6-C7-C8 (the numeration of linker atoms is according to Figure C.4). The length of each replica, in which one of the 5 CVs was biased, was of 20 ns. As in the previous cases, the convergence was checked comparing the mono-dimensional free-energy profiles derived by averaging on the two halves of the simulation. In all the cases the convergence was reached and the difference between the two profiles was found to be always lower than $2k_bT$.
To identify the possible conformational minima of the eight c(CGisoDGRG)-linker conjugates, I analyzed the BE-META trajectories with the Metagui tool,[141] using the procedure described in Section 2.2.3. The same CVs biased during the production run were used for the analysis. From the 4/5 most populated conformational minima herein identified, I saved 25 structures randomly selected, so to obtain a pool of 100 structures for each of the eight input c(CGisoDGRG)-linker conformations to be used for docking calculations.

4.2.4.7 Docking calculations of c(CGisoDGRG)-linker conjugate

The pools of structures identified during the sampling step have been used as ligand input structures for subsequent docking calculations, performed with the software Glide of the Schrödinger suite. The αvβ3 input structure (PDB codes: 1L5G) has been prepared as described above (see Section 4.2.4.2). Additionally, all the same settings used during the docking of the cyclopeptide with αvβ3 were

used, with the following exceptions: i. the grid dimension was enlarged to fit ligands with dimensions of 28 Å, ii. positional constraints were applied to force the cyclopeptide to form the canonical interaction clamp with integrin. In particular the atoms of the Arginine guanidinium group and isoAspartate carboxylic group of c(CGisoDGRG) were forced to occupy a spherical volume of 3 Å centered at the position occupied by the analogous atoms of Arginine and Aspartate of Cilengitide in 1L5G. For each of the input structures, at most three decoy poses, satisfying these constraints, were printed as output.
NMR data from STD experiments have been used to individuate the hydrogens of the linker that are in close proximity to the receptor protons (see Figure 4.14). Only the docking poses in which all the individuated hydrogens were closer than 4 Å to at least one $\alpha v\beta 3$ proton have been selected. In the case of equivalent hydrogens, at least one of the atoms was requested to be closer than the cutoff to the receptor protons. The selected docking poses have been clustered according to geometric criteria as described in the previous Section; herein the RMSD matrix between the docked ligands has been calculated on all the backbone atoms of the cyclopeptide and on the non-hydrogen atoms of the linker. The RMSD cut-off for the cluster analysis was set to 3 Å and only clusters containing more than four structures have been considered.

Chapter 5

Conclusion and Perspectives

The main objective of this PhD work is to investigate how computational tools could be applied to accelerate the design process and the development of therapeutic cyclopeptides. Cyclopeptides are indeed gaining increasing importance as modulators of protein-protein interactions and of other difficult therapeutic targets thanks to their favorable pharmacokinetic properties and to their ability to maximize the interacting surface with the receptor while constraining ligands into pre-organized bioactive protein-binding conformations. In particular, cyclopeptides have been widely and successfully applied in the field of inhibition of integrins, a class of trans-membrane receptors involved in several human diseases including thrombosis, strokes, cancer and angiogenesis. Herein, cyclopeptide affinity and selectivity for the diverse integrins subtypes can be fine-tuned by modulating the residues flanking the recognition motifs (RGD or isoDGR) or by introducing chemical modifications, as acetylation and methylation. In this context, it is well recognized that the three-dimensional conformation adopted, already in solution, by the cyclopeptide, and particularly by the recognition motif, is critical for the selective recognition of different integrins; therefore a strong correlation exists linking cyclopeptide conformation and its integrin affinity. Despite the interest inspired by cyclopeptides, their optimization as drugs delivery agents or therapeutics is still mainly based on empirical approaches, requiring expensive and time-consuming synthesis campaigns. In this framework, computational techniques

could be extremely useful to aid and accelerate the process of cyclopeptides design and optimization. However, the application of these techniques remains challenging because of the difficulties in determining cyclopeptides preferred conformations, due to their constrained geometry, and consequently in predicting their binding affinity.

Within this thesis I have shown that the combination of enhanced sampling molecular dynamic techniques (specifically Bias Exchange Metadynamics, BE-META), docking calculations and rescoring methods (Molecular Mechanics-Generalized Born Surface Area, MM-GB/SA) is a reliable approach to perform structure-activity studies in integrin targeting cyclopeptides. On one side, BE-META allows the identification of the conformations preferentially adopted by the cyclopeptide in solution, as well as the evaluation of their relative equilibrium; on the other side docking calculations, complemented by MM-GB/SA rescoring and cluster analysis, allow to individuate the conformers with favorable binding properties. As a result, a reliable ranking of the cyclopeptides according to their binding affinity and rationalization of their structure-activity relationship is obtained.
The proposed multi-stage computational protocol has been applied in different projects.
Firstly, as described in Chapter 3, it has been employed to investigate the impact of single and multiple *N*-methylations on the equilibrium conformation of five RGD-based cyclic hexapeptides, which displayed highly different affinities for αIIbβ3 integrin. From a structural point of view, I have found that introduction of *N*-methylation in the investigated RGD-based cyclopeptides can modulate their structural preferences by favoring the accessibility to additional conformations, characterized by a $\sim$180° twist of the peptide bond plane preceding the methylated residue with respect to the unmethylated peptide. In particular, I have shown that the diverse pre-organization of the cyclopeptide in solution is responsible of the observed divergences in αIIbβ3 binding strength, rationalizing the interplay between structure and affinity. Of note, both the conformational sampling and the predicted ranking of cyclopeptides according to their binding strength are in excellent agreement with the available experimental data, proving the reliability of the proposed methodology.

As a second project, described in Chapter 4, I have investigated the applicability of the protocol in structure-activity studies of cyclopeptides based on the novel integrin recognition motif isoDGR. In order to verify the accuracy of MD techniques in reproducing the equilibrium properties of molecules containing the non-canonical amino acid isoAspartate, I have systematically tested eight different force fields variants on five isoDGR-based cyclic peptides, for which NMR experiments have been acquired. The results suggest that most of the investigated force fields can properly reproduce the equilibrium J scalar couplings of cyclic peptides, but only two out of the eight force fields considered (i.e. AMBER ff99sb-ildn and AMBER ff99sb*-ildn) are accurate in reproducing the NMR observables characteristic of isoAspartate. I therefore concluded that transferability of force field parameters is not straightforward and that introduction of isoAspartate, and generally of non-standard amino acids, into cyclopeptides sequence could affect the reliability of the simulations. However, using the most accurate force field among the studied ones, I have been able to investigate the activity and selectivity profiles for the two integrins $\alpha v\beta 3$ and $\alpha 5\beta 1$ of three isoDGR-based cyclopeptides, differing for the type of flanking residues and showing suitable tumor-homing properties. Interestingly, previous studies have shown that one of these molecules, c(CGisoDGRG), after conjugation to human serum albumin, can be used as a drug delivery agents targeting tumor vessels by selectively recognizing integrin $\alpha v\beta 3$. The application of the computational protocol allowed to discriminate low-nanomolar integrin binders from weak binders; additionally, the visual inspection of the binding modes, complemented by the analysis of the MM-GB/SA contributions, revealed the structural basis that regulate cyclopeptide selectivity towards $\alpha 5\beta 1$. Herein it has been suggested that bulky hydrophobic residues preceding isoAspartate in isoDGR-based peptides could be exploited to drive the selectivity towards $\alpha 5\beta 1$. Lastly, I have rationalized the findings of previous biochemical studies, in which it was shown that the chemical linker used to conjugate c(CGisoDGRG) with human serum albumin plays an important role in regulating cyclopeptide affinity and selectivity towards $\alpha v\beta 3$. Combining NMR experimental data and computational methods, I proposed two models of the interacting complex well-explaining this improved $\alpha v\beta 3$ affinity and selectivity.

Overall, in this thesis I have shown that the proposed computational protocol is a valuable strategy for the study of cyclopeptides conformations and binding affinity, allowing the investigation of both RGD and isoDGR-based cyclopeptides, and also being able to account for diverse integrin subtypes as receptor. Of note, both the mechanisms influencing affinity modulation have been successfully investigated: in one case the integrin affinity was regulated by tuning the preferred cyclopeptide conformations, in the second case it was influenced by the engagement of additional direct interactions with the receptor. Considering these positive results, we plan to extend and further apply this computational protocol to screen novel integrin-targeting peptides library with the aim to accelerate and aid the process of design and development of therapeutic cyclopeptides.

Specifically, considering that the research field of isoDGR-based integrin inhibitors is relatively recent and new efforts are needed to explore their potential as therapeutics, drug carriers or agents for diagnostic imaging, I aim to apply the computational method to study the impact of chemistry and stereo-chemistry of flanking residues of isoDGR-based cyclopeptides on ligands conformation and integrin binding activity. Experimentally, I plan to systematically screen a library of head-to-tail cyclic pentapeptides of the form c(X_1isoDGRX_2), where X_i can be any of the twenty existing natural amino acids in both L- and D-configuration. Based on the well accepted idea that the stereo-chemistry of flanking residues is fundamental in dictating the backbone conformational preferences of cyclic peptides, while the side-chain only negligibly affects this equilibrium, I will perform the BE-META simulations only on nine reference peptides c(X_1isoDGRX_2), in which X_1 and X_2 can be L-Alanine, D-Alanine or Glycine. Then, the docking calculations will be run on the whole library of molecules, using as input backbone-structures the ones identified within the sampling. In order to accelerate the work-flow, all the steps, including generation of the library, simulation, docking and analysis will be automatized by home-made python/bash scripts, which have been already optimized and tested for the simulations described in Chapter 4. This kind of screening could be significant for the development of new cyclic isoDGR-based $\alpha v\beta 3$ inhibitors.

Appendices

Appendix A

Chapter 3: Supplementary Tables and Figures

Table A.1: **Definition of BE-META minima according to their position in the six-dimensional space.** In Table are listed for each minimum the corresponding lower and upper bound in the six dimensions.

Name	ϕ(R2) [°]		ψ(R2) [°]		ϕ(f5) [°]		ψ(f5) [°]		ϕ(L6) [°]		ψ(L6) [°]	
	LB	UB	LB	UB	LB	UB	LB	UB	LB	UB	LB	UB
A	-180	0	60	180	0	180	-180	-60	-180	0	-60	60
B	0	180	-60	60	0	180	-180	-60	-180	0	-60	60
C	-180	0	-60	60	-180	0	-60	60	-180	0	-60	60
D	0	180	-60	60	-180	0	-60	60	-180	0	-60	60

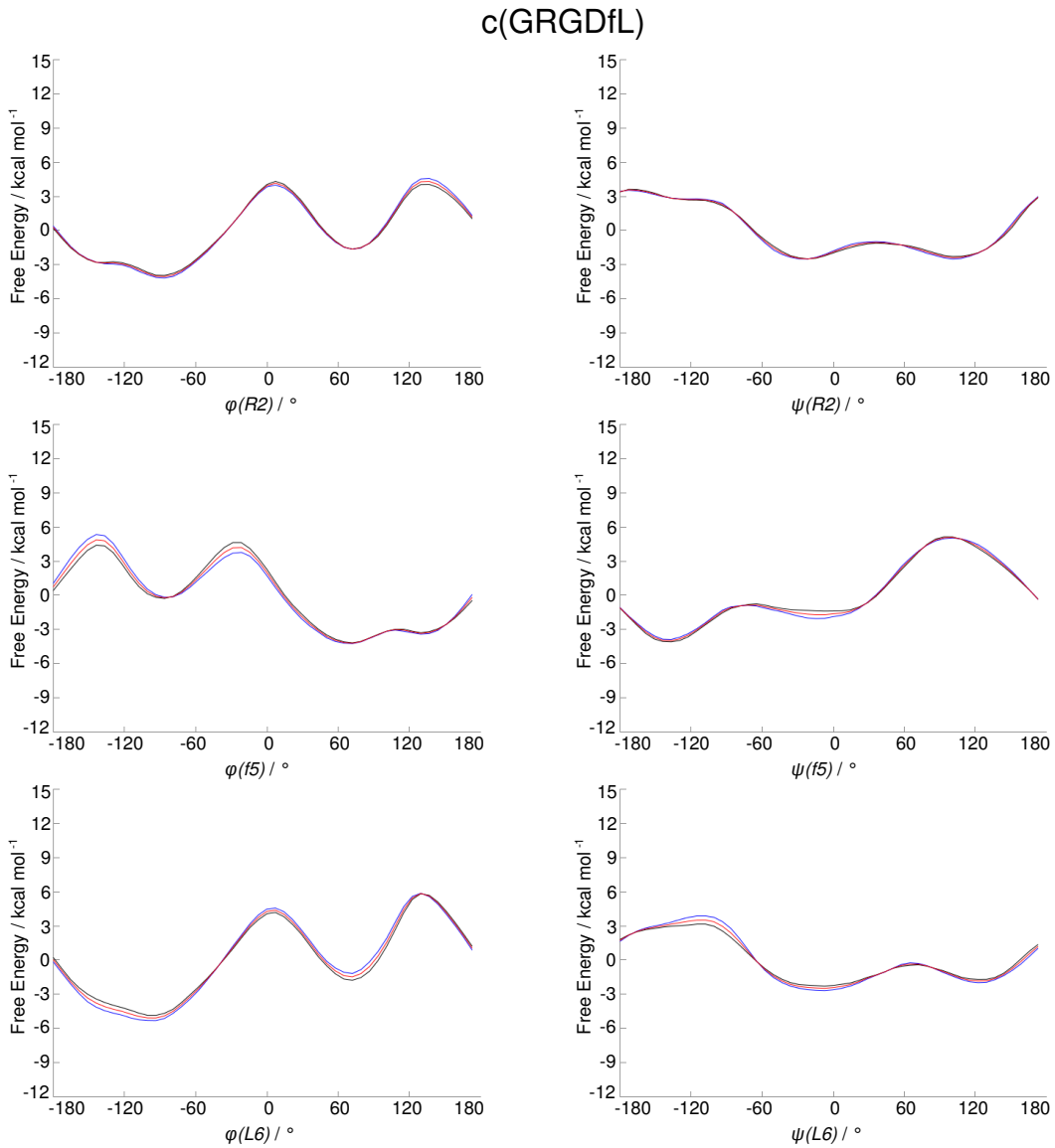
Table A.2: **MM-GB/SA vs Glide Score.** For each cluster, the average Extra Precision Glide Score (GS) and the average MM-GB/SA energy values are reported. Additionally, it is indicated if the poses of the cluster are able (y) or not (n) to summarize canonical-like interactions, bridging the two domains of integrin αIIbβ3. We observed that GS is not able to discriminate poses summarizing canonical interactions from the others: as an example clusters 2 (*good*) and 7 (*bad*) display a similar GS. Conversely, the MM-GB/SA score clearly distinguishes between poses that summarize the canonical interactions (cluster 1- 6), from the others (cluster 7 to 13).

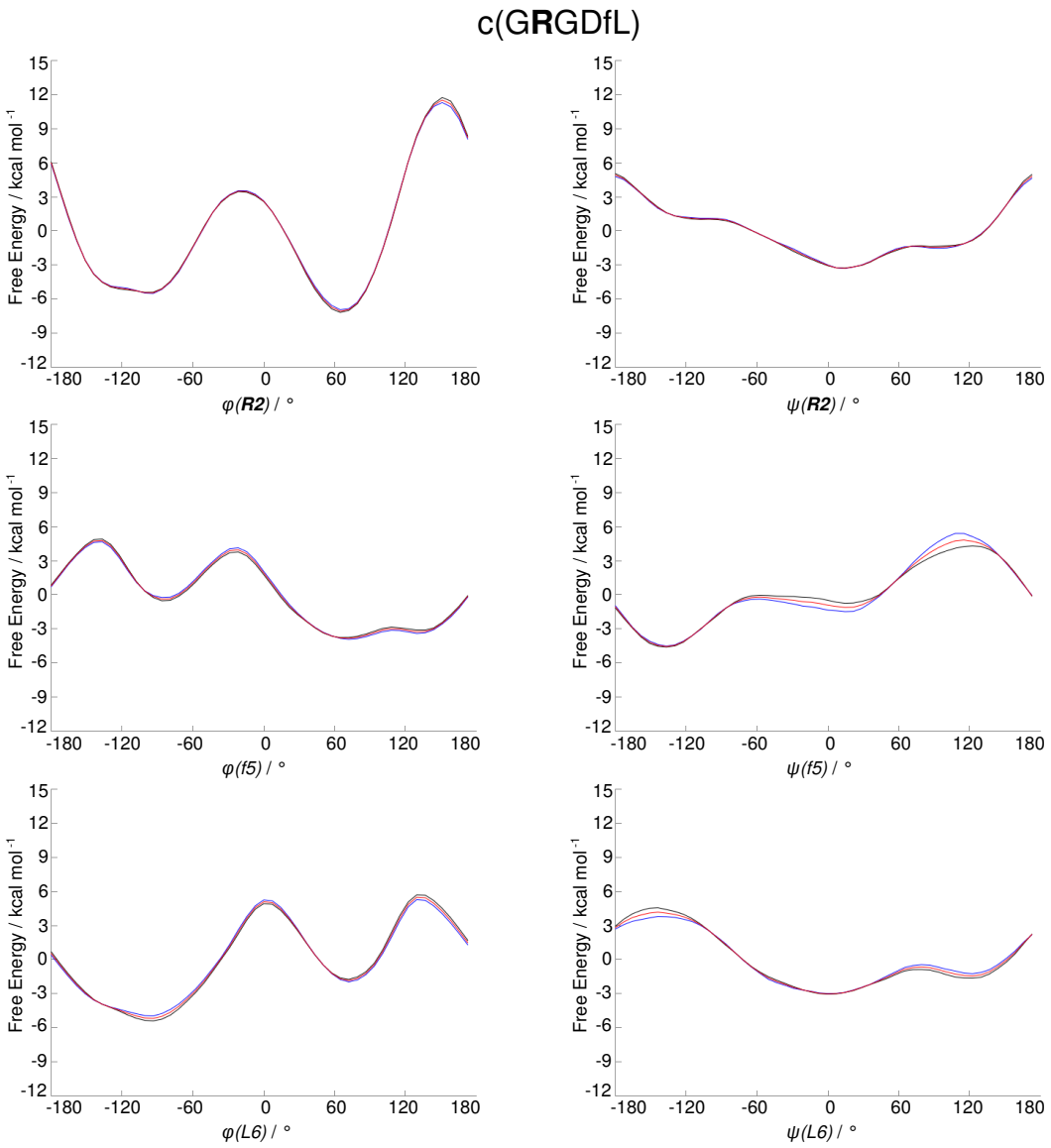
Cluster	$\langle$GS$\rangle$ [kcal mol^{-1}]	$\langle$MM-GB/SA$\rangle$ [kcal mol^{-1}]	Canonical interactions
1	-9.5 ± 0.7	-68.3 ± 6.1	y
2	-9.1 ± 0.7	-57.7 ± 6.5	y
3	-9.3 ± 0.7	-56.3 ± 9.5	y
4	-9.0 ± 0.6	-56.2 ± 7.0	y
5	-9.1 ± 0.6	-53.4 ± 8.4	y
6	-8.7 ± 0.4	-49.9 ± 10.7	y
7	-9.1 ± 0.8	-46.9 ± 8.6	n
8	-9.3 ± 0.9	-43.2 ± 10.1	n
9	-8.6 ± 0.4	-43.1 ± 9.2	n
10	-8.8 ± 0.6	-40.8 ± 7.5	n
11	-8.5 ± 0.4	-36.0 ± 7.7	n
12	-8.6 ± 0.4	-34.8 ± 6.1	n
13	-8.6 ± 0.3	-28.2 ± 5.1	n

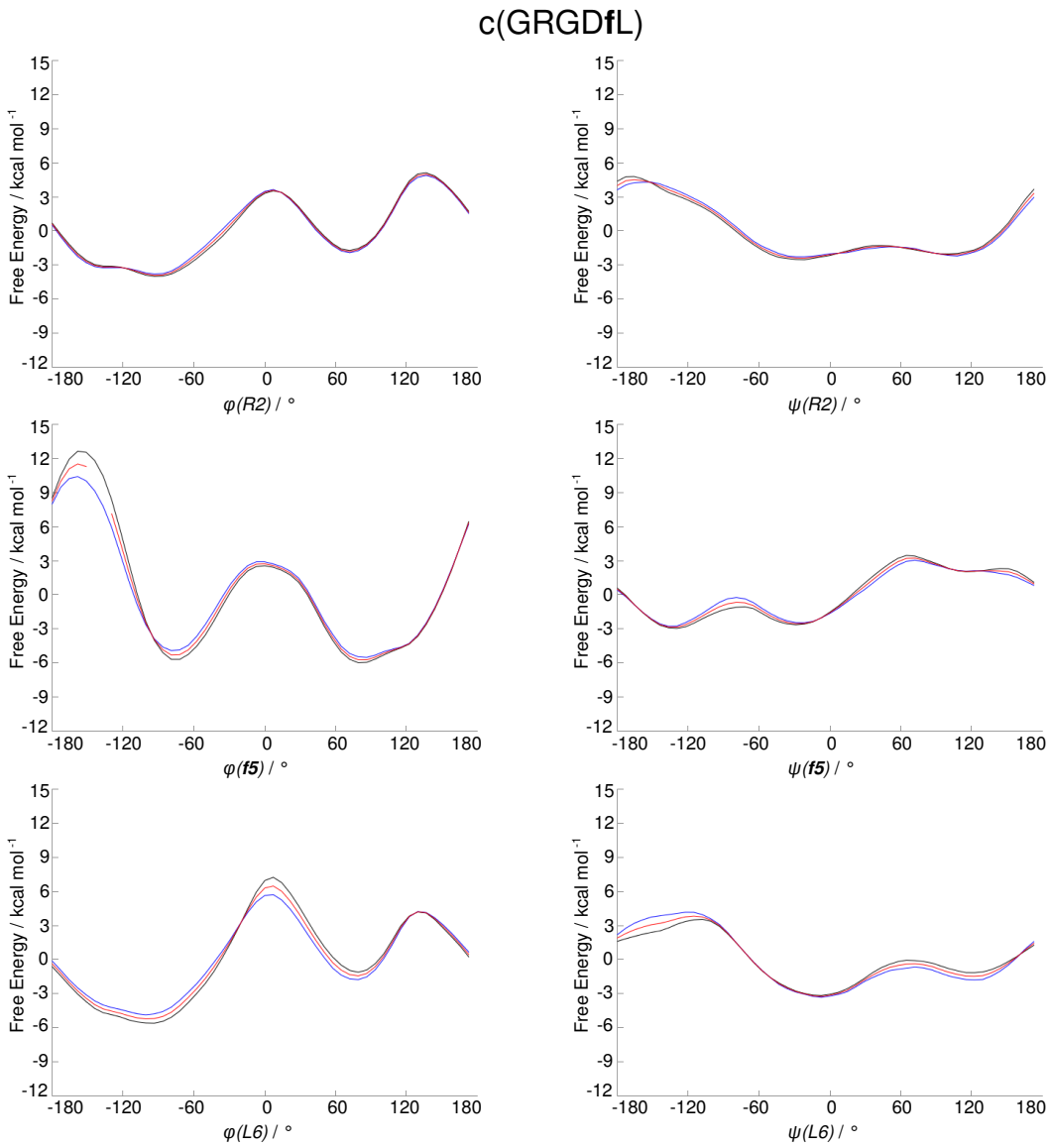
Table A.3: **MM-GB/SA energetic contributions per cluster.** The main MM-GB/SA contributions for each cluster, including Coulomb (Coul), Generalized Born electrostatic solvation energy (GBS), van der Waals (vdW), lipophilic energy (lipo), covalent binding (cov), hydrogen-bonding (hb), π-π packing energy (π-π) and self-contact correction (self) terms are reported. The errors concerning the Polar, Non Polar and Additional contributions are the ones of Table 3.4 in the main text.

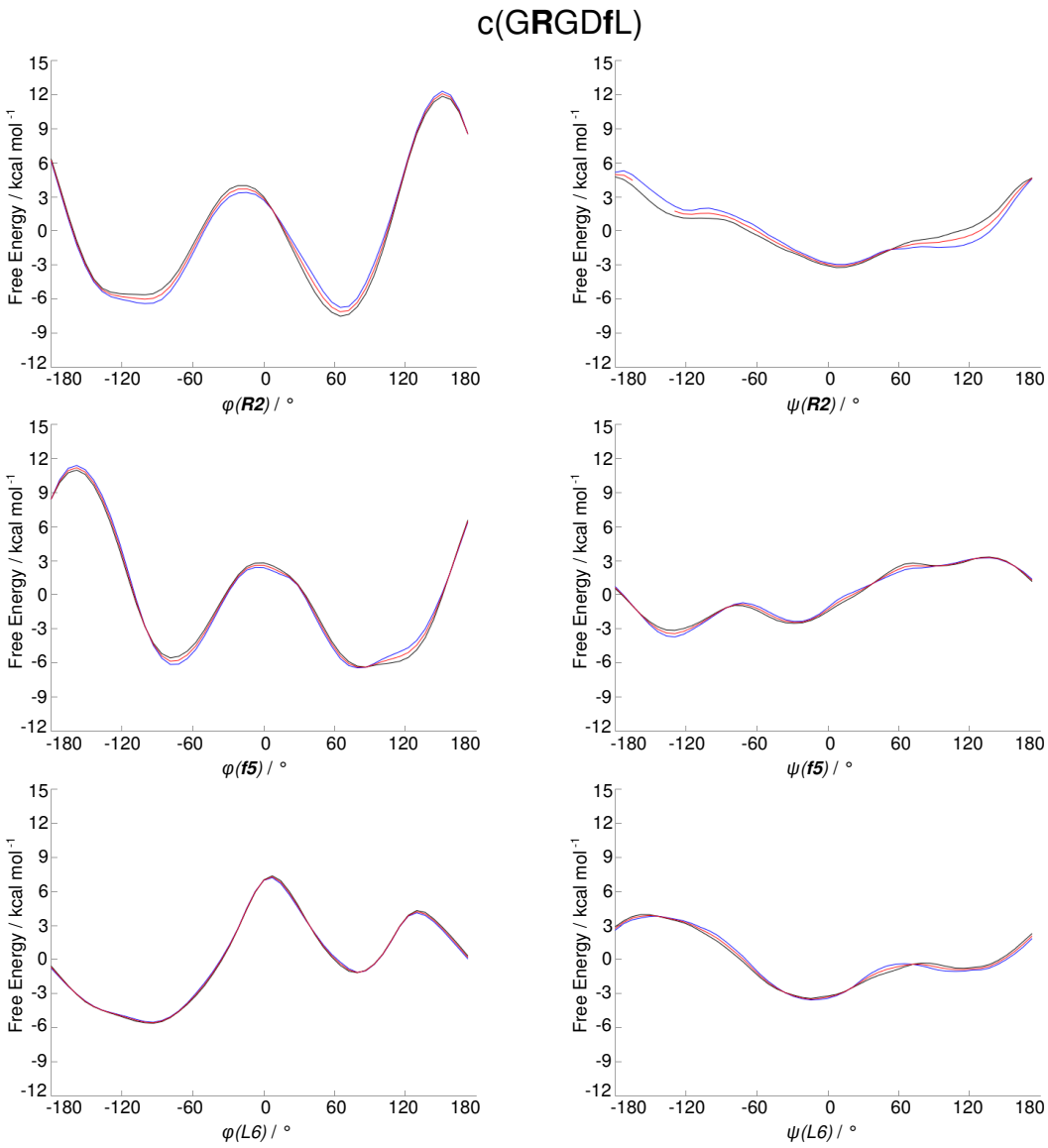
Cluster	Energetic Contributions [kcal mol^{-1}]							
	Polar		Non Polar			Additional		
	⟨Coul⟩	⟨GBS⟩	⟨vdW⟩	⟨lipo⟩	⟨cov⟩	⟨hb⟩	⟨π-π⟩	⟨self⟩
1	-57.8	73.6	-58.8	-30.1	7.2	-3.0	0.8	-0.3
2	-46.5	59.0	-48.2	-25.5	6.0	-3.0	0.6	0.0
3	-55.1	67.8	-47.8	-25.8	8.2	-3.9	0.2	0.0
4	-64.4	79.6	-48.8	-28.9	9.9	-4.0	0.7	-0.3
5	-49.1	65.6	-49.2	-24.6	6.8	-2.9	0.2	-0.3
6	-68.4	87.0	-52.1	-23.6	9.7	-2.5	0.2	-0.2
7	-25.0	59.9	-59.6	-32.4	11.6	-3.9	2.5	0.0
8	-30.6	59.6	-49.6	-30.3	10.4	-4.4	1.7	0.0
9	-20.6	46.1	-48.1	-26.2	6.0	-1.1	1.0	-0.3
10	-47.2	79.0	-52.7	-24.8	5.7	-3.1	2.2	0.0
11	-8.0	40.1	-49.0	-23.2	5.1	-1.2	0.6	-0.3
12	-38.8	67.1	-44.6	-23.7	5.3	-2.4	2.2	0.1
13	-14.9	54.5	-48.4	-23.3	3.6	-1.0	1.4	-0.1

Figure A.1: **Monodimensional free energy profiles.** Convergence of the free energy profiles for each cyclopeptide, as a function of ϕ and ψ dihedral angles of R2, f5 and L6. The black profile represents the time average within $(t_{eq}, (t_{eq} + t_{tot})/2)$, the blue one represents the time average within the second half $((t_{eq} + t_{tot})/2, t_{tot})$; the red line is the average between the black and blue profiles, which are consistent within $2k_bT$. In Figure, FES for $t_{eq} = 1$ ns and $t_{tot} = 40$ ns are reported.[213]




c(GRGDfL)
Free Energy / kcal mol-1
φ(R2) / °
ψ(R2) / °
φ(f5) / °
ψ(f5) / °
φ(L6) / °
ψ(L6) / °


c(GRGDfL)
Free Energy / kcal mol -1
φ(R2) / °
ψ(R2) / °
φ(f5) / °
ψ(f5) / °
φ(L6) / °
ψ(L6) / °


c(GRGDfL)
Free Energy / kcal mol-1
φ(R2) / °
ψ(R2) / °
φ(f5) / °
ψ(f5) / °
φ(L6) / °
ψ(L6) / °

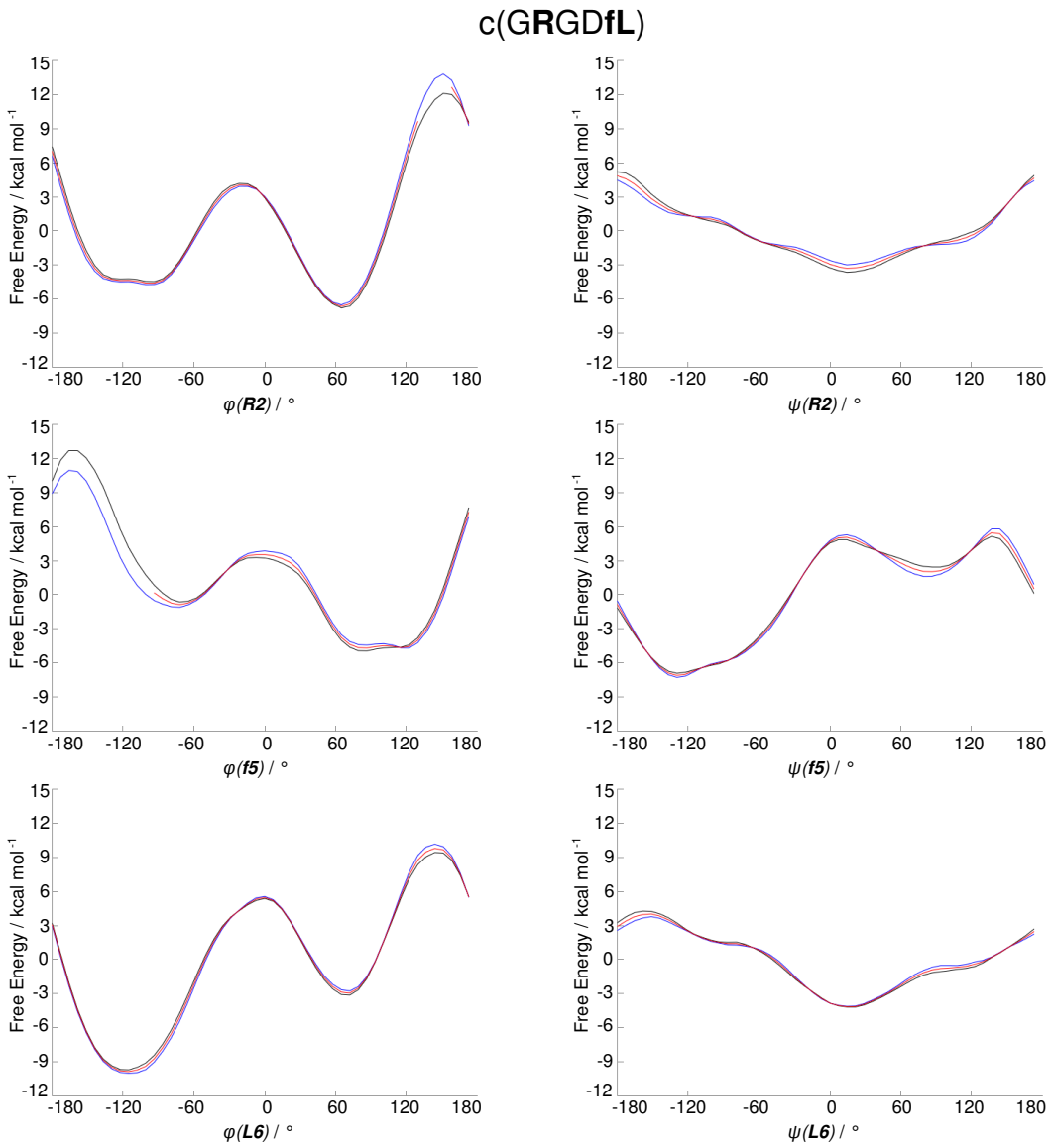
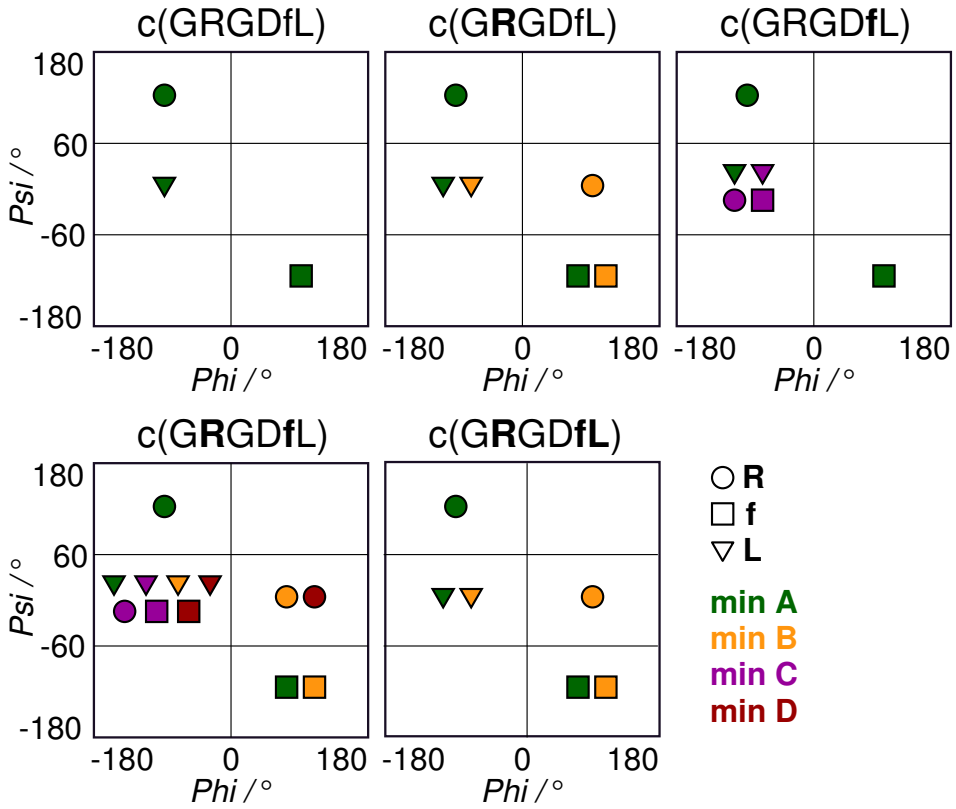

c(GRGDfL)
Free Energy / kcal mol -1
φ(R2) / °
ψ(R2) / °
φ(f5) / °
ψ(f5) / °
φ(L6) / °
ψ(L6) / °

Figure A.2: **BE-META minima in a Ramachandran-like plot.** [213] The coordinates of the main free-energy minima are reported in the six-dimensional space defined by the ϕ and ψ dihedral angles of R2, f5 and L6. Shapes and colors represent residue and minima, respectively.



Appendix B

Section 4.1: Supplementary Tables and Figures

Table B.1: **RESP charges.** In Table are reported the RESP atomic partial charges for Phenylglycine and isoAspartate, derived as described in section 4.1.4.2.

Phenylglycine		isoAspartate	
Atom Names	Partial Charges [e]	Atom Names	Partial Charges [e]
N	-0.3515	N	-0.6589
H	0.2453	H	0.3423
Cα	-0.0660	Cα	0.1271
Hα	0.1388	Hα	0.0597
Cβ	-0.0197	Cβ	-0.0177
Cγ*	-0.0729	Hβ*	0.0031
Hγ*	0.1071	Cγ	0.7753
Cδ*	-0.1717	Oγ*	-0.7695
Hδ*	0.1455	C	0.4700
CZ	-0.1248	O	-0.5650
HZ	0.1392		
C	0.5367		
O	-0.5140		

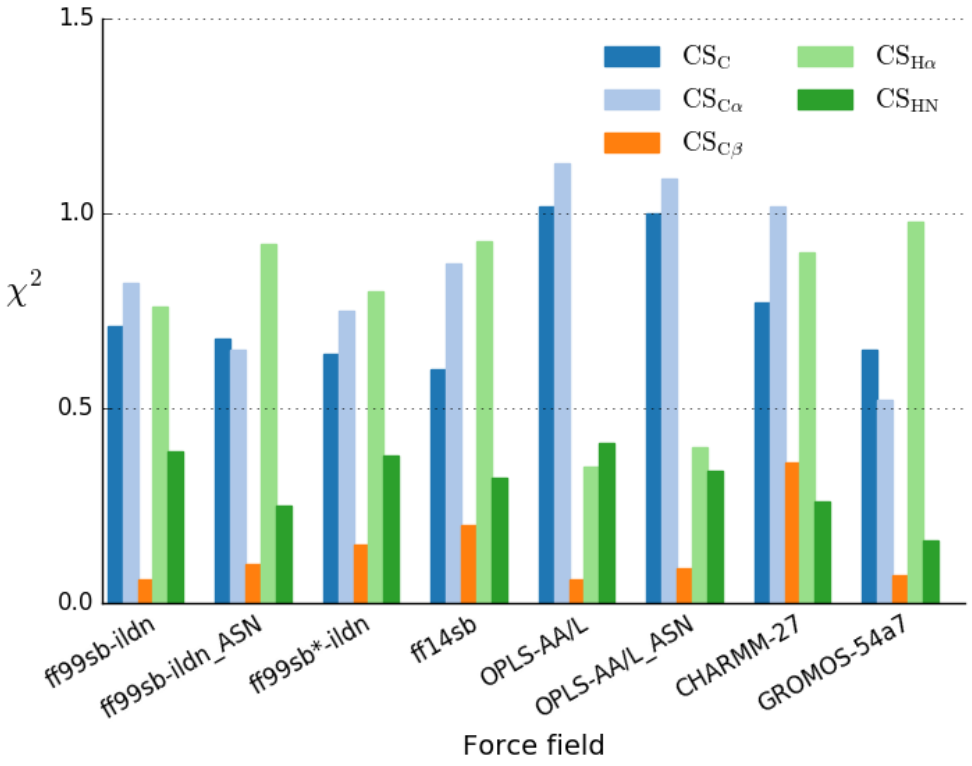


Figure B.1: **Chemical shift χ^2, subdivided per type.** For each force field, the ability to reproduce experimental chemical shift displacements is quantified thorough the χ^2 function. The differences in the prediction of chemical shift displacements for various atoms (C, Cα, Cβ, Hα and HN represented in blue, light blue, orange, light green and green, respectively) are highlighted.

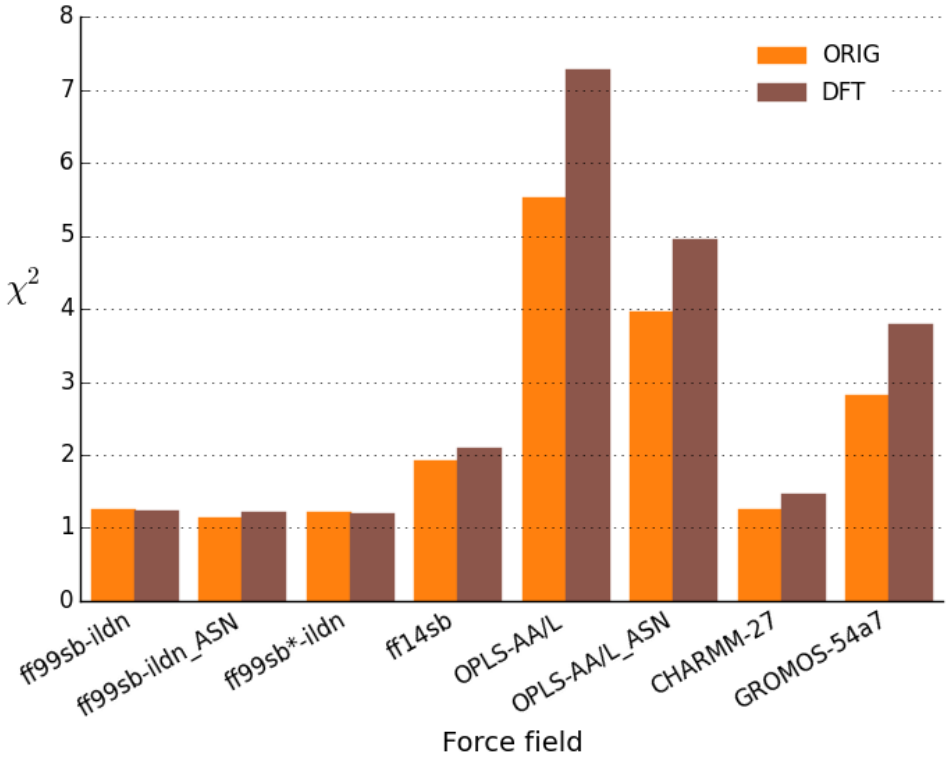


Figure B.2: χ^2 **values, indicating the agreement between experimental and back-calculated** $^3J_{\mathbf{HN,H\alpha}}$**, computed with different sets of Karplus parameters.** The χ^2 values, estimating the ability of each force field to reproduce experimental $^3J_{\mathrm{HN,H\alpha}}$ scalar couplings, obtained with ORIG (orange bars) or DFT (brown bars) Karplus parameters are reported.

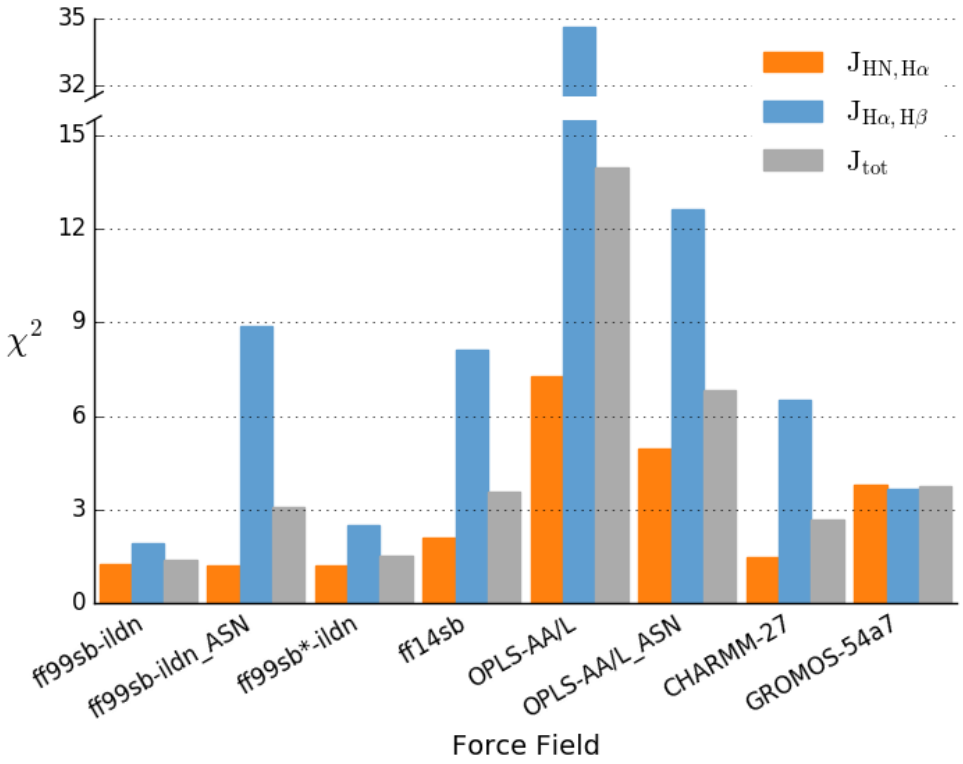


Figure B.3: **Comparison between experimental and computed J scalar couplings.** For each force field, the ability to reproduce experimental J scalar couplings is quantified through the χ^2 function. The χ^2 values obtained, considering alternatively ${}^3J_{\mathrm{HN,H}\alpha}$, ${}^3J^{\mathrm{isoD}}_{\mathrm{H}\alpha,\mathrm{H}\beta}$, or both these variables together, is shown with orange, blue and gray bars, respectively. In order to avoid flattening of the data a gap in the y axis has been introduced. All the ${}^3J_{\mathrm{HN,H}\alpha}$ scalar couplings have been computed using the DFT set of Karplus parameters.

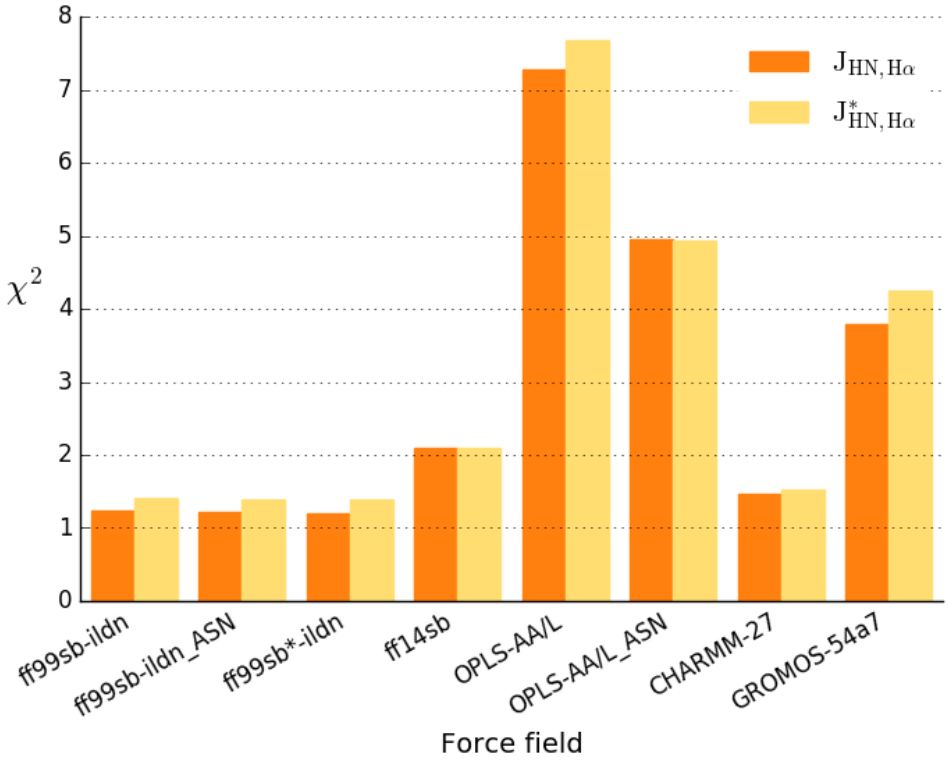


Figure B.4: **Influence of isoAspartate on $^3J_{\mathbf{HN,H\alpha}}$ reproducibility.** For each force field, the ability to reproduce experimental $^3J_{\mathrm{HN,H\alpha}}$ scalar couplings is quantified through the χ^2 function. The χ^2 values, computed including or excluding the coupling constants associated to isoAspartate ϕ', are reported as orange and yellow bars, respectively. All the $^3J_{\mathrm{HN,H\alpha}}$ scalar couplings have been computed using the DFT set of Karplus parameters.

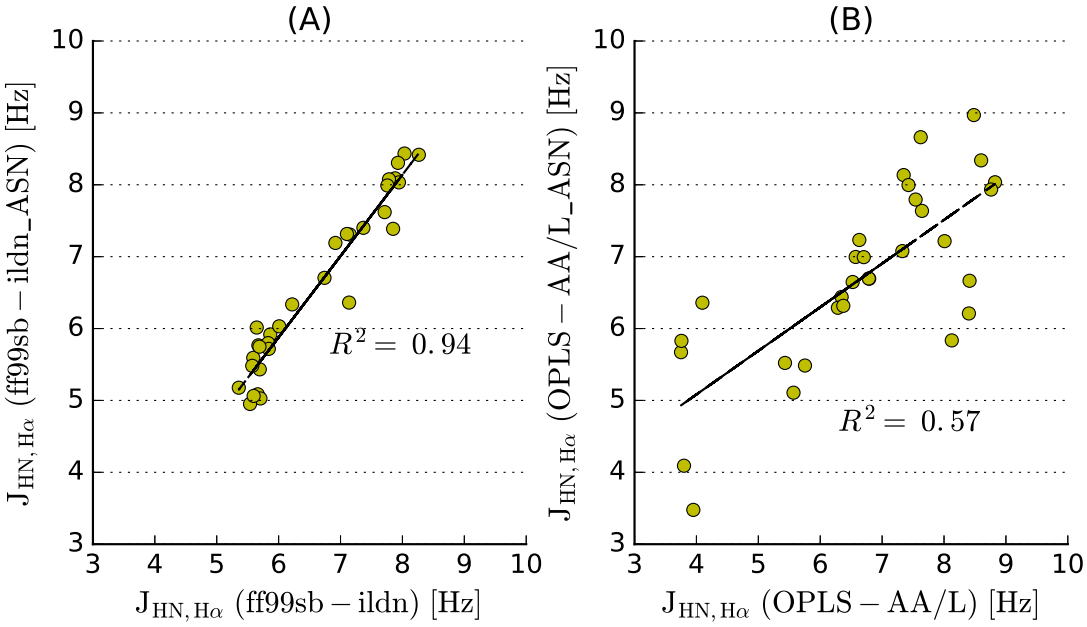


Figure B.5: **Correlation of $^3J_{\mathbf{HN,H\alpha}}$ computed with different force field variants.** Panels show the correlation between $^3J_{HN,H\alpha}$ computed with (A) AMBER ff99sb-ildn force field or its variant ff99sb-ildn_ASN; (B) OPLS-AA/L force field or its variant OPLS-AA/L_ASN. All the $^3J_{HN,H\alpha}$ scalar couplings have been computed using the DFT set of Karplus parameters.

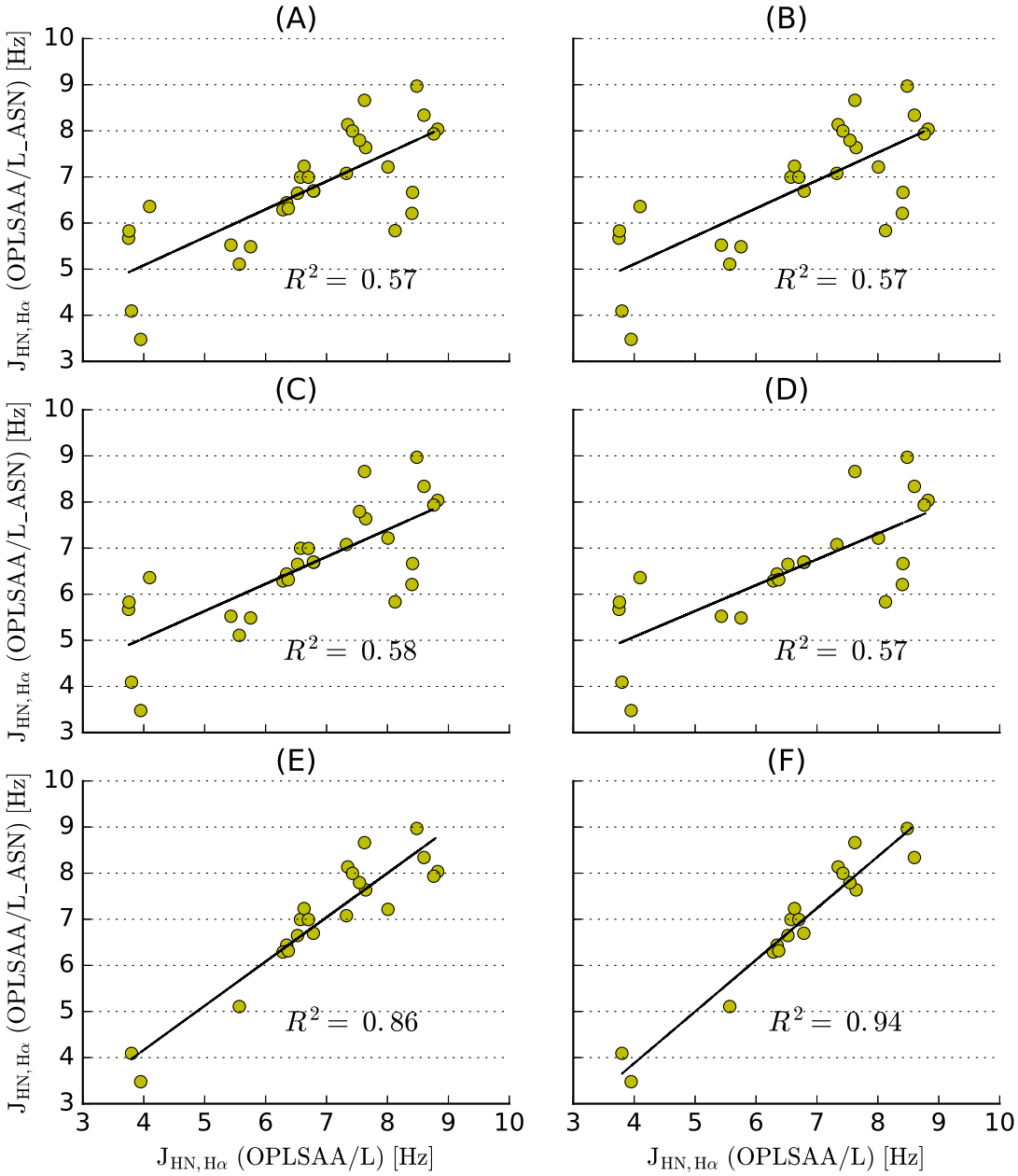


Figure B.6: **Correlation of $^3J_{\mathbf{HN,H\alpha}}$ computed with OPLS-AA/L and OPLS-AA/L_ASN force field variants.** Panels show the correlation between: (A) all $^3J_{\mathrm{HN,H\alpha}}$, (B) all $^3J_{\mathrm{HN,H\alpha}}$ except isoAspartate, (C) all $^3J_{\mathrm{HN,H\alpha}}$ except the residue preceding isoAspartate, (D) all $^3J_{\mathrm{HN,H\alpha}}$ except the two residues preceding isoAspartate, (E) all $^3J_{\mathrm{HN,H\alpha}}$ except the residue following isoAspartate, (F) all $^3J_{\mathrm{HN,H\alpha}}$ except the two residues following isoAspartate, computed with OPLS-AA/L force field or its variant OPLS-AA/L_ASN. All the $^3J_{\mathrm{HN,H\alpha}}$ scalar couplings have been computed using the DFT set of Karplus parameters.

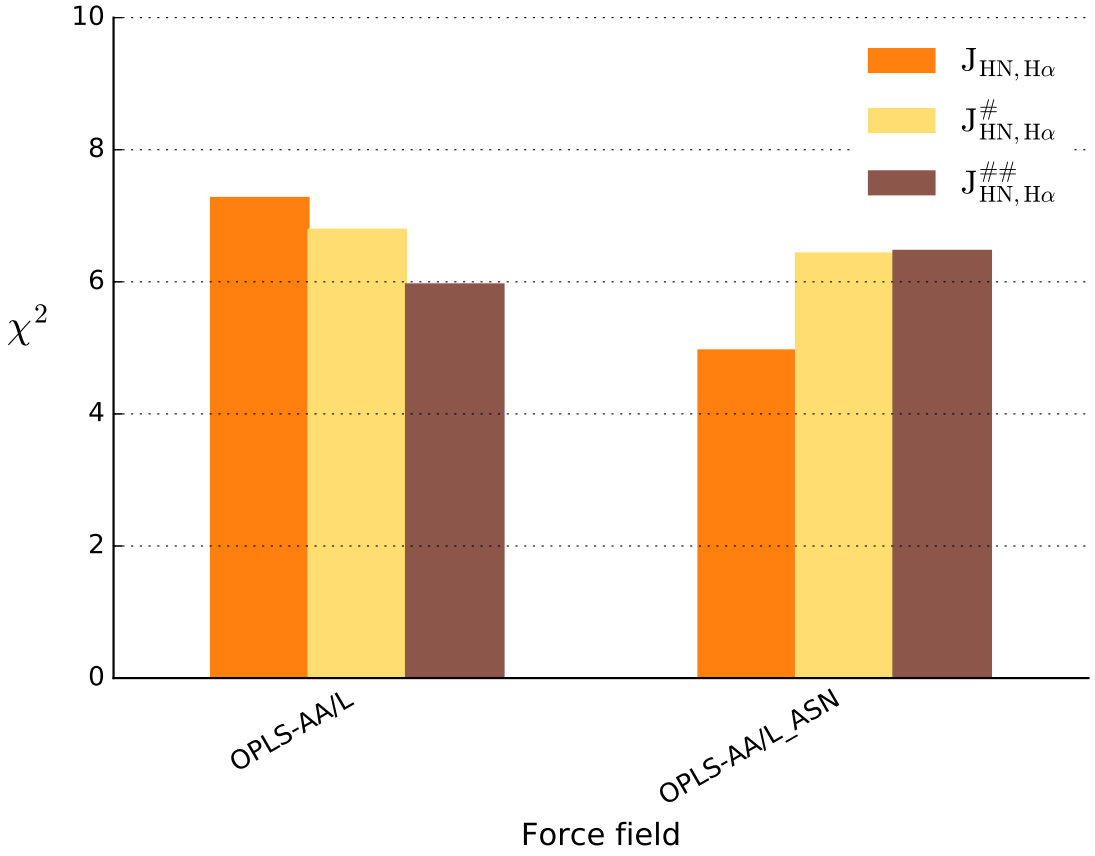


Figure B.7: **Influence of isoAspartate parameterization in OPLS on the $^3J_{HN,H\alpha}$ χ^2.** For OPLS-AA/L and OPLS-AA/L_ASN force fields, are reported the χ^2 values, estimating the ability to reproduce: all the experimental $^3J_{HN,H\alpha}$ scalar couplings (orange bars), all the $^3J_{HN,H\alpha}$ couplings except the ones of the residue following isoAspartate ($^3J^{\#}_{HN,H\alpha}$, yellow bars), all the $^3J_{HN,H\alpha}$ couplings except the ones of the two residues following isoAspartate ($^3J^{\#\#}_{HN,H\alpha}$, brown bars). All the $^3J_{HN,H\alpha}$ scalar couplings have been computed using the DFT set of Karplus parameters.

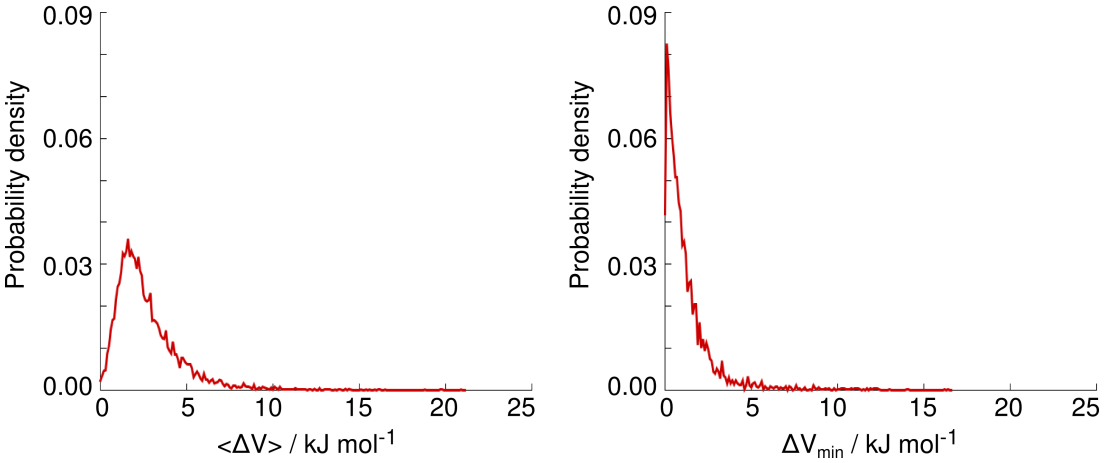


Figure B.8: **Spread in bias potential within bins.** In order to verify how finely the grid follows the free-energy landscape in the CVs subspace, the spread in bias potential (ΔV) between each microstate μ and its 2*N_{dim} neighbor bins has been measured, where N_{dim} indicates the dimensionality of the subspace. Specifically, for each μ the following quantities have been computed: i. the average spread $\langle \Delta V \rangle$, considering all the neighbor bins with defined energy; ii. the minimum spread ΔV_{min}, among the positive ΔV (this was done to avoid redundancies). The distributions of $\langle \Delta V \rangle$ and ΔV_{min}, referred to the simulation of c(CGisoDGRG) with the AMBER ff99sb force field, are shown in the left and right panel, respectively. In approximately 90% of microstates, the average spread is lower than $2k_bT$ (left panel) and μ has at least one neighbor with an energy closer than k_bT (righy panel), indicating that the grid follows the energy landscape pretty finely.

Appendix C

Section 4.2: Supplementary Tables and Figures

Table C.1: **Turns type.** Definition of the α, β and γ turn types, according to the values adopted by the involved dihedral angles. α, β and γ turns are associated to the presence of an intramolecular hydrogen bond between backbone atoms O(i) and HN(i+4), HN(i+3), HN(i+2), respectively. Here are reported only the reference dihedral angles for the turn types individuated in the cyclopeptides under investigation, for a complete list of the turns type see reference[253].

	Dihedral angles [°]					
Turn Type	$\phi(i+1)$	$\psi(i+1)$	$\phi(i+2)$	$\psi(i+2)$	$\phi(i+3)$	$\psi(i+3)$
I-α_{LS}	48	42	67	33	70	32
II-α_{RS}	-59	129	88	-16	-91	-32
II-α_{RU}	54	39	67	-5	-125	-34
βI	-60	-30	-90	0		
βI'	60	30	90	0		
βII'	60	-120	-80	0		
γ	75	-65				

Table C.2: **Chemical Shifts.** Chemical shifts of the carbon atoms for the cyclopeptide c(CGisoGRG) and c(CGisoGRG)-linker conjugate.

Residue ID	Residue Name	Atom	Chemical Shift [ppm] c(CGisoGRG)	Chemical Shift [ppm] c(CGisoGRG)-linker	Difference
1	Cys	C	176.8	175.3	-1.5
1	Cys	Cα	57.7	56.2	-1.5
1	Cys	Cβ	28.3	35.4	7.1
2	Gly	C	173.6	173.6	0.0
2	Gly	Cα	45.9	45.9	0.0
3	isoAsp	C	176.3	176.2	-0.1
3	isoAsp	COO-	179.6	179.3	-0.3
3	isoAsp	Cβ	40.5	40.4	-0.1
3	isoAsp	Cα	54.5	54.6	0.1
4	Gly	C	174.8	174.9	0.1
4	Gly	Cα	45.5	45.5	0.0
5	Arg	C	177.7	177.4	-0.3
5	Arg	Cα	56.8	56.9	0.1
5	Arg	Cβ	30.1	29.9	-0.2
5	Arg	Cδ	43.3	43.3	0.0
5	Arg	Cγ	27.0	27.1	0.1
6	Arg	C	174.8	174.9	0.1
6	Arg	Cα	45.9	45.9	0.0

Table C.3: **RESP charges.** In Table are reported the RESP atomic partial charges for the Linker (LIN) and the Cysteine used for the conjugation to the linker (CYL), derived as described in section 4.2.4.5.

LIN		CYL	
Atom Names	Partial Charges [e]	Atom Names	Partial Charges [e]
C1	-0.4271	N	-0.4157
H1	0.2320	H	0.2719
C2	0.4861	Cα	-0.0289
O2	-0.4913	Hα	0.1215
N3	0.0601	Cβ	-0.0095
C4	0.3980	Hβ*	0.0951
O4	-0.5210	S	-0.1589
C5	-0.0317	C	0.5973
H5*	0.0952	O	-0.5679
C6	-0.1565		
H6*	0.0858		
C7	0.0463		
H7	0.0573		
C8*	-0.0301		
H8*	0.0222		
C9*	-0.0671		
H9*	0.0309		
C10	-0.0334		
H10	0.0144		
C11	0.7356		
O11	-0.5951		
N11	-0.9759		
H11*	0.4111		

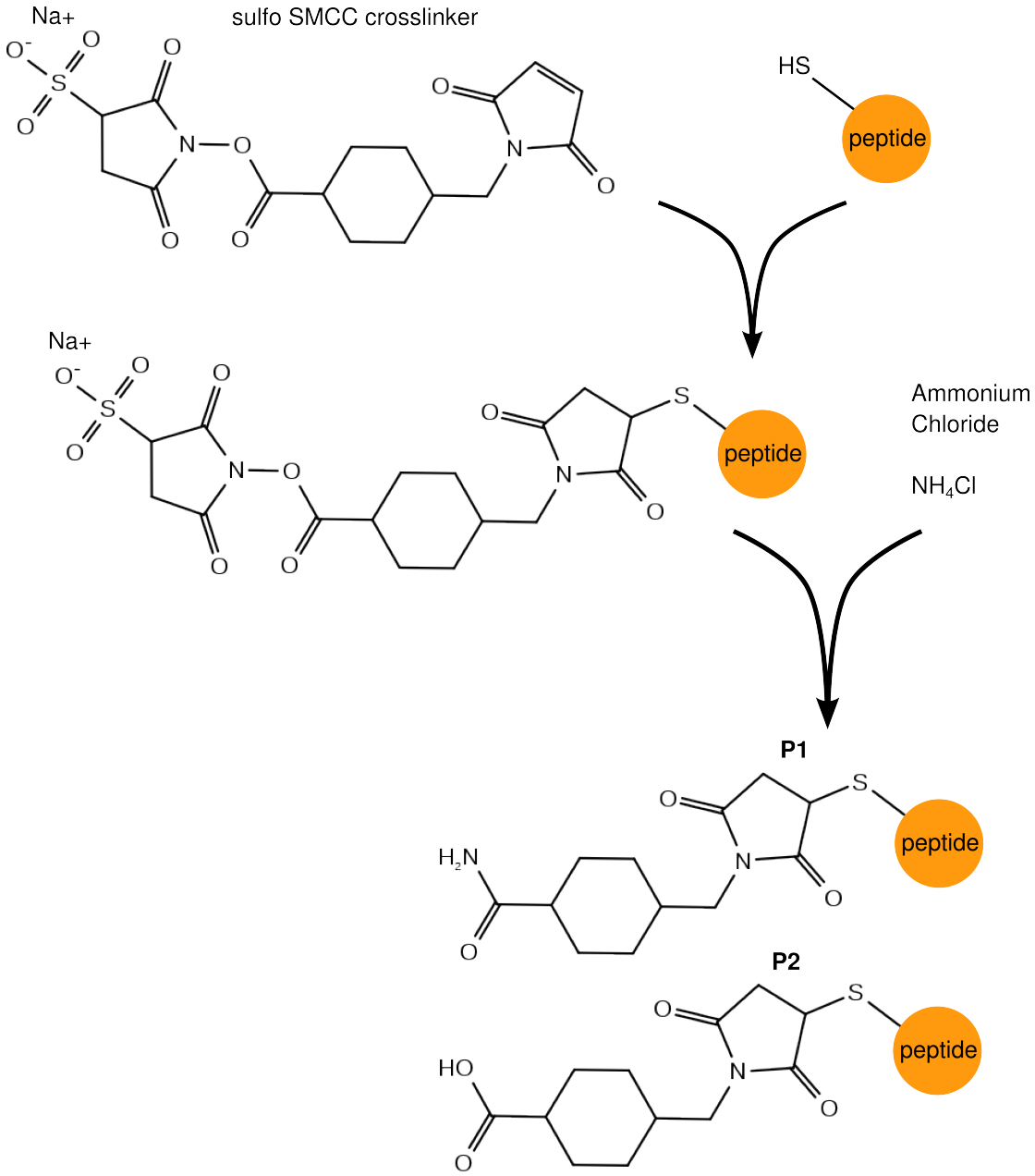
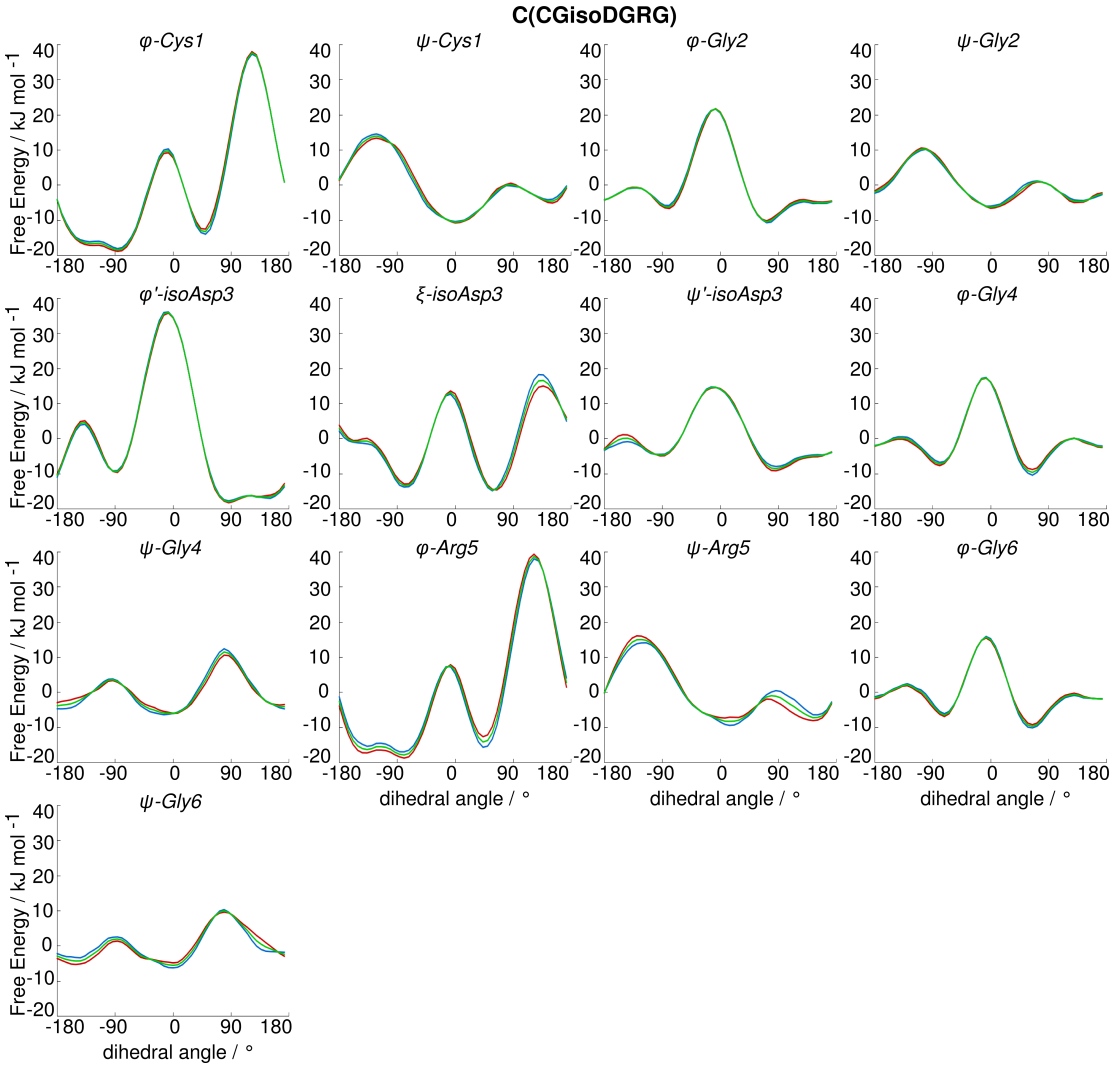
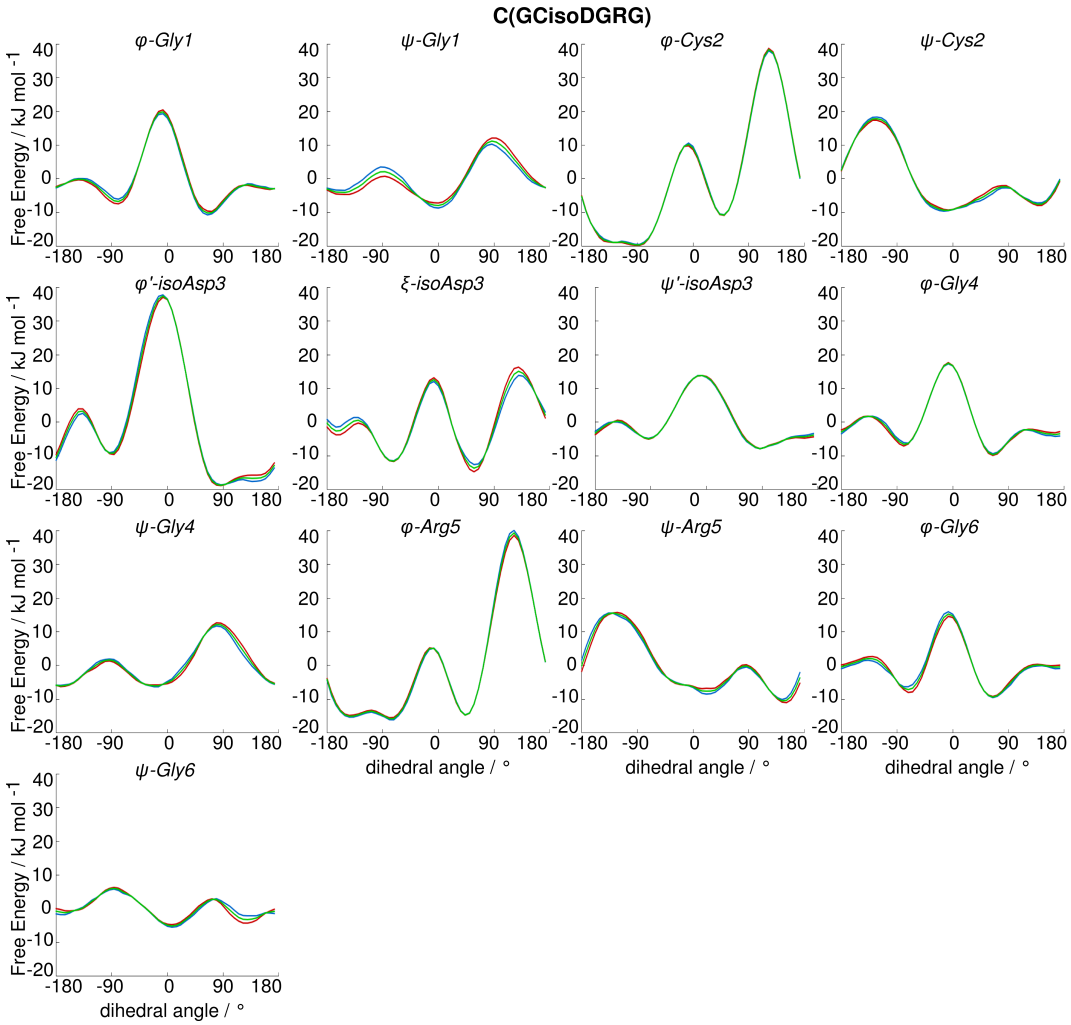
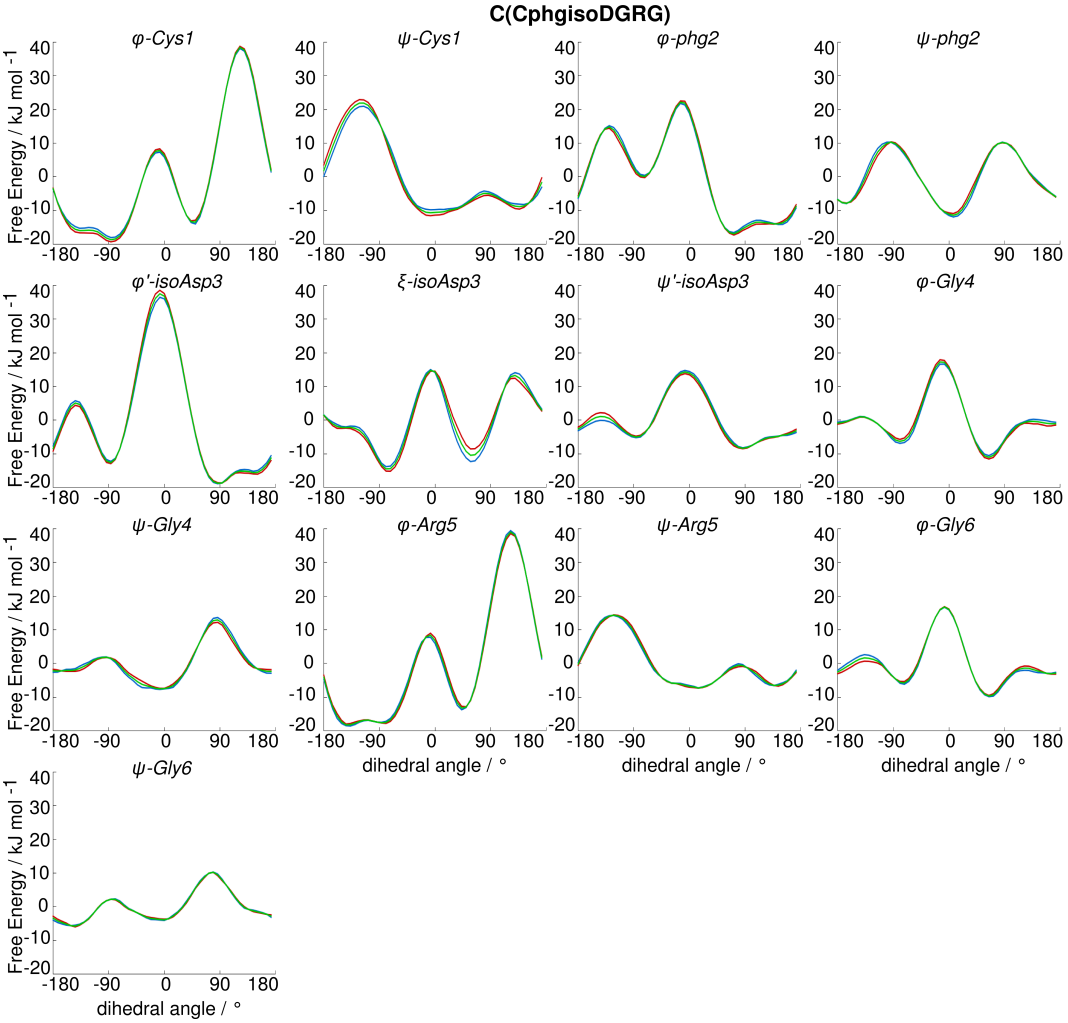


Figure C.1: **Peptide-linker conjugation.** Sulfo SMCC cross-linker and cyclopeptide c(CGisoDGRG) are mixed and then left to react at room temperature for 3 hours, afterward ammonium chloride is added to the solution and left to react for 1 hour at room temperature. The reaction products are the two peptide-linker conjugates, P1 and P2, corresponding to conjugation with an amide or a carboxyl group respectively, likely owing to partial hydrolysis of sulfo-succinimide during the blocking step.[76] The two products are then separated after the reaction. Only peptide-linker conjugate P1 has been investigated within this thesis.

Figure C.2: **Monodimensional free energy profiles.** Convergence of the free energy profiles for the cyclopeptides c(CGisoDGRG), c(GCisoDGRG) and c(CphgisoDGRG) as a function of the biased dihedral angles. The red and blue profiles represent the time averages within $(t_{eq}, (t_{eq} + t_{tot})/2)$ and $((t_{eq} + t_{tot})/2, t_{tot})$, respectively; the green line is the average between the red and blue profiles, which are consistent within $2k_bT$.




C(GCisoDGRG)
φ-Gly1
ψ-Gly1
φ-Cys2
ψ-Cys2
φ'-isoAsp3
ξ-isoAsp3
ψ'-isoAsp3
φ-Gly4
ψ-Gly4
φ-Arg5
ψ-Arg5
φ-Gly6
ψ-Gly6
Free Energy / kJ mol -1
dihedral angle / °
40
30
20
10
0
-10
-20
-180
-90
0
90
180


C(CphgisoDGRG)
φ-Cys1
ψ-Cys1
φ-phg2
ψ-phg2
φ'-isoAsp3
ξ-isoAsp3
ψ'-isoAsp3
φ-Gly4
ψ-Gly4
φ-Arg5
ψ-Arg5
φ-Gly6
ψ-Gly6
Free Energy / kJ mol -1
dihedral angle / °
40
30
20
10
0
-10
-20
-180
-90
0
90
180

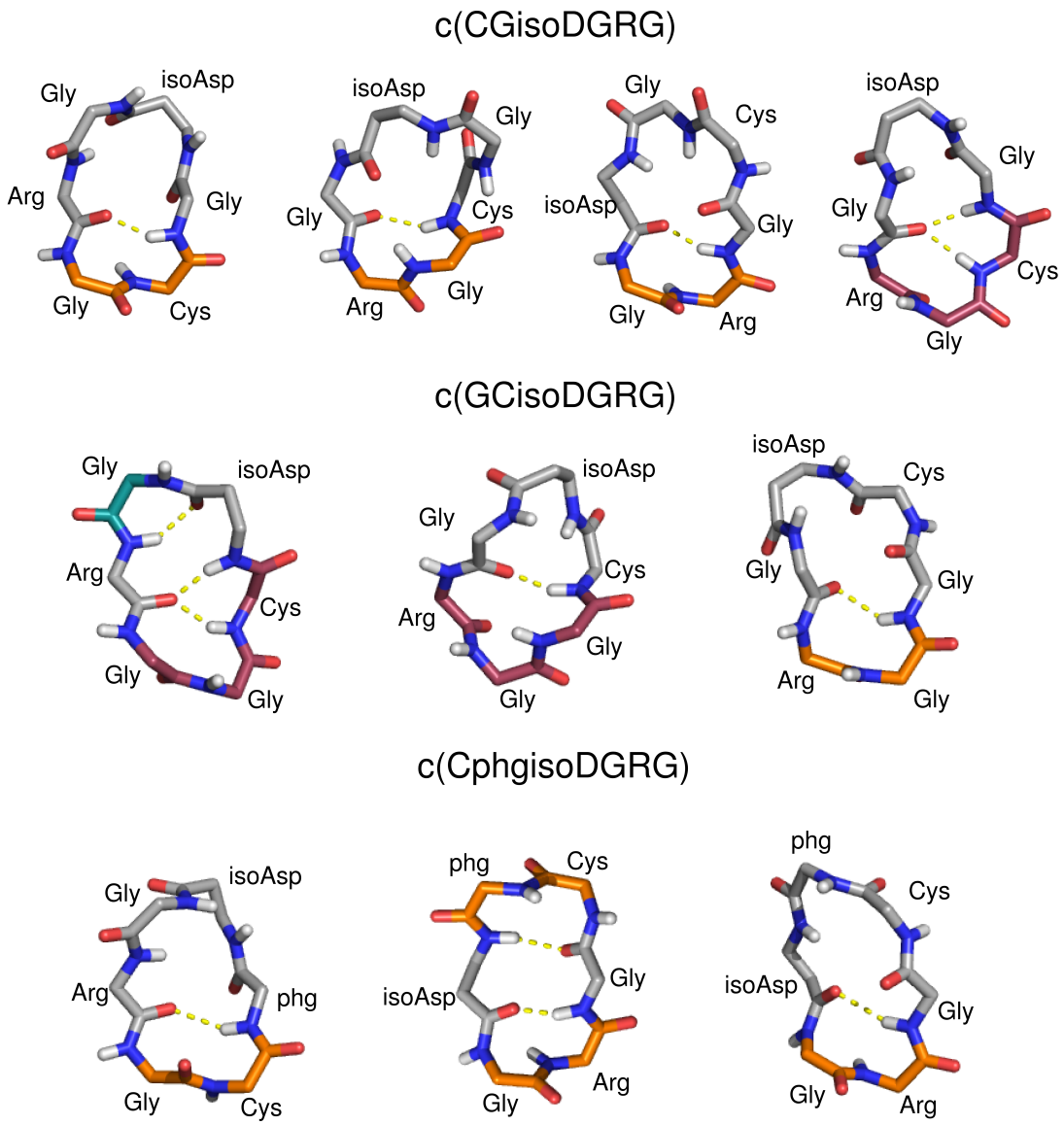
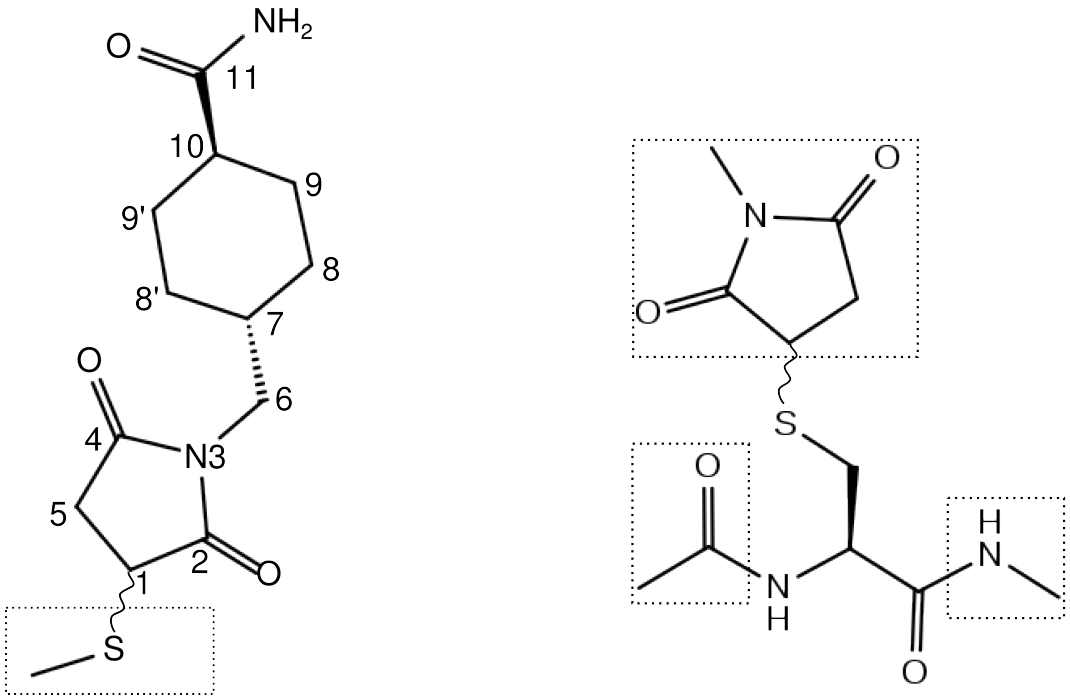


Figure C.3: **Representative conformers of the cyclopeptides c(CGisoDGRG), c(GCisoDGRG) and c(CphgisoDGRG).** For each cyclopeptide a representative backbone conformation of the minima populated more than 10% is reported. Structures are shown in sticks, with carbon, nitrogen, oxygen and polar hydrogen atoms colored in gray, blue, red and white, respectively. The presence of α, β or γ turns is highlighted by coloring the carbon atoms of the involved residues in dark-red, orange and dark-green, respectively. Hydrogen bonds are indicated with yellow dashed lines.



(a) (b)

Figure C.4: **2D-representation of the structures used for the partial charges computation.** The linker LIN capped with a methanethio group and the CYL dipeptide capped with an N-methylsuccinimide group are represented in the left and right panel, respectively. The capping groups, whose total charge has been imposed to be 0, are boxed in dashed lines.

Bibliography

[1] R. O. Hynes. The emergence of integrins: A personal and historical perspective. *Matrix Biol* (2004), 23, 333–340.

[2] J. W. Tamkun, D. W. DeSimone, D. Fonda, R. S. Patel, C. Buck, A. F. Horwitz, R. O. Hynes. Structure of integrin, a glycoprotein involved in the transmembrane linkage between fibronectin and actin. *Cell* (1986), 46, 271–282.

[3] J. S. Desgrosellier, D. A. Cheresh. Integrins in cancer: biological implications and therapeutic opportunities. *Nat Rev Cancer* (2010), 10, 9-22.

[4] I. D. Campbell, M. J. Humphries. Integrin Structure, Activation, and Interactions. *Cold Spring Harb Perspect Biol* (2011), 3, a004994.

[5] G. Niu, X. Chen. Why Integrin as a Primary Target for Imaging and Therapy. *Theranostics* (2011), 1, 30-47.

[6] M. Barczyk, S. Carracedo, D. Gullberg. Integrins. *Cell Tissue Res* (2010), 339, 269-280.

[7] J. D. Humphries, A. Byron, M. J. Humphries. Integrin Ligands. *J Cell Sci* (2006), 119, 3901–3903.

[8] J. P. Xiong, T. Stehle, B. Diefenbach, R. Zhang, R. Dunker, D. L. Scott, A. Joachimiak, S. L. Goodman, M. A. Arnaout. Crystal structure of the extracellular segment of integrin αVβ3. *Science* (2001), 294, 339-345.

[9] J. P. Xiong, T. Stehle, R. Zhang, A. Joachimiak, M. Frech, S. L. Goodman, M. A. Arnaout. Crystal Structure of the Extracellular Segment of Integrin αVβ3 in Complex with an Arg-Gly-Asp Ligand. *Science* (2002), 296, 151–155.

[10] T. Xiao, J. Takagi, B. S. Coller, J-H. Wang, T. A. Springer. Structural basis for allostery in integrins and binding to fibrinogen-mimetic therapeutics. *Nature* (2004), 432, 59-67.

[11] M. A. Arnaout, S. L. Goodman, J. P. Xiong. Structure and mechanics of integrin-based cell adhesion. *Curr Opin Cell Biol* (2007), 19, 495-507.

[12] R. S. Larson, A. L. Corbi, L. Berman, T. Springer. Primary structure of the leukocyte function-associated molecule-1 alpha subunit: an integrin with an embedded domain defining a protein superfamily. *J Cell Biol* (1989), 108, 703-712.

[13] X. Rui, M. Mehrbod, J. F. Van Agthoven, J-P. Xiong, M. R. K. Mofrad, M. A. Arnaout. The α-subunit regulates stability of the metal ion at the ligand-associated metal-ion binding site in β3 integrins. *J Biol Chem* (2014), 289, 23256-23263.

[14] T. A. Springer, J. Zhu, T. Xiao. Structural basis for distinctive recognition of fibrinogen γC peptide by the platelet integrin αIIbβ3. *J Cell Biol* (2008), 182, 791-800.

[15] M. Nagae, S. Re, E. Mihara, T. Nogi, Y. Sugita, J. Takagi. Crystal structure of α5β1 integrin ectodomain: Atomic details of the fibronectin receptor. *J Cell Biol* (2012), 197, 131-140.

[16] W. Xia, T. A. Springer. Metal ion and ligand binding of integrin α5β1. *Proc Nat Acad Sci USA* (2014), 111, 17863-17868.

[17] D. DiCara, C. Rapisarda, J. L. Sutcliffe, S. M. Violette, P. H. Weinreb, I. R. Hart, M. J. Howard, J. F. Marshall. Structure-function analysis of Arg-Gly-Asp helix motifs in αvβ6 integrin ligands. *J Biol Chem* (2007), 282, 9657-9665.

[18] M. Mehrbod, S. Trisno, M. R. Mofrad. On the activation of integrin αIIbβ3: outside-in and inside-out pathways. *Biophys J* (2013), 105, 1304-1315.

[19] D. Bouvard, J. Pouwels, N. De Franceschi, J. Ivaska. Integrin inactivators: balancing cellular functions in vitro and in vivo. *Nat Rev Mol Cell Biol* (2013), 14, 430-442.

[20] M. Millard, S. Odde, N. Neamati. Integrin targeted therapeutics. *Theranostics* (2011), 1, 154-188.

[21] J. P. Xiong, B. Mahalingham, J. L. Alonso, L. A. Borrelli, X. Rui, S. Anand, B. T. Hyman, T. Rysiok, D. Müller-Pompalla, S. L. Goodman, M. A. Arnaout.

Crystal structure of the complete integrin αVβ3 ectodomain plus an α/β transmembrane fragment. *J Cell Biol* (2009), 186, 589-600.

[22] J. Takagi, B. M. Petre, T. Walz, T. A. Springer. Global conformational rearrangements in integrin extracellular domains in outside-in and inside-out signaling. *Cell* (2002), 110, 599-611.

[23] K. R. Legate, S. A. Wickström, R. Fässler. Genetic and cell biological analysis of integrin outside-in signaling *Genes Dev* (2009), 23, 397–418.

[24] K. Holmbäck, M. J. Danton, T. T. Suh, C. C. Daugherty, J. L. Degen. Impaired platelet aggregation and sustained bleeding in mice lacking the fibrinogen motif bound by integrin αIIbβ3. *EMBO J* (1996), 15, 5760–5771.

[25] D. R. Phillips, I. F. Charo, R. M. Scarborough. GPIIb-IIIa: the responsive integrin. *Cell* (1991), 65, 359-362.

[26] S. H. Tam, P. M. Sassoli, R. E. Jordan, M. T. Nakada. Abciximab (ReoPro, chimeric 7E3 Fab) demonstrates equivalent affinity and functional blockade of glycoprotein IIb/IIIa and alpha(v)beta3 integrins. *Circulation* (1998), 98, 1085-1091.

[27] M. S. Dennis, W. J. Henzel, R. M. Pitti, M. T. Lipari, M. A. Napier, T. A. Deisher, S. Bunting, R. A. Lazarus. Platelet glycoprotein IIb-IIIa protein antagonists from snake venoms: evidence for a family of platelet-aggregation inhibitors. *Proc Nat Acad Sci USA* (1990), 87, 2471-2475.

[28] G. D. Hartman, M. S. Egbertson, W. Halczenko, W. L. Laswell, M. E. Duggan, R. L. Smith, A. M. Naylor, P. D. Manno, R. J. Lynch. Non-peptide fibrinogen receptor antagonists. 1. Discovery and design of exosite inhibitors. *J Med Chem* (1992), 35, 4640-4642.

[29] C. J. Avraamides, B. Garmy-Susini, J. A. Varner. Integrins in angiogenesis and lymphangiogenesis. *Nat Rev Cancer* (2008), 8, 604-617.

[30] P. C. Brooks, A. M. P. Montgomery, M. Rosenfeld, R. A. Reisfeld, T. Hu, G. Klier, D. A. Cheresh. Integrin αvβ3 antagonists promote tumor regression by inducing apoptosis of angiogenic blood vessels. *Cell* (1994), 79, 1157-1164.

[31] D. G. Stupack, X. S. Puente, S. Boutsaboualoy, C. M. Storgard, D. A. Cheresh. Apoptosis of adherent cells by recruitment of caspase-8 to unligated integrins. *J Cell Biol* (2001), 155, 459-470.

[32] J. C. Gutheil, T. N. Campbell, P. R. Pierce, J. D. Watkins, W. D. Huse, D. J. Bodkin, D. A. Cheresh. Targeted antiangiogenic therapy for cancer using Vitaxin: a humanized monoclonal antibody to the integrin αvβ3. *Clin Cancer Res* (2000), 6, 3056-3061.

[33] L. B. Nabors, T. Mikkelsen, S. S. Rosenfeld, F. Hochberg, N. S. Akella, J. D. Fisher, G. A. Cloud, Y. Zhang, K. Carson, S. M. Wittemer, A. D. Colevas, S. A. Grossman. Phase I and correlative biology study of cilengitide in patients with recurrent malignant glioma. *J Clin Oncol* (2007), 25, 1651-1657.

[34] R. Stupp, et al. Cilengitide combined with standard treatment for patients with newly diagnosed glioblastoma with methylated MGMT promoter (CENTRIC EORTC 26071-22072 study): a multicentre, randomised, open-label, phase 3 trial. *Lancet Oncol* (2014), 15, 1100-1108.

[35] J. Gong, D. Wang, L. Sun, E. Zborowska, J. K. Willson, M. G. Brattain. Role of α5β1 integrin in determining malignant properties of colon carcinoma cells. *Cell Growth Differ* (1997), 8, 83-90.

[36] S. Kim, K. Bell, S. A. Mousa, J. A. Varner. Regulation of angiogenesis in vivo by ligation of integrin α5bβ1 with the central cell-binding domain of fibronectin. *Am J Pathol* (2000), 156, 1345-1362.

[37] H. Yao, D. M. Veine, Z. Z. Zeng, K. S. Fay, E. D. Staszewski, D. L. Livant. Increased potency of the PHSCN dendrimer as an inhibitor of human prostate cancer cell invasion, extravasation, and lung colony formation. *Clin Exp Metastasis* (2010), 27, 173-184.

[38] F. Schaffner, A. M. Ray, M. Dontenwill. Integrin α5β1, the Fibronectin Receptor, as a Pertinent Therapeutic Target in Solid Tumors. *Cancers* (2013), 5, 27-47.

[39] J. M. Smallheer, C. A. Weigelt, F. J. Woerner, J. S. Wells, W. F. Daneker,

S. A. Mousa, R. R. Wexler, P. K. Jadhav. Synthesis and biological evaluation of nonpeptide integrin antagonists containing spirocyclic scaffolds. *Bioorg Med Chem Lett* (2004), 14, 383-387.

[40] L. Marinelli, A. Meyer, D. Heckmann, A. Lavecchia, E. Novellino, H. Kessler. Ligand binding analysis for human $\alpha5\beta1$ integrin: strategies for designing new $\alpha5\beta1$ integrin antagonists. *J Med Chem* (2005), 48, 4204-4207.

[41] D. Heckmann, A. Meyer, B. Laufer, G. Zahn, R. Stragies, H. Kessler. Rational design of highly active and selective ligands for the $\alpha5\beta1$ integrin receptor. *Chembiochem* (2008), 9, 1397-1407.

[42] E. Martinkova, A. Maglott, D. Y. Leger. D. Bonnet, M. Stiborova, K. Takeda, S. Martin, M. Dontenwill. $\alpha5\beta1$ integrin antagonists reduce chemotherapy-induced premature senescence and facilitate apoptosis in human glioblastoma cells. *Int J Cancer* (2010), 127, 1240-1248.

[43] R. Stragies, F. Osterkamp, G. Zischinsky, D. Vossmeyer, H. Kalkhof, U. Reimer, G. Zahn. Design and synthesis of a new class of selective integrin $\alpha5\beta1$ antagonists. *J Med Chem* (2007), 50, 3786-3794.

[44] B. Delouvrié, K. Al-Kadhimi, J. C. Arnould, S. T. Barry, D. A. Cross, M. Didelot, P. R. Gavine, H. Germain, C. S. Harris, A. M. Hughes, D. A. Jude, J. Kendrew, C. Lambert-van der Brempt, J. J. Lohmann, M. Ménard, A. A. Mortlock, M. Pass, C. Rooney, M. Vautier, J. L. Vincent, N. Warin. Structure-activity relationship of a series of non peptidic RGD integrin antagonists targeting $\alpha5\beta1$: part 1. *Bioorg Med Chem Lett* (2012), 22, 4111-4116.

[45] K. Ley, J. Rivera-Nieves, W. J. Sandborn, S. Shattil. Integrin-based therapeutics: biological basis, clinical use and new drugs *Nat Rev Drug Discovery* (2016), 15, 173-183.

[46] S. L. Goodman, M. Picard. Integrins as therapeutic targets *Trends Pharmacol Sci* (2012), 33, 405-412.

[47] R. O. Hynes. Integrins: bidirectional, allosteric signaling machines *Cell* (2002), 110, 673-687.

[48] D. Cox, M. Brennan, N. Moran. Integrins as therapeutic targets: lessons and opportunities. *Nat Rev Drug Discovery* (2010), 9, 804-820.

[49] K. Fosgerau, T. Hoffmann. Peptide therapeutics: current status and future directions. *Drug Discov Today* (2015), 20, 122-128.

[50] M. Hamzeh-Mivehroud, A. A. Alizadeh, M. B. Morris, W. B. Church, S. Dastmalchi. Phage display as a technology delivering on the promise of peptide drug discovery. *Drug Discov Today* (2013), 18, 1144-1157.

[51] C. A. Lipinski. Drug-like properties and the causes of poor solubility and poor permeability. *J Pharmacol Toxicol Methods* (2000), 44, 235-249.

[52] D. J. Craik, D. P. Fairlie, S. Liras, D. Price. The Future of Peptide-based Drugs. *Chem Biol Drug Des* (2013), 81, 136-147.

[53] F. Albericio, H. G. Kruger. Therapeutic peptides *Future Med Chem* (2012), 4, 1527-1531.

[54] T. Uhliga, T. Kyprianoua, F. G. Martinellia, C. A. Oppicia, D. Heiligersa, D. Hillsa, X. R. Calvoa, P. Verhaerta The emergence of peptides in the pharmaceutical business: From exploration to exploitation. *EuPA Open Proteom* (2014), 4, 58-69.

[55] D. Wu, Y. Gao, Y. Qi, L. Chen, Y. Ma, Y. Li Peptide-based cancer therapy: Opportunity and challenge. *Cancer Lett* (2014), 351, 13-22.

[56] D. S. Nielsen, H. N. Hoang, R-J. Lohman, T. A. Hill, A. J. Lucke, D. J. Craik, D. J. Edmonds, D. A. Griffith, C. J. Rotter, R. B. Ruggeri, D. A. Price, S. Liras, D. P. Fairlie. Improving on Nature: Making a Cyclic Heptapeptide Orally Bioavailable. *Angew Chem, Int Ed* (2014), 5, 12059-12063.

[57] H. Kessler, R. Gratias, G. Hessler, M. Gurrath, G. Muller. Conformation of cyclic peptides. Principle concepts and the design of selectivity and superactivity in bioactive sequences by 'spatial screening'. *Pure Appl Chem* (1996), 68, 1201-1205.

[58] J. Chatterjee, D. Mierke, H. Kessler. N-Methylated Cyclic Pentaalanine Peptides as Template Structures. *J Am Chem Soc* (2006), 128, 15164-15172.

[59] J. Chatterjee, D. Mierke, H. Kessler. Conformational Preference and Potential Templates of N-Methylated Cyclic Pentaalanine Peptides. *Chem- Eur J* (2008), 14, 1508-1517.

[60] J. Chatterjee, C. Gilon, A. Hoffman, H. Kessler. N-Methylation of Peptides: A New Perspective in Medicinal Chemistry. *Acc Chem Res* (2008), 41, 1331-1342.

[61] J. Chatterjee, F. Rechenmacher, H. Kessler. N-Methylation of Peptides and Proteins: An Important Element for Modulating Biological Functions. *Angew Chem, Int Ed* (2013), 52, 254-269.

[62] L. Marinelli, A. Lavecchia, K-E. Gottschalk, E. Novellino, H. Kessler. Docking Studies on αvβ3 Integrin Ligands: Pharmacophore Refinement and Implications for Drug Design. *J Med Chem* (2003), 46, 4393-4404.

[63] S. Cheng, W. S. Craig, D. Mullen, J. F. Tschopp, D. Dixon, M. D. Pierschbacher. Design and Synthesis of Novel Cyclic RGD-ContainingPeptides as Highly Potent and Selective Integrin αIIbβ3 Antagonists. *J Med Chem* (1994), 37, 1-8.

[64] F. Danhier, A. Le Breton, V. Préat. RGD-Based Strategies To Target αvβ3 Integrin in Cancer Therapy and Diagnosis. *Mol Pharmaceutics* (2012), 9, 2961-2973.

[65] K. Temming, R. M. Schiffelers, G. Molema, R. J. Kok. RGD-based strategies for selective delivery of therapeutics and imaging agents to the tumour vasculature. *Drug Resist Update* (2005), 8, 381-402.

[66] C. Mas-Moruno, F. Rechenmacher, H. Kessler. Cilengitide: The First Anti-Angiogenic Small Molecule Drug Candidate. Design, Synthesis and Clinical Evaluation. *Anti-Cancer Agents Med Chem* (2010), 10, 753-768.

[67] H. M. Sheldrake, L. H. Patterson. Strategies to inhibit tumor associated integrin receptors: rationale for dual and multi-antagonists. *J Med Chem* (2014), 57, 6301-6315.

[68] O. L. Chinot. Cilengitide in glioblastoma: when did it fail? *Lancet Oncol* (2014), 15, 1044-1045.

[69] F. Díaz-González, J. Forsyth, B. Steiner, M. H. Ginsberg. Trans-dominant inhibition of integrin function. *Mol Biol Cell* (1996), 7, 1939-1951.

[70] F. Curnis, R. Longhi, L. Crippa, A. Cattaneo, E. Dondossola, A. Bachi, A. Corti. Spontaneous Formation of L-Isoaspartate and Gain of Function in Fibronectin. *J Biol Chem* (2006), 281, 36466-36476.

[71] A. Spitaleri, S. Mari, F. Curnis, C. Traversari, R. Longhi, C. Bordignon, A. Corti, G. P. Rizzardi, G. Musco. Structural Basis for the Interaction of isoDGR with the RGD-binding Site of αvβ3 Integrin. *J Biol Chem* (2008), 283, 19757-19768.

[72] M. Ghitti, A. Spitaleri, B. Valentinis, S. Mari, C. Asperti, C. Traversari, G. P. Rizzardi, G. Musco. Molecular Dynamics Reveal that isoDGR-Containing Cyclopeptides Are True αvβ3 Antagonists Unable To Promote Integrin Allostery and Activation. *Angew Chem, Int Ed* (2012), 51, 7702-7705.

[73] A. Corti, F. Curnis. Isoaspartate-dependent molecular switches for integrin–ligand recognition. *J Cell Sci* (2011), 124, 515-522.

[74] F. Curnis, A. Cattaneo, R. Longhi, A. Sacchi, A. M. Gasparri, F. Pastorino, P. Di Matteo, C. Traversari, A. Bachi, M. Ponzoni, G. P. Rizzardi, A. Corti. Critical Role of Flanking Residues in NGR-to-isoDGR Transition and CD13/Integrin Receptor Switching. *J Biol Chem* (2010), 285, 9114–9123.

[75] A. Spitaleri, M. Ghitti, S. Mari, L. Alberici, C. Traversari, G. P. Rizzardi, G. Musco. Use of Metadynamics in the Design of isoDGR-Based αvβ3 Antagonists To Fine-Tune the Conformational Ensemble. *Angew Chem, Int Ed* (2011), 50, 1832-1836.

[76] F. Curnis, A. Sacchi, R. Longhi, B. Colombo, A. Gasparri, A. Corti. IsoDGR-Tagged Albumin: A New αvβ3 Selective Carrier for Nanodrug Delivery to Tumors. *Small* (2013), 9, 673–678.

[77] A. O. Frank, E. Otto, C. Mas-Moruno, H. B. Schiller, L. Marinelli,

S. Cosconati, A. Bochen, D. Vossmeyer, G. Zahn, R. Stragies, E. Novellino, H. Kessler. Conformational Control of Integrin-Subtype Selectivity in isoDGR Peptide Motifs: A Biological Switch. *Angew Chem, Int Ed* (2010), 49, 9278-9281.

[78] A. Bochen, U. K. Marelli, E. Otto, D. Pallarola, C. Mas-Moruno, F. S. Di Leva, H. Boehm, J. P. Spatz, E. Novellino, H. Kessler, L. Marinelli. Biselectivity of isoDGR Peptides for Fibronectin Binding Integrin Subtypes $\alpha 5\beta 1$ and $\alpha v\beta 6$: Conformational Control through Flanking Amino Acids. *J Med Chem* (2013), 56, 1509-1519.

[79] M. Mingozzi, A. Dal Corso, M. Marchini, I. Guzzetti, M. Civera, U. Piarulli, D. Arosio, L. Belvisi, D. Potenza, L. Pignataro, C. Gennari. Cyclic isoDGR Peptidomimetics as Low-Nanomolar $\alpha v\beta 3$ Integrin Ligands. *Chem- Eur J* (2013), 19, 3563-3567.

[80] J. Meller. Molecular Dynamics. *eLS* (2001), DOI: 10.1038/npg.els.0003048.

[81] J. A. McCammon, B. R. Gelin, M. Karplus. Dynamics of folded proteins. *Nature* (1977), 267, 585–590.

[82] M. Karplus, J. A. McCammon. Molecular dynamics simulations of biomolecules. *Nat Struct Biol* (2002), 9, 646–652.

[83] M. Karplus. Development of Multiscale Models for Complex Chemical Systems: From H+H2 to Biomolecules (Nobel Lecture). *Angew Chem, Int Ed* (2014), 53, 9992–10005.

[84] S. A. Adcock, J. A. McCammon. Molecular dynamics: survey of methods for simulating the activity of proteins. *Chem Rev* (2006), 106, 1589–1615.

[85] M. Ghitti, G. Musco, A. Spitaleri. NMR and computational methods in the structural and dynamic characterization of ligand-receptor interactions. *Adv Exp Med Biol* (2014), 805, 271-304.

[86] M. J. Abraham, D. van der Spoel, E. Lindahl, B. Hess, and the GROMACS development team. GROMACS User Manual version 5.0.5. *www.gromacs.org* (2015).

[87] S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess, E. Lindahl. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit *Bioinformatics* (2013), 29, 845-854.

[88] R. Salomon-Ferrer, D. A. Case, R. C. Walker. An overview of the Amber biomolecular simulation package. *WIREs Comput Mol Sci* (2013), 3, 198–210.

[89] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, K. Schulten. Scalable molecular dynamics with NAMD. *J Comput Chem* (2005), 26, 1781–1802.

[90] L. Verlet. Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys Rev* (1967), 159, 98–103.

[91] W. C. Swope, H. C. Andersen, P. H. Berens, K. R. Wilson. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *J Chem Phys* (1982), 76, 637–649.

[92] R. W. Hockney. The potential calculation and some applications. *Methods Comput Phys* (1970), 9, 136–211.

[93] C. W. Gear. Numerical Initial Value Problems in Ordinary Differential Equations. *Prentice Hall PTR* (1971).

[94] W. F. van Gunsteren, H. J. C. Berendsen. Algorithms for macromolecular dynamics and constraint dynamics. *Mol Phys* (1977), 34, 1311–1327.

[95] B. Hess, H. Bekker, H. J. C. Berendsen, J. G. E. M. Fraaije. LINCS: A linear constraint solver for molecular simulations. *J Comput Chem* (1997), 18, 1463–1472.

[96] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak. Molecular dynamics with coupling to an external bath. *J Chem Phys* (1984), 81, 3684–3690.

[97] S. Nosé. A unified formulation of the constant temperature molecular-dynamics methods. *J Chem Phys* (1984), 81, 511–519.

[98] G. Bussi, D. Donadio, M. Parrinello. Canonical sampling through velocity-rescaling. *J Chem Phys* (2007), 126, 014101.

[99] G. S. Grest, K. Kremer. Molecular-dynamics simulation for polymers in the presence of a heat bath. *Phys Rev A* (1986), 33, 3628–3631.

[100] M. Parrinello, A. Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *J Appl Phys* (1981), 52, 7182–7190.

[101] V. Hornak, R. Abel, C. Simmerling. Comparison of multiple AMBER force fields and development of improved protein backbone parameters. *Proteins: Struct, Funct, Bioinf* (2006), 65, 712–725.

[102] W. L. Jorgensen, D. S. Maxwell, J. Tirado-Rives. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J Am Chem Soc* (1996), 118, 11225–11236.

[103] A. D. MacKerell, D. Bashford, M. Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* (1998), 102, 3586–3616.

[104] C. Oostenbrink, A. Villa, W. F. V. Gunsteren. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53a5 and 53a6. *J Comput Chem* (2004), 25, 1656–1676.

[105] M. J. Robertson, J. Tirado-Rives, W. L. Jorgensen. Improved Peptide and Protein Torsional Energetics with the OPLS-AA Force Field. *J Chem Theory Comput* (2015), 11, 3499–3509.

[106] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, C. Simmerling. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J Chem Theory Comput* (2015), 11, 3696–3713.

[107] F. Jiang, C.-Y. Zhou, Y.-D. Wu. Residue-Specific Force Field Based on the Protein Coil Library. RSFF1: Modification of OPLS-AA/L. *J Phys Chem B* (2014), 118, 6983–6998.

[108] C.-Y. Zhou, F. Jiang, Y.-D. Wu. Residue-Specific Force Field Based on

Protein Coil Library. RSFF2: Modification of AMBER ff99SB. *J Phys Chem B* (2015), 119, 1035–1047.

[109] K. Lindorff-Larsen, P. Maragakis, S. Piana, M. P. Eastwood, R. O. Dror, D. E. Shaw. Systematic Validation of Protein Force Fields against Experimental Data. *PLoS ONE* (2012), 7, e32131.

[110] K. A. Beauchamp, Y.-S. Lin, R. Das, V. S. Pande. Are Protein Force Fields Getting Better? A Systematic Benchmark on 524 Diverse NMR Measurements. *J Chem Theory Comput* (2012), 8, 1409–1414.

[111] F. Martín-García, E. Papaleo, P. Gomez-Puertas, W. Boomsma, K. Lindorff-Larsen. Comparing Molecular Dynamics Force Fields in the Essential Subspace. *PLoS ONE* (2015), 10, e0121114.

[112] S. Li, A. H. Elcock. Residue-Specific Force Field (RSFF2) Improves the Modeling of Conformational Behavior of Peptides and Proteins. *J Phys Chem Lett* (2015), 6, 2127–2133.

[113] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein. Comparison of simple potential functions for simulating liquid water. *J Chem Phys* (1983), 79, 926–935.

[114] M. W. Mahoney, W. L. Jorgensen. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *J Chem Phys* (2000), 112, 8910–8922.

[115] H. J. C. Berendsen, J. P. M. Postma, W. F. V. Gunsteren, J. Hermans. Interaction models for water in relation to protein hydration. In *Intermolecular Forces, ed. B. Pullman* (1981), 331-342.

[116] T. A. Halgren. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J Comput Chem* (1996), 17, 490–519.

[117] M. Clark, R. D. Cramer, N. Van Opdenbosch. Validation of the general purpose tripos 5.2 force field. *J Comput Chem* (1989), 10, 982–1012.

[118] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, A. D. MacKerell Jr. CHARMM

General Force Field: A Force field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Field. *J Comput Chem* (2010), 31, 671–690.

[119] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case. Development and Testing of a General Amber Force Field. *J Comput Chem* (2004), 25, 1157–1174.

[120] J. Wang, W. Wang, P. A. Kollman, D. A. Case. Antechamber: an accessory software package for molecular mechanical calculations *J Am Chem Soc* (2001), 222, 403.

[121] P. J. Steinbach, B. R. Brooks. New spherical-cutoff methods for long-range forces in macromolecular simulation. *J Comput Chem* (1994), 15, 667–683.

[122] T. Darden, D. York, L. Pedersen. Particle mesh Ewald: An $N\,log(N)$ method for Ewald sums in large systems. *J Chem Phys* (1993), 98, 10089-10092.

[123] M. O. Steinhauser, S. Hiermaier. A review of computational methods in materials science: examples from shock-wave and polymer physics. *Int J Mol Sci* (2009), 10, 5135-5216.

[124] K. Teilum, J. G. Olsen, B. B. Kragelund. Functional aspects of protein flexibility. *Cell Mol Life Sci* (2009), 66, 2231–2247.

[125] H. S. Chan, K. A. Dill. Protein folding in the landscape perspective: Chevron plots and non-arrhenius kinetics *Proteins: Struct, Funct, Bioinf* (1998), 30, 2-33.

[126] P. Cossio, F. Marinelli, A. Laio, F. Pietrucci. Folding a 48-Residue LysM Domain Using a Coarse-Grained Model *J Phys Chem B* (2010), 114, 3259–3265.

[127] V. Spiwok, Z. Sucur, P. Hosek. Enhanced sampling techniques in biomolecular simulations. *Biotechnol Adv* (2015), 33, 1130–1140.

[128] S. J. Marrink, D. P. Tieleman. Perspective on the Martini model. *Chem Soc Rev* (2013), 42, 6801–6822.

[129] S. Kirkpatrick, C. D. Gelatt, M. P. Vecchi. Optimization by simulated annealing. *Science* (1983), 220, 671–680.

[130] Y. Sugita, Y. Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chem Phys Lett* (1999), 314, 141–151.

[131] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, P. A. Kollman. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J Comput Chem* (1992), 13, 1011–1021.

[132] A. Laio, M. Parrinello. Escaping free-energy minima. *Proc Nat Acad Sci USA* (2002), 99, 12562–12566.

[133] A. Laio, F. L. Gervasio. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Rep Prog Phys* (2008), 71, 126601–126622.

[134] S. Piana, A. Laio. A Bias-Exchange Approach to Protein Folding. *J Phys Chem B* (2007), 111, 4553–4559.

[135] E. Zurek, W. Grochala. Predicting crystal structures and properties of matter under extreme conditions via quantum mechanics: the pressure is on. *Phys Chem Chem Phys* (2015), 17, 2917–2934.

[136] A. Barducci, G. Bussi, M. Parrinello. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys Rev Lett* (2008), 100, 020603.

[137] A. Laio, A. Rodriguez-Fortea, F. L. Gervasio, M. Ceccarelli, M. Parrinello. Assessing the Accuracy of Metadynamics. *J Phys Chem B* (2005), 109, 6714–6721.

[138] F. Marinelli, F. Pietrucci, A. Laio, S. Piana. A Kinetic Model of Trp-Cage Folding from Multiple Biased Molecular Dynamics Simulations. *PLoS Comput Biol* (2009), 5, e1000452.

[139] D. Granata, C. Camilloni, M. Vendruscolo, A. Laio. Characterization of the free-energy landscapes of proteins by NMR-guided metadynamics. *Proc Nat Acad Sci USA* (2013), 110, 6817–6822.

[140] F. Noé, I. Horenko, C. Schütte, J. C. Smith. Hierarchical analysis of conformational dynamics in biomolecules: transition networks of metastable states. *J Chem Phys* (2007), 126, 155102.

[141] X. Biarnés, F. Pietrucci, F. Marinelli, A. Laio. METAGUI. A VMD interface for analyzing metadynamics and molecular dynamics simulations. *Comput Phys Commun* (2012), 183, 203–211.

[142] M. Xu, M. A. Lill. Induced fit docking, and the use of QM/MM methods in docking. *Drug Discov Today Technol* (2013), 10, e411-e418.

[143] S. Agarwal, R. Mehrotra. An overview of Molecular Docking. *JSM Chem* (2016), 4, 1024.

[144] S. Z. Grinter, X. Zou. Challenges, Applications, and Recent Advances of Protein-Ligand Docking in Structure-Based Drug Design. *Molecules* (2014), 19, 10150–10176.

[145] I. A. Guedes, C. S. de Magalhães, L. E. Dardenne. Receptor–ligand molecular docking. *Biophys Rev* (2014), 6, 75–87.

[146] J. Audie, J. Swanson. Recent work in the development and application of protein-peptide docking. *Future Med Chem* (2012), 4, 1619–1644.

[147] A. Grosdidier, V. Zoete, O. Michielin. SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic Acids Res* (2011), 39, W270–W277.

[148] D. Röthlisberger, O. Khersonsky, A. M. Wollacott, L. Jiang, J. DeChancie, J. Betker, J. L. Gallaher, E. A. Althoff, A. Zanghellini, O. Dym, S. Albeck, K. N. Houk, D. S. Tawfik, D. Baker. Kemp elimination catalysts by computational enzyme design. *Nature* (2008), 453, 190–195.

[149] J. Cai, C. Han, T. Hu, J. Zhang, D. Wu, F. Wang, Y. Liu, J. Ding, K. Chen, J. Yue, X. Shen, H. Jiang. Peptide deformylase is a potential target for anti-Helicobacter pylori drugs: Reverse docking, enzymatic assay, and X-ray crystallography validation. *Protein Sci* (2006), 15, 2071–2081.

[150] R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic,

D. T. Mainz, M. P. Repasky, E. H. Knoll, D. E. Shaw, M. Shelley, J. K. Perry, P. Francis, P. S. Shenkin. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J Med Chem* (2004), 47, 1739–1749.

[151] T. A. Halgren, R. B. Murphy, R. A. Friesner, H. S. Beard, L. L. Frye, W. T. Pollard, J. L. Banks. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J Med Chem* (2004), 47, 1750–1759.

[152] R. A. Friesner, R. B. Murphy, M. P. Repasky, L. L. Frye, J. R. Greenwood, T. A. Halgren, P. C. Sanschagrin, D. T. Mainz. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes. *J Med Chem* (2006), 49, 6177–6196.

[153] C. Dominguez, R. Boelens, A. M. Bonvin. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc* (2003), 125, 1731–1737.

[154] M. Trellet, A. S. J. Melquiond, A. M. Bonvin. Information-driven modeling of protein-peptide complexes. *Methods Mol Biol* (2015), 1268, 221-239.

[155] E. Deplazes, J. Davies, A. M. Bonvin, G. F. King, A. E. Mark. On the Combination of Ambiguous and Unambiguous Data in the Restraint-driven Docking of Flexible Peptides with HADDOCK: The Binding of the Spider Toxin PcTx1 to the Acid Sensing Ion channel (ASIC)1a. *J Chem Inf Model* (2016), 56, 127–138.

[156] S. Vajda, D. R. Hall, D. Kozakov. Sampling and scoring: A marriage made in heaven. *Proteins: Struct, Funct, Bioinf* (2013), 81, 1874–1884.

[157] X. Y. Meng, H. X. Zhang, M. Mezei, M. Cui. Molecular docking: a powerful approach for structure-based drug discovery. *Curr Comput-Aided Drug Des* (2011), 7, 146–157.

[158] L. G. Ferreira, R. N. Dos Santos, G. Oliva, A. D. Andricopulo. Molecular

docking and structure-based drug design strategies. *Molecules* (2015), 20, 13384–13421.

[159] J. -P. Changeux, S. Edelstein. Conformational selection or induced fit? 50 years of debate resolved. *F1000 Biol Rep* (2011), 3, 19.

[160] M. L. Teodoro, L. E. Kavraki. Conformational flexibility models for the receptor in structure based drug design. *Curr Pharm Des* (2003), 9, 1635–1648.

[161] E. Yuriev, J. Holien, P. A. Ramsland. Improvements, trends, and new ideas in molecular docking: 2012-2013 in review. *J Mol Recognit* (2015), 28, 581–604.

[162] A. N. Jain. Scoring functions for protein-ligand docking. *Curr Protein Pept Sci* (2006), 7, 407–420.

[163] S. -Y. Huang, S. Z. Grintera, X. Zou. Scoring functions and their evaluation methods for protein–ligand docking: recent advances and future directions. *Phys Chem Chem Phys* (2010), 12, 12899–12908.

[164] M. Feher. Consensus scoring for protein-ligand interactions. *Drug Discov Today* (2006), 11, 421–428.

[165] P. L. Kastritis, A. M. Bonvin. Molecular origins of binding affinity: seeking the Archimedean point. *Curr Opin Struct Biol* (2013), 23, 868–877.

[166] A. D. J. van Dijk, A. M. Bonvin. Solvated docking: introducing water into the modelling of biomolecular complexes *Bioinformatics* (2006), 22, 2340–2347.

[167] M. Rarey, B. Kramer, T. Lengauer. The particle concept: placing discrete water molecules during protein-ligand docking predictions. *Proteins: Struct, Funct, Bioinf* (1999), 34, 17–28.

[168] M. L. Verdonk, G. Chessari, J. C. Cole, M. J. Hartshorn, C. W. Murray, J. W. M. Nissink, R. D. Taylor, R. Taylor. Modeling Water Molecules in Protein-Ligand Docking Using GOLD *J Med Chem* (2005), 48, 6504–6515.

[169] S. Forli, A. J. Olson. A Force Field with Discrete Displaceable Waters and Desolvation Entropy for Hydrated Ligand Docking. *J Med Chem* (2012), 55, 623–638.

[170] S. Y. Huang, X. Zou. Advances and challenges in protein-ligand docking. *Int J Mol Sci* (2010), 11, 3016–3034.

[171] D. C. Thompson, C. Humblet, D. Joseph-McCarthy. Investigation of MM-PBSA rescoring of docking poses. *J Chem Inf Model* (2008), 48, 1081–1091.

[172] H.-Y. Liu, X. Zou. Electrostatics of ligand binding: Parametrization of the generalized born model and comparison with the Poisson-Boltzmann approach. *J Phys Chem B* (2006), 110, 9304-9313.

[173] S. Huo, I. Massova, P. A. Kollman. Computational alanine scanning of the 1:1 human growth hormone-receptor complex. *J Comput Chem* (2002), 23, 15–27.

[174] R. T. Bradshaw, B. H. Patel, E. W. Tate, R. J. Leatherbarrow, I. R. Gould. Comparing experimental and computational alanine scanning techniques for probing a prototypical protein-protein interaction. *Protein Eng, Des Sel* (2011), 24, 197–207.

[175] M. D. Eldridge, C. W. Murray, T. R. Auton, G. V. Paolini, R. P. Mee. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J Comput-Aided Mol Des* (1997), 11, 425–445.

[176] M. H. Bohari, G. N. Sastry. FDA approved drugs complexed to their targets: evaluating pose prediction accuracy of docking protocols. *J Mol Model* (2012), 18, 4263-4274.

[177] T. Fraczek, A. Siwek, P. Paneth. Assessing molecular docking tools for relative biological activity prediction: a case study of triazole HIV-1 NNRTIs. *J Chem Inf Model* (2013), 53, 3326–3342.

[178] M. P. Jacobson, D. L. Pincus, C. S. Rapp, T. J. F. Day, B. Honig, D. E. Shaw, R. A. Friesner. A Hierarchical Approach to All-Atom Protein Loop Prediction. *Proteins: Struct, Funct, Bioinf* (2004), 55, 351–367.

[179] M. P. Jacobson, R. A. Friesner, Z. Xiang, B. Honig. On the Role of Crystal

Packing Forces in Determining Protein Sidechain Conformations. *J Mol Biol* (2002), 320, 597–608.

[180] J. Du, H. Sun, L. Xi,J. Li, Y. Yang, H. Liu, X. Yao. Molecular modeling study of Checkpoint Kinase 1 inhibitors by multiple docking strategies and Prime/MMGBSA. *J Comput Chem* (2011), 32, 2800–2808.

[181] I. Tubert-Brohman, W. Sherman, M. Repasky, T. Beuming. Improved Docking of Polypeptides with Glide *J Chem Inf Model* (2013), 53, 1689–1699.

[182] S. J. de Vries, M. van Dijk, A. M. Bonvin. The HADDOCK web server for data-driven biomolecular docking. *Nat Protoc* (2010), 5, 833–897.

[183] S. J. de Vries, A. D. van Dijk, M. Krzeminski, M. van Dijk, A. Thureau, V. Hsu, T. Wassenaar, A. M. Bonvin. HADDOCK versus HADDOCK: new features and performance of HADDOCK2.0 on the CAPRI targets. *Proteins: Struct, Funct, Bioinf* (2007), 69, 726–733.

[184] J. P. Rodrigues, E. Karaca, A. M. Bonvin. Information-driven structural modelling of protein-protein interactions. *Methods Mol Biol* (2015), 1215, 399-424.

[185] E. Karaca, A. M. Bonvin. Advances in integrated modelling of biomolecular complexes. *Methods* (2013), 59, 372–381.

[186] *WeNMR*, 18 Mar 2011, Web 7 Oct 2016 <https://www.wenmr.eu/wenmr/generating-necessary-restraint-files-running-haddock-manually>

[187] J. Fernández-Recio, M. Totrov M, R. Abagyan. Identification of protein-protein interaction sites from docking energy landscapes. *J Mol Biol* (2004), 335, 843–865.

[188] E. Karaca, A. M. Bonvin A multi-domain flexible docking approach to deal with large conformational changes in the modeling of biomolecular complexes. *Structure* (2011), 19, 555–565.

[189] A. Vangone, A. M. Bonvin. Contacts-based prediction of binding affinity in protein-protein complexes. *eLife* (2015), 4, e07454.

[190] J. P. Rodrigues, M. Trellet, C. Schmitz, P. L. Kastritis, E. Karaca, A. S. J. Melquiond, A. M. Bonvin. Clustering biomolecular complexes by residue contacts similarity. *Proteins: Struct, Funct, Bioinf* (2012), 80, 1810–1817.

[191] D. Spiliotopoulos, P. L. Kastritis, A. S. J. Melquiond, A. M. Bonvin, G. Musco, W. Rocchia, A. Spitaleri. dMM-PBSA: a new HADDOCK scoring function for protein-peptide docking. *Front Mol Biosci* (2016), 3, 46.

[192] P. L. Kastritis, A. D. J. van Dijk, A. M. Bonvin. Explicit Treatment of Water Molecules in Data-Driven Protein-Protein Docking: The Solvated HADDOCKing Approach. *Methods Mol Biol* (2012), 819, 355-374.

[193] I. S. Moreira, P. A. Fernandes, M. J. Ramos. Protein–protein docking dealing with the unknown. *J Comput Chem* (2010), 31, 317–342.

[194] M. F. Lensink et al. Prediction of homo- and hetero-protein complexes by ab-initio and template-based docking: a CASP-CAPRI experiment. *Proteins: Struct, Funct, Bioinf* (2016), 84, 323–348.

[195] M. F. Lensink et al. Blind Prediction of Interfacial Water Positions in CAPRI. *Proteins: Struct, Funct, Bioinf* (2014), 82, 620–632.

[196] D. Shortle, K. T. Simons, D. Baker. Clustering of low-energy conformations near the native structures of small proteins. *Proc Nat Acad Sci USA* (1998), 95, 11158–11162.

[197] D. Kozakov, K. H. Clodfelter, S. Vajda, C. J. Camacho. Optimal Clustering for Detecting Near-Native Conformations in Protein Docking. *Biophys J* (2005), 89, 867–875.

[198] X. Daura, K. Gademann, B. Jaun, D. Seebach, W. F. van Gunsteren, A. E. Mark. Peptide Folding: When Simulation Meets Experiment. *Angew Chem, Int Ed* (1999), 38, 236–240.

[199] H. Yu, Y. Lin. Toward structure prediction of cyclic peptides *Phys Chem Chem Phys* (2015), 17, 4210–4219.

[200] R. Kasher, D. A. Oren, Y. Barda, C. Gilon. Miniaturized proteins: the backbone cyclic proteinomimetic approach. *J Mol Biol* (1999), 292, 421-429.

[201] T. Weide, A. Modlinger, H. Kessler. Spatial Screening for the Identification of the Bioactive Conformation of Integrin Ligands. *Top Curr Chem* (2007), 272, 1-50.

[202] A. E. Wakefield, W. M. Wuest, V. A. Voelz. Molecular Simulation of Conformational Pre-Organization in Cyclic RGD Peptides. *J Chem Inf Model* (2015), 55, 806-813.

[203] L. Belvisi, T. Riccioni, M. Marcellini, L. Vesci, I. Chiarucci, D. Efrati, D. Potenza, C. Scolastico, L. Manzoni, K. Lombardo, M. A. Stasi, A. Orlandi, A. Ciucci, B. Nico, D. Ribatti, G. Giannini, M. Presta, P. Carminati, C. Pisano. Biological and molecular properties of a new $\alpha v\beta 3/\alpha v\beta 5$ integrin antagonist. *Mol Cancer Ther* (2005), 4, 1670-1680.

[204] T. R. White, C. M. Renzelman, A. C. Rand, T. Rezai, C. M. McEwen, V. M. Gelev, R. A. Turner, R. G. Linington, S. S. Leung, A. S. Kalgutkar, J. N. Bauman, Y. Zhang, S. Liras, D. A. Price, A. M. Mathiowetz, M. P. Jacobson, R. S. Lokey. On-resin N-methylation of cyclic peptides for discovery of orally bioavailable scaffolds. *Nat Chem Biol* (2011), 7, 810-817.

[205] J. Chatterjee, O. Ovadia, G. Zahn, L. Marinelli, A. Hoffman, C. Gilon, H. Kessler. Multiple N-Methylation by a Designed Approach Enhances Receptor Selectivity. *J Med Chem* (2007), 50, 5878-5881.

[206] C. Mas-Moruno, J. G. Beck, L. Doedens, A. O. Frank, L. Marinelli, S. Cosconati, E. Novellino, H. Kessler. Increasing $\alpha v\beta 3$ Selectivity of the Anti-Angiogenic Drug Cilengitide by N-Methylation. *Angew Chem, Int Ed* (2011), 50, 9496-9500.

[207] S. E. Allen, N. V. Dokholyan, A. A. Bowers. Dynamic Docking of Conformationally Constrained Macrocycles: Methods and Applications. *ACS Chem Biol* (2016), 11, 10-24.

[208] S. M. McHugh, J. R. Rogers, H. Yu, Y.-S. Lin. Insights into How Cyclic

Peptides Switch Conformations. *J Chem Theory Comput* (2016), 12, 2480-2488.

[209] K. W. Lexa, H. A. Carlson. Protein flexibility in docking and surface mapping. *Q Rev Biophys* (2012), 45, 301-343.

[210] M. Rubinstein, M. Y. Niv. Peptidic modulators of protein-protein interactions: Progress and challenges in computational design. *Biopolymers* (2009), 91, 505-513.

[211] M. Feher, C. I. Williams. Effect of Input Differences on the Results of Docking Calculations. *J Chem Inf Model* (2009), 49, 1704-1714.

[212] F. Vasile, M. Civera, L. Belvisi, D. Potenza, G. Tiana. Thermodynamically–Weighted Conformational Ensemble of Cyclic RGD Peptidomimetics from NOE Data. *J Phys Chem B* (2016), 120, 7098-7107.

[213] C. Paissoni, M. Ghitti, L. Belvisi, A. Spitaleri, G. Musco. Metadynamics Simulations Rationalise the Conformational Effects Induced by N-Methylation of RGD Cyclic Hexapeptides. *Chem- Eur J* (2015), 21, 14165-14170.

[214] S. Md Abdur Rauf, P. I. Arvidsson, F. Albericio, T. Govender, G. E. Maguire, H. G. Kruger, B. Honarparvar. The effect of N-methylation of amino acids (Ac-X-OMe) on solubility and conformation: a DFT study. *Org Biomol Chem* (2015), 13, 9993-10006.

[215] W. Humphrey, A. Dalke, K. Schulten. VMD: visual molecular dynamics. *J Mol Graphics* (1996), 14, 33-38.

[216] R. B. Best, G. Hummer. Optimized Molecular Dynamics Force Fields Applied to the Helix-Coil Transition of Polypeptides. *J Phys Chem B* (2009), 113, 9004-9015.

[217] A. Barducci, M. Bonomi, M. Parrinello. Linking Well-Tempered Metadynamics Simulations with Experiments *Biophys J* (2010), 98, L44-L46.

[218] T. Hou, J. Wang, Y. Li, W. Wang. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy

calculations based on molecular dynamics simulations. *J Chem Inf Model* (2011), 51, 69-82.

[219] P. A. Greenidge, C. Kramer, J.-C. Mozziconacci, W. Sherman. Improving docking results via reranking of ensembles of ligand poses in multiple X-ray protein conformations with MM-GBSA. *J Chem Inf Model* (2014), 54, 2697-2717.

[220] *Pymol molecular visualization system*, 5 Oct 2016, Web 27 Oct 2016 <http://www.pymol.org/>

[221] M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci, R. A. Broglia, M. Parrinello. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comput Phys Commun* (2009), 180, 1961-1972.

[222] G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, W. L. Jorgensen. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *J Phys Chem B* (2001), 105, 6474-6487.

[223] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen. A smooth particle mesh Ewald method. *J Chem Phys* (1995), 103, 8577-8593.

[224] M. Karplus. Vicinal Proton Coupling in Nuclear Magnetic Resonance. *J Am Chem Soc* (1963), 85, 2870-2871.

[225] J. M. Schmidt. Conformational Equilibria in Polypeptides. II. Dihedral-Angle Distribution in Antamanide Based on Three-Bond Coupling Information. *J Magn Reson* (1997), 124, 310-322.

[226] G. M. Sastry, M. Adzhigirey, T. Day, R. Annabhimoju, W. Sherman. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J Comput-Aided Mol Des* (2013), 27, 221-234.

[227] Z. Yu, M. P. Jacobson, R. A. Friesner. What role do surfaces play in GB models? A new-generation of surface-generalized born model based on a novel gaussian surface for biomolecules. *J Comput Chem* (2006), 27, 72-89.

[228] M. Wu, Å. Strid, L. A. Eriksson. Development of non-standard arginine residue parameters for use with the AMBER force fields. *Chem Phys Lett* (2013), 584, 188-194.

[229] H. Geng, F. Jiang, Y. D. Wu. Accurate Structure Prediction and Conformational Analysis of Cyclic Peptides with Residue-Specific Force Fields. *J Phys Chem Lett* (2016), 7, 1805-1810.

[230] R. B. Best, N.-V. Buchete, G. Hummer. Are Current Molecular Dynamics Force Fields too Helical? *Biophys J* (2008), 95, L07-L09.

[231] O. F. Lange, D. van der Spoel, B. L. de Groot. Scrutinizing molecular mechanics force fields on the submicrosecond timescale with NMR data. *Biophys J* (2010), 99, 647-655.

[232] E. A. Cino, C. Wing-Yiu, M. Karttunen. Comparison of Secondary Structure Formation Using 10 Different Force Fields in Microsecond Molecular Dynamics Simulations. *J Chem Theory Comput* (2012), 8, 2725-2740.

[233] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, D. E. Shaw. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* (2010), 78, 1950-1958.

[234] S. Piana, K. Lindorff-Larsen, D. E. Shaw. How robust are protein folding simulations with respect to force field parameterization? *Biophys J* (2011), 100, L47-L49.

[235] A. D. Mackerell, M. Feig, C. L. Brooks. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J Comput Chem* (2004), 25, 1400-1415.

[236] N. Schmid, A. P. Eichenberger, A. Choutko, S. Riniker, M. Winger, A. E. Mark, W. F. van Gunsteren. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur Biophys J* (2011), 40, 843-856.

[237] J. Wang, P. Cieplak, P. A. Kollman. How well does a restrained electrostatic

potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J Comput Chem* (2000), 21, 1049-1074.

[238] W. F. van Gunsteren, J. R. Allison, X. Daura, J. Dolenc, N. Hansen, A. E. Mark, C. Oostenbrink, V. H. Rusu, L. J. Smith. Deriving Structural Information from Experimentally Measured Data on Biomolecules. *Angew Chem, Int Ed* (2016), doi:10.1002/anie.201601828.

[239] Y. Shen, A. Bax. SPARTA+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. *J Biomol NMR* (2010), 48, 13-22.

[240] S. Zaretsky, J. L. Hickey, M. A. St. Denis, C. C. G. Scully, A. L. Roughton, D. J. Tantillo, M. W. Lodewykc, A. K. Yudina. Predicting cyclic peptide chemical shifts using quantum mechanical calculations. *Tetrahedron* (2014), 70, 7655-7663.

[241] J. Graf, P. H. Nguyen, G. Stock, H. Schwalbe. Structure and Dynamics of the Homologous Series of Alanine Peptides: A Joint Molecular Dynamics/NMR Study. *J Am Chem Soc* (2007), 129, 1179-1189.

[242] D. A. Case, C. Scheurer, R. Brüschweiler. Static and Dynamic Effects on Vicinal Scalar J Couplings in Proteins and Peptides: A MD/DFT Analysis. *J Am Chem Soc* (2000), 122, 10390-10397.

[243] W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides, E. D. Laue. The CCPN Data Model for NMR Spectroscopy: Development of a Software Pipeline. *Proteins: Struct, Funct, Bioinf* (2005), 59, 687-696.

[244] A. T. Brunger. Version 1.2 of the Crystallography and NMR System. *Nat Protoc* (2007), 2, 2728-2733.

[245] F.-Y. Dupradeau, A. Pigache, T. Zaffran, C. Savineau, R. Lelong, N. Grivel, D. Lelong, W. Rosanski, P. Cieplak. The R.E.D. tools: Advances in RESP and ESP charge derivation and force field library building. *Phys Chem Chem Phys* (2010), 12, 7821–7839.

[246] M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, J. A. Montgomery. General Atomic and Molecular Electronic Structure System. *J Comput Chem* (1993), 14, 1347-1363.

[247] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J Am Chem Soc* (1995), 117, 5179-5197.

[248] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, G. Bussi. PLUMED2: New feathers for an old bird. *Comput Phys Commun* (2014), 185, 604-613.

[249] L. Wickstrom, A. Okur, C. Simmerling. Evaluating the Performance of the ff99SB Force Field Based on NMR Scalar Coupling Data. *Biophys J* (2009), 97, 853-856.

[250] V. F. Bystrov. Spin—spin coupling and the conformational states of peptide systems. *Progress in NMR Spectroscopy* (1976), 10, 41-82.

[251] J. R. Allison, W. F. van Gunsteren. A method to explore protein side chain conformational variability using experimental data. *Chem Phys Chem* (2009), 10, 3213-3228.

[252] A. Corti, F. Curnis, G. Rossoni, F. Marcucci, V. Gregorc. Peptide-Mediated Targeting of Cytokines to Tumor Vasculature: The NGR-hTNF Example. *BioDrugs* (2013), 27, 591-603.

[253] K.-C. Chou. Prediction of Tight Turns and Their Types in Proteins. *Analytical Biochemistry* (2000), 286, 1-16.

[254] S. Hubbard, J. Thornton. NACCESS, Computer Program. *Department of Biochemistry Molecular Biology, University College London* (1993).

List of candidate's pubblications

C. Paissoni, M. Ghitti, L. Belvisi, A. Spitaleri, G. Musco.
Metadynamics Simulations Rationalise the Conformational Effects Induced by N-Methylation of RGD Cyclic Hexapeptides.
Chem- Eur J (2015), 21, 14165-14170.

G. Tiana, F. Villa, Y. Zhan, R. Capelli, C. Paissoni, P. Sormanni, E. Heard, L. Giorgetti, R. Meloni.
MonteGrappa: An iterative Monte Carlo program to optimize biomolecular potentials in simplified models.
Comp Phys Comm (2015), 186, 93-104.

C. Paissoni, D. Spiliotopoulos, G. Musco, A. Spitaleri.
GMXPBSA 2.0: A GROMACS tool to perform MM/PBSA and computational alanine scanning.
Comp Phys Comm (2014), 185, 2920-2929.

R. Capelli, C. Paissoni, P. Sormanni, G. Tiana.
Iterative derivation of effective potentials to sample the conformational space of proteins at atomistic scale.
J Chem Phys (2014), 140, 195101.