UNIVERSITY OF NAPLES
"FEDERICO II"

DEPARTMENT OF PHARMACY

Ph.D. Course
in
"Scienza del Farmaco"

XXXIV CICLO

Innovative Computational Approaches in Drug Discovery: Design and Development of Brand New Chemotherapeutic Agents

Candidate:
Vincenzo Maria D'Amore

Tutor:
Prof.: **Luciana Marinelli**

Coordinator:
Prof. **Maria Valeria D'Auria**

CONTENTS

Outline of the Thesis

During my Ph.D., I mainly focused on the applications of classical and advanced computational techniques to the medicinal chemistry field.

Particularly, I tried to exploit, to the best of my ability, the wide armoury of known computational methods to facilitate the identification and the development of new potential drug candidates. During my studies, I also had the opportunity to deal with different kinds of biological targets, ranging from complex cellular surface proteins (i.e., Integrins or G-protein coupled receptors (GPCR)) to non-standard nucleic acid structures, such as G-quadruplexes.

This thesis is divided into four chapters. The first one describes the classical drug discovery pipeline and the advantages offered by computer-aided approaches. In the second one, a theoretical overview of the methodologies applied in this thesis is provided. The final chapters (3 and 4) focus on the four major research works of my Ph.D. These works are indeed divided into two distinct sections, based on the kind of employed methodologies. In detail, in Chapter 3, two studies focused on the ligand binding problem are presented: i) a successful virtual screening campaign targeting the DNA G-Quadruplex structure of the *KRAS* proto-oncogene promoter, and ii) a mechanistic insight into the binding mechanism of small

molecules to FPR2, a GPCR involved in the resolution of the inflammatory process. In Chapter 4, instead, I aimed to highlight the importance of an accurate conformational sampling for rationalizing the activity/selectivity profile of peptide ligands. In the first case study, due to the availability of NMR-derived data, a mixed computational-experimental approach was adopted to investigate the folding and binding properties of a small cyclopeptide, endowed with remarkable antiviral activity against Herpes Simplex Virus 1 (HSV1) infections. On the other hand, in the last project of this thesis, a purely computational approach was employed for studying the binding mechanism of the well-characterized antitumoral nonapeptide iRGD to integrin receptors.

Chapter 1. Introduction

The development of a new drug is a lengthy, difficult, and costly process going from the target discovery and validation, right until the identification of a potential drug candidate or lead compound.[1]

The drug discovery pipeline is conventionally divided into the following steps:

1. *Target identification and Validation*

   The main goal of this task is to identify an endogenous component playing relevant roles in the etiology of a disease. A perfect pharmacological target needs to be "druggable", which means easily modulable through drug-like molecules (small molecules or peptides). At the same time, researchers need to validate the target assuring that its therapeutic modulation is effective against the pathology without significant side effects. In other terms, only targets with a good "therapeutic window" can be selected for the next phase.

2. *Step 2 – Hit identification*

   Finding molecules having the desired effects against the identified targets is one of the key steps of the entire drug discovery cycle. Many different approaches are usually employed to this aim. Big pharma companies usually set up large-scale *high-throughput screening (HTS)*

programs, in which millions of chemical compounds are directly tested on the target by various kinds of pharmacological tests. On the other hand, more time-consuming but cheaper approaches are represented by knowledge-based and computer-aided drug design, which will be better described in the following paragraphs.

3. *Step 3 – Moving from a hit to a lead*

   Once a good number of hits has been found, a chemical refinement procedure is necessary to obtain more potent and selective molecules. At this stage, parallel research programs on structurally different compound series are highly recommended. Some of them, in fact, will likely fail due to series-specific chemical and pharmacological features.

4. *Step 4 – Lead Optimization*

   In this task, researchers aim at preserving the desired pharmacological properties of the lead compound but trying to still ameliorate its activity. This is achieved by identifying potential deficiencies in the lead structure, and by improving them accordingly to obtain a preclinical drug candidate. Moreover, a pharmacokinetic and toxicologic analysis is often performed to find out whether the drug candidate is metabolized in the right area of the body, or whether it can carry any concerning side effects. For this task, an integrated multidisciplinary approach is thus recommended. The combination of

specialists in computational chemistry, medical chemistry, drug metabolism, and other areas can provide unique insights fundamental for a *tout-court* optimization of the lead compound.

1.1. Computational methods in Drug Discovery

As briefly described in the previous paragraph, bringing a new drug to the market is a demanding process in terms of time, economical efforts, and manpower. It has been estimated that about $1.8 billion is required for an entire drug discovery campaign, and that the average attrition rate of the process is as high as 96%.[2] The main reasons of the poor success percentage are related to both pharmacodynamical and pharmacokinetic problems affecting the new drug candidates. Often, in the latest preclinical or clinical steps researchers are forced to abandon the development of a particular chemical series due to poor drug efficacy, deficient absorption/metabolism/secretion or high toxicity.[3] *High-throughput screening (HTS)* can partially compensate for the first part of the problem, giving a reliable estimate of the drug efficacy. In HTS automated equipment is used to rapidly test thousands to millions of samples for biological activity at the model organism, cellular, pathway, or molecular level. In its most common form, HTS is an experimental process in which 10^3–10^6 small molecules of known structure are screened in parallel. However, this traduces in great

economical efforts. Both the synthesis of wide chemical libraries and the automatization of the pharmacological tests make HTS affordable only to big pharma companies.[4] Moreover, the lack of full understanding of the molecular mechanism behind the activities of HTS-identified hits can hamper their optimization and conversion to a lead compound.[5,6]

In the last three decades computer-aided drug design (CADD) came to the rescue, introducing a different, more rational, knowledge-based approach. CADD employs a wide range of techniques deriving from computational chemistry and bioinformatics to preliminary assess *in silico* the pharmacodynamics and pharmacokinetics of the designed molecules, minimizing the efforts spent in the following experimental tasks. In a typical CADD based workflow, only compounds showing promising properties *in silico* are selected, synthesized, and tested against the target. This way, not only the number of biological tests is minimized, but also the success rate of the entire process is significantly improved by preliminary filtering out potentially inefficient and toxic compounds.[7,8] Some of the most revolutionary drugs approved in the last two decades are fruits of computer-driven design. Among these are some human immonudefiency virus (HIV-1)-inhibiting agents (atazanavir,[9] saquinavir,[10] indinavir, and ritonavir[11]), anti-cancer drugs (raltitrexed)[12], antibiotics (norfloxacin)[13] or antihypertensive compounds (captopril)[14].

However, what are the main tasks that can be accomplished by CADD? First, starting from information collected over the years about gene expression, proteomics, and transgenic phenotype, bioinformatical techniques can be used to identify genes involved in the pathogenesis or the interactions between infectious pathogens and hosts. In turn, proteins or enzymes codified by these genes can be potentially used as drug targets.[15]

Second, given a target, computational chemistry can help in the identification of potential hit candidates. Under the theoretical point of view, computational techniques are mathematical models coming from classical mechanics and quantum-chemistry incorporated in computer algorithms to simulate events of physical and biological interests. In the case of drug design, the main object under study is the interaction between the potential drug candidate and its pharmacological target.

Based on the availability of structural information on the target, CADD can be divided into structure-based drug design (SBDD) and ligand based drug design (LBDD).[7] Most of the methods employed in this thesis belong to the SBDD class, so they will be quickly introduced here and discussed into theoretical details in Chapter 2. For SBDD a 3D structure of the target, coming from experiments such as X-ray crystallography, NMR or Cryo-Electron Microscopy, is needed. If neither one is available, a model system can be predicted *in silico* by homology with other proteins of the same family

(*homology modelling*) or *ab initio*.[16] The most important goal of SBDD is to predict the binding modalities and, in turn, the affinity of potential drug candidates to the desired macromolecule. At a ground level, this objective is achievable by the employment of *Molecular Docking*[17] (2.1), that can also be used to test big chemical databases on the same target in the so-called *Virtual Screening* approach[6] (2.2).

Later, structure-based techniques can be very useful and effective also in the optimization of the hit / lead compounds discovered in the previous screening phases. For instance, the prediction of the ligand binding modalities by molecular dockings allows to quickly evaluate how specific chemical modifications could affect the affinity of the investigated molecules. This way, the researchers can rationalize the structure-activity relationships (SARs) of specific chemical series and run the following optimization step based on these. However, fast methods (e.g., docking calculations) suffer of intrinsic limitations, like the neglection of protein flexibility and solvent effects. Since these factors can be fundamental determinants in the ligand binding event, more accurate calculations are often required to investigate in depth the mechanism of action of a drug. In this context, valuable tools are also biosimulations, both in their standard form of Molecular Dynamics (MD)[18] and as advanced free energy methods (2.3, 2.4).

On the other hand, when the target experimental structure is not available and it is not possible to predict a high-quality model using *in silico* methods, LBDD represents a valuable alternative. This approach requires prior information of known active molecules on the target protein and assumes that structurally related compounds can similarly interact with biological targets, exerting comparable pharmacological effects. Usually, a set of known ligands active and inactive against a relevant target is taken to identify common structural and physicochemical properties (molecular descriptors) responsible for the observed biological activity. The most common LBDD techniques are quantitative structure–activity relationships (QSARs) and pharmacophore-based methods.

Chapter 2. Theoretical Background

2.1. Molecular Docking

Molecular docking is a computational method, discovered in the early 1980s, aimed at foreseeing the interaction modalities of two molecules.[19] In medicinal chemistry, molecular docking is widely employed for predicting the conformation and orientation (posing) of a potential drug candidate (i.e. small organic ligands or peptides) within the binding cleft of a given pharmacological target.[20] However, advanced docking algorithms to investigate protein-protein and nucleic acid-protein interaction mode are also currently available.[21–24] The first step of a docking calculation is represented by the generation of different ligand conformations which are then posed in the receptor binding site, through the so-called *search algorithm*. Later, the results of the search phase are energetically evaluated and ranked by a *scoring function.* Notably, important approximations are applied both in the search and in the scoring phases, which are responsible for the high time efficiency of docking as well as for its limited accuracy. The most important approximations concern the employment of simplified force fields, restrictions of the search space and the neglection of the target conformational flexibility

and solvent effects.[25] All these issues will be discussed in the following sections.

2.1.1. Search Methods

The search methods are implemented in docking programs to explore the ligand conformational space upon the binding event. These can be divided into three major classes:[26]

1. ***Systematic methods:*** algorithms that attempt to explore all the degrees of freedom of a molecule to generate all its possible conformations. The simplest implementation of systematic algorithms is the *exhaustive search*, in which the software regularly rotates every ligand bond at a given interval. The major drawback of this approach is that the computation time exponentially increases with the number of rotatable bonds of the ligand. Therefore, geometrical/chemical constraints are usually applied in the first steps of docking to provide an initial guess of the binding pose.[26] Then, the filtered conformations are optimized and refined in the later stages hierarchically. The docking program mostly employed over this thesis, namely Glide,[21,22] works with a similar scheme. Unlikely, other software divides the ligand into a rigid core and flexible branches. The ligand conformation is thus generated by primary docking the rigid core and then incrementally adding the

flexible parts. This approach is usually referred to as *growing search*.[27] A last type of conformational search is represented by *ensemble methods*,[28] in which an ensemble of pre-generated ligand conformations is rigidly docked. Then, the ligand conformational flexibility is considered by evaluating different binding modes coming from distinct docking runs, which are collected and ranked according to their binding energy scores.

2. ***Stochastic methods*:** also known as *random search* algorithms, they sample the possible binding poses introducing at each docking step random changes in the ligand conformational and roto-translational space. The acceptance or the rejection of each change in the ligand configuration is ruled by a probabilistic criterion. The two most famous types of random search algorithms are the Monte Carlo (MC) methods and the Evolutionary algorithms (EA). In MC search, the random perturbation introduced to the initial ligand binding conformations are accepted or discarded according to the Metropolis-Hastings criterion.[29] On the other hand, EAs, also known as genetic algorithms (GA), follows the basic idea of natural selection in biological systems. GAs treat docking as an optimization problem, in which each binding pose candidate has a set of properties that can be mutated to find the best solution. These parameters represent the "chromosomes" of each

model and are subjected to stochastic mutations (*generation phase*). The newly generated chromosomes are then evaluated by a fitness function, which selects the best intermediates to submit to the next generation steps. This iterative process usually terminates when either a maximum number of generations has been produced, or a satisfactory fitness level has been reached for the population.[30]

3. ***Simulation methods***: algorithms based on molecular mechanics force-fields. The most common of these methods is energy minimization, often employed to refine poses obtained by a variety of docking programs.

2.1.2. Receptor molecular representation

Different approaches have been employed over the years to describe the ligand binding partner in docking calculations. These can be divided into three major categories:[20,31]

1. *The atomistic representation* of the target is too computational expensive to be employed during all the steps of the calculations. For this reason, atomistic details are usually evaluated by a potential energy function only during the final ranking procedures. [32]
2. The *surface representation* is often used in rigid-body, protein-protein or protein-DNA docking. These methods attempt to align points on

surfaces by minimizing the angle between the surfaces of opposing molecules. In turn, the receptor surface can be described in different methods, like Van der Whaals surface, Connoly surface and so on.[33,34]

3. Most of the contemporary docking programs use a macromolecule representation which embeds in a surface description (important to achieve geometric complementary between the ligand and the target) features mimicking the receptor energetic contribution. As described in the seminal paper by Goodford,[35] contour surfaces are built in the three-dimensional space on grid points, each of them usually storing two types of receptor potential energy terms: electrostatic and Van der Waals. Particularly, the Van der Waals term is computed at each grid point by estimating the interaction energy between the protein and a probe atom based on a Lennard-Jones potential. Notably, probe atoms are selected to mimic all the possible atom types present in the investigated ligands. On the other hand, the Coulomb equation is applied for the calculation of the electrostatic contribution, usually employing as probe point charges of 1.60219×10^{-19} C. Then, both the electrostatic and Van der Waals terms are used in the scoring phase to evaluate the stability of the proposed docking pose.

2.1.3. Protein Flexibility

One of the biggest limitations of docking is that most programs treat the target macromolecule as a rigid body. Due to this, many shortcuts have been developed over the years to reproduce the receptor flexibility. Some of the most famous approaches are:

1. ***"Ensemble docking",*** where the ligand is docked in multiple three-dimensional conformations of the receptor. In the best case these can be experimentally derived structures, otherwise they can also be low-energy conformations sampled *in silico,* using for instance molecular dynamics simulations;[36,37]
2. ***"Energy-weighted"*** schemes, describing an ensemble of receptor conformations as an average potential energy grid of the underlining stuctures;[38,39]
3. ***"Soft docking"***, where the criterion for steric fit between the ligand and the receptor is softened. In particular, more tolerant scoring function are employed to reduce steric hindrance or clashes upon the interaction of the binding partners;[40]
4. ***"Induced-Fit"*** docking procedures, which treat the side chains of a user-defined portion of the binding site as flexible. This restricted type of conformational sampling can be performed through systematic search, MC or MD based algorithms.[23]

2.1.4. Scoring

At the end of the *posing* phase, docking calculations come to the final step, the *scoring*. Its aim is to rank the solutions proposed by the posing, according to the predicted binding affinity for the receptor. This is a crucial aspect of the entire docking process because, even when the real binding mode is predicted, a docking result could be useless or misleading if correct poses are not differentiated from incorrect ones. The physical quantity best describing the ligand-target affinity is the binding free energy (ΔG_b). State-of-art computational methods for an accurate quantitative prediction ΔG_b are free energy calculations (2.4), which however are time-demanding and unpractical in the screening of many molecules. On the other hand, the functions used in the docking scoring phase provide only coarse-grained approximations of ΔG_b and can be divided into three major classes:

1. ***Force-field based scoring.*** Molecular mechanics force fields treat the energetics of a protein-ligand complex as the sum between the two partners interaction energy and the ligand internal energy. The protein internal energy is usually omitted because most software only considers a single receptor conformation, drastically simplifying the calculation. The description of the two energy terms is based on the application of a Coulomb's law (with a distance-dependent dielectric

permittivity) and a Lennard-Jones potential for electrostatic and Van der Waals interactions, respectively. Standard force-field based scoring functions, however, suffer of major limitations due to the lack of solvation and entropic terms.

2. ***Empirical scoring functions.*** The development of this class of functions is based on the idea that ΔG_b can be approximated by a sum of individual uncorrelated terms fitted to reproduce experimental data.[41] The coefficient of the various terms are computed by regression analysis from experimental binding free energies or X-ray structural information. Usually, empirical functions are simpler to evaluate than the force-field based ones. The main drawback of this approach is that they are strongly dependent on the molecular data sets used to perform regression analyses and fitting. Moreover, terms accounting for non-bonded interactions, like hydrogen-bonds, and non-enthalpic contributions can be included, although they can provide only incomplete descriptions of these effects on protein–ligand binding.[20]
3. ***Knowledge-based scoring functions***. These are designed to reproduce experimental structures rather than binding energies. Simple atomic interaction-pair potentials are used in these models for describing protein-ligand interaction. In particular, based on the different molecular environment, atom-type interactions are defined. As for the

empirical scoring, the main disadvantage is that their derivation is based on information implicitly encoded in limited sets of protein–ligand complex structures.

A possible approach to improve the performance of docking scoring is to use multiple functions in a *consensus* workflow.[42] Notably, the combination of information coming from distinct scores could balance errors affecting the single ones. However, this is not the case when the used scoring functions are made of strongly correlated terms that could, in turn, lead to an error amplification.

2.1.5. The Glide docking program

Glide is a widely employed docking program, originally developed in 2005.[21,22] It follows a workflow based on the hierarchical applications of filters to predict the optimal ligand conformation in the binding site (Figure 2.1).

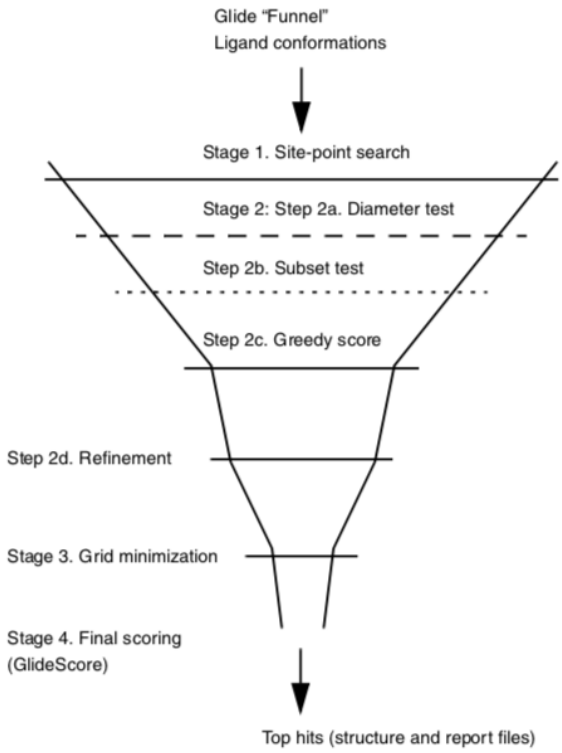


Figure 2.1 The glide docking hierarchy from Ref. 22

First, the ligand conformational flexibility is treated with an *exhaustive search* refined by the employment of a heuristic screen to rapidly discard unsuitable ligand conformations. In practice, the ligand is divided into a *core* region and *rotamers*. From every *rotamer*, a set of *rotamer states* is enumerated. Conformations made by different arrangements of the *core* regions plus all the possible *rotamer states* are docked as single objects. The posing begins with the selection of "Site-points" (Step 1 in Figure 2.1) in an equally spaced 2 Å grid describing the protein active site. To make this selection, the software compares precomputed distances from the site point to the receptor surface (evaluated at a series of prespecified directions and binned in 1 A˚ ranges) with

binned distances from the ligand centre to the ligand surface. Then, Glide positions the ligand centre at the site point if there is a good enough match of the histograms of binned distances but skips over the site point if there is not. The user can define the search area of possible site-points for the ligand by defining the dimension of the grid, by default 12 Å per cartesian dimension.

In stage 2, the placement of atoms laying within a given distance from the major ligand axis is evaluated. Particularly, if there are too many steric clashes with the receptor, the orientation is discarded (step 2a, Figure 2.1). Next, the ligand is rotated around its axis, looking for possible hydrogen bonds with the receptor, which are here evaluated according to a discretized version of the ChemScore[43] function (step 2b, Figure 2.1). If the resulting score is good, then all the interactions of the pose are evaluated (*greedy scoring* - step 2c, Figure 2.1). The final step in stage 2 consists in rescoring the top greedy-scoring poses via a "refinement" procedure (step 2d, Figure 2.1) in which the entire ligand is allowed to rigidly move by 1 Å in the Cartesian directions.

Only a few of the scored poses coming from stage 2 are then energy minimized on precomputed electrostatic and Van der Waals grids using the force field OPLS-AA[44]. The interaction of the ligand with each vertex of the grid is evaluated using trilinear interpolation formulas for a cube. The accuracy of the method is enhanced by the fact that the smaller the distance of the grid point from the receptor surface, the higher the resolution of the box used. The

Coulomb/van der Waals grid is initially built using large boxes, typically 3.2 Å along each side, and is then refined hierarchically into boxes of 1.6, 0.8, or 0.4 Å depending on the distance of the box to the van der Waals surface of the protein.

Finally, the minimized poses are ranked using Schrödinger's proprietary *GlideScore* scoring function. GlideScore is based on ChemScore, but it includes a steric-clash term, adds other rewards and penalties such as buried polar terms (to penalize electrostatic mismatches), amide twist penalties, hydrophobic enclosure terms, and excluded volume penalties, and has modifications to other terms:

$$GScore = 0.05vdW + 0.15Coul + Lipo + Hbond + Metal + Rewards + RotB + Site \quad (2.1)$$

2.2. Virtual Screening

Virtual Screening has become a popular computational technique used in the *hit identification* phase of drug discovery. Basically, VS methods can be divided into *structure-based (SBVS)* and *ligand-based (LBVS)*. In the first case, a 3D experimental structure or a model of the target receptor is available and multiple docking calculations are performed to screen big database of chemical compounds. On the other hand, LBVS is based on the concept of ligand similarity and requires a set of already known active compounds

against the target. In the following, I will focus on the only VS approach employed in this thesis: SBVS.

2.2.1. Structure-based Virtual Screening

The main goal of SBVS campaigns is to identify novel potential hits to test against a given pharmacological target through a preliminary computational evaluation of large chemical databases.[45] A typical SBVS workflow can be divided into the following phases:

1. Choice and preparation of ligand libraries
2. Selection of the target structure
3. Analysis of the results and hit selection

2.2.1.1. Choice and preparation of ligand libraries

The selection of the ligands to screen against the desired target can be made from a large number of web-available databases such as PubChem (https://pubchem.ncbi.nlm.nih.gov), eMolecules (www.emolecules.com) and ZINC.[46]

Often large databases can be preliminarily filtered and divided into smaller subsets preserving a wide chemical diversity but reducing the computational time required for docking it against the target. This procedure can be

performed following different guidelines. An example is to discard from the original library all the molecules not endowed with drug-like properties, according to the well-known Lipinski's rule of five[47] This rule, indeed, relies on the most common chemical features of FDA-approved drugs; in particular, it states that, generally, an orally active drug should have: i) at least 5 hydrogen bond donors and 10 acceptors; ii) a molecular weight (M.W.) lower than 500 Da; and iii) a $\log P_{octanol/water}$ coefficient lower than 5. Notably, a certain tolerance should be permitted since many marketed drugs (i.e. peptides) do not totally fit these criteria.

Besides the pre-filtering stage, before the docking, all the possible tautomeric and protonation states of each compound should be generated, to certainly take into account the optimal chemico-physical state for the interaction with the receptor. However, recent studies have shown that, in some cases, the use of the most probable tautomer/protomer can be better than docking the entire ensemble, since the scoring functions are generally not enough accurate to discriminate among the different ligand binding states.[48]

2.2.1.2. Selection of the target structure

As already mentioned, a fundamental pre-requisite for SBVS is the availability of a 3D structure of the target macromolecule which can derive both from experiments and from computational/bioinformatics approaches. Particularly,

most of the 3D structures of proteins and nucleic acids deposited in the Protein Data Bank (PDB) are results of structural studies performed through X-ray crystallography, Cryo-Electron Microscopy (Cryo-EM), and NMR. On the other hand, in the lack of any experimental set of coordinates for the desired target, the user can resource to homology modelling (HM). HM is a bioinformatic technique aimed at reconstructing the structure of a protein, based on its sequence similarity with related macromolecules whose 3D arrangement has already been resolved.

Notably, the ideal condition for a SBVS is when the active site of the target is well-defined. In these cases, it is desirable to select a 3D structure with an inhibitor bound (*holo*), to have a receptor conformation that is more suitable for the binding of new compounds. Of course, this in not always possible; thus, unbound target molecules (*apo*) sets must often be selected for VS calculations. In these cases, issues of flexibility and protonation state or errors in modelling could affect VS results and thus must be addressed prior to starting docking calculations. In some other cases, researchers can have no *a priori* knowledge of ligand binding sites target molecule. In these conditions they can do a blind docking onto the entire protein, to identify putative ligand binding sites.

2.2.1.3. Analysis of results and hits selection

In structure-based VS, two aspects are critical to discriminate between active and inactive compounds. First, the calculations should find a relevant binding conformation for each of the investigated ligands. To validate the docking protocol adopted during the VS, "redocking" experiments are conventionally performed. In other words, if complexes of the target macromolecule with a known ligand are available, these are separated and the compound is redocked to evaluate the capability of the program to reproduce the experimental binding mode. This also allows to fine-tune the docking parameters and to validate the target preparation procedure. Moreover, if a series of active compounds against the target exists, this could be docked into the receptor to take the minimum energy score of these molecules as a threshold in the evaluation of VS results.

Secondly, the scoring function should be accurate enough to ensure that top-ranked compounds behave as effective binders when tested experimentally. In this perspective, researchers usually carry out a process of "enrichment", where the set of compounds that are predicted to bind tightly are enriched in compounds that show strong binding upon testing. Conceptually, the enrichment factor metric is simply the measure of how many actives are found within a defined "early recognition" fraction of the ordered list relative to a random distribution, as follows:

$$E_f = \frac{N_{experimental}x\%}{N_{expected}x\%} / \frac{N_{active}}{N_{total}}, \quad (2.2)$$

where $N_{experimental}^{x\%}$ is the number of experimentally found active structures in the top x% of the sorted database, $N_{expected}^{x\%}$ is the number of expected active structures from VS, $N_{active}$and N_{total} are the number of experimentally active structures and the total number of compounds contained in the database, respectively.

The most straightforward way to classify the VS results is to employ the predicted binding score for each compound. However, an additional suggestion is to look at the convergence of the calculation towards a particular solution, or in other words, the recurrence of a particular docking pose. This could be achieved by clustering the solutions according to the RMSD of the ligand's coordinates.[49] A further measure is the ligand efficiency, defined as the docking score normalized per non-hydrogen atoms in the ligand. This parameter is employed to balance the strong bias of scoring functions towards large compounds which are usually predicted with higher binding affinities than low M.W. ligands.[50]

Notably, the standard error of docking methods is estimated at around ±2 kcal/mol. Thus the estimated free energy value should not be taken as an only reference criterion for the selection of potential hits. Therefore, visual inspection is a needed practice to increase the success rate. Several aspects of the docked conformation may be used to filter the set of compounds, such as

the presence of key contacts with critical residues in the target, similar to known positions of ligands or waters in the active site, or the presence of unpaired hydrogen bond donors or acceptors in the ligand-receptor complex.[45]

2.3 Molecular Dynamics

Over the last decades, Molecular Dynamics (MD) simulations have become a widely employed instrument to get mechanistic insights into the functioning of complex biological processes. In the medicinal chemistry field, researchers use biosimulation for studying a large variety of phenomena ranging from the identification of cryptic drug binding sites to the prediction of important properties, such as drug resistance.[18]

From the theoretical point of view, MD solves the motion of many-body systems of *N* particles under the action of forces acting in certain conditions of temperature, pressure, and volume. In detail, the movement of particles/atoms in real life is replicated by deterministically solving Newton's equation of motions:

$$F_i = m_i \frac{\delta^2 r_i}{\delta t^2} \qquad i = 1 \dots N \tag{2.2}$$

Where *m* is the mass and *r* is the position of each *N* particle of index *i*, while *t* is the time. To better understand the theoretical bases of MD, some

clarifications need to be done. First, standard MD simulations obey the laws of classical mechanics, indeed no quantum-mechanical effects are taken into consideration. According to the *Born-Oppenheimer approximation* the electrons in an MD simulation are always considered in their ground state and their position is directly dependent on the motions of the nuclei, due to their opposite charge and much different size. As a by-product, biological processes based on the differential motion between electrons and nuclei cannot be correctly reproduced. This level of resolution is thus well-accepted for big molecular systems and as long as the degrees of freedom of the system (i.e. atomic vibrations, ν) do not overcome the environmental energy:

$$h\nu = k_B T \tag{2.3}$$

where h is the Plank's constant, k_B is the Boltzman constant, and T is the temperature. Notably, it is impossible to replicate a biological environment in its entirety, thus researchers are forced to restrict the simulation area to a portion of the system under study (generally up to 1 million particles). The latter is cut off from the rest of the environment and positioned within a simulation box. To avoid the margin of the system getting in contact with the void and to minimize the introduction of physical artifacts, MD simulations are set up in the so-defined *Periodic Boundary Conditions (PBC)*. This procedure consists in the replication of the simulation box along all Cartesian axes, approximating an infinite medium.

2.3.1. Defining the potential energy

A crucial point in MD methods is related to how the forces are computed for the integration of Newton's equation of motion. In *atomistic* MD, the interaction potentials between particles are described based on a set of analytical functions called *force fields*. On the other hand, in *ab initio* calculations electrons' behaviour is treated explicitly and obtained from first principles by using a quantum-mechanical method of choice (i.e. *density functional theory, DFT*). If the latter methods are particularly rigorous and able to describe phenomena like building/breaking reactions or excited (transition) states transitions, atomistic simulations are, in turn, more computationally affordable and suitable for the investigation of large and complex biomolecular systems. All the simulations presented in this thesis were performed at atomistic resolution. For this reason, in the following paragraphs, a quick overview of force fields' structure and functioning will be carried out.

2.3.2. Force Fields

In molecular mechanics, the expression "Force Fields (FF)" refers to the tabulated data set of parameters and the functional form used for the computation of the potential energy of every single particle in a simulation. The FFs parameters can either be fitted from *ab-initio* calculations or extracted from experimental data. Force fields can have different degrees of resolution. We can distinguish three major FF classes:

1. *All-atoms force fields* include parameters for all the atoms of the system, including the hydrogens
2. *United-atoms force fields* treat the hydrogen and carbon atoms in methyl groups and methylene bridges as one interaction centre
3. *Coarse-grained force fields* are usually employed for long time-scale simulations of very big macromolecules. They sacrifice chemical details for higher computing efficiency by mapping from 2 to 4 atoms in a single entity called a *bead.*

All the simulations discussed in this thesis adopted a popular *all-atoms* force field for proteins and nucleic acids: the AMBER force field. Its master equation is the following:

$$V_i = \sum V_{bonds} + \sum V_{angles} + \sum V_{dih} + \sum V_{imp} \sum V_{vdW} + \sum V_{Coulomb} \quad (2.4)$$

where the total potential energy V acting on the atom *i* is the sum over all the bonded (bonds, angles, dihedral, and improper torsions) and non-bonded (van der Waals and Coulomb) terms. The accuracy of an MD simulation is

strictly related to the precision of the FF because all the forces acting on the atoms are derived from their potentials according to the formula:

$$F_i = \frac{\partial V_i}{\partial r_i} \tag{2.5}$$

where V is the potential computed with the master equation of the chosen FF (i.e., eq. 2.4) and "r" is the position of the atom of index "i".

2.3.2.1. Bonded interactions

Bonded interactions typically include 4 different terms:

1. *bonds*, connecting two atoms sharing a chemical bond
2. *angles*, recognized by three atoms linked by two adjacent bonds
3. *proper dihedrals*, describing the angle between two planes identified by a set of four adjacent atoms
4. *improper dihedrals*, geometrically like proper dihedral but ensuring the planarity of certain chemical groups (e.g., peptide bonds, benzene rings, etc.).

Starting with the bonds, the covalent interaction between two atoms *i* and *j* is treated as a harmonic restraint with the equation:

$$V_{ij} = \frac{1}{2} k_{ij} (r_{ij} - r_{ij}^*)^2 \tag{2.6}$$

where *V* is the potential for the bond, *k* is the harmonic constant determining the strength of the interaction, *r* is the distance between atoms *i* and *j* provided by the simulation step, and r^* is the equilibrium distance for the bond according to the FF.

Similar to bonds, also angles potential is described by a harmonic function:

$$V_{ijk} = \frac{1}{2} k_{ijk} (\theta_{ijk} - \theta^*_{ijk})^2 \quad (2.7)$$

where *k* is the constant regulating the strength of the harmonic angle restraint (tabulated in the FF), θ is the angle among atoms *i*, *j*, and *k* provided by the simulation and θ^* is the tabulated equilibrium value for the given angle.

Dihedrals, together with angles, determine the correct geometry of a molecule. In detail, dihedral angles regulate rotations along with bonds, providing orientations among groups and accounting for conjugation if necessary. They are generally defined with a cosine function (or combination of cosine functions):

$$V_{ijkl} = \sum_z k_{z,ijkl}[1 + \cos (n_z \phi_{ijkl} - \phi^*_z)] \quad (2.8)$$

where *k* is the constant regulating the strength of the torsion, *n* controls the period of the cosine function, ϕ is the angle between the planes formed by atoms *i*, *j*, and *k* and *j*, *k*, and *l*, ϕ* is the phase.

Finally, improper dihedrals are different from the other bonded interactions because they are not declared for every possible combination of four bonded

atoms like in normal dihedrals. An improper dihedral is applied on four specific atoms only on some structures to impede the out-of-plane motion of one of them. Its definition follows a harmonic restraint:

$$V_{ijkl} = \frac{1}{2} k_{ijkl} (\zeta_{ijk} - \zeta^*_{ijk})^2 \quad (2.9)$$

where k is the constant regulating the force of the restraint, ζ is the angle formed by two planes composed by the atoms i, j, k, and j, and $\zeta*$ is generally defined as π radians.

2.3.2.2 Non-bonded interactions

All the interactions formed between atoms not sharing covalent bonds are qualified as *non-bonded.* The computation of this part of the force field master equation is the most time-consuming procedure in an MD simulation. In fact, each atom can potentially interact with all the others in the simulation box, thus the number of calculations theoretically scales with $N*(N - 1)/2$. To reduce the computational cost while preserving a good degree of precision, the problem is generally divided into two macro areas: short and long-range interactions. These are delimited by an arbitrary value defined as a cutoff radius (r_c).

Short-range interactions are considered as pair-based interactions encompassing repulsion, dispersion, and electrostatic terms among atoms in

the r_c threshold. Two functions are usually employed for the description of short-range communications:

1. *Lennard-Jones (LJ) potential* - it treats the van der Waals (vdW) forces using the following form:

$$V_{LJ} = 4\varepsilon[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6}] \quad (2.10)$$

   where *i* and *j* are the two atoms used to compute the potential, *r* is their distance, σ is the van der Waals radius of the *i* and *j* atoms while ε is the strength of such interaction;

2. *Coulomb potential* - it reproduces the electrostatic interaction between a pair of atoms using the following formula:

$$V_C = \frac{q_i q_j}{4\pi\varepsilon_0\varepsilon_r r_{ij}} \quad (2.11)$$

   where *q* is the charge of atoms *i* and *j*, ε_0 is the vacuum permittivity, ε_r is the relative permittivity, and *r* is a distance.

However, the research of the *non-bonded* pairs (between which the forces are computed) still needs to be performed at each simulation step over all the atoms of the system. To reduce the dimensionality of this N^2 problem, usually MD engines resort to *neighbour lists.* For instance, in the *Verlet list* work scheme,[51] a secondary threshold r_v is created, which must be greater than r_c. Looking for the interaction partners of given atom *i*, a list is created with respect to *i* with all the other atoms circumscribed by r_v. If the diffusion of

atoms is lower than r_v - r_c, in the following time step, the pre-generated list might be used to save time. Even if the creation of the list scales with N^2, the successive calculations will only depend on the N atoms in the list, and this complexity order is applicable until the Verlet list must be updated. In addition, some MD engines prefer to shift the potential field projected on i by a constant to obtain 0 upon reaching the r_c limit.

Outside of the range of r_c, *long-range interactions* are computed. The easiest solution to reduce the computational load of long-range interactions would be to completely ignore all contributions outside of the r_c range. Depending on the value of r_c, this is possible only for potentials that go to 0 very fast (rate of decay faster than r^{-3}).[52] Thus, it is valid for vdW interactions, but not for the electrostatic component. In modern MD simulations, the most employed method for the treatment of long-range electrostatic interaction is the *Particle Mesh Ewald (PME)* algorithm.[53] The latter divides the long-range interaction into two parts: a short-range contribution, calculated in the real space, and a long-range computed using a Fourier transform. Notably, the electrostatic potentials are derived from a distribution of charges deposed on a grid (*mesh).* The PME algorithm scales as $N\log(N)$ and is optimal for up to 10^5 atoms.

2.3.3 Integration of Newton's equation of motion

After having calculated all the potential energy terms, these can be converted into forces and applied on every single atom for the integration of Newton's equation of motion. This procedure is iterated several times with a user-defined stride, generally referred to as *timestep* (Δt).

MD engines implement a wide selection of integrators for this task. Although an accurate description of how these algorithms work is out of the scope of this thesis, it is important to know that the change in position of a particle is generally defined as a Taylor series expansion with the form:

$$r(t + \Delta t) = r(t) + v(t)\Delta t + \frac{F(t)}{2m}\Delta t^2 + \frac{\Delta t^3}{3!}r^{\cdots} \ldots \quad (2.12)$$

where r is the position of an atom, t is the time, v is the velocity of a particle, $F(t)$ is the forces applied on the particle with respect to its current position r and time t, and Δt is the time step of the MD simulation. Whatever algorithm one might use to integrate this equation, the maximum time step to be used safely is determined by the frequency of the fastest motion in the system. For molecular systems these are the stretching vibrations involving the hydrogen atoms, leading to time steps in the order of femtoseconds. Such a small time step ensures energy conservation and good accuracy for the sampling of the statistical ensemble; on the other hand, this requires a huge number of steps needed to sample the whole phase space. Indeed, for proteins and nucleic acids the time scale of stretching vibrations are orders of magnitude smaller

compared to large motions, such as conformational changes or folding events, which occur in the range of microseconds to seconds.

2.3.4. The problem of rare events

During an MD simulation, the system visits several microscopic configurations that can be used to measure different observables. Let us consider a simulation performed at a constant number of particles *N*, temperature *T*, and volume *V* (canonical or *NVT* ensemble). An observable can be defined as the ensemble average of a function $s = s(\boldsymbol{R})$ of the microscopic coordinates:

$$\langle s \rangle = \int ds\, s\, P(s) \tag{2.13}$$

where P(s) is the probability distribution of s in the canonical ensemble. This can be written as:

$$P(s) = \langle \delta(s - s(\boldsymbol{R})) \rangle = \frac{\int d\boldsymbol{R} \delta(s - s(\boldsymbol{R})) e^{-\beta U(\boldsymbol{R})}}{Z} \tag{2.14}$$

being $\beta = 1/(k_B T)$, k_B the Boltzmann constant and Z the partition function:

$$Z = \int d\boldsymbol{R}^{-\beta U(\boldsymbol{R})} \tag{2.15}$$

The probability distribution $P(s)$ is related to the Helmholtz free energy by the relation:

$$F(s) = -\frac{1}{\beta}\ln P(s) + C \tag{2.16}$$

where C is an irrelevant additive constant. Since in the NVT ensemble the configurations are distributed according to the Boltzmann weight $e^{-\beta U(R)}$, the chance to extract the probability distribution from an MD run is reliable only if the MD run is long enough for the system to visit all the energetically relevant configurations or, in other words, if the system is ergodic in the time scale of the simulation. In real-world simulations this is condition is rare, because the most relevant configurations of the simulated systems are usually separated by high free-energy barriers which require very long simulations times to be overcome. As a result, plain MD simulations can hardly reproduce important biophysical processes in their entirety such as folding/unfolding, protein conformational changes, or ligand binding events. These kinds of phenomena are usually defined as "rare events" and can be sampled through two main different strategies: i) running very long MD simulations to collect enough statistics on the process under investigation,[54,55] or ii) employing the so-called enhanced sampling methods. In such approaches, the system experiences an additional potential acting on a selected number of degrees of freedom, often referred to as collective variables (CVs). Thus, if $\boldsymbol{S}$ is a set of d functions of the microscopic coordinate $\boldsymbol{R}$ of the system, the additional time-dependent potential $V(S(\boldsymbol{R}), t)$, modifies Newton's equation of motions as:

$$M\ddot{R} = -\frac{\partial U(\boldsymbol{R})}{\partial \boldsymbol{R}} - \frac{\partial V(S(\boldsymbol{R}),t)}{\partial \boldsymbol{R}} \quad (2.17)$$

Here, we will specifically focus on one of these methods, called metadynamics,[56] which, in the last decades, has emerged as a valuable tool to investigate complex biological processes in a reasonable computational time.

2.4. Metadynamics

In metadynamics (MetaD) the sampling is boosted by the deposition of a history-dependent bias potential deposited along some user-defined degrees of freedom of the system, the collective variables (CVs). In detail, the bias potential is built as a sum of Gaussian functions centred on the previously visited configurations of the system in the CV space. Let $\boldsymbol{S}$ be a set of d functions of the microscopic coordinate $\boldsymbol{R}$ of the systems:

$$\boldsymbol{S}(\boldsymbol{R}) = \left(\boldsymbol{S_1}(\boldsymbol{R}),\ldots,\boldsymbol{S_d}(\boldsymbol{R})\right) \quad (2.18)$$

The external bias potential $V_G = V_G(\boldsymbol{S}, t)$ function of the CVs is then added to the Hamiltonian of the system, directly acting on its microscopic coordinates. This potential can be analytically described by the following formula:

$$V_G(\boldsymbol{S},t) = \int_0^t dt' \omega \exp - \sum_{i=1}^{d} \frac{\left(S_i(\boldsymbol{R}) - S_i(\boldsymbol{R}(t'))\right)^2}{2\sigma_i^2} \quad (2.19)$$

where V_G is the total deposed bias, S_i is the value of the i^{th} CV, σ_i is the width of the Gaussian function and ω is the rate at which the bias is deposited. The energy rate is constant and is usually expressed in terms of a Gaussian height W and a deposition stride τ_G:

$$W = \frac{\omega}{\tau_G} \tag{2.20}$$

In figure 2.2 a graphical illustration of the time evolution of a system under the effects of a one-dimensional bias potential V_G is presented. In this example, the time t is measured by counting the number of Gaussians deposited. The simulation starts with the system localized in the central local minimum. As time goes by, gaussians are deposited in position S along the selected CV (x), gradually filling the energy basin and pushing the system toward another local minimum (t=20). The natural and more convenient escape route is passing above the lowest barrier to fall into the left basin. The system is then trapped in this configuration until the bias deposition makes the underlying free-energy basin filled ($t \approx 120$). Starting from t = 160 the system can easily access also the third minimum region on the right. Finally, when also this basin is compensated by the bias potential (t = 320), the system can move in a random walk on the flat free energy surface (FES).

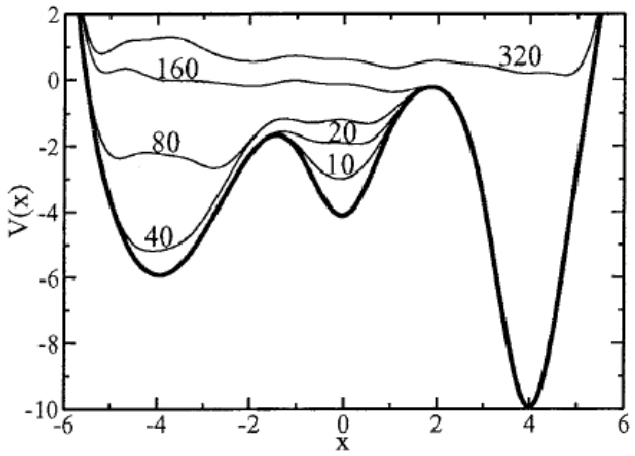


Figure 2.2. Time evolution of the sum of a one-dimensional model potential $V(x)$ and the accumulating Gaussian terms of Eq. 2.14. The dynamic evolution (thin lines) is labeled by the number of Gaussian added. The starting potential (thick line) has three minima and the dynamics is initiated in the second minimum. The figure is adapted from Ref. 56.

This example clearly shows the major advantages of metadynamics. The sampling of rare events is significantly enhanced, discouraging the system to stay in previously visited configurations. Thus, the exploration of new reaction pathways is also favoured because the system tendentially evolves in the direction where lower free barriers are encountered. Different from many others free energy methods, metadynamics does not require prior knowledge of the underlying free energy landscape. Notably, once all the energy basins are filled by the bias potential and the system assumes a semi-diffusive behaviour in the CVs space, the simulations can be considered as "converged". This allows estimating the free energy surface (FES) according to the following analytical formula:

$$\lim_{t\to\infty} V_G(\boldsymbol{S}, t) = -F(\boldsymbol{S}) + C \tag{2.21}$$

Where $F(\boldsymbol{S})$ is the free energy and C is an irrelevant additive constant.

However, we can identify two main drawbacks related to metadynamics:

1. It is difficult to determine the optimal moment to stop a metadynamics simulation because V_G does not tend to the exact value of *F(s)* but oscillates around it. Therefore, the additive bias can overfill the free energy basins pushing the systems towards configurations not physically meaningful. To solve this issue, a variant of metadynamics called *Well-tempered* metadynamics[57] has been developed (2.4.1 for details);
2. The choice of the CVs is far from trivial and must be done before the simulation.

An ideal CV is a physical descriptor of the system which should be able to discriminate between initial, final, and intermediate states of the biophysical process under investigation. To this aim, it must include all the slow modes of the system. We can define as "slow" all the degrees of freedom that, at a given temperature, present free energy barriers too high to be overcome in classical MD time scales. Notably, if a slow mode is neglected in the CVs choice, the system could experience a hysteretic behaviour in which, once the system moves from one basin to another, it has difficulty in coming back to the initial metastable state. When this happens, the bias potential usually does not

converge to *F(s)*. Moreover, since the time required to fill the underlying free energy surface exponentially increase with its dimensionality, the selected number of CVs must be limited. In most practical applications, the employed number of CVs goes up to 3. In order to alleviate these limitations, different combinations of Metadynamics with other enhanced sampling methods (i.e. Parallel Tempering, 2.5.2) have been developed for those cases in which some slow modes can not be easily included in the set of CVs.[58–60]

2.4.1. Well-tempered Metadynamics

The most intriguing novelty introduced in metadynamics by the *Well-tempered* (WT-MetaD) formalism is that the deposition rate ω is exponentially rescaled over time based on how much potential has already been added on the same region of the CV phase space:

$$W = \omega_0 \tau_G e^{-\frac{V_G(\boldsymbol{S},t)}{k_B \Delta T}} \tag{2.22}$$

where W is always the Gaussian height, k_B is Boltzmann's constant, ω_0 is the initial energy rate, τ_G is the Gaussian deposition stride, ΔT a temperature and $V_G(\boldsymbol{S}, t)$ is the bias potential accumulated in $\boldsymbol{S}$ over time t. Notably, in WT-MetaD the bias does not fully compensate the FES but it converges to:

$$V_G(\boldsymbol{S}, t \to \infty) = -\frac{\Delta T}{\Delta T + T} F(\boldsymbol{S}) \tag{2.23}$$

where T is the temperature of the system. In other words, at convergence the CVs are sampled at a (fictitious) higher temperature $T + \Delta T$:

$$P(\boldsymbol{S}, t \to \infty) \propto e^{-\frac{F(\boldsymbol{S})}{k_B(T+\Delta T)}} \tag{2.24}$$

Well-tempered metadynamics solves two of the major problems of standard metadynamics. The first one is the lack of convergence. In standard metadynamics, the bias is continuously added to the system, also after compensating the underlying FES. This causes the estimate of the free energy to oscillate around the correct value. In well-tempered metadynamics the amount of bias added decreases in time, its variation going to zero as $1/t$ while V_G converges to a fraction of the FES (Eq. 2.22). The second problem, strictly related to the first one, is overfilling. In well-tempered metadynamics, the exploration can be restricted to low free-energy regions by properly tuning ΔT. In particular, standard MD is recovered for $\Delta T \to 0$, traditional metadynamics for $\Delta T \to \infty$. This possibility of regulating the extent of exploration is particularly useful for saving computational time in case a large number of CVs is used.[61]

2.4.2. Funnel Metadynamics

Funnel Metadynamics (FM) is a variant of Metadynamics developed in 2013[62] to replicate the entire ligand binding process in an affordable simulation time. FM combines the sampling acceleration due to the deposition of the metadynamics potential with a funnel-shaped restraint to limit the ligand exploration of the phase space only to the target binding site. In detail, the funnel is composed of a cone restraint, set as to include the binding site, and a cylindrical part which, in turn, points toward the solvent. Therefore, the ligand sampling is focalized to the states meaningful for the binding process (conical region). In parallel, the cylindrical part allows the ligand to reach the unbound state without losing too much time in the bulk water. This way, many binding-unbinding recrossing events can easily be simulated, allowing for fast convergence of the calculation and a quantitatively well-characterized binding free energy surface (BFES). Notably, the ligand experiences a repulsive potential only when it tries to leave the area identified by the funnel, whereas is completely

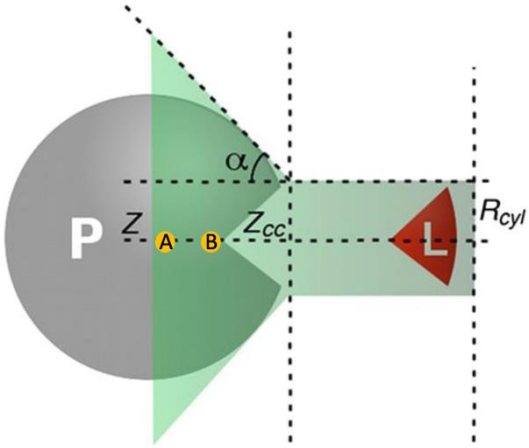


Figure 2.4 Graphical representation of the FM parameters. The image was re-elaborated from Ref. 62

untouched when inside its volume. The funnel can be set up on a given pharmacological target through the definition of a few parameters:[63]

1. ***Linepos (z-axis)*** is the axis of the funnel and can be identified as the straight line passing for two user-defined points *A* and *B;*
2. ***Zcc*** is the value of the switching point between the conical and cylindrical sections;
3. $\boldsymbol{\alpha}$ is the value of the angle determining the cone width;
4. $\boldsymbol{R_{cyl}}$ is the radius of the cylindrical part (usually set to 1Å).

FM is a valuable tool for a thorough thermodynamic characterization of the ligand binding event. In fact, at the end of the calculation, the real ligand binding pose is identified as the lowest free energy minimum in BFES and an accurate estimation of the absolute binding free energy ΔG_b^0 can be computed as the difference between the bound and unbound state. It is important to underline how ΔG_b^0 is a state of function and so only depends on the free energy values of the two states taken into consideration, independently from the path connecting them. Analytically ΔG_b^0 can be computed as:

$$\Delta G_b^0 = -k_B T ln(C^0 K_b) \tag{2.25}$$

where C^0 = 1/1660 $Å^{-3}$ is the standard concentration of 1 M for all reacting molecules, k_B is the Boltzmann constant, and T is the temperature of the system in Kelvins. K_b is the equilibrium binding constant defined as:

$$K_b = \frac{\int_{bound} dr e^{-\beta W(r)}}{e^{-\beta W(r^*)}} = \int_{bound} dr\, e^{-\beta(W(r)-W(r*))} \quad (2.26)$$

where r is a position in the chosen CV, $W(r)$ is the potential of mean force (PMF, is equal to the Free Energy expressed as a function of a collective variable) of the bound states and $W(r^*)$ the PMF at a reference unbound state. The latter is chosen in FM as a region of the phase space where the ligand is far enough from the protein (high values on the z axis of the funnel) to not feel any kind of influence from it. In detail, the reference state is defined as all the points in the surface delimited by the horizontal slice of the cylinder at the chosen unbound value of z. Thus, equation 2.26 must be corrected for a term πr_{cyl}^2 equal to the area of this section:

$$K_b = \pi r_{cyl}^2 \int_{bound} dr\, e^{-\beta(W(r)-W(r*))} \quad (2.27)$$

Then, inserting equation 2.27 into 2.25, we obtain the analytical formula for computing a correct and accurate estimate of ΔG_b^0 from FM calculations:

$$\Delta G_b^0 = \Delta G_b - \frac{1}{\beta} ln(C^0 \pi r_{cyl}^2) \quad (2.28)$$

where C^0 = 1/1660 Å^{-3} and ΔG_b is a summation over the free energy differences between all the bound poses in the simulation and the selected reference unbound state.

2.5. Multiple replicas simulations

The concept of "replica" is fundamental in molecular simulations. For replicas, we usually refer to multiple identically parameterized copies of the same system. Theoretically, according to the central limit theorem and ergodic hypothesis, two replicas simulated for infinite timescales should provide identical distributions. However, in real-life cases, this condition is rarely satisfied. Moreover, even for systems initialized from the same velocities, the resulting trajectories are slightly diverging due to bit errors, differences in floating-point precision on different machines, the number and type of processors, compiling options, system-specific random number generators, and/or dynamical load balancing.

Thus, replicas can be variously employed for different aims such as enlarging the statistics collected during the simulation, attempting the reproducibility of the results, or accelerating rare events. For instance, independent multiple replicas are often simulated starting from the same coordinates to provide different unbiased insights into the system under study. This approach is particularly convenient when High Performing Computer (HPC) infrastructure is available because the real time required for collecting the same amount of statistics is significantly reduced compared to a single longer trajectory. Some other times, replicas are not independent one from the other, but they share specific properties of the system. Two important examples could be the Multiple Walker (MW) approach in Metadynamics or the Replica-

Averaged Molecular Dynamics (RAMD) simulations. In MW, different copies of the system are simulated sharing the same MetaD bias. In such a way, the underlying FES is contemporary filled by the bias deposition of multiple agents (walkers), thus reducing the time needed to reach the convergence of a factor equal to the number of employed replicas. On the other hand, in RAMD, structural restraints are applied to make observables' averages (computed over the ensemble of configurations provided by replicas) match experimental measurements (2.5.1 for details).

Conversely, in parallel-tempering (PT), also known as replica-exchange, copies of the system are simulated at different temperatures. Periodic coordinates' exchanges are attempted between adjacent replicas. The acceptance or the rejection of a given exchange is usually determined by an energetic evaluation performed with the Metropolis-Hastings criterion. The general idea is that high-temperature replicas can sample large volumes of phase space, whereas low-temperature systems, whilst having precise sampling in a local region of phase space, may become trapped in local energy minima during the timescale of a typical computer simulation. PT thus enhances the sampling by allowing the systems at different (close) temperatures to exchange complete configurations. In fact, the inclusion of higher temperature systems ensures that the lower temperature replicas can access a representative set of physiological regions of phase space.

2.5.1 Replica-averaged Molecular Dynamics

Replica-averaged MD, also known as RAMD, is a valuable tool for incorporating experimental measurement as structural restraints in the framework of molecular dynamics.[64] In principle, a perfect molecular simulation, with an ideal force field and an infinite time, can quantitatively reproduce the results of experiments. Often, only a qualitative agreement is reachable due to the limitations of force fields and accessible time scales. A strategy to increase the precision of force fields in a system-dependent manner is to add to the Hamiltonian a term based on the agreement with known experimental data. Notably, applying the experimental data as a simple restraint would force the system to be always in agreement with the reference value. However, this condition is not faithfully reproducing what happens during experiments. In fact, all the equilibrium experimental data (e.g. NMR spectra) derive from measurement performed either over an ensemble of structures or over time. Thus, the RAMD solution to this problem is to apply the experimental restraint over a collective variable computed as the average of the given observable over multiple parallel simulations of the same system. In this way, every single replica does not need to be always in agreement with the experimental data, but these must be respected on average. Cavalli et al.[65] demonstrated that RAMD simulations can generate structural ensembles in

accordance with the maximum entropy principle providing, in turn, an accurate approximation of the unknown Boltzmann distribution of the system.

2.5.2. Parallel-tempering in the Well-tempered ensemble

Parallel-tempering in the well-tempered ensemble (PT-WTE) is a powerful enhanced sampling method derived by the combination of other two techniques: Metadynamics (MetaD) and Parallel Tempering (PT). When the potential energy is used as collective variable in WT-MetaD simulations, a well-defined distribution known as well-tempered ensemble (WTE) is sampled. This condition allows observing transitions between states that otherwise would have been impossible to study in standard MD conditions. Notably, in WTE the user does not need to define a system-specific set of CVs, but all the degrees of freedom of the systems are contemporary accelerated by the MetaD bias. Afterwards, once the simulation is converged, is also possible to reconstruct the FES as a function of some CV, taking advantage of the preferred reweighting scheme.[66,67] On the other hand, in PT n replicas of the system at the temperatures β_i, with i=1…n, are sampled and a MC procedure is used to attempt ex- changing configurations between replicas. Colder replicas are prevented from being trapped in local minima by the exchange with the higher-temperature ones. The quality of a PT simulation is related to

the ability of each replica to diffuse across the entire range of temperatures β. This is measured in terms of round-trip time t_γ, which is the time needed for a configuration in the coldest replica to reach the hottest temperature and come back, and exchange ratio between adjacent replicas, which is usually acceptable above 20%. To minimize t_γ and increase the exchange ratio, it is important to properly distribute the replicas over the chosen range of temperatures.[68] Often, this causes to employ a very big number of replicas and computational resources. Given the special properties of WTE, it is tempting to combine it with PT, giving birth to the PT-WTE formalism. This is achieved by simulating different replicas at growing temperatures like in PT, whilst a metadynamics potential is added to each replica on the potential energy of the system, to define the WTE environment. Since WTE causes enhanced energy fluctuations it can greatly facilitate the exchange processes in PT, drastically reducing the number of replicas required to span the desired temperature range and thus saving computational resources. The speedup provided by WTE to the PT simulation is regulated by the WT-MetaD tunable parameter γ, also known as bias factor:

$$\gamma = \frac{T + \Delta T}{T} \tag{2.29}$$

where T is the temperature of the system and $T + \Delta T$ is the fictitious temperature at which the CV is sampled. As already shown in equations 2.23 and 2.24, the parameter γ regulate the rate at which the gaussian heights are

re-scaled in a WT-MetaD simulation. It has been shown that in PT-WTE the energy fluctuations are enhanced by a factor equal to the square root of the bias factor, allowing for a direct rescaling of the number of replicas needed to reach the convergence.[59,60]

Chapter 3. Deepening into ligand binding

In this chapter, we will focus on two case studies highlighting the impact that an accurate investigation of the ligand-target interactions can have in pharmaceutical research. The first one is a typical computer-aided drug discovery project aimed at identifying new binders of the G-quadruplex structure in the *KRAS* proto-oncogene as potential anticancer agents. In the second case, the employment of advanced simulative methods allowed a thorough mechanistic description of the binding mechanism of small organic molecules to Formyl Peptide Receptors. These GPCRs have recently come to the limelight for their involvement in many physiological and pathological processes such as inflammation or cancer. Interestingly, FPR2 agonists with good selectivity over the FPR1 receptor are emerging as promising resources in the field of regenerative medicine. For this reason, the elucidation of the molecular requirements for binding and selectivity between these proteins represents an important milestone that will pave the way for new and more focused drug discovery campaigns.

3.1. Targeting the KRAS oncogene: novel DNA G-quadruplex binders

3.1.1. G-quadruplexes

DNA can arrange in multiple higher-order structures, playing important roles both in biological and pathological functions.[69–72] G-quadruplexes (G4) are non-canonical DNA conformations formed by guanine-rich sequences organized in tetrads stabilized by Hoogsteen hydrogen bonds and monovalent cations (Figure 3.1).[73] Since many key genome regions have been proved to possibly arrange into G4s, the interest in these structures as drug targets is constantly growing. For instance, recent studies reported how the expression of oncogenes' promoters such as *MYC*, *VEGF*, *BCL2*, *KIT*, and *KRAS* can be downregulated by G4-stabilizing molecules.[74–78]

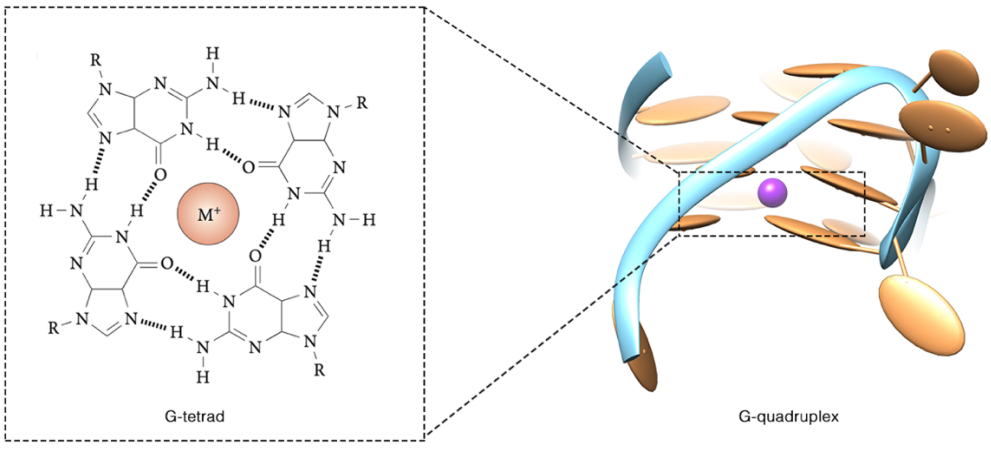


Figure 3.1 An illustration of the interactions in a G-quartet (left), and G-quadruplex (right). M^+ denotes a monovalent cation. The figure is adapted from ref. 73

3.1.2. *KRAS*

A significant percentage of all human cancers (about 30%), including pancreatic ductual adenocarcinoma and colorectal cancer, is characterized by the overexpression of the *KRAS* gene.[79–81] The latter encodes for a 21 kDa membrane GDP/GTPase (Uniprot: P01116) whose activation is directly dependent on GTP/GDP homeostasis. Missense mutations often occur at codons 12 and 13 of the gene, causing constitutive activation of KRAS and of its downstream cellular pathways (e.g. RAF, MEK, or PI3K).[82] Although the direct targeting of KRAS at protein level resulted unfruitful for a long time,[83–86] some promising results have been recently shown by inhibitors of the constitutively active G12C *KRAS* mutant.[87,88] In this context, a convenient alternative is represented by the downregulation of KRAS at a gene level. This can be achieved by the stabilization of the G4 conformation usually adopted by the guanine-reach sequence featuring the nuclease-hypersensitive element (NHE) within the *KRAS* P1 promoter. Over the years, different classes of compounds have been developed to this aim: porphyrins, acridines, anthraquinones, phenanthrolines, perylenes, and quinolines.[89–96] These molecules share similar chemical properties granting selectivity over the duplex DNA, such as polycyclic and heteroaromatic moieties that can strengthen the π-π stacking interactions with the guanine tetrads. However, more difficult is gaining selectivity over a large set of G4 structures. In fact,

most of these molecules are poorly selective, causing off-target effects, and lack basic drug-like properties.

3.1.3. Targeting *KRAS* G4

The main goal of this project was to discover novel selective stabilizers of the *KRAS* P1 promoter G4.[97] To this aim, we carried out a structure-based VS on the NMR conformation of this motif. The VS identified compounds were first evaluated through thermal melting experiments, then the best hits were submitted to a lead optimization program. Again, the newly synthesized compounds were tested for their stabilizing activity and selectivity toward both diverse G4 topologies and duplex-DNA. Then, the binding affinity, stoichiometry, and modalities to *KRAS* G4 of the most promising derivative were evaluated through an extensive biophysical and computational characterization. Finally, pharmacological analysis allowed gaining insight into the effects exerted by this compound on the *KRAS* expression and tumor cell viability.

3.1.4. Results

3.1.4.1 Virtual Screening

For our VS campaign, we used as target the *KRAS* G4 NMR structure (PDB code: 5I2V)[98], characterized by two wide stacking surfaces and four grooves of medium size, all representing potential drug binding sites. The ligand dataset employed for these calculations came from two different sources: i) the commercial Asinex Platinum Collection library and ii) an in-house chemical library including all the molecules developed over the years in our laboratories toward different biological targets such as kinases, integrins, chemokine receptors, and nucleic acids. This latter library indeed offers the advantage to employ compounds already available in stock or easily re-synthesizable. After having carefully prepared the two databases (3.1.6.1 for details), these were subjected to a filtering procedure to discard all the neutral and negatively charged compounds as well as all the molecules featuring less than two aromatic rings. These chemical features, indeed, are well-known molecular requirements to establish favorable contacts with the DNA phosphate backbone and with the G-tetrads stacking surfaces or possibly the G4 grooves, respectively.[99,100] The resulting subsets of ligands were then submitted to docking calculations. The top 15% of the ranked solutions from each database was selected to obtain a subset of molecules, all showing a

docking score < −7.0, feasible for the following examination. In fact, these compounds underwent a careful visual inspection in order to evaluate their predicted interaction mode with KRAS G4. In particular, the formation of π-stacking or polar interactions (i.e. salt bridges) with the target DNA was investigated. Finally, the selected compounds were inspected for good chemical binding geometry. Based on these criteria we chose 12 compounds (Chart 3.1): 4 were picked from the in-house database (1–4), while 8 were taken from the Asinex Platinum Collection library and then purchased from the vendor (5–12).

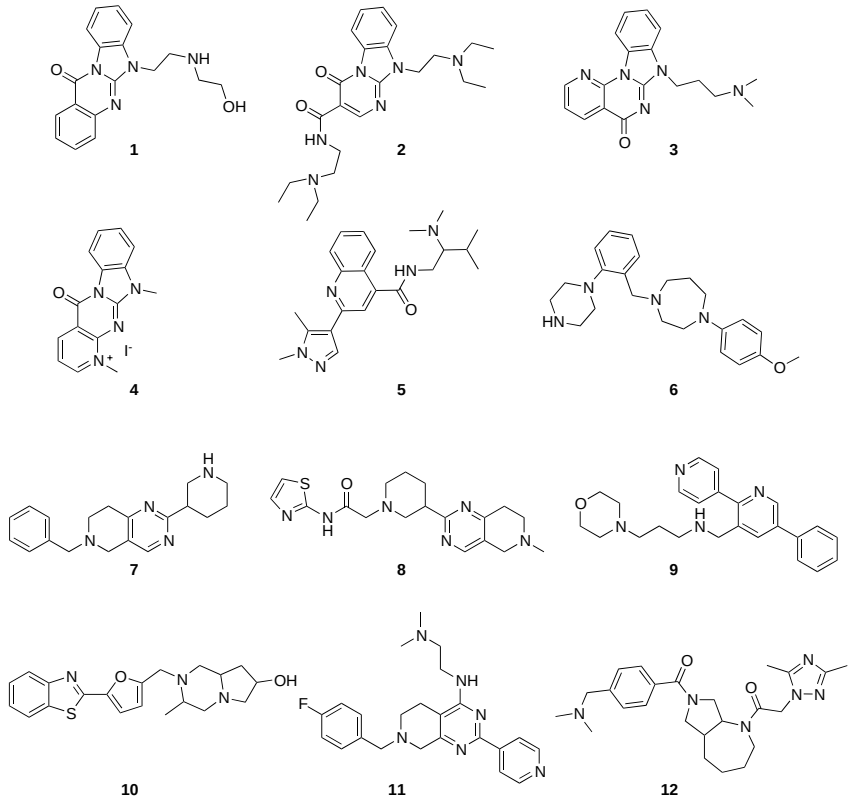


Chart. 3.1 Chemical structures of VS hits 1–4 from the in-house library and 5–12 from the Asinex Platinum Collection. Taken from Ref. 97

Interestingly, the docking poses of the best-ranked compounds from each library, **1** and **9**, suggested two alternative binding modes, either at the G4 stacking surface (Fig. 3.1A) or at the DNA grooves (Fig. 3.1B).

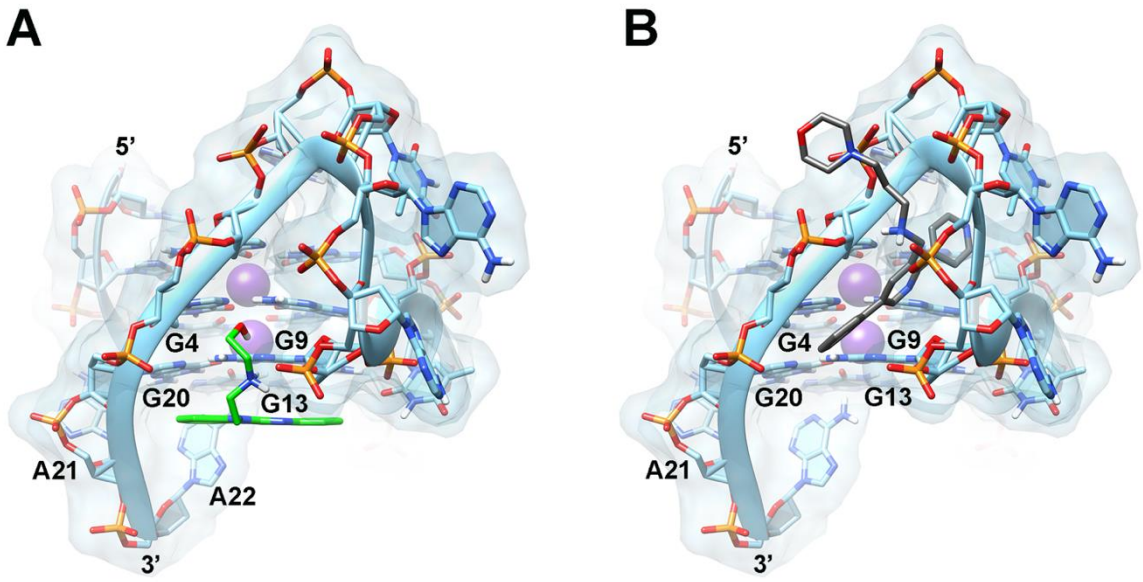


Figure 3.1 Docking-predicted poses of **1** (left panel, green sticks) and **9** (right panel, grey sticks) at the NMR structure of the *KRAS* G-quadruplex (PDB code: 5I2V). DNA is shown as cyan cartoons and transparent surface. Nucleotides are highlighted as sticks; aromatic rings are filled with thin slabs. The color code for the heteroatoms is: blue for nitrogen, red for oxygen. K^+ ions are depicted as purple spheres. Nonpolar hydrogens are omitted for clarity. Figure taken from Ref. 97

3.1.4.2. CD experiments

All the compounds were then tested in biophysical assays to verify their G4 binding properties. First, we checked the KRAS G4 conformation by CD spectroscopy, which provided a spectrum profile typical of parallel-stranded quadruplexes (a positive band at 264 nm and a negative band at 240 nm).[101] Then, we assessed the stabilizing effect of each of the selected hits (**1**–**12**) on *KRAS* G4 measuring the ΔT_m by CD thermal melting experiments at 264 nm (Fig. 3.2). Good results, although still not fully satisfactory, were shown by compounds **1** and **4** (ΔT_m = +4.5 °C and ΔT_m = +7.5 °C, respectively; Fig. 3.2). Notably, the overall folding of the *KRAS* proto-oncogene was not altered since no change was observed in the CD spectrum upon ligand binding. Thus, we decided to also evaluate the selectivity of **1** and **4** against the 20-mer hairpin duplex DNA. To this aim, we carried out other CD thermal melting experiments at 280 nm, which highlighted no duplex stabilizing effects for **1** and **4**, thus prompting their selection as hit compounds for chemical optimization.

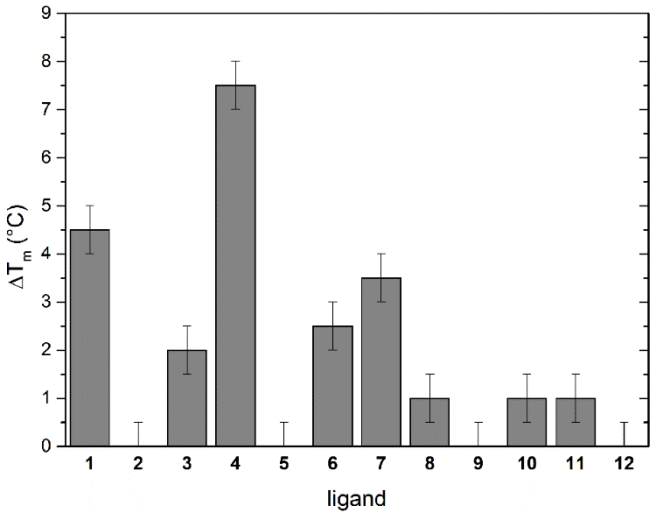


Figure 3.2 ΔT_m of KRAS G4 upon interaction with VS hits. Figure taken from Ref. 97.

3.1.4.3. Hit optimization

The aim of this optimization task was to improve the stabilizing effects of **1** and **4** on the *KRAS* G4. We investigated, in first place, the influence of side chains of different sizes and nature on the ligand binding and selectivity by selecting from our in-house library four analogues of **1** (compounds **13**-**16**, Chart 3.2), endowed with a terminal amino group instead of the hydroxyl function.

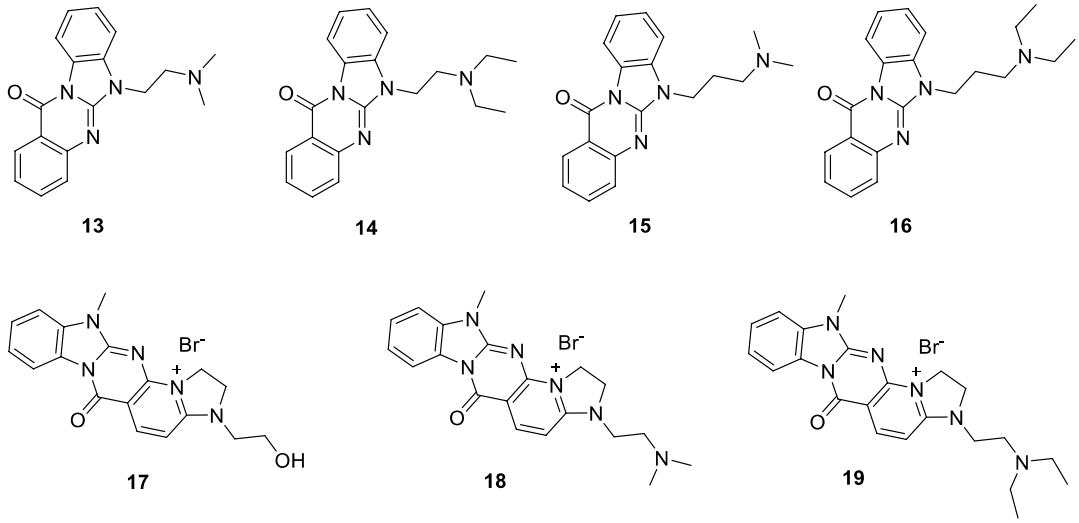
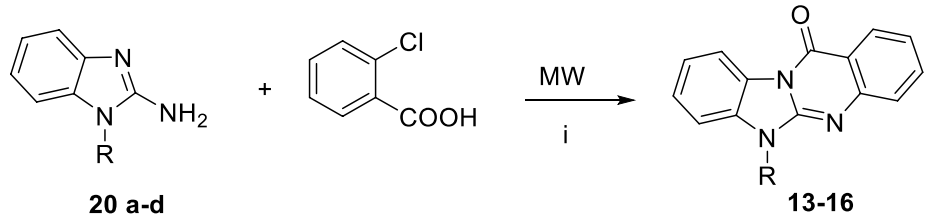


Chart. 3.2 Chemical structures of derivative **13-19**. Taken from Ref. 97.

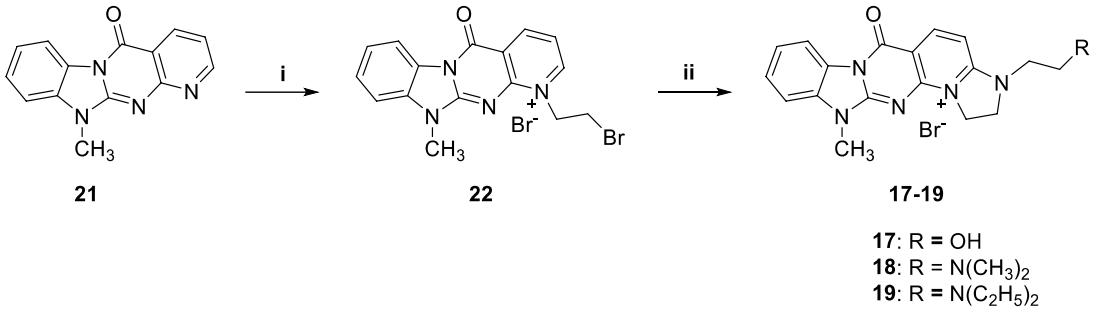
Notably, the original synthesis of **13–16**[102] has been here improved using microwaves (MW), with higher yields and shorter reaction time (12 min versus 5–7 h, Scheme 1).



20a, 13: R = $(CH_2)_2N(CH_3)_2$
20b, 14: R = $(CH_2)_2N(C_2H_5)_2$
20c, **15**: R = $(CH_2)_3N(CH_3)_2$
20d, **16:** R = $(CH_2)_3N(C_2H_5)_2$

Scheme 1. Synthesis of compounds **13-16**. Reagents and conditions: (i) K_2CO_3, CuBr, KI, DMF, 150 °C, P = 10 bar, power = 40-150 W, *ramp* t = 2 min, t = 12 min. Taken from Ref. 97.

In particular, an MW- assisted Ullmann condensation reaction between the appropriate 1- dialkylaminoalkyl-2-aminobenzimidazole **20a-d**[102] and the commercially available 2-chlorobenzoic acid (Scheme 1) was performed. After cooling, the reaction mixture was poured into ice and the precipitated crude 6-(aminoalkyl)benzo[4,5]imidazo[2,1-b]quina- zolin-12(6H)-ones **13–16** were collected and purified by recrystallization from ethanol. In the case of **4**, we decided to introduce differently functionalized protonable pendant side chains at the positively charged pyridine nitrogen since positively charged moieties are recognized to establish favourable contacts with the phosphate backbone of DNA.[100,103] In this vein, the synthetic procedure previously developed by us for the preparation of **4**[104] was applied starting from the 11-methylbenzo[4,5]imidazo[1,2-a]pyrido[2,3-d]pyrimidin-5(11H)-one,[104] with few modifications (Scheme 2).



Scheme 2. Synthesis of compounds **17-19**. Reagents and conditions: (i) 1,2-dibromoethane, 90 °C; (ii) $H_2N(CH_2)_2R$, ethanol, reflux. Taken from Ref. 97.

The reaction of compound **21** with an excess of 1,2-dibromoethane at 90°C for 40 h furnished **22**, which was suspended in ethanol and added with an excess of the appropriate amine. The reaction mixture was refluxed for 4 h and then filtered to give crude products, finally purified by recrystallization from ethanol. A deep analysis of the spectral data allowed us to unequivocally assign to the obtained compounds **17–19** the structures outlined in Scheme 2, featuring a fused pentacyclic system. Although the obtained derivatives **17–19** possessed a rather unexpected structure, they were considered for biological evaluation, as their structural features still met the requirements to act as G4 stabilizing molecules, that is a positively charged pendant chain and an extended aromatic system.

3.1.4.4. Biophysical Assays

We tested compounds **13–19** on *KRAS* G4 and duplex hairpin DNA, evaluating ΔT_m by CD thermal melting experiments (Fig. 3.3 – 3.4).

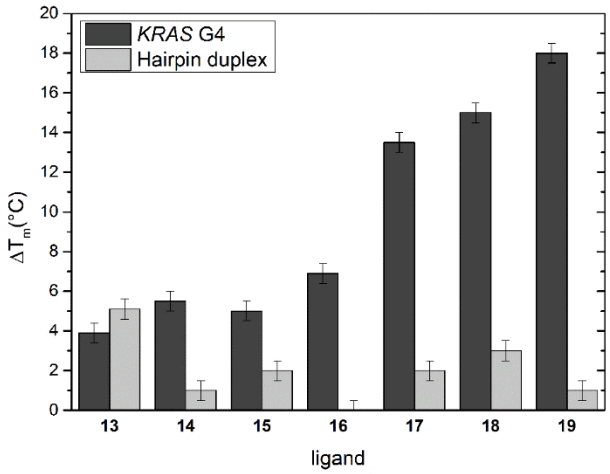


Figure 3.3 ΔT_m of KRAS G4 upon interaction **16-19.** Taken from Ref. 97.

Interestingly, comparing the **13-16** subset with the parent hit **1**, we noticed improved *KRAS* G4 stabilizing properties and good selectivity over the duplex DNA for all the compounds (Fig. 3.3). However, the most interesting results were observed for the pentacyclic derivatives **17–19,** which all showed markedly improved activity (ΔT_m > 10 °C) compared to their parent hit **4** (Fig. 3.2, ΔT_m = 7.5 °C) but still having a good *KRAS* G4/duplex DNA selectivity ratio (Fig. 3.3). In view of these data, **17–19** were further tested against other G4 topologies such as the parallel *KIT* G4 and the hybrid-1 Tel23. Although these ligands showed preferential binding to the parallel G4s (Table 3.1), all of them turned out to increase the *KRAS* G4 thermal stability to a larger extent than the other G4s.

Table 3.1. Comparison of **17**, **18,** and **19** ΔT_m with *KRAS* G4 and control sequences.

ligand	*KRAS* G4 ΔT_m ± 0.5 °C	*KIT* G4 ΔT_m ± 0.5 °C	Tel23 ΔT_m ± 0.5 °C
17	+ 13.5	+ 7.0	+ 2.5
18	+ 15.0	+ 10.0	+ 5.5
19	+ 18.0	+ 10.5	+ 7.5

In particular, **19** was the most promising compound of the series (Fig. 3.4) and was thereby selected for further tests.[105,106]

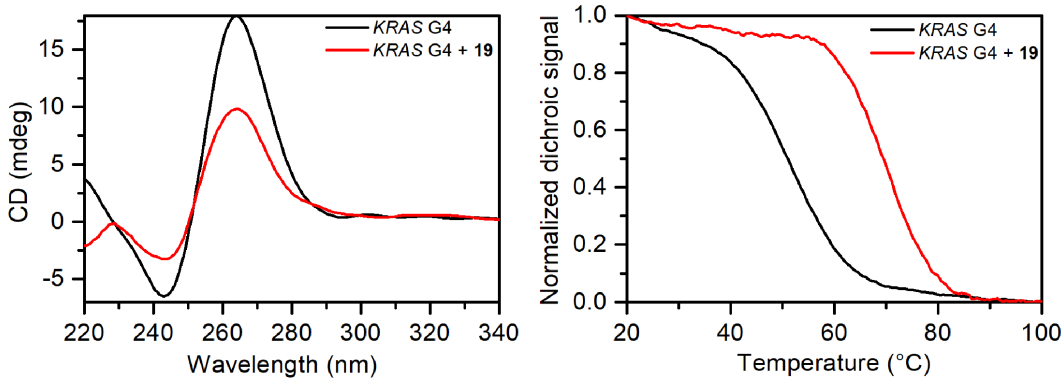


Figure 3.4 CD spectra (left) and thermal melting profiles (right) of *KRAS* G4 without and with (black/red line) **19**. Figure taken from Ref. 97.

Fluorescence analysis (Fig. 3.5) showed a 1:1 stoichiometry ratio in the *KRAS* G4/**19** complex with a binding constant (K_b) of 7.2 ± 0.4 * 10^6 M^{-1}. Analogous fluorescence experiments were also performed with *KIT* G4, revealing a 1:1 interaction in the *KIT* G4/**19** complex with K_b = 2.3 ± 0.5 10^6 M^{-1}. These data strengthen the CD results (Table 3.1) confirming the selectivity of **19** for *KRAS* with respect to *KIT*, where its binding constant K_b resulted threefold lower compared to that of *KRAS* G4. In parallel, we investigated the behaviour of the *KRAS* G4/**19** complex upon the addition of different amounts of **19** in a non-denaturing gel electrophoresis (PAGE) experiment. Here, *KRAS* G4 moves as a single band in the gel, thus suggesting the absence of high-order DNA structures. Notably, the addition of **19** to the *KRAS* G4 did not affect the G4 mobility at all investigated ratios, in agreement with the stoichiometry revealed by the fluorescence assay.

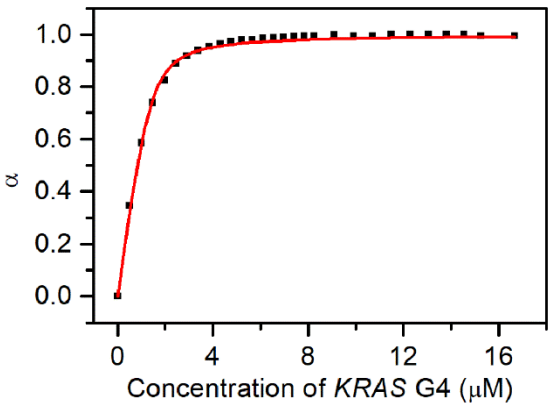


Figure. 3.5 Fluorescence titration of **19** with *KRAS* G4. Figure taken from Ref. 97.

3.1.4.4 Computational Studies

In order to investigate at an atomic level the binding mode of the newly synthesized derivatives to *KRAS* G4, we performed extensive molecular

modelling studies on the most promising compound of the series, **19**. First, molecular docking of this molecule was performed on the entire *KRAS* G4 NMR structure (PDB code: 5I2V). As result, two similar binding modes were predominately predicted, **A** and **B** (Fig. 3.6), in which the ligand is located at the 3′ region of the target DNA to interact with residues G9, G13, and A22. A relevant difference in the two poses is given by the arrangement of the ligand planar scaffold which is mutually flipped by 180°. Thus, the stability and the energetics of these poses were investigated through more accurate calculations. In particular, we submitted both the docking-predicted **19**/DNA complexes to 1.5 μs molecular dynamics (MD) simulations in explicit solvent to fully take into account the receptor flexibility and the water and ions effects.

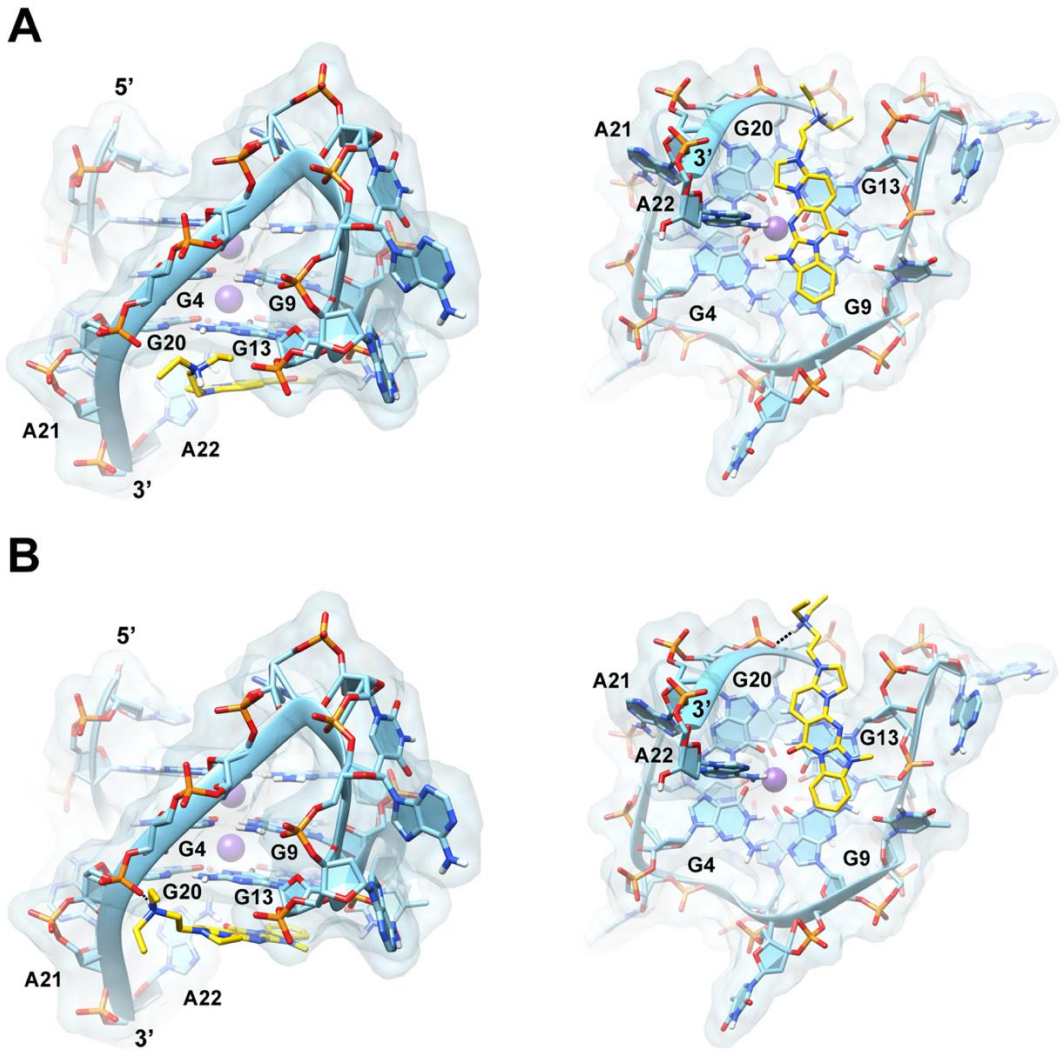


Figure 3.6. Front (left) and bottom view (right) of the two docking-predicted poses of **19** (yellow sticks) at the NMR structure of the *KRAS* G-quadruplex (PDB code: 5IV2). DNA is shown as cyan cartoons and transparent surface. Nucleotides are highlighted as sticks, aromatic rings are filled with thin slabs, while K^+ ions are depicted as purple spheres. Nonpolar hydrogens are omitted for clarity. Hydrogen bonds are shown as dashed black lines. Figure taken from Ref. 97.

Comparing the ligand RMSD plots over the two MD trajectories it is possible to notice how only one of the two docking solutions (pose A) evolves in a stable binding conformation over the simulated time scale (Fig. 3.7). In fact, starting from this pose, in the first half of the MD run the ligand experiences a

slight rearrangement to reach a well-defined binding mode, which is conserved throughout the rest of the simulation.

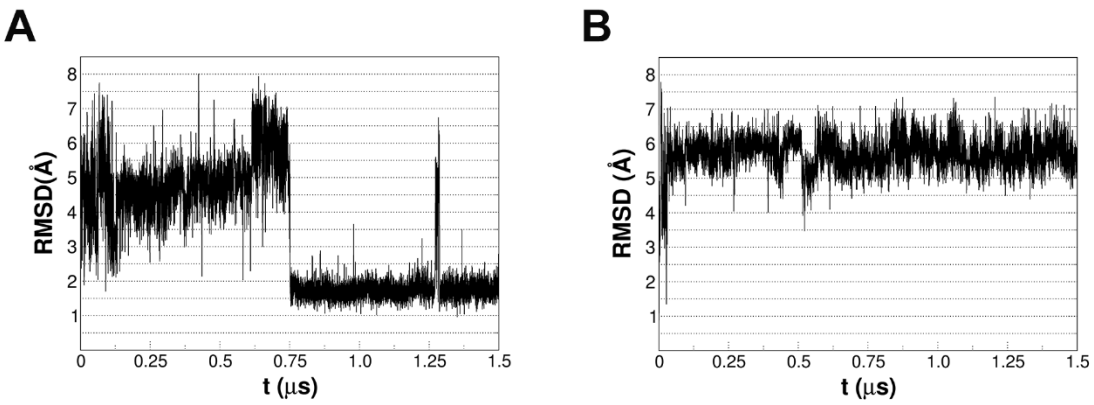


Figure 3.7. RMSD plots of the polycyclic scaffold heavy atoms of **19** along the MD simulations on docking poses A and B. Prior to RMSD calculations, trajectory frames were aligned on the DNA guanine stacks heavy atoms. Figure taken from Ref. 97.

The good stability of this binding mode is due to tight interactions with the DNA. Specifically, **19** lies at the G4 3′ end where it engages favorable π-stacking with G9, G13, and G20 (Fig. 3.8).

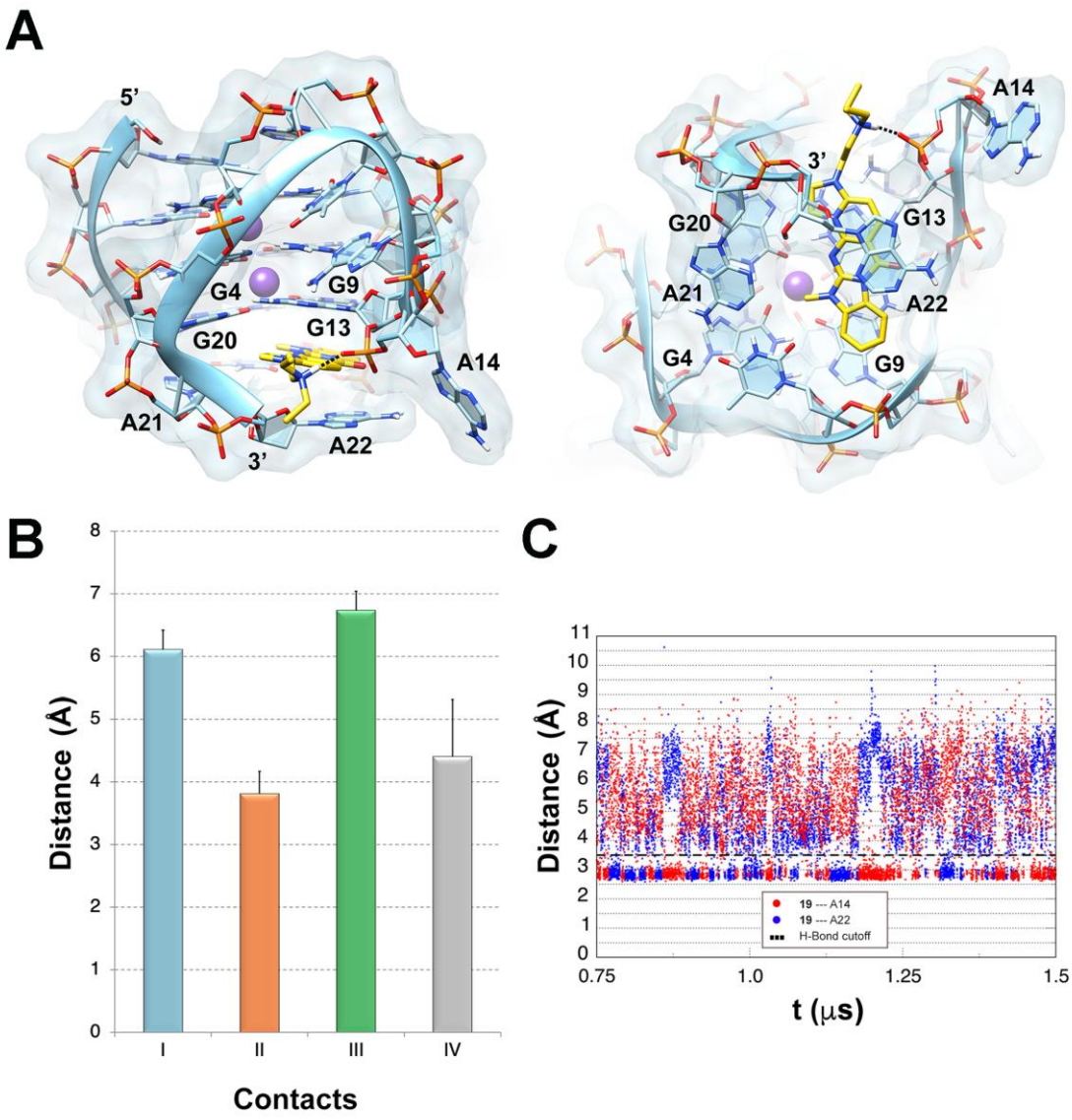


Figure. 3.8 (**A**) Front (*left panel*) and bottom (*right panel*) view of the MD predicted binding pose of **19** (yellow sticks) at the NMR structure of the *KRAS* G4 (PDB code: 5IV2[36]). DNA is shown as cyan cartoons and transparent surface. Nucleotides are highlighted as sticks, aromatic rings are filled with thin slabs, while K^+ ions are depicted as purple spheres. Hydrogen bonds are represented as dashed black lines. (**B**) Interatomic distances (mean ± S.D.) representative of the ligand (polycyclic nucleus - center of mass)/DNA residues (aromatic ring centroid) stacking interactions along the second half of the MD calculations on docking pose A: (I) G9, cyan bar; (II) G13, orange bar; (III) G20, green bar; (IV) A22, gray bar. (**C**) Distance between the ligand diethylamino group (N) and representative A14 and A22 phosphate oxygens along the MD simulations. Figure taken from Ref. 97.

Interestingly, a salt bridge is observed between the ligand's *N,N*-(diethyl)aminoethyl alternatively with the DNA phosphate backbone of either A14 (Fig. 3.8) or A22 (Fig. 3.8C). At this regard, we report that in MD calculations on pose B this kind of ligand-DNA interaction is conversely much less conserved along the whole simulation time (Fig. 3.9).

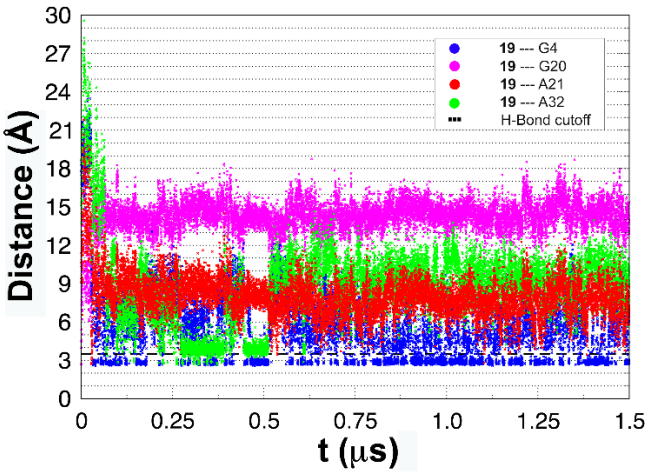


Figure 3.9 Distance between the ligand diethylamino group (N) and representative G4, G20, A21 and A32 phosphate oxygens along the MD simulations. Figure taken from Ref. 97.

It is also important to note that **19** can form additional stacking interactions with the terminal A22, which shifts from its initial position to be packed against the ligand aromatic scaffold. Remarkably, these contacts stabilize not only the ligand binding conformation but also the overall KRAS G4 architecture. This effect can be appreciated by looking at Fig. 3.10 where the RMSD plot of the G4 backbone along the MD trajectory is shown.

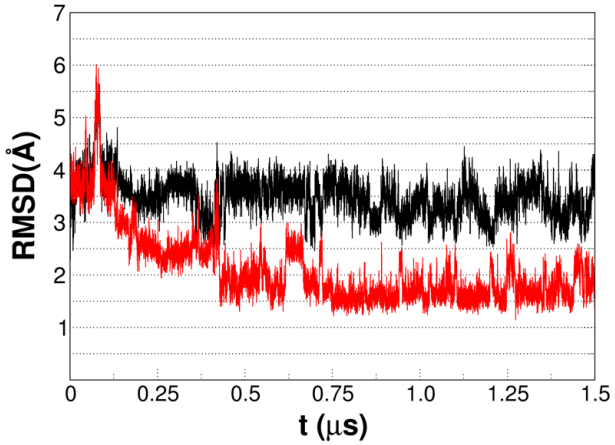


Figure 3.10. Standard (black) and averaged (red) RMSD plots of the *KRAS* G4 backbone heavy atoms along the MD simulation. Trajectory frames were aligned on the DNA guanine stacks heavy atoms. Figure taken from Ref. 97.

3.1.4.6. Biological Experiments

Finally, we estimated the biological impact of our research by testing the capability of **19** to reduce the *KRAS* expression in tumour cells. First, we treated HCT116 colorectal cancer cells with **19** at 2 μM concentration (IC_{50} of the compound computed by viability assay). Notably, the effect on *KRAS* expression was evaluated at both gene and protein levels. QRT-PCR analysis (Fig. 3.11A) revealed that treatment of HCT116 colorectal cancer cells with **19** (2 μM for 24, 48 or 72 h) reduced the mRNA levels of KRAS up to 40%, when compared to their untreated counterpart. Subsequently, the effect of **19** on KRAS was validated also in terms of protein expression. As evidenced in Fig. 3.11B, treatment of the cells for 72 h with 2 μM of 19 determined the reduction of *KRAS* levels of about 30%. Remarkably, in contrast to other G4 ligands,

treatment of cells with 2 μM of **19** for 24 h did not produce any increase in the phosphorylation levels of histone H2AX (γH2AX, Fig. 3.11B), a hallmark of DNA double-strand breaks, reinforcing the idea that the selected compound can be selective for the G4 structure present at the gene promoter level.

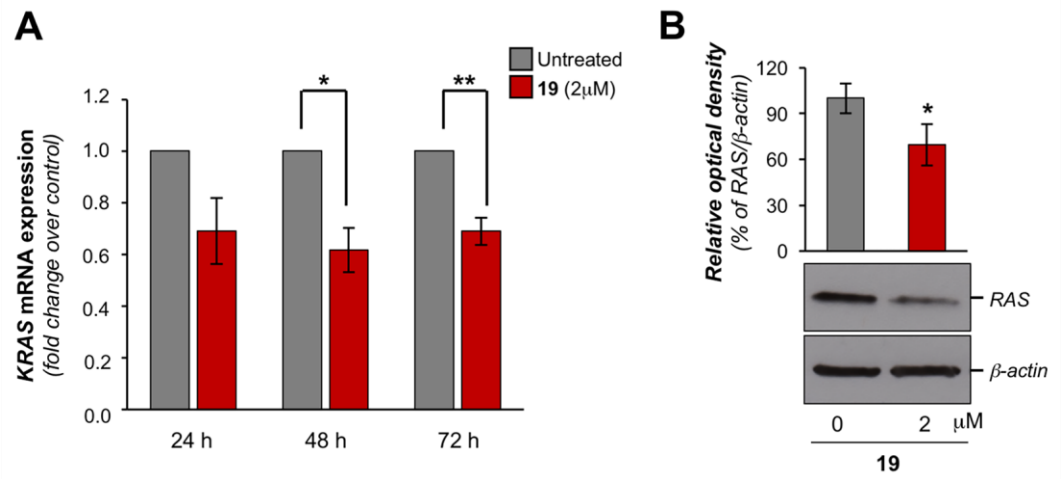


Fig. 3.11. (A) Gene expression of *KRAS* was evaluated by qPCR in HCT116 cells untreated or treated with 2 μM of **19** for the indicated times (24, 48 and 72 h). Results are expressed as fold change of mRNA levels in treated cells over their controls, after β-actin normalization. Histograms are shown as mean ± S.D. (*p ≤ 0.05, **p ≤ 0.01; Student's t-test). (B) Protein expression was evaluated by Western Blot (WB) analysis. *Upper panel*, histogram showing the relative optical density of KRAS expression evaluated by Image-J quantification tool and normalized for β-actin. The graph shows the mean ± S.D. (*p < 0.05; Student's t-test). *Lower panel*, representative WB images of KRAS, γH2AX and β-actin. Figure taken from Ref. 97.

Since KRAS is considered a driver oncogene, we also evaluated whether **19** was able to affect the viability of two *KRAS* isogenic tumour cell lines HK2-6 and HKE-3, derived from the HCT116 cells and carrying the mutated (KRASG13D/G13D) or wild-type (KRASwt/wt) gene alleles, respectively. Briefly, the two cell lines were treated with different concentrations of **19** (ranging from 1 to 10 μM) and the number of viable cells was evaluated by crystal violet assay (Fig. 3.12). Remarkably, these experiments clearly evidenced that the cytotoxic effect of **19** on the HKE-3 cells (IC50 > 10 μM) was lower than that produced on HK2-6 cells (IC50 = 2.78 μM). Since the oncogenic potential of KRAS increases in the presence of hotspot mutations,[84] it is possible to conclude that the efficacy of **19** grows together with the activity of KRAS as a driver oncogene. These data indicate that **19** would represent the prototype of a new class of compounds that, inhibiting the expression of the *KRAS* gene, might in principle counteract *KRAS*-mutated tumors that are refractory to treatment with anti-EGFR

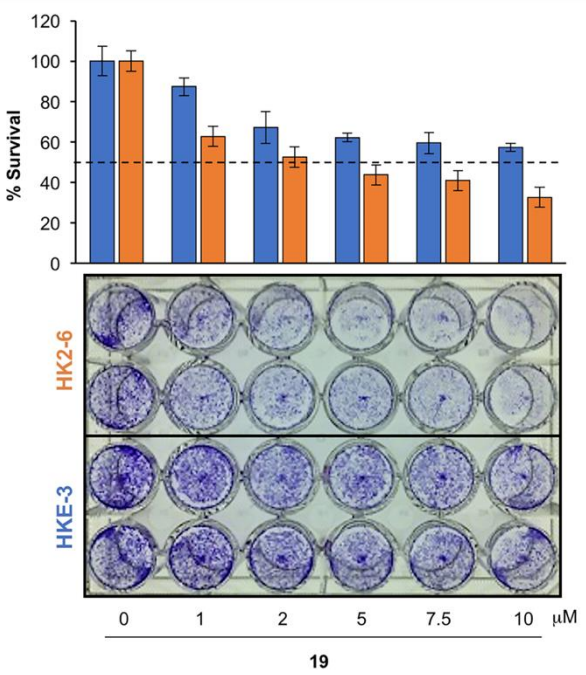


Figure 3.12 HK2-6 and HKE-3 cells were treated with compound **19** at the indicated doses for 72 h. Viable cell number was determined by colorimetric crystal violet assay. Histogram shows the mean

antibodies (e.g. cetuximab and panitumumab), so far devoid of valid therapeutic treatments.[107]

3.1.5. Conclusions

The stabilization of G4 DNA motifs in oncogene promoters by small molecules has emerged in the last years as a promising strategy to control aberrant protein expression in cancer cells.[72] Among the druggable oncogenes is KRAS which codifies for the homonymous protein and is mutated and overexpressed in a high percentage of tumors.[82,108] Here a VS campaign led to the discovery of new chemotypes able to recognize and stabilize a G4 from the KRAS P1 promoter. The chemical optimization of the identified molecules resulted in a set of derivatives that were extensively characterized for their G4 stabilizing properties. Particularly one of these analogues, namely **19**, showed a high affinity for the KRAS G4 with remarkable selectivity against duplex DNA. Subsequent fluorescence titration experiments on the *KRAS* G4/**19** complex showed a 1:1 stoichiometry ratio and a ligand/DNA binding constant of about 7.2 10^6 M^{-1}. Molecular dynamics simulations provided not only structural insights into the binding mode of **19** to KRAS G4 but also the molecular bases for the stabilization of the target DNA by this compound. Finally, biological assays demonstrated that **19** exerts cytotoxic effects, at low micromolar concentration, in tumour cells expressing constitutively active

forms of mutated KRAS. In conclusion, our data indicate that **19** can represent the prototype of a novel class of antitumoral drugs able to inhibit the expression of the KRAS driver oncogene, which has long been considered undruggable. In this perspective, the development of a new class of anti-KRAS molecules would represent an important curative opportunity for a subclass of patients for whom effective therapies are still missing.

3.1.6. Methods

3.1.6.1 Virtual Screening and Molecular Docking

For our study, an in-house virtual database of 5,858 compounds and the commercially available Asinex Platinum Collection library (http://www.asinex.com) of 9,216 molecules were selected. All the possible tautomeric and protonation states in the pH range 7.4 ± 1.5 were generated for each compound using Epik[109,110] for a total amount of 7,924 and 17,500 structures, respectively, for the in-house and the Asinex libraries. A filtering procedure was then applied to provide a more focused set of ligands. Specifically, all the molecules having at least two aromatic rings and a positive total charge were retained, resulting in two final subsets containing, respectively, 875 and 14,280 structures. Concerning the target macromolecule,

we selected the NMR structure of the KRAS G4 structure (PDB code: 5IV2)[98] that was prepared through the Protein Preparation Wizard implemented in Maestro Suite 2019.[111] During the preparation, all water molecules were deleted, hydrogen atoms were added, and the complex was minimized. The docking search area was set on the centre of mass of the macromolecule to enclose the entire G4. The interaction grids were thus computed through the grid generation tool of Glide 6.7.[21,22] The OPLS 2005 force field was employed for docking. The best 500 initial poses per ligand were retained for post-docking energy minimization. Otherwise, default parameters were applied. The results from each set of calculations were evaluated and ranked based on the Glide SP scoring function.[21,22] Thus, the top-ranked compounds (the best 15 %) of each subset were visually inspected for their binding modes and for good chemical geometry. Compounds **1**-**4** were retrieved from our in-house library, while compounds **5**-**12** were purchased from the Asinex vendor. Compounds purity (> 95 %) was determined by HPLC, according to the procedure described in the chemistry section. Docking simulations on compound **19** were performed applying the same protocol followed in VS calculations. Prior to docking, the ligand three-dimensional structure was generated with the Maestro Build Panel, and its tautomeric and protonation states at physiological pH (7.4 ± 1.5) were then predicted by Epik.[109,110]

3.1.6.2. Molecular Dynamics

The docking predicted **19**/*KRAS* G4 complexes were solvated in a 12.0 Å layer cubic water box using the TIP3P water model parameter.[112] 19 K^+ cations were used to neutralize each system, with two of these ions placed at the centre of the G-tetrads. Further 2 K^+ and 2 Cl^- ions were added to reach the standard 150 mM KCl concentration. The *parmbsc1*[113] and *gaff*[114] Amber force fields were used to parameterize the nucleic acid and ligand, respectively. Amber charges were applied to the DNA and water molecules, whereas ligand charges were computed using the restrained electrostatic potential (RESP) fitting procedure.[115] The ESP was first calculated by means of the Gaussian package [116] using a 6–31G* basis set at B3LYP level of theory, and then the RESP charges were obtained by a two-stage fitting procedure using Antechamber [117]. The NAMD 2.13 [118] code was used to perform the simulations. A cutoff of 10 Å was used for short-range interactions. The long-range electrostatic interactions were computed by means of the particle mesh Ewald method using a 1.0 Å grid spacing in periodic boundary conditions. Each system was minimized and heated up to 300 K while putting harmonic constraints, which were gradually released along the thermalization process. Then, production runs were performed in the NPT ensemble at 1 atm and 300 K.

3.2. Structural insights into ligand binding to FPR receptors

3.2.1 G protein coupled receptors (GPCRs)

The largest family in the human membrane proteome is represented by G-protein coupled receptors (GPCRs), which are codified by more than 800 genes. These receptors are usually classified into five main clusters based on sequence homology and phylogenetic analyses: rhodopsin family (class A), secretin family (class B), glutamate family (class C), frizzled family (class F), and adhesion family.[119,120] GPCRs are membrane integral proteins made of seven transmembrane (TM) α-helices domains linked by 3 extracellular (ECL) and 3 intracellular (ICL) loops. The N-terminus part of the receptor points towards the extracellular matrix, whereas the C-terminus extends in the intracellular side and can be folded into an additional helix (Helix 8), especially in the Class A subgroup.[121] The 7 TMs are arranged in a barrel shape, forming a cavity that spans the entire plasma membrane. Interestingly, most of the physiological GPCRs binders are hosted in a pocket located at the outer part of this TM bundle, between a high conserved amino acid $W^{6.48}$ (Wenstein-Ballesteros numeration used) and the ECL2, known as orthosteric binding site.[122] One of the most intriguing aspects of GPCRs physiology is the

multitude of biological actions which they can mediate in response to several stimuli induced by different signalling molecules such as peptides, hormones, and neurotransmitters.[123] Notably, signals transduction requires the GPCR coupling to heterotrimeric associated G proteins and the following release of second messengers. These can, in turn, modulate the activity of several enzymes controlling a plethora of cellular functions. In addition, the GPCRs signalling can be mediated also by second effectors alternative to G proteins, such as kinases (GPCR kinases, or GRKs) and arrestins (i.e. β-arrestin).[124] Recently, it was observed that some GPCRs are endowed with basal activity, and thus can activate their downstream pathways also in absence of agonists. Such discovery suggested that ligands can exert their pharmacological functions by changing the receptor states distribution - from inactive to active state - in relation to their intrinsic activity. Indeed, it is supposed that agonists increase the percentage of receptors in activated states, inverse agonists proportionally stabilize receptors in the inactive conformations, while antagonist competitively inhibit agonists without affecting the conformational equilibrium. In addition, ligands targeting the same GPCR can also elicit alternative cellular signals; this phenomenon is known as biased agonism.[125] In fact, different rearrangements in the GPCR structure can be observed upon binding of distinct ligands. This can lead either to the recruitment of diverse effector proteins (e.g. G proteins vs Arrestin), or to an alteration in the

GTP/GDP exchange rate of the G protein, which can trigger the activation of diverse downstream pathways.[125] All of these aspects make GPCRs one of the most fascinating and investigated molecular targets in biology and pharmacology so that 30% of the FDA approved drugs act on a GPCR (41 drugs approved in the last 5 years).[126]

3.2.2 Formyl Peptide Receptors (FPRs)

The Formyl Peptide Receptors (FPR) is a family of chemoattractant Class A GPCRs clustered on chromosome 19q13.3 and codifying for three different proteins: FPR1, FPR2 and FPR3.[127] While FPR3 is only found in monocytes, FPR1 and FPR2 are expressed also in neutrophils, with FPR2 showing a wider distribution pattern including astrocytoma, epithelial, and malignant glioma cells. Although these receptors share an overall high sequence identity and overlapping functions, they differ in ligand recognition and activation mechanisms. Since their first discovery, FPRs have been described as fundamental players in innate immune response through recognition of pathogen-associated molecular patterns (PAMPs) and damage-associated molecular patterns (DAMPs). In fact, their name derives from the ability to bind highly conserved N-formyl methionine-containing protein and peptide sequences of bacterial and mitochondrial origin.[128] Later, the identification of

a wide variety of FPRs endogenous ligands clarified that the biologic function of these receptors extends well beyond the regulation of leukocyte trafficking. In fact, while the physiological role of FPR3 is still quite unknown, FPR1 and FPR2 were found to exert important regulatory effects in a broad range of both physiological and pathological conditions, including atherosclerosis, inflammation, neo-angiogenesis, chronic obstructive pulmonary disease and cancer.[129] Particularly, FPR1 can initiate inflammatory responses such as neutrophil chemotaxis, degranulation, respiratory burst, and cytokine release when activated with ligands such as N-formyl-methionyl-leucyl-phenylalanine (fMLF) produced by bacteria. On the other hand, ambiguous effects on inflammation are observed upon binding with a glucocorticoid-related protein, namely Annexin A1 (AnxA1). While at high concentration AnxA1 induces FPR1-mediated pro-inflammatory effects, low levels of this protein only partially activate the receptor, leading to desensitization of neutrophils; also, in such condition, the migration of these cells triggered by other chemoattractant agents is reduced.

An even more intricate physiological and pharmacological puzzle is offered by FPR2. The latter is the most promiscuous receptor of its family, able to interact with a wide variety of structurally diverse ligands including proteins, peptides and lipids.[130] In agreement with a typical GPCR biased signalling, opposite downstream effects are reported in literature upon binding of FPR2

with distinct modulators, due to ligand-specific receptor conformational changes.[131] For instance, while the interaction of FPR2 with N-formylated peptides or with the Serum Amyloid A (SAA) can stimulate the massive production of proinflammatory cytokines,[132–134] marked anti-inflammatory effects are induced by Annexin-A1 (protein), Lipoxin A4 (LxA4) and resolvin D2 (RvD2).[135] Notably, Lipoxin A4 and resolvin D2 are both endogenous lipidic molecules belonging to the class of specialized pro-resolving mediators (SPM) that are collecting growing attention for their potential applications in the treatment of chronic inflammatory diseases.[135,136] The biological versatility of FPR2 is in fact reflected by the range of its possible downstream pathways. FPR2 usually couples to an inhibitory G protein, thus increasing IP3 levels and, in turn, the intracellular calcium fluxes together with the involvement of MAPKs. However, recent shreds of evidence have reported that FPR2 can engage distinct G proteins, hence activating different signalling cascades, presumably in a target cell-specific manner.[137] Congruently, it was observed that the administration of aspirin-triggered resolvin D1 can induce FPR2-dependent cAMP activation, presumably following engagement with a G protein containing a G_s alpha subunit.[138]

3.2.3. Formyl Peptide Receptors Ligands

In recent times, due to the severe drawbacks (i.e. strong immunosuppressive effects) related to the use of canonical anti-inflammatory drugs (i.e. glucocorticoids) in chronic inflammatory diseases, growing attention is being paid to SPMs receptors with the aim to develop new pharmaceutically affordable pro-resolving agents. In this context, a number of peptides and organic ligands targeting FPR1 and especially FPR2 have been discovered in the past few years. The first peptides were developed based on the structure of the endogenous anti-inflammatory protein AnxA1. Particularly, the N-terminal sequence of AnxA1 was used as a template to design AC2-26, a compound able to activate and desensitize all the three FPR family members at a similar concentration in HEK-293 cells.[139] The limited potency of AC2−26 on FPR2, combined with a lack of selectivity against FPR1, spurred the development of alternative AnxA1-derived peptides such as AnxA12−50, which showed improved potency for FPR2 and suitable anti-inflammatory and pro-resolving effects in various animal models.[140]

Diversely, the synthetic compound WKYMVM-NH_2 was initially identified through a peptide library screening in which the formation of inositol phosphate was evaluated.[141] Then, the replacement of the C-terminal methionine with a D-amino acid yielded WKYMVm-NH_2 (*Wpep*, **1** – Chart 3.2) that is highly active on both FPR1 and FPR2. Particularly, the compound

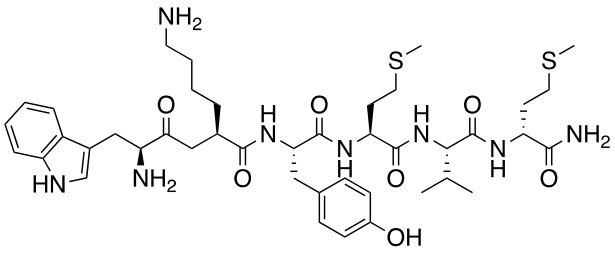
showed promising results in biological tests such as, for example, the inositol phosphate hydrolysis and the cAMP accumulation assays.[141,142] For this reason, **1** is widely employed as the reference compound during the biochemical assessment of potential drug candidates modulating the downstream pathways of FPR1 and FPR2.

In parallel, many efforts have been paid to develop also FPRs-oriented small molecules with variable intrinsic activity.[134,143,144] Representative agonists include pyrazolone ureas (*compound 43*, **2** - Chart 3.2),[145,146] benzimidazoles,[147] bridged spiro[2.4]heptanes,[148] and amino-triazoles (*ACT-389949*, **3** - Chart 3.2)[149]. Among these, a particular mention is deserved by compounds **2** and **3**. The first is a dual FPR2/FPR1 ligand which has been widely evaluated in *in vivo* studies for the preservation of cardiac function and prevention of adverse remodeling in rodent heart failure models.[146,150] Interestingly, this study opened a new scenario regarding the therapeutic applications of FPR1/FPR2 modulators, focusing on their employment in the regenerative medicine field. Compound **3**, instead, has reached phase 1 clinical trials, showing positive changes in the levels of plasma biomarkers (e.g., IL-10 and leukocyte levels) in humans after one dose; however, tachyphylaxis was observed in the multiday dosing study. Notably, due to the high sequence homology of FPR1 and FPR2, most of these ligands are active towards both receptors. However, the lack of selectivity hampers the complete evaluation of both the function and the

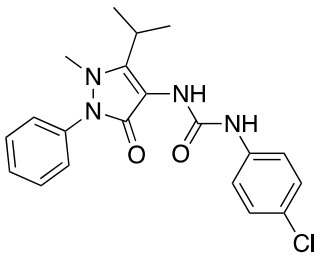
clinical potentialities of FPRs receptors. Thus, the development of isoform-selective chemical tools remains an important open point in the field.

In this scenario, the most promising results are being provided by *BMS-986235* (also known as *LAR-1219*, **4** - Chart 3.2) a potent, selective, and orally bioavailable agonist of FPR2, currently under phase 1 clinical evaluation.[151] Initial studies with **4** higlighted its potential to improve cardiac structure when given post myocardial infarction (MI) in the mouse. Recently, it was further proven that **4** can stimulate the resolutory activities of macrophages, induce neutrophil apoptosis and clearance, and preserve cardiac function post MI.[152]

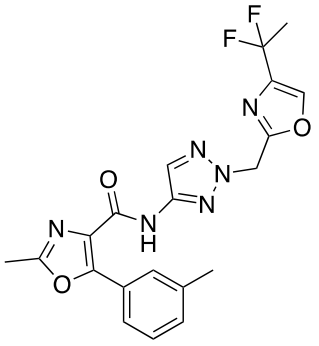
From a drug-design perspective, it is important to underline that no information about the binding site, the binding mode and binding mechanism of the known FPRs-targeting small molecules is yet available. In fact, only the X-ray and Cryo-EM complexes of FPR2 bound to the hexapeptide **1** have been solved so far.[153,154] With this in mind, different computational approaches have been here employed to fully characterize the binding mechanism, including binding modes and thermodynamic properties, of the most potent and selective known FPR2 organic agonist: *BMS-986235* (**4,** Chart 3.2). In fact, the rationalization of the molecular requirements for developing potent and selective FPRs small molecule ligands might have a strong impact on future FPR1/2-directed drug discovery campaigns.



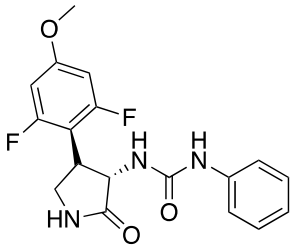
WKYMVm ,"Wpep" - **1**



Compound 43 - **2**



ACT-389949 - **3**



BMS-986235 - **4**

Chart 3.2 Structures of representative FPR2 ligands

3.2.4. Results

3.2.4.1 Ligand binding assay

In order to elucidate the binding mechanism of **4**, we first tried to determine whether the ligand acts either as a typical GPCR orthosteric binder or as an allosteric agent. Thus, we performed an *in cell* competitive binding assay between **4** and the well-characterized orthosteric FPR2 agonist **1**. Following the example of Chen et al.,[153] we incubated U937 monoblastic cells with fluorescein isothiocyanate (FITC)-conjugated-**1**. Then, we measured by flow

cytometry the changes in the mean fluorescence intensity of the system after treatment with increasing concentrations of **4**.

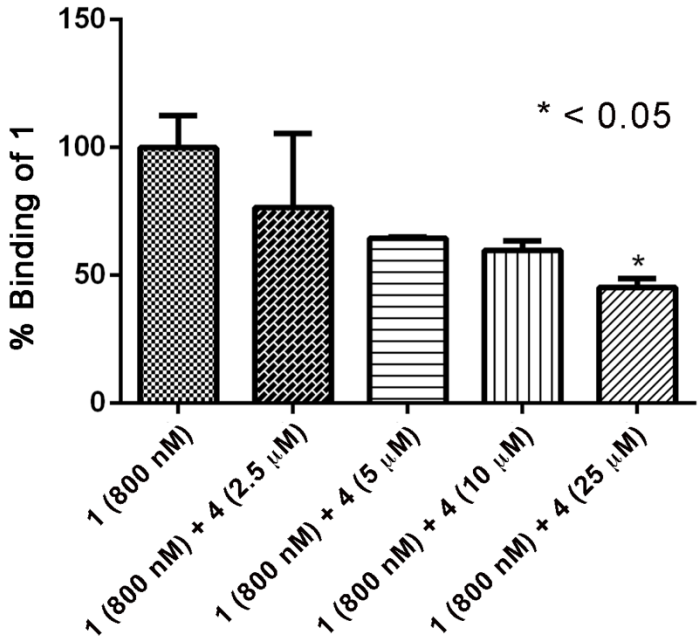


Figure 3.13 Competitive binding assay between compounds **1** and **4**.

The preliminary data in Figure 3.13 clearly show that the fluorescence signal due to the binding of FITC-**1** to FPR2 linearly decreased with growing concentrations of **4**. This demonstrates that a competition mechanism occurs between the two ligands; hence, that **4** binds at the orthostheric binding of the receptor. Notably, **1** was here tested at higher concentrations than those reported by Chen et al. due to the different FPR2 expression levels on the cell membrane. In fact, the U937 cells physiologically express moderate amounts of the receptor, while the HEK293F cells employed in ref.[153] were transfected with the FPR2 plasmid in order to overexpress the receptor.

3.2.4.2. Binding mode studies

To disclose at the atomistic level the binding path of **4**, a set of different computational approaches with increasing levels of accuracy were implemented. First, molecular docking and standard MD were employed to obtain an initial guess of the possible poses of **4** at the FPR2 orthosteric binding site. Then, taking advantage of the information collected with these calculations, we set up a Funnel Metadynamics (FM) simulation to characterize the free energy landscape of the entire binding event.

The highest-resolution (2.8 Å) x-Ray structure of FPR2 in complex with the reference compound **1** (PDB code: 6LW5)[153] was selected as the receptor's 3D model. A large search area enclosing the entire GPCR orthosteric binding site was set for docking simulations with two different docking software, namely Glide and Autodock. In spite of the small size of compound **4** compared to the co-crystalized ligand, calculations converged towards a single predominant binding pose (Figure 3.14). In this pose, the ligand is buried in the bottom of the orthosteric binding site, where it interacts with two hydrophobic clusters of residues. In detail, its 2,6-difluoro-4-methoxyphenyl group is accommodated in a groove identified by the side chains of L33$^{1.36}$, L81$^{2.60}$, M85$^{2.64}$, V105$^{3.32}$, F257$^{6.51}$ and F292$^{7.43}$, while its unsubstituted phenyl ring points toward the inner part of the TM bundle where it contacts L109$^{3.36}$, F110$^{3.37}$ and W254$^{6.48}$, a well-known key regulatory residue for Class A GPCRs activation.[155]

In fact, the side chain of this residue usually adopts different conformations in agonist- and antagonist-bound GPCR structures. The ligand's binding mode is further stabilized by two sets of polar interactions.

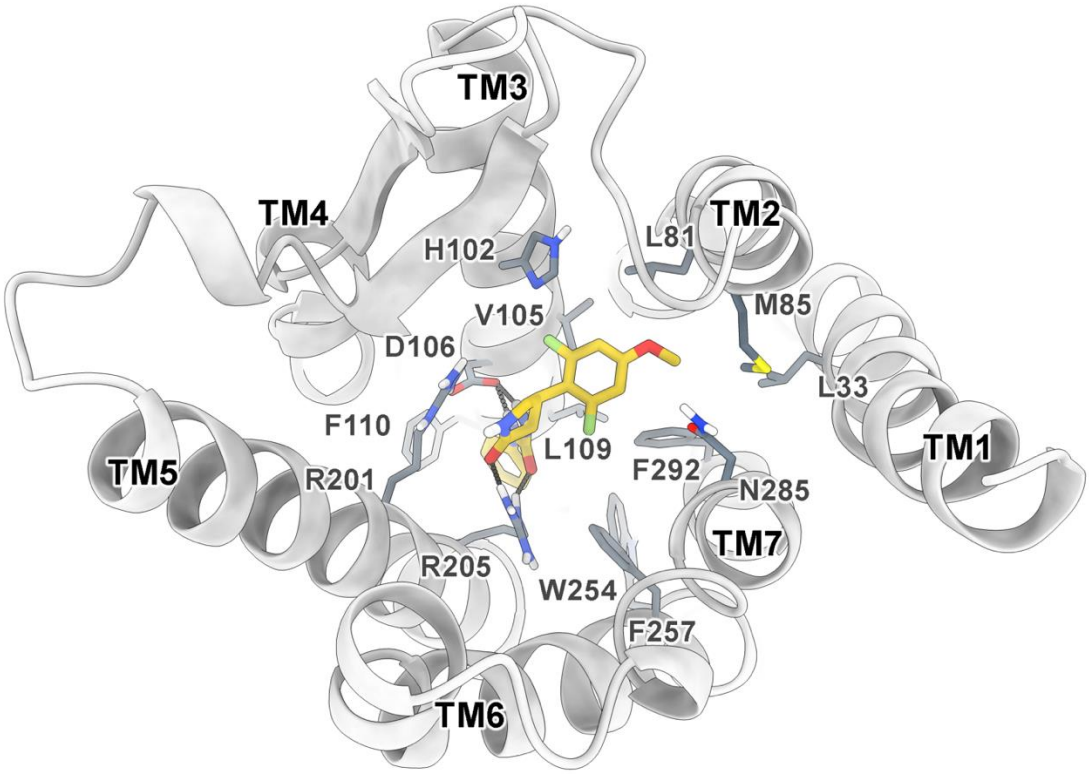


Figure 3.14 Docking-predicted binding pose of BMS-986235 at the FPR2 ortostheric binding site. The receptor is shown as silver cartoon, the ligand, and the residues important for its binding are highlighted as gold and dark ray sticks, respectively. Non-polar hydrogens are omitted for sake of clarity, while H-bonds are depicted as black dashed lines.

In fact, two H-bonds are detected between the nitrogen atoms of the ligand's ureidic moiety and the aspartic acid of D106$^{3.33}$, and other two between the R205$^{5.42}$ side chain and the carbonyl oxygens of the pirrolydinone and ureydic groups of **4**, respectively. This interaction scheme would thus suggest that the ligand is involved in an extended polar network with D106$^{3.33}$, R201$^{5.38}$ and

R205$^{5.42}$ on TM3 and TM5, likely stabilizing a salt-bridge between the former two residues observed in the FPR2 crystal structure.

To evaluate the stability of this binding mode, also considering protein motion and solvent effects, we submitted the docking complex to extensive (5 μs-long) MD simulations in explicit solvent and membrane. As shown in Figure 3.15A-B, while the ligand explores the inner part of the binding cavity, its interaction pattern remains quite conserved over the simulated timescale. In fact, high frequency of occurrence during the simulation (> to 75% of the collected MD frames, Figure 3.15B) is observed for most of the contacts characterizing the binding pose. Notably, many of the residues involved in these interactions (eg. V105$^{3.32}$, D106$^{3.33}$, L109$^{3.36}$, R201$^{5.38}$, R205$^{5.42}$, F257$^{6.51}$) correspond to amino acids crucial for the binding of the reference peptide **1** by previous mutagenesis studies.[153]

However, the RMSD plot of the ligand's heavy atoms shows some rearrangement of its binding conformation with respect to the starting pose (Figure 3.15C). In particular, we can discriminate three long-living states (Figure 3.15 E,F,G), in which, although **4** preserves most of the already described ligand-protein interactions, its position in the binding cleft is slightly translated.

The main reasons for this phenomenon are to be found in a movement of the D106$^{3.33}$- R201$^{5.38}$-R205$^{5.42}$ triad (Figure 3.15D). As already mentioned, at the

beginning of the simulation the ligand is inserted between R205$^{5.42}$ and the ionic lock formed by D106$^{3.33}$ and R201$^{5.38}$ (Figure 3.15E). After about 0.5 μs of MD simulation, R205$^{5.42}$ also approaches D106$^{3.33}$ (Figure 3.15D), forcing the ligand to partly rearrange to reach the second long-living state. Here, (Figure 3.15 F) the hydrogen bond between the ureidic carbonyl oxygen of **4** and the side chain of R205$^{5.42}$ is lost, while the carbonyl oxygen of the pyrrolidinone ring, which in the former state only contacted R205$^{5.42}$, can H-bond also with R201$^{5.38}$. On the other hand, in the last state, the ligand's ureidic moiety is pushed slightly farther from the polar triad causing the disruption of its H-bonds with R201$^{5.38}$.

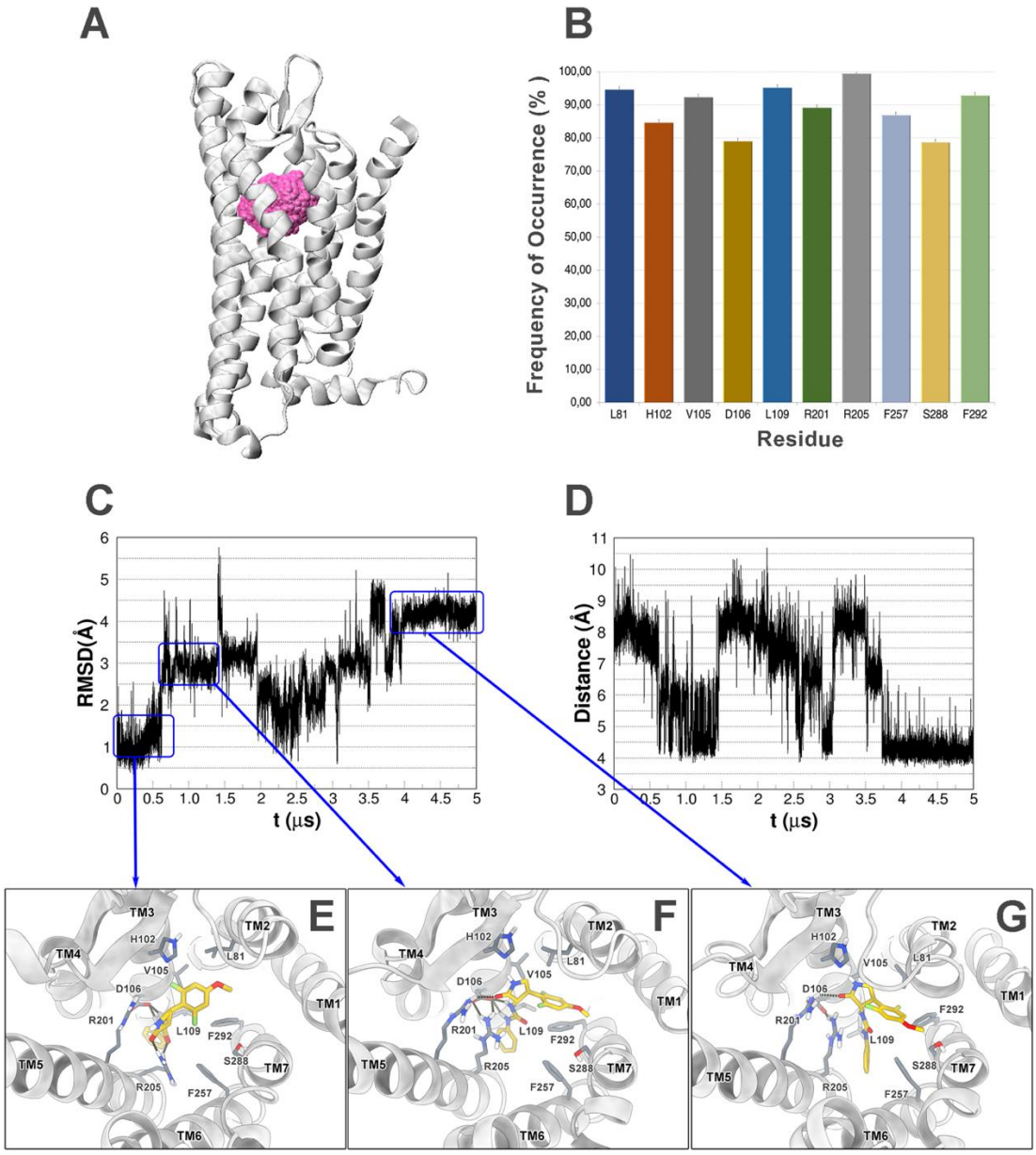


Figure 3.15 A) Position occupied by the ligand's centre of mass (purple spheres) along the MD simulation; B) Frequency of occurrence of the most relevant ligand-protein contacts over the MD trajectory. Contacts are computed considering all the residues in a range of 3 Å from the ligand. C) RMSD plot of ligand's heavy atoms (the trajectory was aligned to the Cα of the receptor's helices prior to calculation). D) Distance between D106 (Cγ) and R205 (Cζ) over the MD trajectory. E-F-G) Representative binding conformation assumed by the ligand at the FPR2 orthosteric site. The receptor is shown as silver cartoon, the ligand and the residues important for its binding are highlighted as gold and dark ray sticks, respectively. Non-polar hydrogens are omitted for sake of clarity, while H-bonds are depicted as black dashed lines.

Thus, to better elucidate the role of the amino acids involved in the binding of **4**, especially the D106$^{3.33}$- R201$^{5.38}$-R205$^{5.42}$ triad, we performed 5 μs-long MD simulations both on the crystallographic **1**-FPR2 complex and on the *apo* form of the receptor.

In agreement with what is observed in the crystal structures and described in previous SARs studies,[153] our calculations indicate that **1** mainly interacts with FPR2 through its C-terminal sequence (Met^{4}-Val^{5}-D-met^{6}-NH_2).

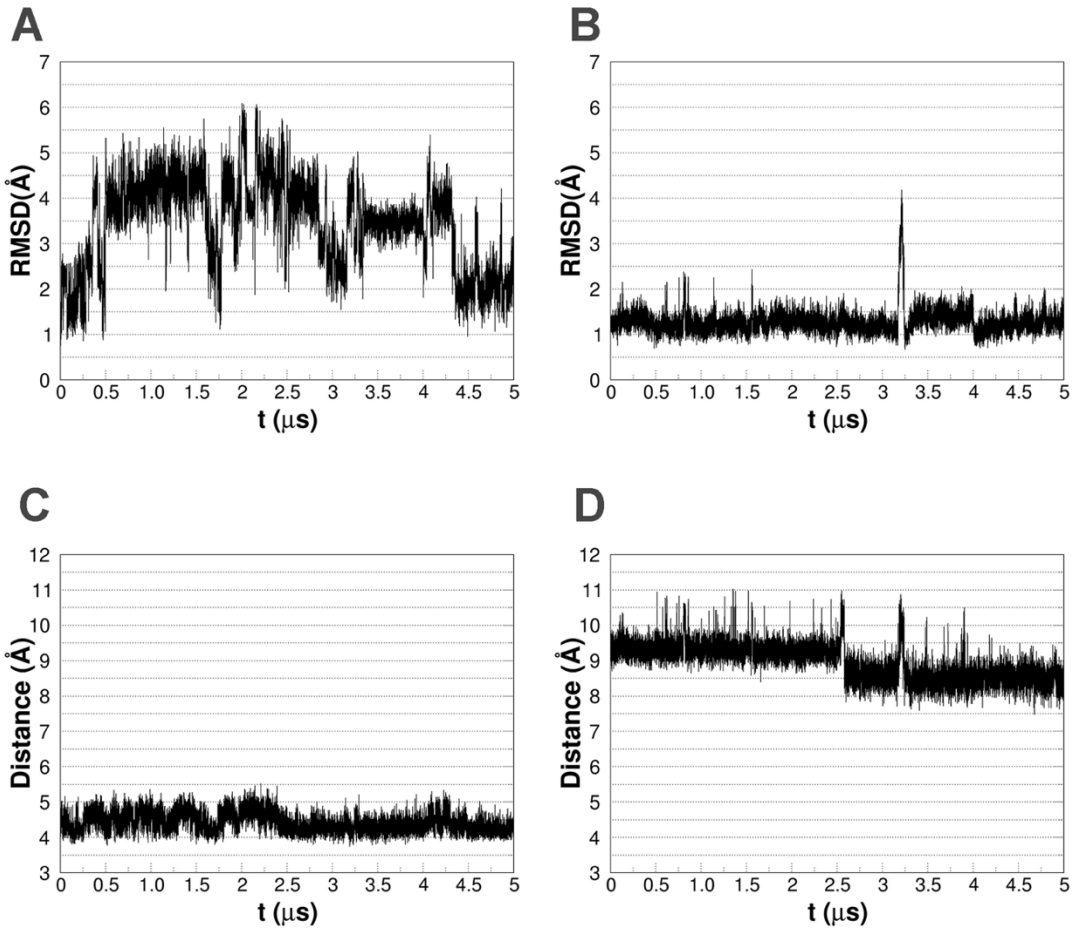


Figure 3.16 A) RMSD plot of N-terminus (A) and C-terminus (B) side of **1** over the MD trajectory. The RMSD was computed on the heavy atoms and the trajectory was aligned to the Cα of the receptor's helices prior to calculation. C) Distance between D106 (Cγ) and R201 (Cζ) over the MD trajectory. D) Distance between D106 (Cγ) and R205 (Cζ) over the MD trajectory.

In fact, the RMSD value assumed by this motif is very low (Figure 3.16B), at variance with its N-terminal residues (Trp^{1}-Lys^{2}-Tyr^{3}) which fluctuates a lot during the entire trajectory (Figure 3.16A). Interestingly, the C-ter amino acids participate in a hydrogen bond network with D106$^{3.33}$, R201$^{5.38}$ and R205$^{5.42}$, similar to what was observed in the case of **4**. In detail, the main chain carbonyl of Met^{4}, the backbone nitrogen of D-met^{6} and the peptide's C-terminal amide contact D106$^{3.33}$ and R201$^{5.38}$, while two additional H-bonds are formed by the main chain carbonyls of Met^{4} and Val^{5} with R205$^{5.42}$. Notably, these interactions appear more stable than those formed by **4**, keeping the polar triad in a conformation where only D106$^{3.33}$ and R201$^{5.38}$ are in direct contact (Figure 3.16C), whereas the distance between R205$^{5.42}$ and the D106$^{3.33}$-R201$^{5.38}$ ionic lock is constantly much higher than the traditional H-bond threshold (Figure 3.16D).

On the other hand, in the simulation on the *apo* receptor R205$^{5.42}$ showed a certain tendency to leave its initial position. In fact, its side chain gets closer to D106 so as to form an additional salt bridge with this residue, as transiently observed also in the simulation on the **4**-FPR2 complex (Figure 3.17).

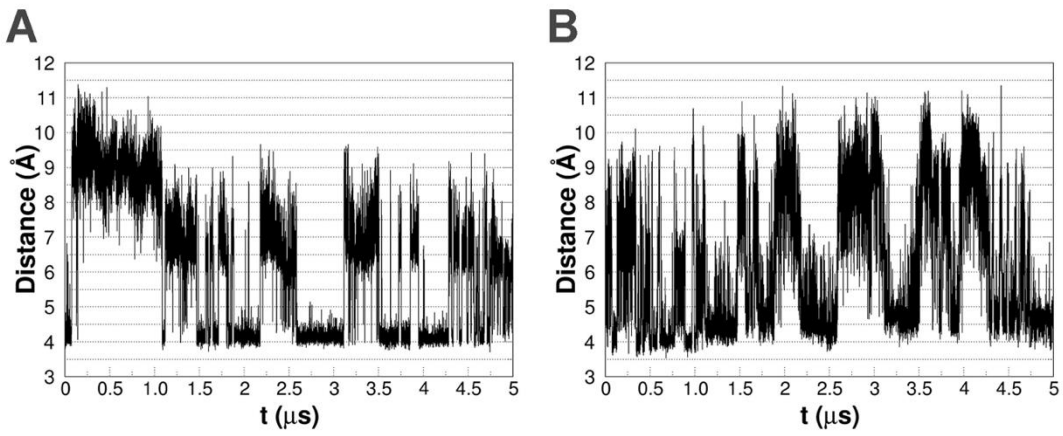


Figure 3.17 A) Distance between D106 (Cγ) and R201 (Cζ) over the MD trajectory. B) Distance between D106 (Cγ) and R205 (Cζ) over the apo-MD trajectory.

In conclusion, our simulations show that the D106$^{3.33}$-R201$^{5.38}$-R206$^{5.42}$ polar triad might profoundly influence - or be dependent on - the ligand binding modalities. Such hypothesis is also supported by mutagenesis studies in which the mutation to Ala of any of these three amino acids dramatically impaired the binding of the co-crystalized ligand **1**.[153] Moreover, bioinformatic analysis was here performed, showing that no other GPCR possesses, at the 3.33 and 5.38 (Weinstein-Ballesteros numbering) positions, salt bridge forming residues, which are thus peculiar of FPRs (Figure 3.18).

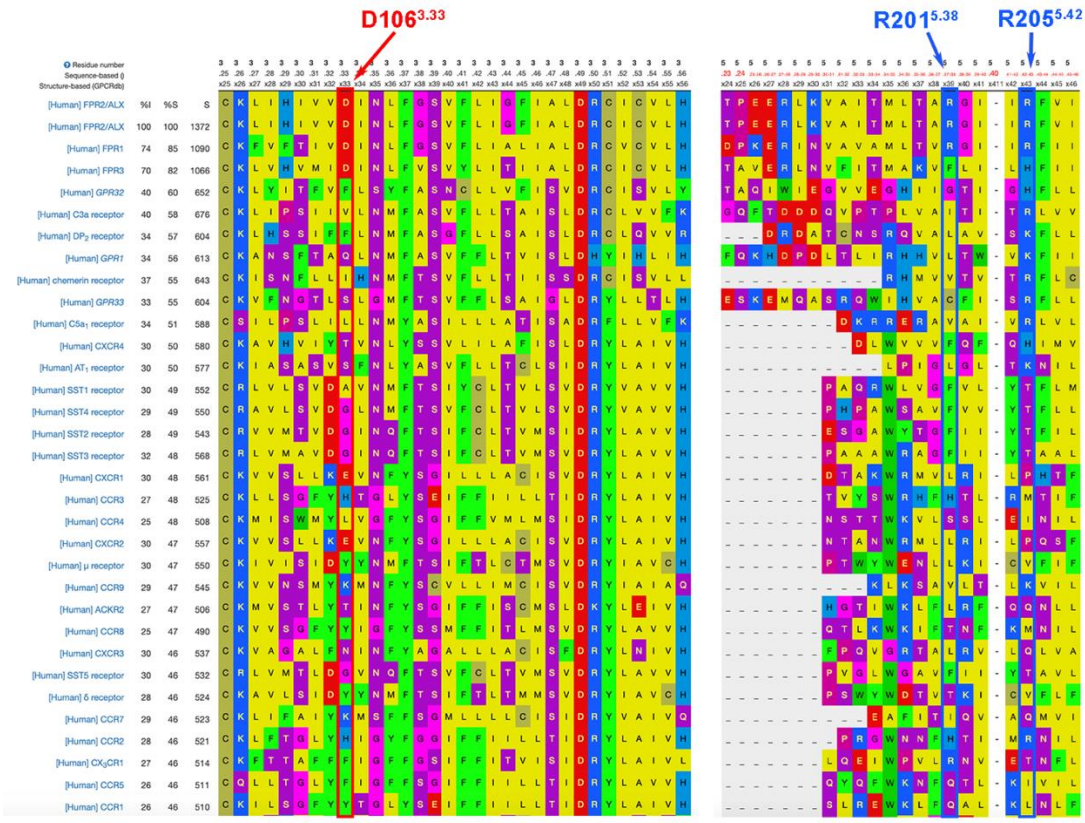


Figure 3.18 Conservation of D106$^{3.33}$-R201$^{5.38}$-R205$^{5.42}$ triad in the 30 Class A GPCRs most similar to FPR2.

For all these reasons, it is likely that the D106$^{3.33}$-R201$^{5.38}$-R206$^{5.42}$ triad may take part both in the ligand binding and in the molecular switches responsible for the activation of these GPCRs. These aspects were explicitly taken into consideration in the following thermodynamics evaluation of the binding event of **4**. To this aim, we designed an *ad hoc* collective variable (CV) for Funnel Metadynamics (FM), able to discriminate between the two main different states assumed by the polar triad during our unbiased MD simulations:

i) *open*: only the D106$^{3.33}$-R201$^{5.38}$ salt bridge is formed

ii) *closed*: an additional salt bridge between R205$^{5.42}$ and D106$^{3.33}$ is established

3.2.4.3. Funnel Metadynamics: Ligand Binding Mechanism

As briefly reported in the previous paragraphs, current literature on ligand binding to FPR2 only refers to peptide molecules. The two experimental structures of FPR2 in complex with **1** revealed that this compound adopts a peculiar binding mode, extending in a deeper region of the transmembrane bundle compared to other peptide-binding GPCRs.[153,154] On the other hand, a comprehensive elucidation of the binding mechanism of small organics molecules is still missing. In fact, small-sized ligands can establish other interaction patterns than peptides, especially in large and amphipathic pockets like in FPR2. In this perspective, an extensive *in silico* investigation was here engaged with the aim to disclose the binding mechanism of the representative small FPR2 agonist **4**. Indeed, although preliminary docking calculations predicted this compound to bind in the same area of peptide **1**, its pose resulted not completely stable over very long MD-timescales likely due to specific protein's motions. For this reason, a more rigorous approach was adopted, based on free energy calculations, particularly on Funnel Metadynamics (FM). With this technique, the ligand can explore the entire volume of the orthosteric binding site, allowing to identifying the favourite

ligand binding mode and accurately estimating the absolute protein-ligand binding free energy. This represents an important advance compared to standard MD-based approaches, in which the region sampled by the ligand is limited to the proximity of its starting position due to timescale limitations.

FM applies a funnel-shaped restraint to define the receptor's region of interest, accelerating the recrossing events between the ligand bound and fully solvated unbound state. The funnel potential is applied in a traditional WT-MetaD framework, where a crucial step is the definition of properly chosen CVs. These are presumably slow degrees of freedom of the system (CVs) with whom an adaptive bias potential is built to enhance the sampling of the phenomenon under investigation. In the present project, a distance *(d)* and a dihedral angle (torsion - φ) CV were chosen, as they are widely employed in ligand-binding investigations. Moreover, an *ad-hoc* variable was defined to take into account the conformational movement of the protein's polar triad $D106^{3.33}$-$R201^{5.38}$-$R206^{5.42}$ *(cmap)*. In detail, *d* was computed during the simulation as the distance between the centre of mass of the FPR2's orthosteric binding site and the centre of mass of the ligand. This CV is fundamental to allow the ligand to explore both the bound and the unbound state, thus giving a reliable estimation of ΔG_b. The φ CV was meant to sample all the possible orientations of the ligand with respect to the binding site. It was defined as the dihedral angle between the major inertia axis of the ligand and the plane

identified by two properly selected atoms in the protein binding cleft. For a more detailed description of the employed CVs, the reader is referenced to paragraph 3.2.6.3. The FM calculation was performed in a multiple walker scheme, in which 10 parallel replicas of the system were simulated starting from different regions of the phase space and sharing the same metadynamics bias. Each walker was simulated for 700 ns, for a total simulation time of 7 μs. Looking at Figure 3.19A, we can appreciate how the different walkers well cover all the regions in the CV (distance – d) space, although we still aim to obtain more bound-unbound recrossing events and a faster CV diffusion (known indicators of a perfectly converged MetaD simulation). In fact, the simulation is still running in order to fully reach the convergence in the estimation of the BFES and of the ΔG_b. At this regard, promising signals are provided by the plot of the ΔG_b as a function of the simulation time (Figure 3.19B), where the fluctuations in the estimated free energy difference between the ligand bound and the unbound states progressively decrease toward a

mean value of -11.7 kcal/mol. Notably, this prediction is in good agreement with the low nanomolar experimental IC_{50} of **4** for FPR2 .[151]

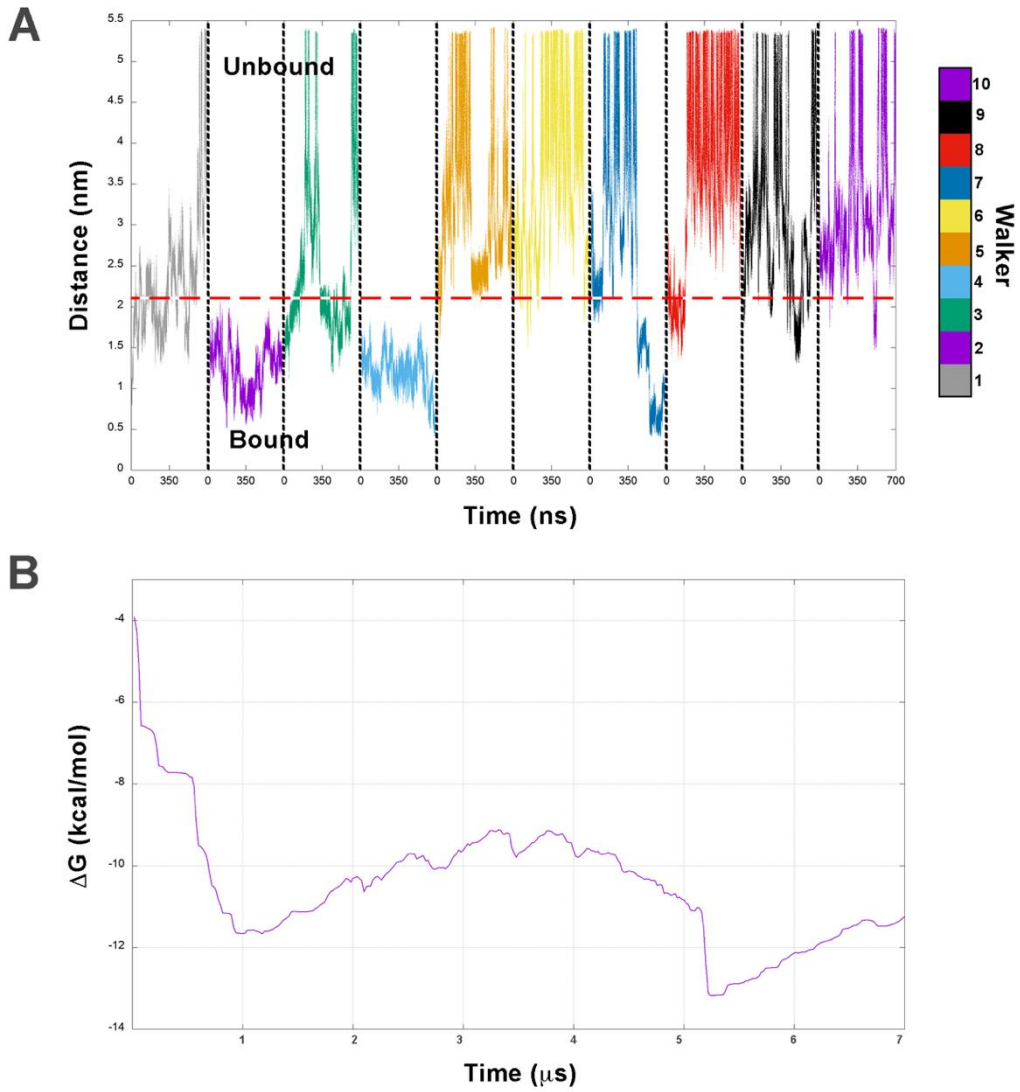


Figure 3.19 A) Diffusion of the CV distance in function of simulation time. Each walker is represented with a different color. B) Plot of the evolution of ΔG_b over time simulation time.

For a qualitative description of the binding mechanism, we computed the BFES as function of the *d* and φ CVs (Figure 3.20), which can accurately describe the position and the orientation of the ligand with respect to the

receptor's pocket. In the computation of the BFES in Figure 3.20, the bias accumulated during the simulation on the *cmap* CV was integrated out.

At a first glance, the BFES immediately gives the idea of a complex binding pathway, characterized by multiple metastable states. In fact, starting from the fully unbound state (d > 4 nm) the first decrease in terms of ΔG is observed when the ligand comes in proximity of the receptor (2 nm < d < 2.5 nm). In this region, the BFES has isoenergetic values at different degrees of φ, meaning that the ligand does not adopt a single well-defined conformation; conversely, it transiently interacts with the FPR2's extracellular part assuming different orientations. This phenomenon, which could be referred as *rolling*, has been already described for other GPCRs. For instance, a study on the unbinding of alprenolol from the β2-adrenoreceptor showed that this compound can variously contact the outer part of ECL2 and ECL3 before reaching the fully solvated state.[156] Coming to the bound portion of the BFES, 4 different energy basins, namely **A**, **B, C** and **D** can be recognized. These four minima have a difference in terms of ΔG of about 2.5-3 kcal/mol. Bearing in mind that the intrinsic error associated with force fields-based simulations is estimated around the 2-2.5 kcal/mol and that our FM calculation is still not completely converged, it is reasonable to consider the four minima isoenergetic and, consequently, as equally probable states.

First, an important consideration is that except for minimum **A**, which corresponds to the pose described in paragraph 3.2.4.2, none of the other three basins was predicted by docking. A description of these minima will here follow, starting from those where the ligand is bound to the external part of the binding site. In the first energy basin (**D**, Figure 3.20) the ligand is hosted by residues belonging to the upper side of TM5, TM6, and TM7. This pose is mainly stabilized by a water-mediated polar network with the side chains of $D281^{7.32}$ and $N285^{7.36}$ and by hydrophobic contacts formed by the ligand's phenyl ring and 2,6-difluoro-4-methoxyphenyl with $L198^{5.35}$, $F257^{6.51}$, $L268^{ECL3}$, $D281^{7.32}$ and $V284^{7.35}$. In both the second and third minima (**B** and **C**) **4** occupy the hydrophobic subpocket located between TM2, TM3, and TM7.

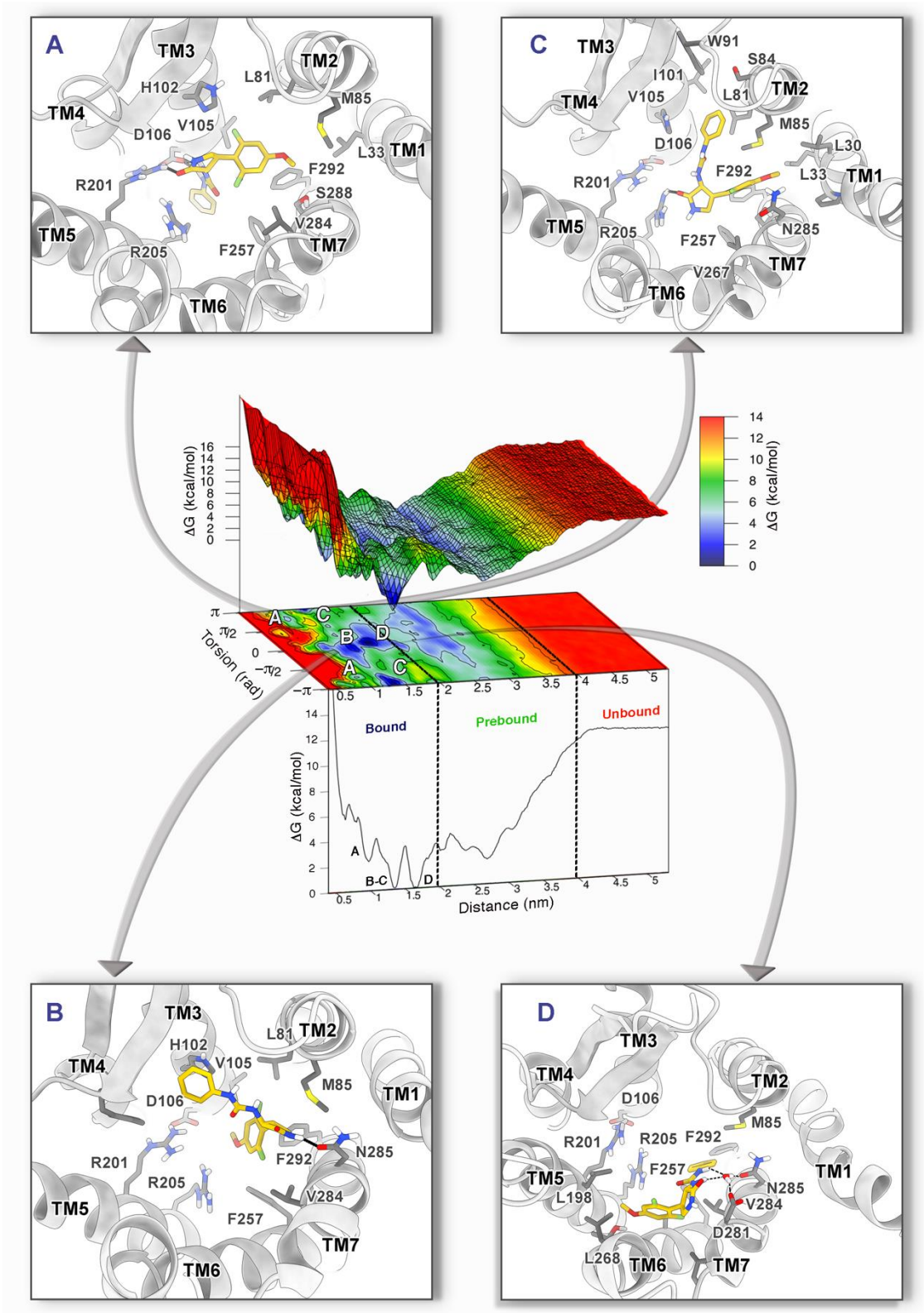


Figure 3.20 One dimensional (as function of the *distance* CV) and two-dimensional (as function of *distance* and *torsion* CVs) Binding Free Energy Surface (BFES). Isosurfaces of 2.5 kcal/mol are used. The representative structure of the most significant energy minima are shown as insets. FPR2 is displayed as grey sticks and cartoon, while the ligand is shown as gold sticks. The oxygens of the water molecules are depicted as red spheres. Non-polar hydrogens are omitted for clarity. Hydrogen bonds are highlighted as black dashed lines.

Notably, these basins are separated by very low energy barriers, and thus they might in principle convert one into the other in standard MD timescales. Notably, in **basin C** the ligand engages the polar triad D106$^{3.33}$-R201$^{5.38}$-R206$^{5.42}$ with its pyrrolidinone ring. This binding conformation could represent an important intermediate state visited by the ligand before reaching **basin A** in the innermost region of the binding pocket. Here, the ligand binding conformation is very close to that predicted by docking and MD calculations. In fact, we clusterized this minimum and compared the mostly occurring structures with the three long-living states identified from unbiased MD (paragraph 3.2.4.2 – Figure 3.15). Notably, most of the analysed poses are very similar to the first MD state (Figure 3.15E), where the *cmap* assumes an open conformation and the ligand is inserted between the ionic lock D106$^{3.33}$-R201$^{5.38}$ and the side chain of R205$^{5.42}$. We can thus conclude that the other two (Figure 3.15F-G) could be relatively higher energy conformations encountered by the ligand in its transition from/to **basin A.**

Overall, FM simulations are providing an accurate description of the entire binding event of compound **4** to FPR2, together with the estimation of the underlying free energy surface (BFES). Indeed, a minimum free energy pathway (MFEP) from the unbound to the bound state can be defined, where the ligand passes through **basin D** and **basin C/B** to finally reach **basin A**. The

existence of multiple metastable states can overall contribute to the marked potency of compound **4** towards FPR2, providing a complete picture of the molecular requirements for high-affinity small organic ligands of this receptor.

3.2.5. Conclusions and future perspectives

Formyl Peptide Receptors, especially FPR1 and FPR2, are peculiar class A GPCRs involved in a plethora of physiological and pathological functions, including host defense and inflammation. For these reasons, the medicinal chemistry community is focusing on the discovery of new ligands capable of modulating the activity of FPRs. These efforts resulted in a number of peptide and organic ligands active on both FPR1 and FPR2. While the structural basis of peptides' interaction with FPR2 has been recently discovered by X-ray and Cryo-EM studies, no information is yet available on the molecular requirements of high-affinity small organic compounds. In this perspective, we here performed a thorough characterization of the binding mechanism of the most promising known FPR2 agonist, namely *BMS-9862335* (compound **4**). First, a competitive *in vitro* binding assay was carried out in presence of the reference peptide ligand **1**, demonstrating that **4** behaves as a pure orthosteric binder. Starting from these data, an *in silico* investigation of its binding mechanism was engaged. Particularly, unbiased MD and free energy simulations revealed that the ligand could assume multiple metastable

binding modes within the large amphipathic pocket of FPR2. Indeed, the binding free energy surface (BFES) highlighted the presence of four energy minima in the bound region, which can all contribute to the marked potency of **4** toward FPR2. Remarkably, only one of these wells, namely **basin A** (Figure 3.20), corresponded to the pose previously predicted also by standard docking calculations. This data suggests the importance of rigorous computational approaches to fully elucidate the dynamics and the energetics of the ligand binding event. As proof of the reliability of these predictions, the estimation of ΔG_b for compound **4** in FPR2 is fluctuating around -11.7 kcal/mol, which agrees with the experimental EC_{50} value in the low nanomolar range (EC_{50} = 0.91 nM).[151] However, the BFES is not completely converged and simulations are still ongoing.

An intriguing feature of FPR1 and FPR2 is their high sequence homology; most binders of these receptors are not selective compounds. This can represent an important drawback affecting both the complete understanding of the biological role of FPRs and the therapeutic potentialities of their ligands. In fact, FPR1 seems to mediate both pro-inflammatory and anti-inflammatory signals, whereas FPR2 is critically involved in the resolution of inflammation. In this context, the high-resolution data afforded by molecular simulations on the **4**-FPR2 complex provide valuable hints to develop selective FPRs ligands. Particularly, in the lack of any experimental structure of FPR1, the

representative conformations of the energy basins identified in paragraph 3.2.4.3 will be used as templates for building multiple homology models of this receptor. Then, the molecular reasons for the minor affinity of compound **4** to FPR1 will be elucidated by performing accurate comparative binding mode studies between each of these receptor's conformations and the FPR2 isoform. This way, we will provide an exhaustive picture of all the possible chemical strategies to achieve both high affinity and selectivity for either FPR1 or FPR2, paving the way for a new phase of the FPRs-oriented drug discovery, more rational and knowledge based.

3.2.6. Methods

3.2.6.1 Molecular Docking

Docking calculations of **4** were performed in the X-ray crystal structure of FPR2 in complex with **1** (PDB code 6LW5).[153] The ligand's tridimensional structure was generated with the 2D build panel in the Maestro suite;[157] then its protonation and tautomeric states were predicted by the Epik software.[109,111] On the other hand, the receptor was prepared with the aid of the *Protein Preparation Wizard*[111] tool implemented in Maestro, assigning to the protein the correct bond orders, adding missing hydrogens and heavy atoms. Then, the ionization and tautomerization states of the side chains were predicted using

Epik,[109] so that of the receptor H-bond network was optimized. Finally, the hydrogen atoms' positions were minimized. Docking calculations were performed with two different software: Glide 8.5[21,22] and Autodock 4.1.[23] In both cases, the search area was defined as a 30 Å x 30 Å x 30 Å virtual box comprising the whole FPR2 orthosteric binding site. In the calculations performed with Glide, we applied the standard SP protocol and the OPLS3A force field. In the set of computations carried out with Autodock, 100 separate docking runs were performed, each consisting of 10 million energy evaluations using the Lamarckian genetic algorithm local search (GALS) method. Finally, the predicted binding conformations were clustered based on their RMSD (tolerance = 2.0 Å) and ranked according to the Glide SP and Autodock scoring function, respectively.

3.2.6.2. Molecular Dynamics

Prior to MD simulations, the receptor N-terminal residues (6-18) were cut, since in the original paper of the PDB structure[153] their conformation was reported as affected by the fusion with the thermostabilizing protein bRIL. Thus, the resulting first N-terminal (S19) and last C-terminal (E324) amino acids were capped with acetyl and N-methyl groups, respectively. The three simulated complexes (**1**-FPR2, **4**-FPR2, and *apo* FPR2) were then embedded in a 100 Å x 100 Å (along x and y axes) pre-equilibrated 1-palmitoyl -2-

oleoylphosphatidylcholine (POPC) – cholesterol (7:3 molar ratio) bilayer and solvated with a 18 Å water layer using the TIP3P model with the aid of the membrane-builder tool of CHARMM-GUI.org (http://www.charmm-gui.org). The receptor and the lipidic bilayer were treated according to the Amber ff14SB and lipid17 force fields, respectively.[158] While for **1** we employed the same protein force field (Amber ff14SB), the organic small molecule **4** was parameterized following the protocol of the generalized amber force field (GAFF)[114]. In detail, the bonded and van der Waals parameters were directly taken from the force field, whereas the atomic partial charges were obtained from quantomechanical calculations. At this regard, the ligand's geometry was first optimized in a two-step QM procedure performed with the Gaussian 16 software[159] at the Hartree-Fock level of theory: a preliminary guess with the 3-21G* basis set, followed by a more refined calculation with the 6-31G* basis set. During this second step we also computed the electrostatic potential (ESP), which was then fitted onto atomic partial charges thanks to the two-stages restrained electrostatic potential (RESP)[115] fitting procedure implemented in Antechamber.[117] The topology files of the systems were obtained with the tleap program of AmbertTools19 and then converted into Gromacs format by means of ParmEd. The simulations were carried out with the Gromacs 2020.6[160] code. The cutoff employed for the computation of the short-rage interactions was of 12 Å, whereas the Particle Mesh Ewald[53] method (with a 1.0 Å grid spacing in

periodic boundary conditions) was used for the treatment of long-range ones. A 2 fs integration time step was allowed by constraining bonds with the non-iterative LINCS algorithm[161]. The systems were equilibrated according to the following protocol. Three steps of energy minimization were needed to solve all the steric clashes. Then, each complex was heated up to 300K, alternating NPT and NVT cycles (for a total of 30 ns) with the Berendsen coupling bath and barostat,[162] while applying progressively decreasing harmonic constraints on the heavy atoms' positions of membrane, protein, and ligands. Finally, production runs were performed with the leap-frog integrator in the NPT ensemble; the pressure of 1 atm and the temperature of 300 K were kept constant with the stochastic velocity rescaling[163] and Parrinello-Rahman[164] algorithms, respectively.

3.2.6.3 Funnel Metadynamics

FM simulation was carried out using the Gromacs 2020.6 code patched with Plumed 2.7.0[165,166] and following the same equilibration and simulation protocol of the previous unbiased MD.

The funnel shaped restraint potential was set to include the entire ortostheric binding site, using *Zcc* = 32 Å, α = 0.45 rad and R_{cyl} = 1 Å. The *z* axis of the funnel can be identified in the space x,y,z by two points, one corresponding approximately to the centre of mass of R201Cα, F257Cα, and L81Cα, and the other to the center of mass of V207Cα, D291Cα, and H102Cα; the first one also defines the origin of the z axis (Figure 3.21).

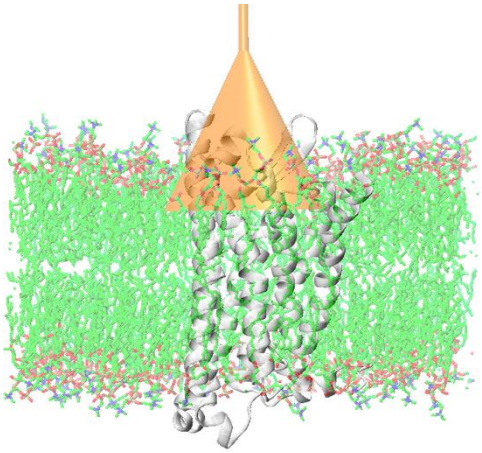


Figure 3.21 Graphical representation of the funnel restraint potential applied during FM simulation

The estimation *F(s,t)* at time *t* of the free-energy surface *F(s)* as a function of the selected CVs was determined by metadynamics[56] in its well-tempered variant, using the following formula:

$$F(s,t) = -\frac{T + \Delta T}{\Delta T} V(s,t)$$

where *V(s,t)* is the bias potential added to the system and T is the temperature of the simulation. ΔT is the difference between the fictitious temperatures of

the chosen CVs and the system, respectively. The bias potential is made up of the sum of the gaussians deposited on the selected CVs. The exploration of the CVs space can be increased by tuning ΔT through the *biasfactor* parameter. In our case, we deposited gaussians of initial height equal to 1.5 kJ/mol, gradually decreased based on a biasfactor of 25, corresponding to a ΔT of 7200 K.

The simulation was performed employing three different CVs (Figure 3.22). Two of them were chosen to describe the different ligand conformations with respect to the binding pocket: i) the distance (*d*) between the centre of mass of the ligand and that of the receptor's orthosteric (atoms: L109-Cα, F110-Cα, G209-Cα, W254-Cα, N294-Cα) ; (ii) the dihedral angle (torsion - φ) defined by the major inertia axis of the ligand and two Cα, V113-Cα and N294-Cα, selected in a sufficiently rigid region of the TM bundle (Figure 3.20B).

An additional CV (*cmap*) was designed to accelerate the movement of the receptor's polar triad D106$^{3.33}$-R201$^{5.38}$-R205$^{5.42}$, described in paragraph 3.2.4.2. The CV was defined as the sum of the following switching function:

$$s_{ij} = \frac{1 - (\frac{r_{ij} - d_0}{r_0})^n}{1 - (\frac{r_{ij} - d_0}{r_0})^m}$$

where *i* and *j* are the atoms used to identify the contacts. In our case, we only considered two interactions: i) D106$^{5.38}$-R205$^{5.42}$ (Cγ- Cζ) and ii) R201$^{5.38}$-R205$^{5.42}$ (Cζ- Cζ). Thus, the CV can assume values tending to 2 when the triad is in its closed conformation (with R205$^{5.42}$ packed against the ionic lock D106$^{5.38}$-

R201$^{5.38}$) and values tending to 0 in its open conformation. As concern the other parameters, the d_0 distance was set to 0, while the n and m exponentials were modulated to 12 and 30, respectively; the r_0 value was set to 7 for contact i) and 5.75 Å for contact ii). Gaussian widths of 0.1 Å, 0.05 rad and 0.1 were used for the d, φ and *cmap* CVs, respectively.

Notably, FM was performed following the multiple walker approach (2.5) with 10 parallel replicas initialized by different regions of the CVs phase-space. To avoid any artifact caused by the MetaD potential, FPR2 was constrained with a harmonic potential on the root mean square deviation (RMSD) of its TM helices (Cα), allowing only for movements up to 2.2 Å. This value was observed as the maximum displacement showed by the receptor in unbiased conditions.

The absolute binding free energy (ΔG_b) of compound **4** was computed according to the equation 2.27 and 2.25. As bound region we selected all the states sampled at value of $d < 2\ nm$, whereas the reference unbound state was chosen as the isoenergetic region at $d > 4.75\ nm$.

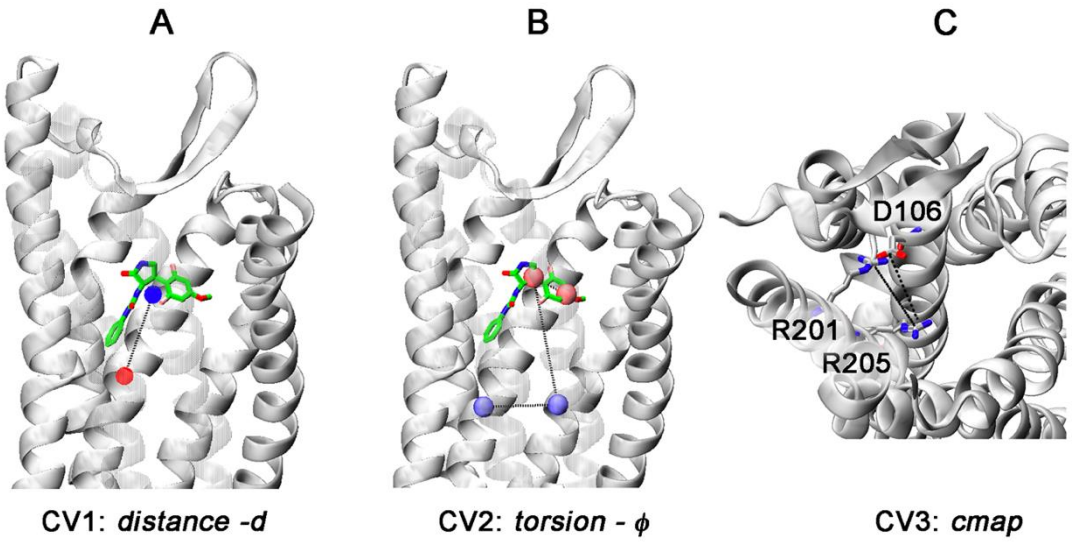


Figure 3.22 Graphical representation of the selected CVs. A) *Distance CV*: red sphere corresponds to the centre of mass of the binding site, whereas the blue sphere represents the centre of mass of the ligand. B) *Torsion CV*: the two atoms of the binding site (blue) and the two atoms of the ligand (red) selected for computing the dihedral angle are shown as spheres. C) *Cmap CV*: the tree residues forming the polar triad are shown as sticks, the two contacts used for the computation of the CVs are depicted as black dashed lines

Chapter 4. Peptides conformational Sampling

In medical sciences, we usually refer to peptides as molecules made of 2-50 aminoacids, which play pivotal roles in human physiology acting as hormones, neurotransmitters, growth factors, or antibacterial agents. Peptides drug design is often considered a complex area due to their low stability in solution, poor permeability through cellular membranes and physiological barriers, such as the blood–brain barrier (BBB). However, in the last decades this sector has experienced a revival testified, for instance, by the 28 new noninsulin peptides approved since 2000, with several of them achieving significant market success. In comparison to small molecules, peptides offer both important advantages and disadvantages that must be carefully considered in a drug discovery campaign. Indeed, despite important pharmacokinetic limitations (e.g., biological instability or poor membrane permeability), they are better suited to retain the physicochemical properties of bioactive proteins/polypeptides needed to target the desired receptor. In fact, their conformational plasticity facilitates the interactions with larger and more shallow surfaces compared to the typically cryptic binding pockets targeted by small molecules, encouraging their use in mimicking/disrupting protein-protein interactions (PPIs).

The rational design of peptides able to exploit the structural characteristics of PPIs can be valuably driven by computational methods. However, the

investigation of the peptide-receptor binding mechanism is a non-trivial task for which, alongside 3D information on the target macromolecule (coming from X-ray crystallography, NMR or Cryo-EM experiments), a deep characterization of the intrinsic peptide conformational landscape is necessary. In fact, since peptides are highly flexible entities, their conformational entropy and ability to assume a well-defined active structure is a determinant factor during the binding event. In this section, two examples will be presented of how to deal *in silico* with the conformational sampling of peptides and, in turn, with their binding mode prediction. In detail, we will focus on two peptides, namely [RGD-Chg-(*N*Me)E]-$CONH_2$ (4.2) and iRGD (4.3), endowed with remarkable antiviral and anticancer properties, respectively. Notably, both the compounds are cyclic peptides targeting the same important family of membrane receptors, RGD integrins (4.1). The main difference in the technical approach to the study of these compounds came from the availability or lack of experimental information on their folding properties. As for [RGD-Chg-(*N*Me)E]-$CONH_2$ (4.2), taking advantage of NMR data it was possible to adopt a combined experimental-computational strategy to predict first the peptide conformation and then its binding mode to integrins. On the other hand, due to the lack of any experimental data on the bioactive structure of iRGD (4.3), we had to resort to a more complex computational approach based on PT-WTE metadynamics.

4.1. Integrins

The integrin family is a group of 24 transmembrane divalent-cationic-dependent cell-adhesion receptors, differently assembled from 18 α and 8 β subunits. Each integrin monomer has an extracellular domain, able to bind extracellular matrix (ECM) components, a single transmembrane region, and a cytoplasmic tail (Figure 4.1). These receptors mainly act as transmembrane linkers between their ECM ligands and the cytoskeleton, modulating, in turn, various signalling pathways essential in the biological functions of most cells.[167] Among them, the members of the RGD-integrin subfamily can recognize, with different extent of selectivity, ECM ligands featuring the RGD tripeptide motif including fibrinogen (Fbg), fibronectin (Fn), vitronectin (Vn), and the latency-associated peptide (LAP) of the transforming growth factor (TGF)-β. RGD ligands typically bind in a site located at the interface between the α and the β subunit. Particularly, the aspartic acid of the RGD sequence is required to chelate the divalent cation present in the MIDAS site of β, whereas the guanidinium moiety of RGD arginine engages a salt bridge with a conserved aspartate in the α subunit (Asp218 in αv isoforms) (Figure 4.1). Besides, in the upper part of the β subunit, a lipophilic cavity is delimited by the so-called specificity determining loop (SDL). SDL is a non-phylogenetically conserved region distinctive of each integrin isoform. Considering the difficulty to reach chemical selectivity over very similar

proteins like the integrins isoforms, the SDL cavity has thus represented a fundamental anchor point for medicinal chemists to develop selective integrin binders. The latter task has remarkable pharmacological applications in very different pathologies, due to the wide range of physiological functions mediated by the different integrins receptors. For instance, most of these play a key role in various steps of tumor angiogenesis, migration and invasiveness, intensively cross-talking with many oncogenes and growth factors, and thus are since long considered prognostic markers and valuable targets for anticancer therapy.[168] In parallel, RGD integrins are involved in the infectious cycle of many viruses, prompting the researchers to take them into account also as antiviral targets.[169] Notably, the two different integrin-focused projects presented in this chapter will discuss both the possible anticancer and antiviral applications of integrin ligands.

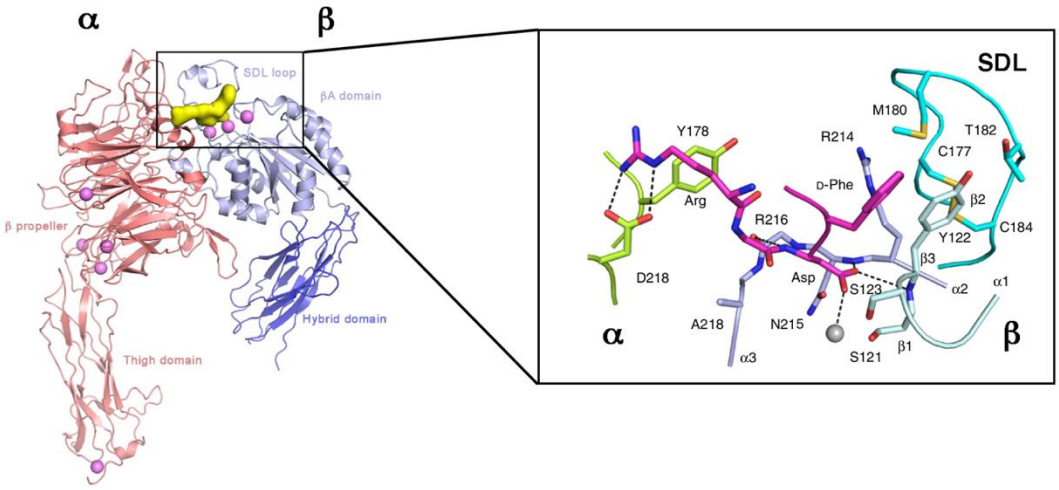


Figure 4.1 3D structure of integrins extracellular domain, with magnification of the ligand binding site and typical RGD interaction scheme.

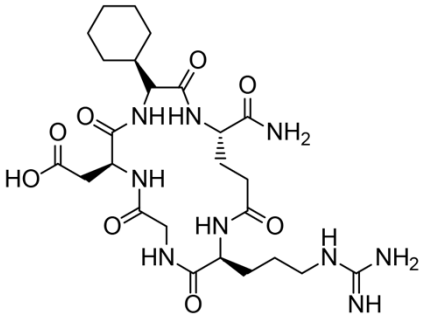
4.2 A dual αvβ6/αvβ8 Ligand against Herpes Simplex Virus-1 infections

4.2.1 Introduction

The Herpesviridae family is a group of heterogenous pathogens including[170] the varicella-zoster virus (VZV), human cytomegalovirus (HCMV), Epstein-Barr virus (EBV) or Kaposi's sarcoma–associated herpesvirus (KSHV) and human herpesviruses 6 and 7 (HHV-6/-7), foot and mouse disease virus (FMDV) and herpes simplex virus (HSV). Interestingly, many of these viruses evolved to take advantage of different integrins subtypes to enlarge their cellular tropism. An important example is HSV, a wide-spread two-members class of pathogens that generally infects either the human oropharyngeal mucosa (HSV-1) or the genital tract (HSV-2), followed by neuronal latency in the peripheral ganglia.[171] The HSV cell entry occurs by a fusion of the viral envelope with the plasma membrane synchronized with an endocytic uptake in neutral/acidic vescicles.[170,172–175] In detail, the fusion is a multistep process ruled by four essential envelope glycoproteins: gD, the heterodimer gH/gL, and gB.[176] The tropism of the virus is governed by the recognition of gD by two cognate receptors on host cells, namely nectin1 and HVEM (herpesvirus entry mediator). This interaction induces conformational changes that shift the gD ectodomain into its functional state.[176–178] In this active form, gD is now capable

of recruiting/stimulating the gH/gL heterodimer, thus triggering the switch of gB into a membrane-permeable fusogenic state leading, ultimately, to cellular fusion.[179] In this context, recent studies have demonstrated that αvβ6 and αvβ8 integrins can play a pivotal role in facilitating the virus penetration into cells. Particularly, the two receptors alternatively interact with gH/gL and stimulates the release of gL from the parent heterodimer. This way, the activation of gH and, eventually, of gB is accelerated.[180] Moreover, it seems that integrin-mediated activation of gH and gB can serve as a trigger checkpoint to ensure that the fusion machinery is not prematurely activated until endocytosis takes place.[181] Accordingly, recent studies have proved that the contemporary inhibition of αvβ6 and αvβ8 either by cell exposure to subtype-selective mAbs or through siRNA transfection can cause a significant drop in HSV infectivity.[181] Starting from these data, the development of dual αvβ6 and αvβ8 inhibitors appear a promising strategy for the discovery of brand-new anti-HSV therapeutic agents. Diversely from what happen with different integrins isoforms like αIIbβ3, αvβ3 and α5β1, still few binders selective for αvβ6 and αvβ8 are known.[182–187] In this context, our research group recently discovered a potent and selective αvβ6-directed cyclic pentapeptide, namely [RGD-Chg-E]-$CONH_2$ (**1**) (Chart 1).[183] In a following work, this compound was also successfully converted into an effective probe for molecular imaging.[186] However, bearing in mind that HSV employs both αvβ6 and αvβ8 as entry co-

receptors, a more efficient antiherpetic agent would be obtained by simultaneous targeting both these receptors. Thus, in this project we engaged a systematic *N*-methylation scan of the backbone amide bonds of **1** to broaden its selectivity profile also to αvβ8. Notably, *N*-methylation is a well-established strategy that has frequently succeeded in tuning subtype specificity of RGD peptides and improving their bioavailability or tolerance to enzymatic degradation.[188–193] Therefore, a small library of five new peptides (**2-6**) was tested for its binding affinity on the integrins of interests. Notably, [RGD-Chg-(*N*Me)E]-$CONH_2$ (**6**) resulted as a potent binder of αvβ6 and αvβ8 with good selectivity over other closely related RGD integrins. Then, we evaluated in extensive cell biological assays the ability of **6**, and its parent peptide **1**, to impair the HSV-1 entry process through an integrin-dependent mechanism of action. Then, a combined approach based on Nuclear Magnetic Resonance (NMR) spectroscopy and molecular modeling studies were fundamental to elucidate the binding mode of **6** to αvβ6 and αvβ8 integrins and to rationalize the molecular basis for its renewed selectivity profile, providing valuable hints for the future design of dual or subtype-specific RGD integrin targeting agents.



1
[Arg-Gly-Asp-Chg-Glu]$CONH_2$

Chart 4.1 Chemical structure of **1**

4.2.2 Results

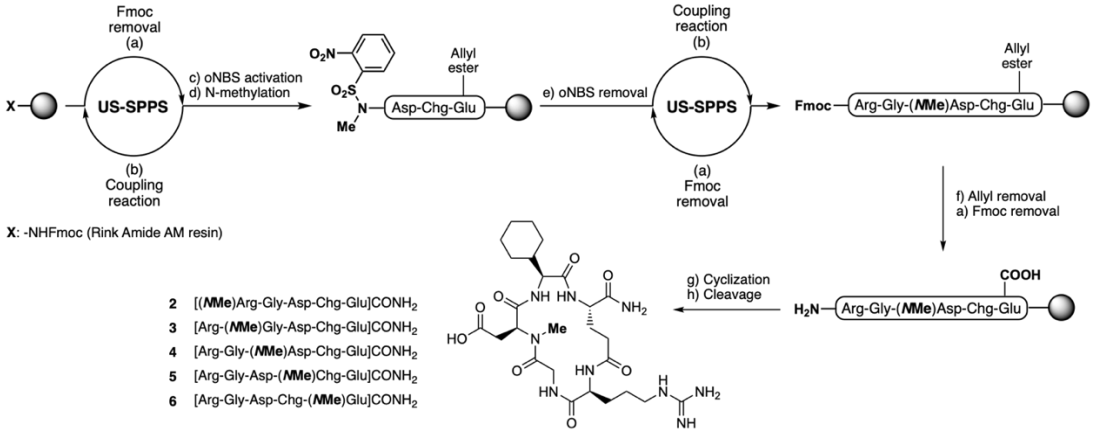
4.2.2.1 Synthesis

The synthesis of peptides **2-6** was performed on a solid support following a Fmoc/*t*Bu approach and an ultrasound-assisted solid-phase protocol (US-SPPS) previously reported by some of us (in Scheme 1, the synthesis of **4** is reported). [194] Notably, the methylation step was obtained by activating the primary amine as *ortho*-nitrobenzensulfonylamide (*o*NBS-amide) and then alkylating with dimethylsulfate (DMS) and DBU as base in NMP. The release of the secondary amine was then achieved thanks to a treatment with mercaptoethanol and DBU as scavenger mixture.

Cyclization step was then carried out on solid support in standard conditions and previously removing allyl and Fmoc protective groups, from glutamate

side chain and N-terminal of the sequence respectively. Cleavage from the Rink amide AM resin in acidic conditions afforded the crude mixture that was purified by reverse phase preparative HPLC. According to the aminoacidic position to methylate, the same protocol was employed for the five new peptides.

Scheme 4.1 Synthetic strategy for compounds **2-6**. Taken from Ref. 195.



Conditions and reagents: a) Piperidine 20% in DMF, 2 × 1 min, US irradiation; b) Fmoc-AA-OH, HBTU, HOBt, DIPEA, DMF, 5 min, US irradiation; c) *o*NBS chloride, TEA, dry DCM, rt, 2 × 30 min; d) Dimethylsulfate, DBU, dry NMP, room termperature, 2 × 30 min; e) Mercaptoethanol, DBU, dry DMF, room temperature, 3 × 15 min; f) TetrakisPd^0, DMBA, DCM/DMF 2:1, 2 × 60 min g) PyAOP, HOAt, DIPEA, DMF, room temperature, 6 h; h) TFA/TIS 95:5, room temperature, 3 h.

4.2.2.2. Binding affinities

A competitive ELISA assay based on immobilized ECM protein and soluble integrin[191] was employed to quantify the binding affinities of the newly synthesized peptides **2-6** and of the stem peptide **1**. All the compounds were

tested toward αvβ6 and αvβ8, whereas only compounds **1** and **6** also on αvβ3 and α5β1 integrin receptors.

Table 1. Evaluation of the binding affinities of **2–6** plus the stem peptide **1** for αvβ6 and αvβ8 integrin subtypes. Compounds **1** and **6** were also tested on αvβ3 and α5β1. Taken from Ref. 195.

	IC_{50} [nM]				
	Sequence	**αvβ6**	**αvβ8**	**αvβ3**	**α5β1**
1	[Arg-Gly-Asp-Chg-Glu]-$CONH_2$	1.3 ± 0.1	174 ± 31	364 ± 96	105 ± 11
2	[(*N*Me)Arg-Gly-Asp-Chg-Glu]-$CONH_2$	105 ± 8	2252 ± 89	n.d.	n.d.
3	[Arg-(*N*Me)Gly-Asp-Chg-Glu]-$CONH_2$	211 ± 26	3319 ± 122	n.d.	n.d.
4	[Arg-Gly-(*N*Me)Asp-Chg-Glu]-$CONH_2$	> 5000	4687 ± 570	n.d.	n.d.
5	[Arg-Gly-Asp-(*N*Me)Chg-Glu]-$CONH_2$	> 5000	> 5000	n.d.	n.d.
6	[Arg-Gly-Asp-Chg-(*N*Me)Glu]-$CONH_2$	1.6 ± 0.1	60 ± 2	1199 ± 121	112 ± 26
cilengitide[a]		n.d.	n.d.	1.4 ± 0.1	22 ± 1
RTDLDSLRT[b]		38 ± 7	122 ± 38	n.d.	n.d

[a]cilengitide was used as an internal reference compound in αvβ3 and α5β1 ELISA assays.

[b]RTDLDSLRT was used as an internal reference compound in αvβ6 and αvβ8 ELISA assays.

Interestingly, the *N*-methylation of amino acids in the parent pentapeptide **1** turned out to be mostly detrimental for the binding to αvβ6, with the sole

exception of the Glu^5 to (*N*Me)Glu^5 modification (**6**), which allowed this compound to maintain an IC_{50} value (1.6 nM) comparable to that of **1** (1.3 nM). Similarly, each *N*-methylation resulted in a drop of ligand binding affinity for αvβ8 except for **6** which proved to be three times more potent than **1** (60 *vs.* 174 nM). Considering the increased αvβ8 affinity, the selectivity profile of **6** towards the structurally related integrins αvβ3 and α5β1 was also evaluated. Interestingly, **6** displays no significant binding affinity for αvβ3, whereas a residual binding for α5β1, comparable to that of the parent compound **1** (112 *vs.* 105 nM), was detected. Thus, through the *N*-methylation of Glu^5, we were able to transform the αvβ6-monoselective peptide **1** in a novel αvβ6/αvβ8 dual ligand, albeit still endowed with a slight binding preference for the former receptor.

4.2.2.3 Conformational Studies

To identify the *in solution* conformation of **6**, we here employed a combined NMR-MD approach. Particularly, a solution of **6** in DMSO was used to acquire the following NMR spectra: 1D ^1H, 2D ^1H -^1H TOCSY (TOtal Correlation SpectroscopY, tmix= 60ms), 2D ^1H-^1H ROESY (Rotating-frame Overhauser Effect SpectroscopY, tmix= 300 ms, spin-lock at 2.8 kHz), 2D ^1H -^1H NOESY (Nuclear Overhauser Effect SpectroscopY, tmix= 200 ms). Indeed, the NOE

cross-peaks volumes were converted to ^{1}H-^{1}H internuclear distances using the linear approximation method with the geminal protons distance of (*N*Me)Glu^{5} Hβa-Hβb (fixed at 1.75 Å) taken as reference. These calculated distances, together with the $^{3}J_{HN\text{-}H\alpha}$ scalar couplings obtained from monodimensional spectra, were incorporated as structural restraints in a 5 μs-long replica-averaged molecular dynamics (RAMD) simulation. As thoroughly explained in paragraph 2.5.1, RAMD allows for an accurate description of the underlying structural ensemble of peptides and proteins by averaging experimental restraints over multiple parallel replicas of the system, according to the maximum entropy principle.[64,65,196,197] As for peptide **6**, RAMD calculations identified the presence of a β II' turn-like motif centered around Gly^{2}-Asp^{3} as proved by the analysis of the dihedral angles ((φ^{i+1},ψ^{i+1},φ^{i+2},ψ^{i+2}) = (63.4, -143.46, -83.98, -1.25)) of these residues (Figure 4.2). Interestingly, monitoring the interatomic distance between Arg^{1}-CO and Chg^{4}-NH (averaged over the multiple replicas at each step of RAMD, Figure 4.3), we noticed that no stable intramolecular hydrogen H-bond is formed between these two groups, contrary to what is observed in the NMR conformation of the stem peptide **1.** These data are in agreement with

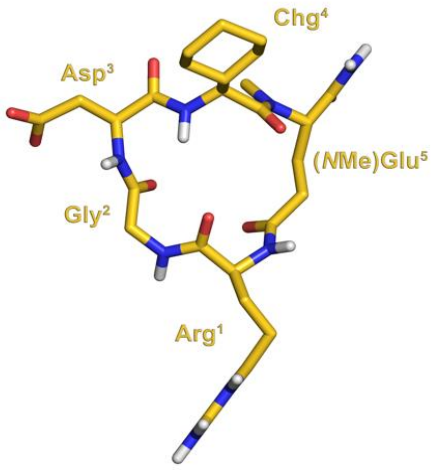


Figure 4.2 NMR-derived conformation of **6.** Figure taken from Ref. 195.

the lower temperature coefficient calculated for Chg[4]-NH of **6** (-7.0 ppb/K) compared to **1** (-0.3 ppb/K), indicating a higher solvent accessibility of the amide group of this residue in the newly synthesized peptide.

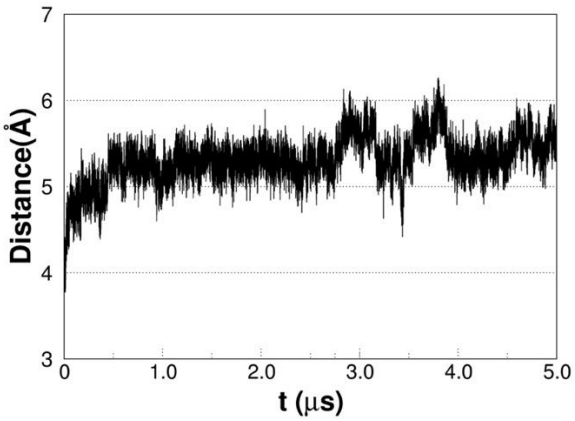


Figure 4.3. Ensemble-averaged interatomic distance between Arg[1]-CO and Chg[4]-NH over the RAMD simulation. Figure taken from Ref. 195.

Moreover, if we compare the tridimensional arrangements of **1** and **6** (Figure 4.4) we can observe how the *N*-methylation of the Glu[5] amide backbone induces changes in the dihedral space of the adjacent Chg[4] residue, which shift from (φ,ψ) = (-89.8, 8.9) in **1** to (φ,ψ) = (-127.2, 80.8) in **6**. These differences result into a shift in the Chg[4] side chain orientation in the newly synthesized peptide whose impact on the peptide binding conformation will be better analysed in the next paragraph.

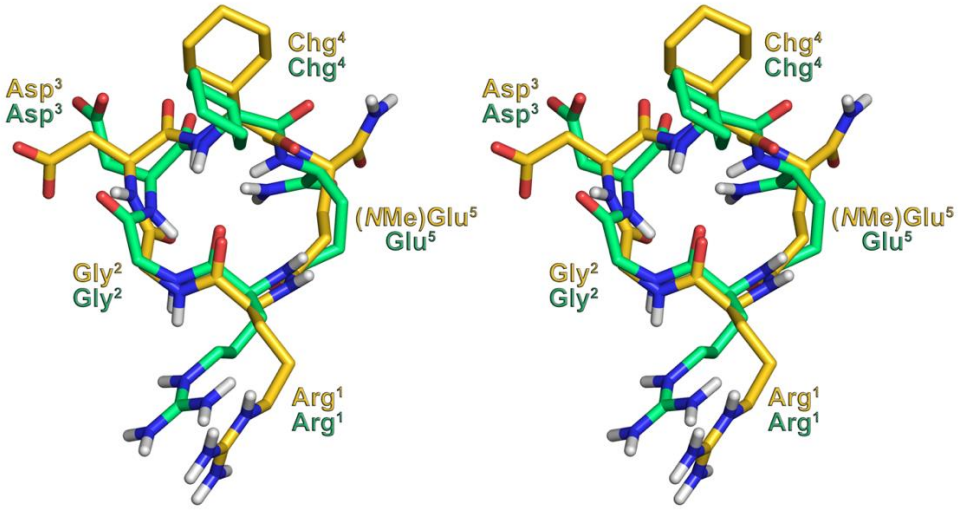


Figure 4.4. Stereo-view of the superposition between the NMR-derived structures of **1** (green sticks) and **6** (gold sticks). Non-polar hydrogens are omitted for sake of clarity. Figure taken from Ref. 195.

4.2.2.4 Molecular Modelling

At this point, we investigated the molecular basis for the integrin activity and selectivity profile of **6** by performing docking calculations in the crystal structures of $\alpha v\beta 6$[198] and $\alpha v\beta 8$[199]. Since recent evidence suggests that the use of the solution NMR structure in docking can improve its ability to reproduce the receptor-bound conformation of small RGD cyclopeptide ligands, we selected as starting point for our calculations the RAMD-predicted structure of **6**.[200] The best docking pose of **6** in both $\alpha v\beta 6$ and $\alpha v\beta 8$ shows a typical RGD mimetic interaction pattern (Figure 4.5). In detail, the divalent cation at the receptor MIDAS site is chelated by the carboxylic acid of the ligand Asp^3, which also forms H-bonds with the backbone of (β6)-N209 and (β8)-N207 in $\alpha v\beta 6$ and $\alpha v\beta 8$, respectively (Figure 4.5); in parallel, a side-on tight salt-bridge

is established by the Arg[1] side chain of **6** with the conserved αv-D218 residue. In addition, the Chg[4] side chain is well-hosted in the pocket identified by the specificity-determining-loop (SDL) loop, where it forms multiple lipophilic contacts with residues of (β6)-A117, (β6)-L174, (β6)-Y176, (β6)-A208 and (β6)-I210 in αvβ6, and (β8)-A115, (β8)-Y172, (β8)-L174, and (β8)-I208 in αvβ8, respectively. Indeed, these binding modes appear coherent with the the low-mid nanomolar IC_{50} values exhibited by **6** toward αvβ6 and αvβ8.

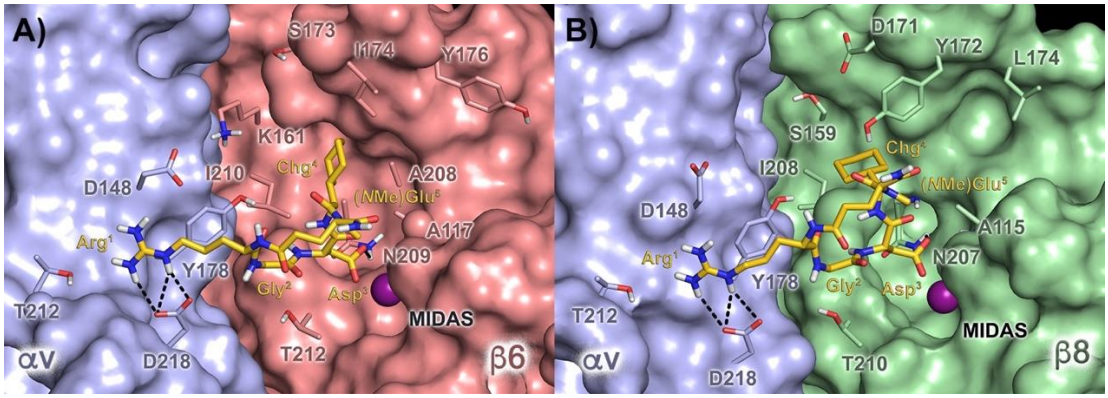


Figure 4.5 Docking poses of **6** (gold sticks) at the (A) αvβ6 (PDB code: 5FFO)[198] and (B) αvβ8 (PDB code: 6OM2)[199] integrins. The αv, β6 and β8 subunits are depicted as light blue, red and green surfaces, respectively. The amino acid side chains important for the ligand binding are represented as sticks. The metal ion at the MIDAS is represented as a purple sphere. Hydrogen bonds are shown as black dashed lines. Figure taken from Ref. 195.

Then, we attempted to disclose the structural bases for the improved αvβ8 affinity of **6** with respect to its parent peptide **1**. First, we performed an accurate comparison between the RGD binding site of αvβ6 and αvβ8 to detect similarities and discrepancies in the two pockets. Notably, the SDL cavity of αvβ8 appears more steric hindered then in αvβ8 due to the replacement of (β6)-I174 with (β8)-Y172. Indeed, this residue might clash with the cyclohexyl

ring of **1** not allowing this peptide to properly accommodate in the β8 subunit, as suggested by the superposition of the **1**/αvβ6 docking complex with the αvβ8 X-ray structure (Figure 4.6). Conversely, in the **6**/αvβ8 docking complex this clash does not occur due to the different orientation assumed by the Chg[4] side chain in the *N*-methylated compound, as also revealed by NMR analysis. In conclusion, our interaction studies suggest that the changes in the peptide conformation induced by the Glu[5] *N*-methylation, together with single point mutations in the SDL cavity of the β6 and β8 subunits, are responsible for the different selectivity profile of compounds **1** and **6**. This, in turn, further proves how minimal chemical modifications in small cyclic peptides can account for large differences in their binding affinities.

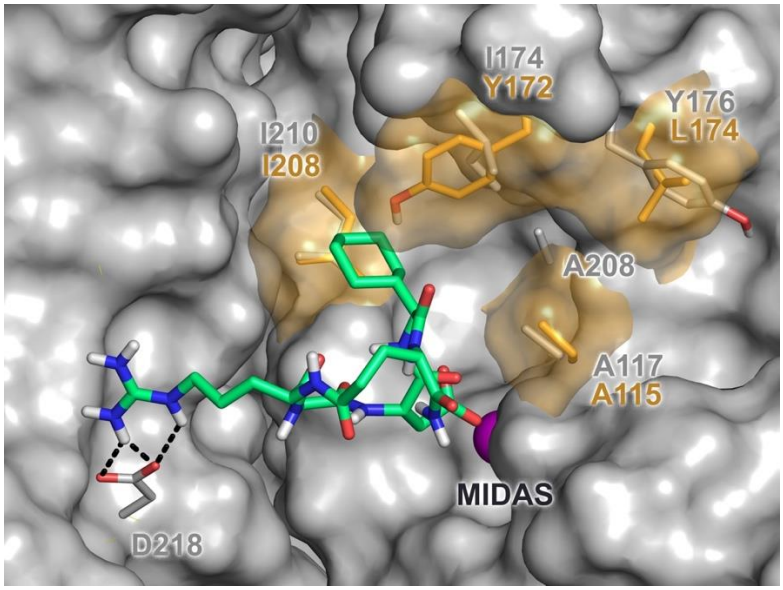


Figure 4.6. Superposition of the **1**/αvβ6 (PDB code: 5FFO)[198] docking complex with the αvβ8 X-ray structure (PDB code: 6OM2)[199]. The peptide is shown as green sticks. The αvβ6 and αvβ8 receptors are represented as gray and orange surfaces, respectively. In both integrin subtypes, residues that are important to selectivity are highlighted as sticks and transparent surfaces for αvβ6 and αvβ8, respectively. The metal ion at MIDAS is shown as a purple sphere. Figure taken from Ref. 195.

4.2.2.5 Biological Evaluation

As described in paragraph 4.2.1, the HSV entry into cells is favored by the interaction of HSV-1 gH/gL to either αvβ6 or αvβ8, which results, in turn, in the dissociation of gL from the parent heterodimer.[180] Thus, we performed in cell experiments to evaluate if our dual αvβ6/αvβ8 ligand **6** is able to impair the HSV-1 infectivity. In these tests we put in comparison **6** with the αvβ6 mono-selective stem peptide **1** and with the well-characterized αvβ3/αvβ5 integrin ligand *cilengitide,* using as positive control R1.302, a nectin1 neutralizing mAb. In the first set of assays, we treated 293T cells, expressing both αvβ6 and αvβ8, alternatively with increasing concentrations of our integrin-binding peptides and R1.302. The cellular exposure to the ligands was performed before and during the infection by a recombinant HSV-1 strain, namely R8102, carrying a lacZ reporter gene under the control of the α27 promoter, whose expression analysis allows an easy quantification of the infectivity as measurement of the β-galactosidase activity.[201]

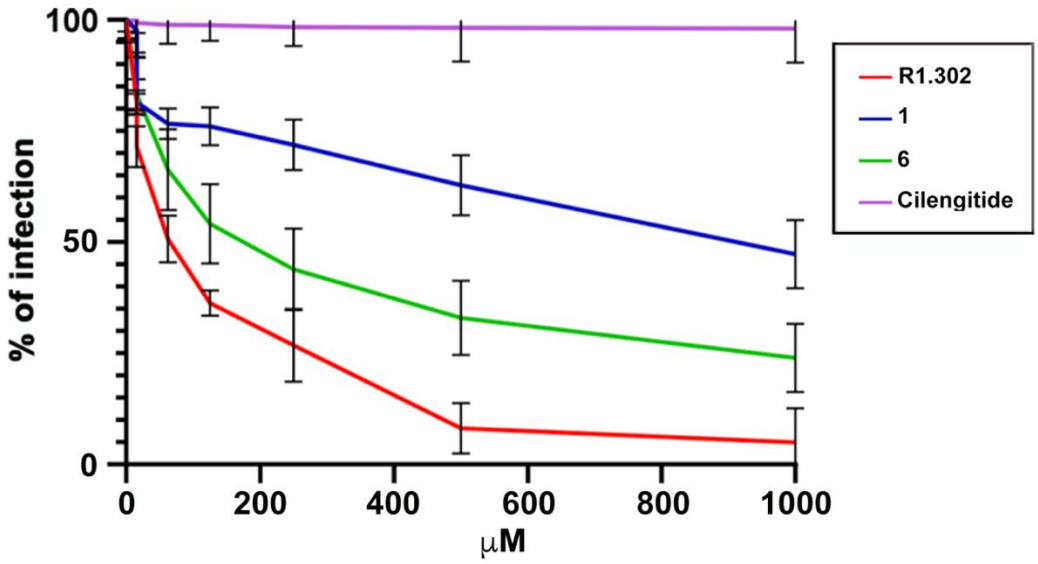


Figure 4.7 Inhibition of HSV-1 infection. 293T cells were exposed to increasing concentrations of the indicated peptides for 1 h prior to infection and during virus attachment for other 90 min. The infection was induced using the recombinant R8102 HSV-1 strain, and measured after 8 h as β-galactosidase activity, using ONPG as substrate. The assays were run in triplicate. Bars show standard deviation (SD). Figure taken from Ref. 195.

Although both **1** and **6** were found able to inhibit HSV-1 infection in a dose-dependent manner (Figure 4.7), **6** (from 70% to 80% at 500 to 1000 μM peptide concentration) was significantly more effective than the parent compound **1** (50% maximum inhibition at 1000 μM). Moreover, the αvβ6/αvβ8-related inhibitory activity of **1** and **6** was testified by the fact that no alteration in the HSV entry process was detected after cells' treatment with *cilengitide*. These preliminary results were in agreement with the known interchangeable and additive roles played by αvβ6 and αvβ8 integrins upon HSV-1 infection.[181] However, since 293T cells express both αvβ6 and αvβ8, we needed further experiments to validate our hypothesis that **1** and **6** exerted different effects

on HSV infection by selectively binding to αvβ6 (**1**) and by simultaneously targeting αvβ6 and αvβ8 (**6**), respectively. To this aim, we set up a second set of experiments where we employed J cells, which do not physiologically express gD receptors. In fact, J cells can be infected by HSV-1 only if gD cognate receptors (i.e. nectin1) are transgenically expressed. Similarly, in these cells, endogenous hamster integrins are present at low levels; thus they can be selectively engineered with αvβ6 or αvβ8 to evaluate their role in the viral infection. Indeed, we transfected J cells with low amounts of nectin1 plasmid, plus either αvβ6 or αvβ8 integrin plasmid or both, and then infected with the recombinant K26GFP HSV-1 strain. Notably, we measured the viral infection by quantifying the cellular presence of the viral capsid protein ICP26, which, in this strain, is fused to the enhanced green fluorescent protein (EGFP) [202]. Particularly, we evaluated the K26GFP cellular penetration by fluorescent microscopy (Figure 4.8, A-T) and the EGFP expression trough flow cytometry as mean fluorescence intensity (MFI) of gated cells (Figure 4.8, U). The assays in J cells expressing nectin 1 alone (Figure 4.8, A-E) confirmed the integrin specificity of **1** and **6**. Here, in fact, only treatment with R1.302 (Figure 4.8, B) prevented HSV-1 infections. On the other hand, in J cells engineered with nectin1 plus both αvβ6 and αvβ8 integrins (Figure 4.8, P-T), the treatment with R.1302 (Figure 4.8, Q), **6** (Figure 4.8, S), and, only partially, compound **1** (Figure 4.8, R) reduced the HSV-1 infectivity.

In parallel, the anti-HSV activity of our peptides was also assessed in J cells expressing nectin1 plus either αvβ6 or αvβ8 alone. Interestingly, when only the αvβ6 plasmid was transfected (Figure 4.8, F-J), both **1** and **6** altered the virus entry (Figure 4.8, H and I); conversely, in J cells exclusively expressing nectin1 plus αvβ8 (Figure 4.8, K-O), inhibitory effects were detected only in presence of **6** (N). We also remark that in all the examined samples the R1.302 mAb blocked HSV entry, whereas no effect was exerted by *cilengitide*. Altogether, from these experiments we can conclude that **1** and **6** can block the HSV-1 cellular penetration by hampering the interaction of viral gH with αvβ6 and αvβ8, and that a simultaneous targeting of both these receptors, like in the case of compound **6**, can bring significant advantages in terms of drug efficacy.

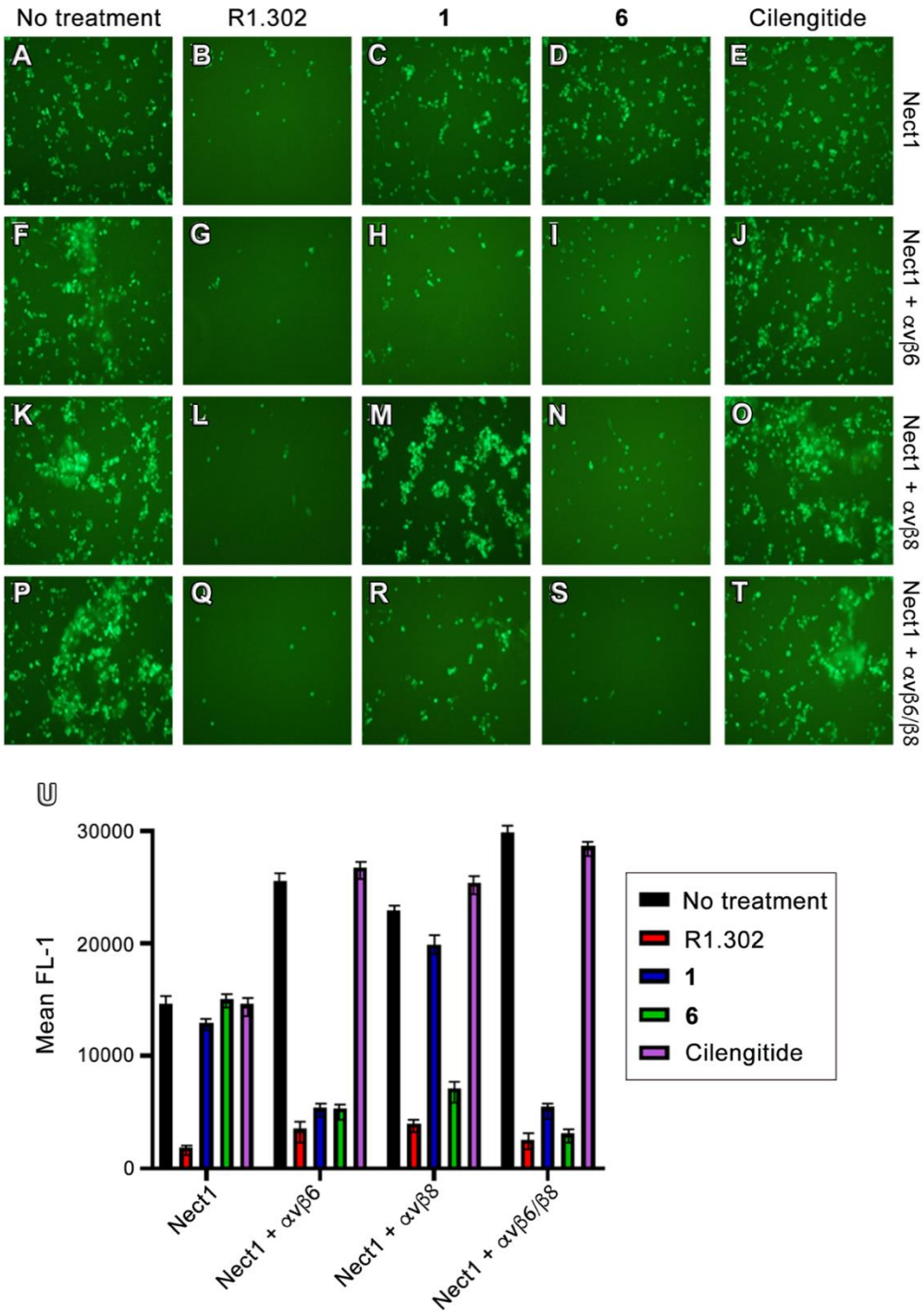


Figure 4.8. Inhibition of HSV-1 infection by peptides. J cells were transfected with low amount (75 ng DNA/24 well) of nectin1 alone (A-E), or with the same amount of nectin1 plus αvβ6 integrin (300 ng DNA/24 well) (F-J), or plus αvβ8 integrin (300 ng DNA/24 well) (K-O) or plus both the integrin receptors (P-T). 48 h after transfection, cells were exposed to 700 μM of peptides (**1**, **6** and cilengitide) for 1 h prior to infection and for 90 min during virus attachment. Cells were infected with K26GFP. Non penetrated virus was inactivated by an acid wash. Infectivity was measured at 16 h after infection as EGFP

expression. A-T panels show the EGFP expression in each sample for a typical experiment. (U) K26GFP infection was quantified as EGFP protein expression in flow cytometry assay as mean fluorescence intensity (MFI) of gated cells. Histograms represent the average of triplicates ± SD. Figure taken from Ref. 195.

At this point of the project, a question regarding the mechanism of action of our peptides was raised by an accurate comparison of our data with the literature. Indeed, it is known that while gH directly interacts with αvβ6 through its RGD domain, thus competing with our peptides for the integrin binding, αvβ8 does not contact gH by recognizing its RGD triad.[38] *Therefore, which alternative mechanism is responsible for the antiviral activity of* ***6*** *on J cells expressing αvβ8 integrin alone and for its higher protective effects on the αvβ6/β8 positive cells?* In fact, we hypothesized that **6** could prevent HSV-1 entry into J cells by inducing the internalization of integrins. To validate this idea, we engineered J cells with nectin1 and the two integrin αvβ6 and αvβ8, and then we incubated these cells with **1** or **6** for 60 min at 37°. The surface expression of integrin β6 and β8, in the presence or absence of peptides, was evaluated by flow cytometry using the FAB4155A (mAbβ6) and FAB4775A (mAbβ8) monoclonal antibodies that recognize other integrin regions than the RGD binding domain.

As shown in Figure 4.9, while both the peptides were able to significantly reduce the membrane expression of integrin β6, only **6** decreased the presence of β8 on the cellular surface, in agreement with what already observed in

Figure 4.7 and 4.8. Notably, the surface expression of nectin1 is not altered following treatment with either peptide. These outcomes indicate that both **1** and **6**, once bound to the RGD binding domain of the targeted integrin, determine the receptor internalization. This results in a lower expression of the two integrins at the cell surface and, accordingly, in a reduced probability to be used as receptors or co-receptors by HSV.

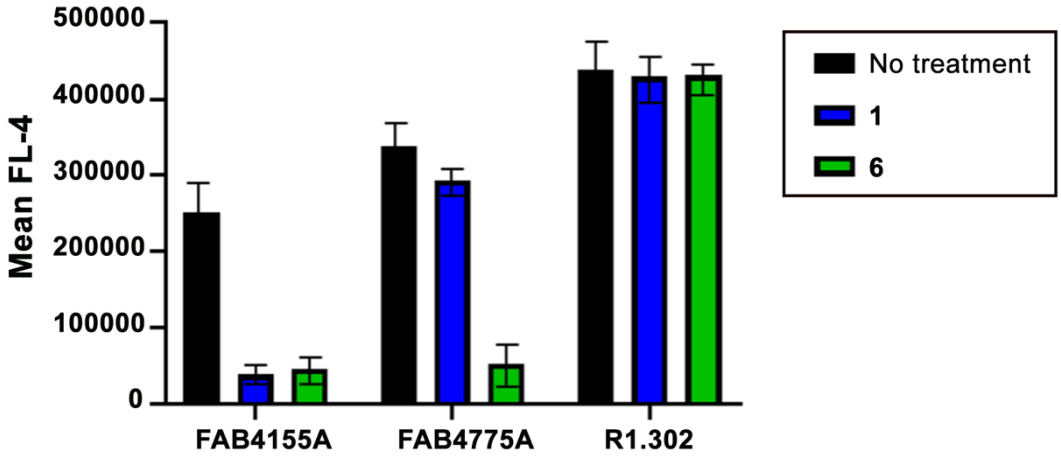


Figure 4.9. Integrins internalization assay. J cells were transfected with low amount (75 ng DNA/24 well) of nectin1 plus αvβ6 integrin (300 ng DNA/24 well) and αvβ8 integrin (300 ng DNA/24 well). 48 h after transfection, cells were exposed to 700 μM of peptides for 1 h at 37° C. Cells derived from samples treated or not treated with peptides were incubated for 1 h at 4 °C with the FAB4155A (mAbβ6), FAB4775A (mAbβ8) and R1.302 (mAbnectin1) mAbs. Samples incubated with R1.302 were washed and subsequently incubated 1 h at 4 °C with APC Mouse secondary antibody. Integrin and nectin1 surface expression was quantified as flow cytometry expression of APC mean fluorescence intensity (MFI) of gated cells. Histograms represent the average of triplicates ± SD. Figure taken from Ref. 195.

To further prove this hypothesis, we tested if our peptides could prevent the interaction between a soluble form of gH (gH_{solST}) and the cellular membrane in a binding assay on J cells transfected with nectin1 and the two integrins αvβ6

and αvβ8. The control used in this assay to confirm the integrin specificity of **1** and **6** was a soluble form of gB, whose binding to the membrane does not depend on αvβ6 and αvβ8 but only on the heparan sulfate. Figure 4.10 shows that the binding of gH_{solST}, but not of gB_{solST}, is strongly reduced in cells treated with **6** and only partially reduced in cells treated with **1**, indicating that integrin internalization induced by our peptides determines a weak capacity of binding gH which, in turn, attenuates HSV-1 infection.

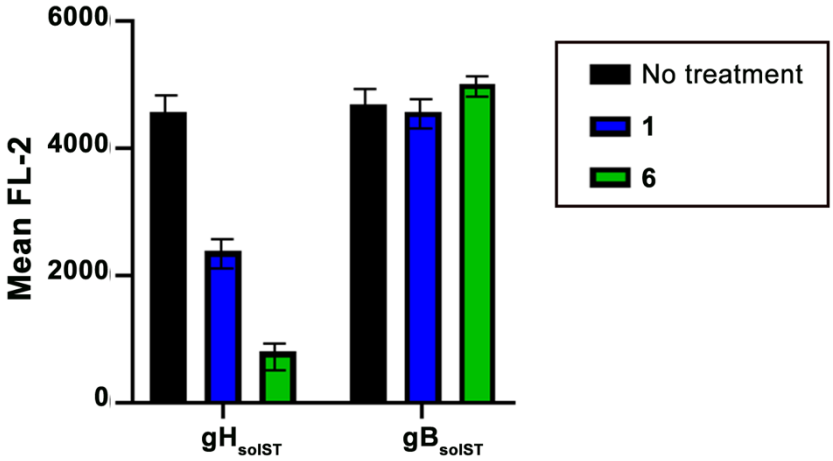


Figure 4.10. gH binding assay. J cells were transfected with low amount (75 ng DNA/24 well) of nectin1 plus αvβ6 integrin (300 ng DNA/24 well) and αvβ8 integrin (300 ng DNA/24 well). 48 h after transfection, cells were exposed to 700 μM of peptides for 1 h at 37° C. Cells derived from samples treated or not treated with peptides were incubated for 1 h at 4 °C with gH_{solST} or gB_{solST}. Samples were washed and subsequently incubated 1 h at 4 °C with PE-conjugated MAb to the One-STrEP tag (Strep-Tactin). gH_{solST} and gB_{solST} binding to cell surface was quantified as flow cytometry expression of PE mean fluorescence intensity (MFI) of gated cells. Histograms represent the average of triplicates ± SD. Figure taken from Ref. 195.

4.3. Conclusions

In this work, we proved that HSV infection can be impaired through an affordable pharmaceutical approach, based on the use of small, αvβ6/αvβ8 dual, RGD-containing cyclic pentapeptides. First, we developed a small library of *N*-methylated derivatives of the αvβ6 specific ligand [RGD-Chg-E]-$CONH_2$ (**1**) recently discovered by us[183] , testing their affinity and selectivity on a selected RGD integrin panel. Among the newly synthesized peptides, [RGD-Chg-(*N*Me)E]-$CONH_2$ (**6**) displayed an increased αvβ8 affinity compared to the parent ligand, representing one of the most potent αvβ6/αvβ8 dual ligand discovered so far. Extensive computational studies, adjuvated by NMR experiments, were fundamental to disclose the molecular bases of the increased αvβ8 potency of **6** compared to the stem peptide **1**. Furthermore, **1** and **6** underwent an extensive biological evaluation, which attested the anti-herpetic properties of both compounds. Nonetheless, **6** showed a higher efficacy than **1** in preventing HSV cellular penetration, highlighting the importance of simultaneously targeting both αvβ6 and αvβ8 to increase the antiviral activity. Moreover, we discovered a peculiar mechanism of action of our ligands, based on the stimulation of receptor internalization rather than competing with the viral glycoproteins for the binding to the canonical RGD site. We expect that this work can have a broad impact in the medicinal chemistry community, since the present drug therapy to counteract HSV-1 is mainly based on the use of Aciclovir (or its pro-drugs) which acts by blocking

viral replication. This treatment, albeit effective, does not hinder the entry of the pathogen into the host and therefore carries frequent side effects, such as periodic virus reactivation events and the insurgence of resistance phenomena.[203] Thus, the use of small peptides as adjuvant inhibitors of the HSV-1 entry might compensate for some of the drawbacks of current therapeutic regimens. Altogether, our outcomes encourage the development of $\alpha v\beta 6/\alpha v\beta 8$ dual compounds as a new potential therapeutic approach to the HSV disease and, at the same time, open up to the possibility of employing small ligands of RGD integrins as weapons against a wide range of viruses that employ this class of receptors as gateways to invade host cells.

4.2.4. Methods

4.2.4.1 Replica Averaged Molecular Dynamics

^{1}H-^{1}H internuclear distances and ^{3}J scalar couplings data coming from NMR experiments were incorporated in the framework of molecular dynamics as structural restraints averaged over ten parallel replicas of the system, starting from randomly generated conformations. The simulations were performed using the GROMACS 2018.8[160] code patched with PLUMED 2.5.3[165,166]. The peptide was built with the Maestro Suite 2019[157] and then solvated in a 12 Å layer cubic DMSO box. The ff14SB[158] Amber force field was used to parametrize the peptide, whereas the parameters for the solvent box were

taken from a previous work by Fox & Kollman.[204] Atom types and bonded parameters for the non-natural Chg amino acid were taken by homology from the Amber force field, whilst its atomic partial charges were predicted using the two-stages restrained electrostatic potential (RESP)[115] fitting procedure implemented in Antechamber.[117] Prior to the RESP fitting, the electrostatic potentials (ESP) were computed with the aid of the quantomechanical package Gaussian16.[116] A geometric optimization procedure was performed in two steps at Hartree-Fock level of theory:

a first calculation with the 3-21G basis set, followed by a more accurate refinement with the 6-31G* basis set, after which the ESP potentials were computed. The topology files of the systems were generated with the tleap program of AmbertTools19 and then converted into Gromacs format with the ParmEd tool. During the simulations, a time step of 2 fs was employed, while the bonds were constrained using the non-iterative LINCS algorithm.[205] A cutoff of 12 Å was chosen for the evaluation of the short-range non-bonded interactions, whereas the long-range electrostatics ones were treated through the particle mesh Ewald[53] method, using a 1.0 Å grid spacing in periodic boundary conditions. The system was minimized through 10,000 steps of the steepest descent algorithm. Then, the simulation box was equilibrated and heated up to 300 K, alternating NPT and NVT cycles with the Berendsen[162] coupling bath and barostat. Finally, 500 ns long production runs were

performed for each replica in the NPT ensemble, resulting in a total simulation time of 5 μs. During the production runs pressure of 1 atm and temperature of 300 K were kept constant using the stochastic velocity rescaling[163] and the Parrinello-Rahman[164] algorithms, respectively. Finally, the trajectories were clustered based on the peptide backbone RMSD and the centroid of the largest population was selected as representative structure of the NMR ensemble.

4.2.4.2 Molecular Dockings

The NMR-predicted conformation of **6** was docked in the crystal structures of either αvβ6 or αvβ8 receptor in complex with proTGF-β (PDB code: 5FFO and 6OM2, respectively)[198,199]; the cyclic peptide backbone was treated as rigid, whereas the side chains were kept flexible. The peptide and the receptors were prepared with the aid of the Protein Preparation Wizard tool as in previous papers.[111,206] Missing hydrogen atoms were added and all the water molecules were deleted from the receptor structure. Regarding the metal cofactors at the protein MIDAS, ADMIDAS (only in αvβ6) and LIMBS sites, the co-crystalized divalent cations Mg^{2+}, Mn^{2+} and Ca^{2+} were retained during calculations. Then, we predicted the side chains ionization and tautomeric states using Epik.[109,110] Prior to docking, the receptor was refined optimizing its hydrogen-bonding network and minimizing the position of the hydrogens. As for the grid generation, a virtual box of 25 Å × 25 Å × 25 Å, centered on the integrin binding

site, was computed through the Receptor Grid Generator tool of Glide 8.5.[21,22] Finally, docking calculations were performed using the Glide SP-peptide default parameters and the OPLS3A force field.[44] The docking solutions were then analyzed based on the Glide docking score leading to the selection of the two top-ranked poses (Gscore: -8.320 in αvβ8 and 7.855 in αvβ6).

4.3 Elucidating the folding and binding properties of the iRGD peptide

4.3.1 Introduction

Today, the development of effective anticancer treatments is frequently hampered by either the poor tumour penetration or the lack of selectivity over healthy cells of chemotherapeutic agents.[207,208] In fact, anticancer drugs often need to be administered at high doses to exert relevant pharmacological effects, with the rise of serious adverse reactions limiting their employment in large-scale therapeutic regimens.[209] A feasible solution to improve tissue penetration is represented by pharmaceutical carriers which can vehicle the desired drug as cargo to the extravascular cancer tissue. In recent years, drug carriers of different nature have been developed such as gold nanoparticles,[210–212] liposomes,[213,214] polymer micelles[215] or peptides.[216] In this context, Ruoslahti's group identified a cyclic nonapeptide, namely iRGD (*internalizing RGD*, CRGDKGPDC, **1** - Chart 4.2), endowed with remarkable selective tumour-homing activity.[217,218] Notably, this peptide has been conjugated with a wide range of chemotherapeutic agents to improve their anticancer effects.[219–223] Particularly, the combination of **1** with nabpaclitaxel and gemcitabine has shown very positive results in phase I clinical trial for the treatment of Metastatic Pancreatic Cancer.[224]

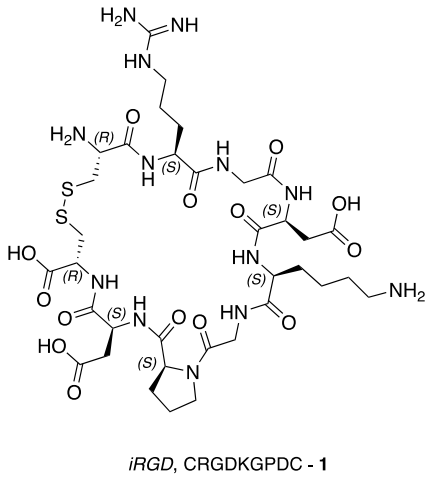


Chart 4.2 2D structure of iRGD.

Under the chemical point of view, **1** is characterized by the basic recognition pattern of RGD integrins, which is the typical arginine-glycine-aspartate (RGD) sequence. As already described in the previous paragraphs of the thesis, these receptors are overexpressed in many forms of cancers, playing critical roles both in vascular angiogenesis and in the tumour progression.[167,168,225] Therefore, RGD integrins have been widely exploited over the years to develop peptides, peptidomimetics, and small molecules for theranostic purposes. It is interesting to report that **1** shows unusual but striking properties compared to all the other known linear and cyclic RGD peptides. Particularly, it outperforms other RGD ligands as tumour-homing agent since it undergoes a multistage internalization process where the binding to integrins represents only the first step (Figure 4.11).[217] In fact, whilst the peptide is still anchored to the receptor surface, it experiences a

proteolytical cleavage which results in the exposure and the following release of the cryptic C-terminal CRGDK sequence. The latter peptide is recognized by Neuropilin-1 (NRP-1), a co-receptor to some tyrosine kinases such as the vascular endothelial growth factor receptor (VEGFR), and is then internalized together with this membrane protein.[226] The strict requirement for the C-terminal exposure of (R/KXXR/K) motif and for its binding to NRP-1 was first described by Ruoslahti and coworkers who termed this phenomenon as the C-end rule (CendR).[226] Indeed, a number of physiological ligands of NRP-1 like VEGF-A165 possess a C-terminal CendR sequence that interacts with this receptor triggering cellular internalization. In the case of iRGD, the interaction of the CendR motif with NRP-1 is not only responsible for its cellular uptake but also endows the peptide with intrinsic anticancer properties.[227] On the other hand, the tropism and selectivity of **1** for cancer tissues both depend on the peptide's binding to RGD integrins.[217] Notably, **1** was initially reported as a pan αv-integrins binder.[217] However, recent studies have identified the αvβ5 isoform as the mostly involved in its tumour-specific delivery. Particularly, in

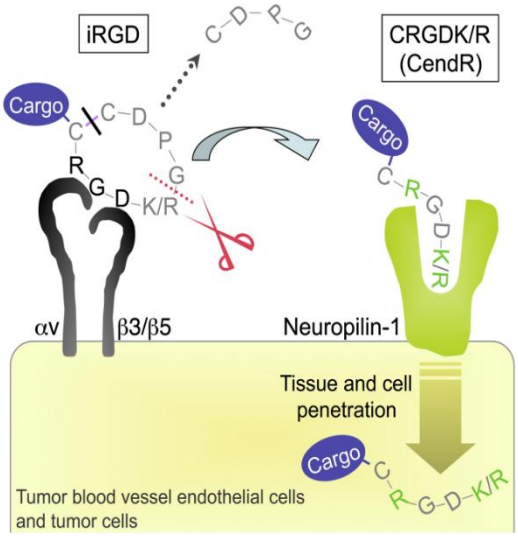


Figure 4.11 Schematic representation of the mechanism of action of **1** (iRGD). Figure taken from Ref.[217]

pancreatic desmoplastic dual adenocarcinoma (PDAC) the tumour-penetration of **1** linearly increased with the surface expression of αvβ5 on carcinoma-associated fibroblasts.[228] In fact, the authors suggested that **1** first interacts with αvβ5 positive connective cells, stimulating the release of soluble factors like TGF-β. The latter, in turn, increases the αvβ5 expression also on the neoplastic tissue, which then becomes more sensitive to **1**. Noteworthy, αvβ5 integrin governs critical events during tumour development and progression, such as angiogenesis, by activating the focal adhesion kinase–steroid receptor coactivator pathway.[229]

Despite the huge relevance of **1** in the field of anticancer drugs, the molecular basis of the peptide's interaction with integrin receptors are still completely missing. However, the elucidation of these aspects might allow fine-tuning or, more generally, further improving its tumour-homing properties. In this project, an advanced computational approach, combining bioinformatics and biosimulations, was thus employed to elucidate the interaction mode of iRGD to integrin receptors. Considering the recent outcome of αvβ5 as a major iRGD receptor in vivo, the study has been initially focused on this integrin isoform. Since no experimental 3D structure of this receptor is currently available, a reliable homology model has been initially built. Then, a core part of the work has consisted in the identification of the putative bioactive structure of **1** through extensive conformational studies. At variance with the previous case

study (4.2), a pure *in silico* approach, based on PT-WTE metadynamics, was here adopted. The obtained results strongly support the reliability of this technique in predicting the bioactive conformation of medium size peptides, especially in absence of any experimental support. Finally, docking and MD studies were performed to characterize the binding pose of **1** to its target receptor. The final results of this work, which is still ongoing, could drive the rational design of a new class of internalizing RGD integrins ligands. In particular, given the wide range of diseaes in which distinct integrin subtypes are involved, the understanding of how to tune the selectivity profile of iRGD-like molecules would be of great impact, paving the way for focused discovery campaigns of tumour-selective homing peptides.

4.3.2 Results

4.3.2.1 Conformational Sampling

In order to predict the putative bioactive conformation of **1,** we performed a PT-WTE metadynamics simulation which allows the sampling of events occurring on a long timescale such as peptide folding in an affordable computational time. This technique does require no prior knowledge of the system. In fact, the sampling is boosted by the combination of a typical parallel tempering scheme with a metadynamics (MetaD) bias potential deposited on

the potential energy of the system (WTE ensemble) (paragraph 2.5.2). Only afterwards, the user may want to define some collective variable (CV), to compute the free energy surface (FES) of the investigated event. In our case, 6 parallel replicas were employed to span the temperature range going from 300 to 450 K in the WTE ensemble. Each replica was simulated for 140 ns, for a total simulation time of 840 ns. At the end of the calculation, the MetaD bias of the replica at physiological temperature (300 K) was reweighted according to the *Tiwary-Parrinello*[67] algorithm and the FES was computed as a function of two CVs specifically selected for describing the folding event. First, the CV *Dihedral Correlation – Dih*$_{cor}$ was used to discriminate between the different conformational states assumed by the peptide backbone. Second, we designed a CV (*H-bond)* to account for any intramolecular backbone-backbone hydrogen bond (see paragraph 4.3.4.1 for details).

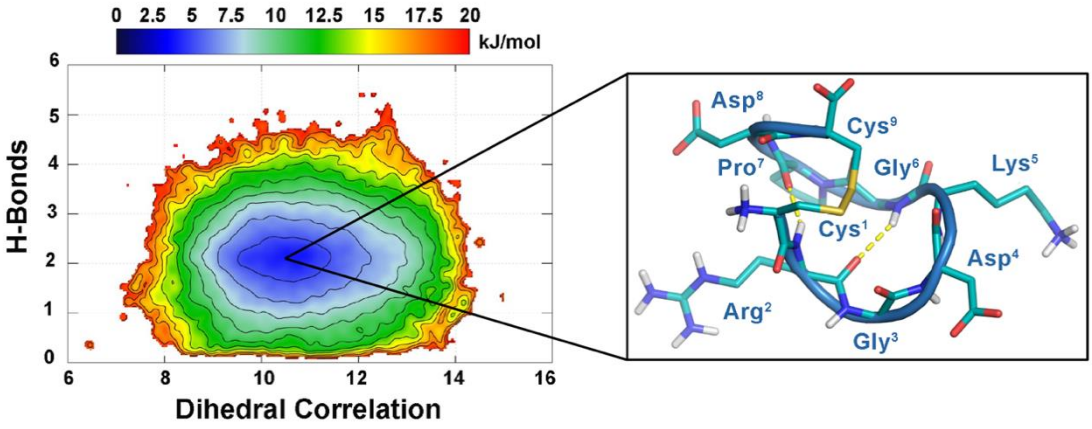


Figure 4.12 Free energy surface (FES) for the folding of 1 as a function of the Dih_{cor} and H_{bond} CVs with

isosurfaces displayed every 1.5 kJ/mol. The conformation representing the main free-energy minima is shown as inset.

Looking at the resulting FES (Figure 4.12), a single energy minimum can be identified. The structures contained in this energy basin were clustered, indicating the presence of a single predominant conformation (> 85% of occurrence). In the latter, compound **1** is folded in a peculiar *horseshoe-like* shape, stabilized by two hydrogen bonds formed by i) the carbonyl oxygen of Arg^{2} with the amide nitrogen of Gly^{6} and by ii) the carbonyl group of Pro^{7} with the amide nitrogen of Arg^{2}. The reliability of the PT-WTE-predicted conformation of **1** was assessed by checking the convergence reached by the simulation. First, the computation of the FES at regular time intervals (Figure 4.12A) highlighted that after the first 80 ns (per replica) of simulation the overall shape of the free energy landscape is conserved.

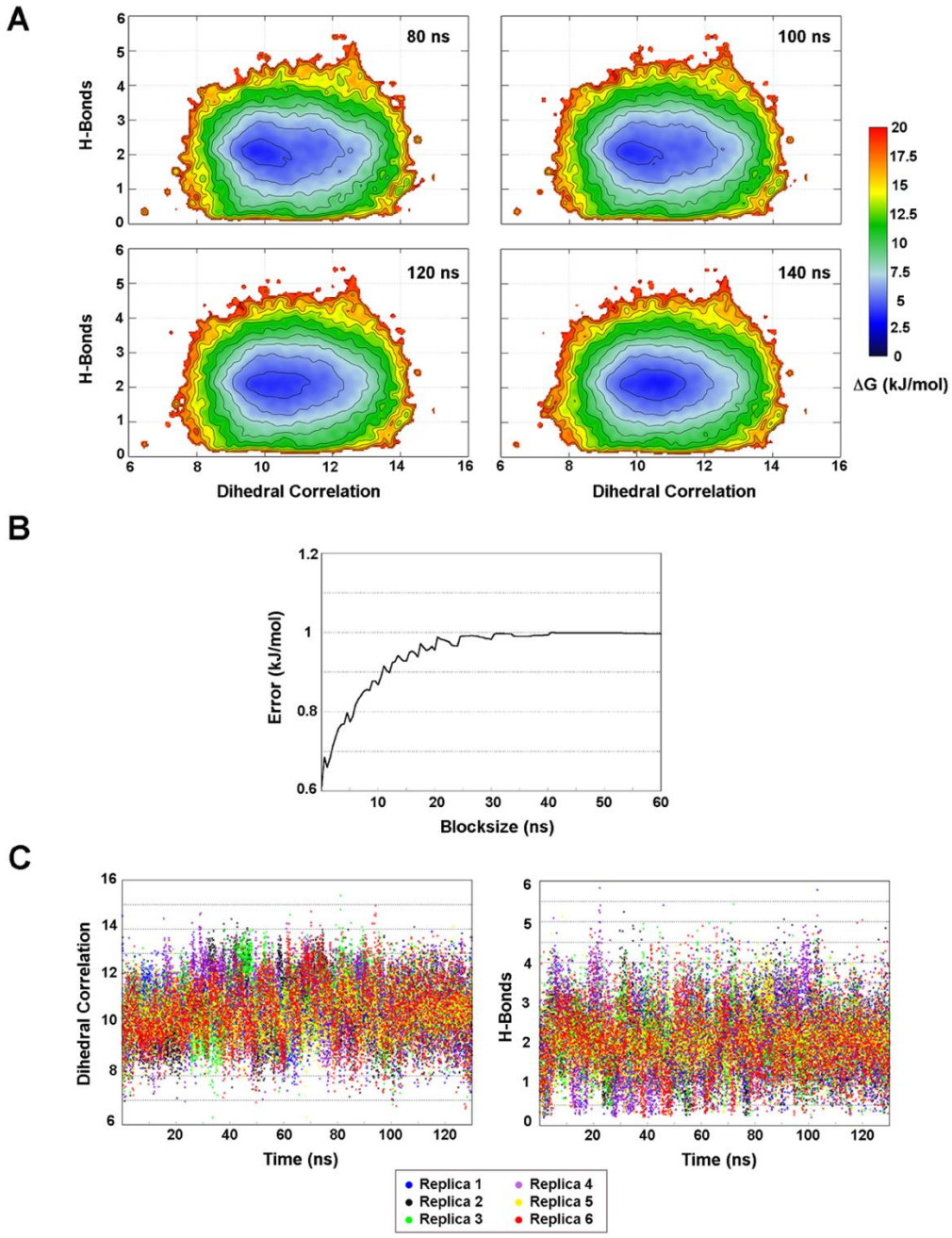


Figure 4.13 Convergence of PT-WTE calculation. A) Time evolution of the FES during the last 60 ns of simulation. B) Quantitave assessment of the error associated with the FES calculation trough block averages analysis. C) CVs diffusion in the six demuxed (continuous) trajectories.

Then, a block averaging analysis estimated the error associated with the ΔG computation (Figure 4.13B) in the acceptable value of ≈ 1 kJ/mol. Finally, the convergence of the parallel tempering was evaluated. Since the average

exchange acceptance ratio between neighbour replicas was 25%, all the replicas well diffused over the entire selected temperature range. To further prove the correctness of the adopted protocol, we monitored the values of the CVs (Figure 4.13C) employed for reweighting the FES in the continuous trajectories (demuxed) of each replica. Notably, from Figure 4.12C we can appreciate how in all the replicas the two CVs have a quite diffusive behaviour, suggesting that no simulation was stuck in a particular region of the phase-space thanks to the boost coming from PT and WTE. This way, the PT-WTE simulation allowed an effective sampling of the entire folding phenomenon, with many recrossing events between the different peptide conformational states.

4.3.2.2 Binding mode studies

The solution conformation of **1** predicted by PT-WTE was then used as starting point for extensive binding mode studies in αvβ5. For this task, a computational protocol based on preliminary docking calculations followed by microseconds MD simulations was employed.

Prior to docking, a homology model of αvβ5 was built due to the lack of any experimental 3D structure of this receptor in the PDB. In particular, the β5 subunit was modelled since the αv counterpart had already been solved in

many X-ray structures of the αvβ3, αvβ6 an αvβ8 isoforms.[198,199,230–232] To build the model, a multiple sequence alignment of all the human RGD-binding β subunits, namely β1, β3, β5, β6, β8, was first carried out, showing that β5 shares the highest homology rate with β3 and β6. Unsurprisingly, all the major mutations occurred at the specificity-determining-loop (SDL), which is at least two residues longer in β5 than in any other RGD integrin. Given the importance of the SDL for the ligand binding to the integrins orthosteric site (Figure 4.1, paragraph 4.1), particular attention was paid to the modelling of this protein region. Thus, a further sequence alignment restricted to the SDL primary sequences of the selected proteins was performed (Figure 4.14), which led to the selection of β6 (PDB: 4UM9) as template for the β5 model.[233] To refine the model an *ab initio* conformational prediction was executed for the SDL residues comprised between the disulfide bridge formed by the cysteine residues C176 and C185. Indeed, according to the SDL alignment, the insertion of the 2 a.a. in β5, absent in the β6 template, occurs at this level, which needed to be accurately considered.

```
CLUSTAL O(1.2.4) SDL sequence alignment

sp|P05556|ITB1_HUMAN      VMPYISTT-PAKLRNPCTSEQ---NCTSPFSY      28
sp|P18084|ITB5_HUMAN      ISPFSYTA-PRYQTNPCIGYKLFPNCVPSFGF      31
sp|P05106|ITB3_HUMAN      VSPYMYISPPEALENPCYDMKT--TCLPMFGY      30
sp|P26012|ITB8_HUMAN      VSPYISIH-PERIHNQCSDYNL--DCMPPHGY      29
sp|P18564|ITB6_HUMAN      VSPFVKTT-PEEIANPCSSIPYF--CLPTFGF      29
                          : *:     *    * * .     *   ..:
```

Figure 4.14 SDL sequence alignment of all the human RGD-binding β subunits.

Then, molecular docking of the PT-WTE-predicted conformation of **1** in the homology model of αvβ5 was performed. In these preliminary calculations the peptide backbone was kept fixed, while the possible side chains' orientations were sampled. As result, docking converged towards a predominant pose (Figure 4.15A) in which **1** adopts a typical RGD binding pattern. In particular, the carboxylate group of the ligand's Asp^4 chelates the Mg^{2+} cation at the protein MIDAS, while Arg^2 forms a salt bridge with the conserved αv residue D218 and a cation-π interaction with the phenol ring of αv-Y178. Besides the RGD motif, the docking pose is further stabilized by two H-bonds between the C-terminal carboxylic acid of **1** and the side chains of β5-T321 and β5-N323, respectively. The docking predicted **1**-αvβ5 complex was then submitted to 2 μs-long MD simulation. Notably, in this case, a refinement of the binding pose was particularly required to optimize potential clashes or small artifacts due to both the use of a homology model and the restraints applied to the ligand's peptide backbone upon docking. In fact, long MD trajectories can allow the system to escape from relative energy minima in which it might be trapped, also taking into account the solvent and entropic contributions which are instead neglected during docking calculations. Looking at the RMSD plot of the ligand's heavy atoms with respect to their initial positions (Figure 4.15B, black line), one can observe that over the first few tens of nanoseconds **1**

rearranges to reach a binding conformation which remains very stable for the rest of the simulation (Figure 4.15B, black line).

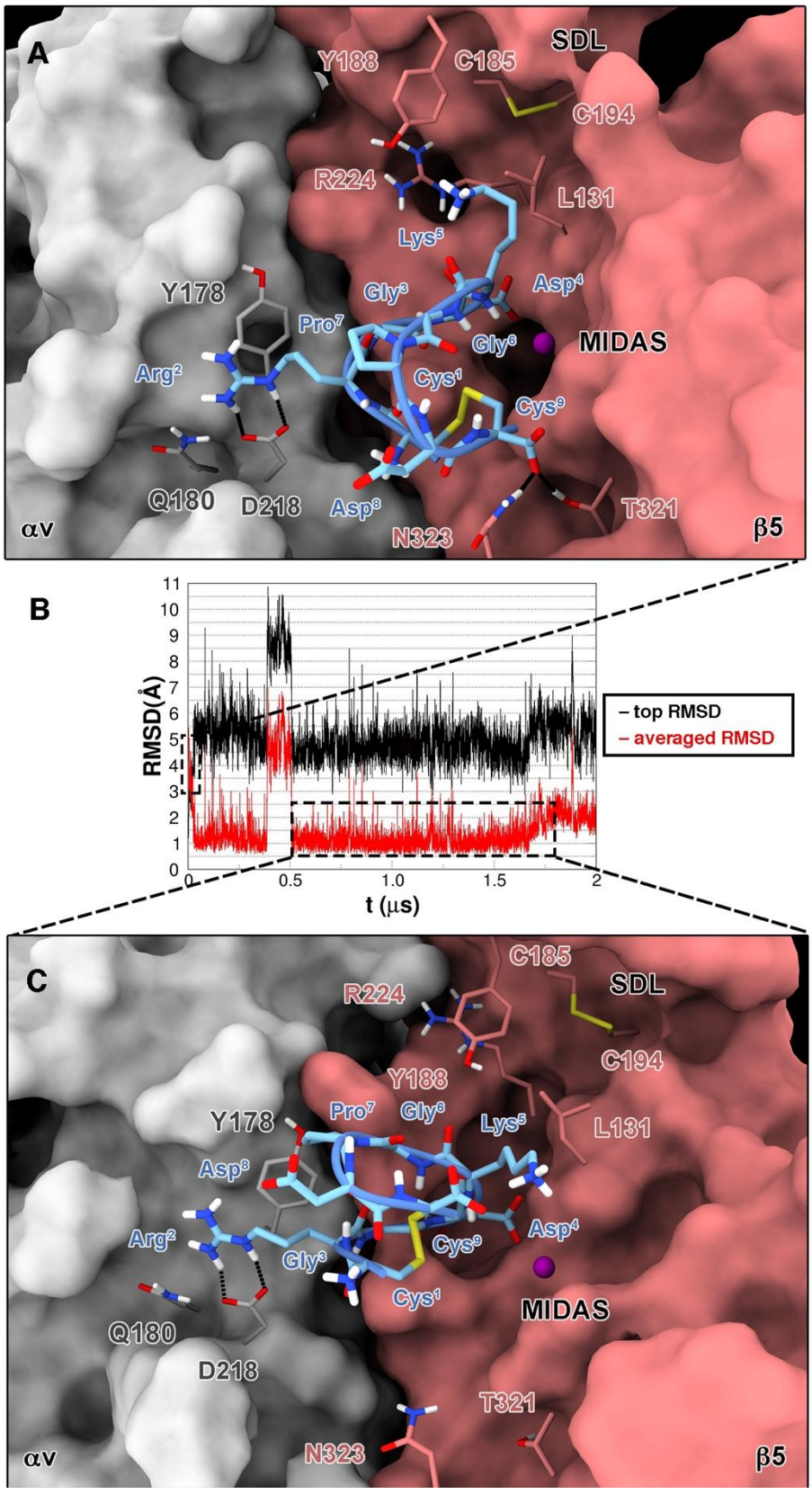


Figure 4.15. A) Docking pose of **1** at the αvβ5 binding site. B) RMSD of the ligands heavy atoms with respect to its initial (black) and average (red) conformation over the MD trajectory. C) Most representative binding pose of **1** during MD. The αv and β5 subunits are depicted as grey and red surfaces, respectively. The ligand (blue) and the residues important for peptide's binding are highlighted in sticks. Hydrogen-bonds are shown as black dashed lines, whereas non-polar hydrogens are omitted for sake of clarity.

The stability of this pose is even more clear if we consider the very low values assumed by the ligand's RMSD with respect to its average position (Figure 4.15B, red line). However, small changes in the peptide's orientation occur throughout the MD simulations. In particular, the interactions of the ligand's C-terminus with β5-T321 and β5-N323 are lost (Figure 4.16), allowing the residues flanking the RGD motif to get closer to the SDL cavity of the receptor (Figure 4.15C).

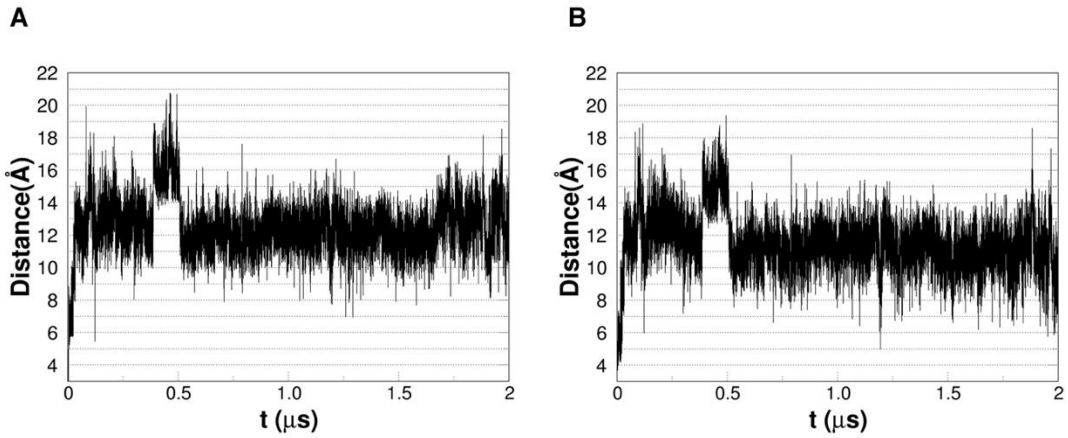


Figure 4.16. Interatomic distances between the ligand C-ter carboxylic carbon with T321-Oγ1 (A) and N323-Cγ (B).

This can provide valuable hints for the design of novel iRGD-derived peptides since the SDL region is generally targeted to modulate subtype affinity and selectivity of RGD ligands. It is also interesting to note that in the MD-refined

both the ligand's Cys^1-Cys^8 disulfide bridge and the Lys^5-Gly^6 residues are not particularly involved in any ligand-receptor interaction position (Figure 4.15C). These are important findings supporting the reliability of the proposed model. In fact, the solvent exposure of Cys^1 and Lys^5 can explain why the functionalization of these residues with bulky groups, for either therapeutic or diagnostic purposes, is permitted. Furthermore, the predicted **1**-αvβ5 complex is compatible with the proteolytic cleavage that the peptide must undergo to release the NRP-1-recognizing CendR sequence. In fact, this process is catalysed by some unknown enzyme that breaks the amide bond between Lys^5 and Gly^6 whilst **1** is still bound to αvβ5. Notably, in the proposed model the Lys^5-Gly^6 bond is solvent-exposed and prone to be cleaved.

Also, we analysed the behaviour of the backbone conformation of **1** during the MD simulation. Interestingly, although the ligand slightly translates during the simulation, its overall backbone arrangement is conserved. This is testified by both the low peptide's mainchain RMSD fluctuations (Figure 4.17A) and the stability of the two aforementioned intramolecular H-bonds (Figure 4.17B-C) along the entire simulated timescale. Thus, the MD results strengthened the goodness of the PT-WTE predicted backbone conformation of **1**, proving that an accurate evaluation of the *in solution* folding properties of cyclic RGD peptides can represent a valuable starting point for obtaining high-resolution binding models.

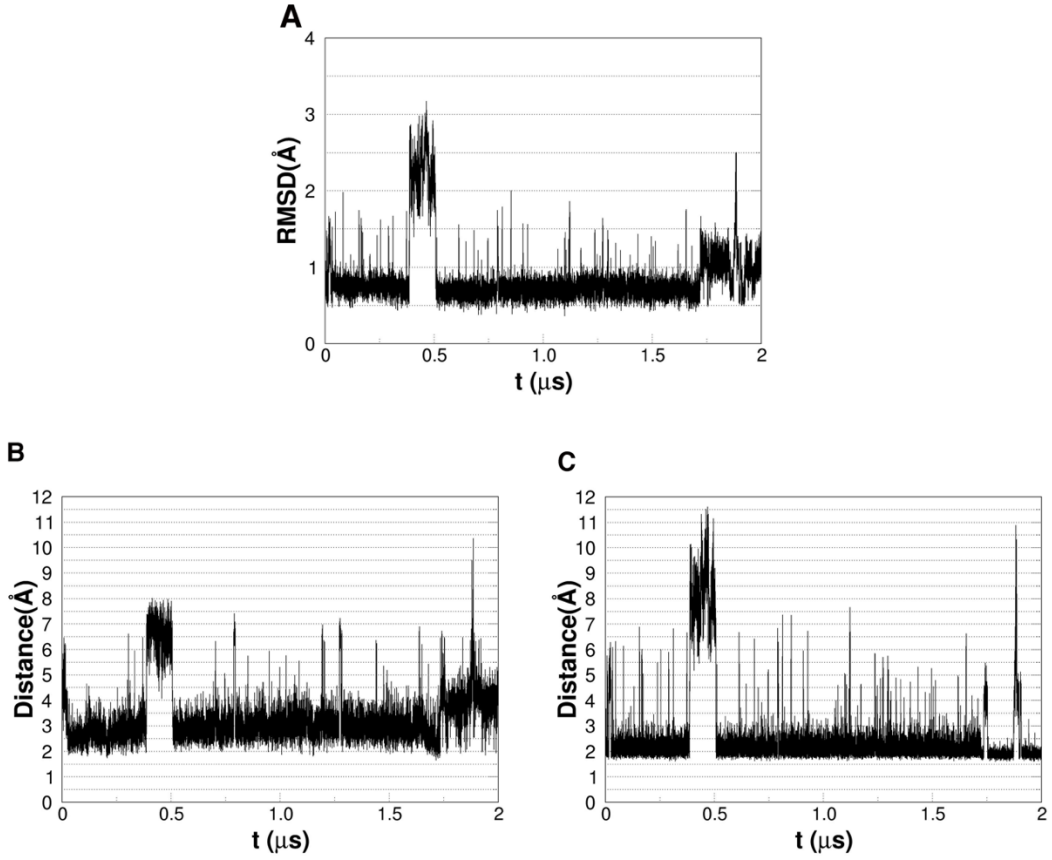


Figure 4.17. A) RMSD plot of the backbone atoms of **1** with respect to the PT-WTE-predicted conformation of the peptide. B-C) Interatomic distances between Arg^2 (C-O)-Gly^6(N-H) and Arg^2 (N-H)-Pro^7(C-O), respectively.

4.3.3 Conclusions and future perspectives

The iRGD (**1)** peptide recently come to the limelight for its remarkable tumour-homing properties.[217,218] Indeed, its functionalization with a plethora of chemotherapeutic agents has established this compound as an efficient tool for improving drug delivery in anticancer treatment.[218,219,221–223,228] For instance, the conjugation of **1** with gemcitabine and nab-paclitaxel has provided very positive results in phase I clinical trials for the treatment of pancreatic dual

adenocarcinoma.[224] The main reasons for the striking biological properties of **1** are to be found in its multistep internalization process, which first requires the peptide's binding to RGD integrins and then the exposure by proteolytic cleavage of the NRP-1 recognizing CEndR sequence; indeed, the interaction of the latter motif with NRP-1 finally leads to cellular internalization.[217] Here, an extensive computational investigation was performed to disclose at atomic level the first stage of this complex phenomenon, namely the binding mechanism of iRGD to the αvβ3/αvβ5 integrins. Specifically, PT-WTE metadynamics was used to predict the solution conformation of **1**, which was then employed as starting point for studying the peptide's binding mode at the integrin receptors through docking and MD simulations. Particular attention was given to αvβ5, since there is recent evidence that this integrin can drive the penetration of the peptide in specific cancer cell types such as PDAC.[228] The obtained **1**-αvβ5 interaction model is in agreement with the peptide's cleavage mechanism, which occurs at the solvent exposed Lys^5-Gly^6 amide bond. Also, it allows explaining why the bioconjugation of bulky molecules with either Cys^1 or Lys^5, which are not directly involved in the interaction with αvβ5, cannot impair the binding of iRGD to integrin receptors. Further experimental and computational and studies are now ongoing to evaluate the binding affinity and the interaction mode of **1** to other RGD integrins, especially αvβ6. In fact, this receptor is expressed at low level

in normal tissues, while it is upregulated in many cancer types and fibrosis.[225,234,235] Also, the SDL region of αvβ5 is wider and lipophilic than in αvβ3/αvβ5, and might be specifically targeted by properly designed bulky hydrophobic moieties. These studies will hopefully provide valuable hints to modulate the potency and selectivity of iRGD towards distinct integrin subtypes, paving the way for the development of novel tumour-homing peptides with finely-tuned specificity for integrin overexpressing cancers.

4.3.4 Methods

4.3.4.1 PT-WTE Simulation

Gaussians were deposited every 0.5 ps with a width of 145 kJ/mol and an initial height of 2.5 kJ/mol, gradually decreased based on a bias factor $\gamma = 24$. Then, 6 replicas were distributed according to the formula proposed by Prakash et al.[68] to span the temperature interval 300-450 K. Each replica was simulated for 140 ns in the NVT ensemble using the stochastic rescaling thermostat.[163] The coordinates' exchanges were attempted every 0.5 ps, obtaining an average acceptance probability of 25% between all the neighbour replicas. A further advantage of the WTE ensemble is that the canonical energy average is conserved, and all the other canonical observables can be estimated *a posteriori*. Thus, the *Tiwary-Parrinello* reweighting scheme[67] was employed to

computed the FES associated to the folding of **1** as function of two selected CVs. First, the *Dihedral Correlation* (*Dih*$_{Cor}$) between all the torsion angles of the peptide backbone, also including the peptide disulfide bridge:

$$Dih_{cor} = \frac{1}{2}\sum_{i}[1 + \cos(\phi i - \psi i)]$$

where the ϕi and ψ values are the instantaneous values for the torsion angles of interest. This function measures the degree of similarity between adjacent dihedral angles and, if extended to the entire backbone, can take into account global conformational changes. On the other hand, the *H*$_{bonds}$ CV estimates the number of intramolecular backbone-backbone H-bonds. This CV was computed as the sum of switching functions with the formula:

$$H_{bonds} = \frac{1 - (\frac{r_{ij} - d_0}{r_0})^n}{1 - (\frac{r_{ij} - d_0}{r_0})^m}$$

where *i* and *j* are defined as all the possible combinations between the amide hydrogen and oxygen atoms of the peptide backbone; d_0 and r_0 distances were set to 0 and 2.5 Å, while the *n* and *m* exponentials were modulated to 10 and 26, respectively. The GROMACS 2018.8[160] code patched with the PLUMED 2.5.6 plugin[165,166] was used to perform PT-WTE simulations. The peptide was parameterized using the ff14SB Amber force field,[158] and then solvated in a 12.0 Å layer rhombic dodecahedron box using the TIP3P water model

parameters.[112] Prior to metadynamics simulations, the system was equilibrated through 5 ns MD under NPT conditions at 1 atm and 300 K. A time step of 2 fs was used. All covalent bonds were constrained to their equilibrium value using the LINCS algorithm[205] Lennard-Jones potential was used to compute atom-pair interactions, with a cutoff of 10.0 Å. The simulation was carried out in periodic boundary conditions and using the Particle Mesh Ewald (PME) to treat long range electrostatic (grid spacing = 1.0 Å).

4.3.4.2 Homology Modelling

Since the αv subunit has already been solved in many X-ray structures of the αvβ3, αvβ6 an αvβ8 isoforms,[198,199,230–232] we here performed a homology model only of the β5 subunit. As first, a multiple sequence alignment between the headpieces (region corresponding to β3 residues 109-353) of all the human RGD β-subunits (β1, β3, β5, β6, β8) was performed with ClustalOmega. As result, two isoforms, namely β3 and β6, showed the highest similarity and identity rate with the human β5 (65% and 58% respectively). However, given that all the few gaps in the alignment occurred at the level of the SDL region, this preliminary sequence analysis was refined with a local alignment between the SDL loop of all the selected isoforms. Based on this further comparison, the crystal structure of the αvβ6 integrin in complex with the LAP peptide of

the TGF-β (PDB code: 4UM9) was chosen as template,[233] due to its higher identity value (44%) in the SDL region compared to β3 (41%). Then, the knowledge-based method implemented in Prime[236] was used to build the 3D receptor model. Furthermore, a refinement was carried out for loops carrying amino acids with missing coordinates (i.e. not coming from the template) by means of the Maestro "Refine Loops" panel.[157] Specifically, short loops were refined using default sampling rates, whereas the folding of the SDL residues comprised between the conserved disulfide bridge C176-C185 was refined using the implemented Extended protocol. Finally, the coordinates of all the non-conserved side chains were optimized using an energy cutoff of 10 kcal/mol.

4.3.4.3 Molecular Docking

Docking of the PT-WTE-predicted conformation of **1** was performed in the homology model of the αvβ5 receptor. Both the ligand and the receptor were prepared using the Protein Preparation Wizard tool, implemented in the Maestro Suite 2019.[111] Correct bond orders were assigned, missing hydrogen atoms added and all the water molecules deleted from the receptor structure. Then, protonation and tautomeric states at pH 7.4 were assigned to the side chain using Epik.[109,110] Finally, the positions of all the hydrogens were minimized. A virtual box of 20 Å × 20 Å × 20 Å, surrounding the typical RGD

binding site, was selected as search area by the means of the Receptor Grid Generator tool of Glide 8.5.[21,22] Docking calculations were performed employing the Glide SP-peptide protocol and the OPLS3A force field.[44] The peptide backbone was kept fixed in order to preserve the conformation obtained from PT-WTE simulations, while all the other parameters were kept to default values. Thus, the obtained solutions were clustered based on the ligand RMSD (cutoff = 2.0 Å) and ranked according to the Glide SP scoring function.[21,22]

4.3.4.4. Molecular Dynamics

Both the protein and the peptide were parametrized using the *ff14SB* Amber force field.[158] The divalent cations present in the integrins structures were treated with the parameters developed by Panteva et al.[237] The AMBER 18[238] code was then used to perform the simulations. A cutoff of 10 Å was used for short-range interactions. The long-range electrostatic interactions were computed through the particle mesh Ewald method[53] using a 1.0 Å grid spacing in periodic boundary conditions. The iterative SHAKE algorithm[239] was applied to constraint all bonds containing hydrogens, allowing for a 2 fs integration time step. In order to solve all the steric clashes, each system underwent 30,000 steps of mixed steepest descent/conjugated gradient energy minimization. Then, each complex was equilibrated and heated up to 300 K,

alternating NPT and NVT cycles (125,000 steps each) with the Langevin coupling bath[240] and the Berendsen barostat,[162] while applying gradually decreasing harmonic constraints on the heavy atoms of protein and ligand. Finally, a production run of 2 μs was performed in the NPT ensemble with target pressure and temperature of 1 atm and 300 K, respectively.

5. Conclusions

Pharmaceutical sciences, and the research in general, constantly offer scientists problems of different natures, for which no unique solution is available. For this reason, during my Ph.D. I tried to best exploit the wide range of computational techniques, to face the more different challenges in the drug design and development field.

My Ph.D. thesis deals with four main projects, in which very different computational approaches were adopted. The first two highlighted the importance of studying the binding interaction between new potential drugs and their pharmaceutical targets. This task can be addressed at different levels of resolutions, based on the aim of the project and the prior target knowledge. If the molecular requirements for the ligand-receptor binding are already known, *Virtual Screening* campaigns can lead to the identification of new hit compounds toward the desired target. Similarly, in Chapter 3.1, I reported a successful VS towards the *KRAS* proto-oncogene, which resulted in the identification of a new class of G4 stabilizers, with remarkable and promising anticancer properties. On the other hand, if the target is less explored and validated, advanced computational techniques can help to define the molecular bases for the receptor-ligand recognition process. For instance, in Chapter 3.2 a computational approach based on Molecular Dynamics and Funnel Metadynamics allowed reconstructing the entire binding event of

small organic molecules to Formyl Peptide Receptors 1 and 2. From this complete picture, valuable information can be derived about the structural basis of high affinity and selectivity FPRs ligands, which will pave the way for future rational drug design campaigns towards these receptors.

The second part of the thesis is rather focused on the study of the folding and binding properties of complex peptide ligands. In the first case study (Chapter 4.2), I had at my disposal experimental information about the peptide folding *in solution*, coming from NMR spectra. Thus, I could combine experimental data with molecular simulations to predict first the bioactive conformation and then the binding modalities of an antiviral RGD cyclopeptide to its receptors, namely αvβ6 and αvβ8. On the contrary, in Chapter 4.3, I proved that the employment of more accurate computational techniques can partially compensate for the lack of experimental support, even for complex molecules like the nonapeptide iRGD. In detail, a combination of PT-WTE, Molecular Dockings, and MD simulations provided a reliable interaction model of iRGD with its cognate integrin receptors. Given the high biological relevance of this compound, I expect that such information can have big impact on the scientific community, particularly in the development of new integrin-directed tumour-homing agents.

Acknowledgments

First of all, I thank my tutor Prof. Luciana Marinelli for her guidance and for all the opportunities she gave me to broaden my scientific knowledge.

A special thank goes to Dr. Francesco Saverio Di Leva. Most of the credit for the young passionate researcher I am trying to become is due to his patience, wise advice, and constant support.

I am also very grateful to Prof. Vittorio Limongelli for giving me the opportunity to join his group in Lugano. This visiting period was an important experience that I will always carry with me.

Finally, I want to thank all my lab mates and friends , especially Dr. Diego Brancaccio, Dr. Pasquale Russomanno, Dr. Stefano Raniolo, Dr. Paolo Conflitti, Dr. Simone Aureli, Dr. Valeria La Pietra, Federica Santoro and Bianca Fiorillo. The company of special people like them was fundamental to make the many hours spent in the lab easier and more enjoyable.

Abbreviations and Symbols

ΔG_b	Ligand binding free energy
β	Thermodynamic beta, $\beta = 1/k_B T$
$\varepsilon_{(r)}$	Sigmoidal distance-dependent dielectric constant
σ_i	Gaussian width for the ith CV
ω	Metadynamics energy rate
τ_G	Gaussian deposition stride
k_B	Boltzmann constant
K_A	Association constant
K_i	Inhibition constant
$\boldsymbol{R}$	Coordinates of the system, $\boldsymbol{R} = (\boldsymbol{R}_1, \ldots, \boldsymbol{R}_N)$
$\boldsymbol{S}$	Coordinates of the system, $\boldsymbol{S} = (S_1(\boldsymbol{R}), \ldots, S_d(\boldsymbol{R}))$
T_m	Melting temperature
V_G	Metadynamics bias potential, $V_G = V_G(\boldsymbol{S}, t)$
W	Gaussian height
AnxA1	Annexin A1
BFES	Binding Free Energy Surface
CADD	Computer-aided drug design
CD	Circular Dicroism
Cryo-EM	Cryo-Electron Microscopy
CV	Collective Variable
DAMP	Damage associated molecular patterns
DFT	Density Functional Theory

EA	Evolutionary algorithms
EBV	Epstein-Barr Virus
ECL	Extracellular Loops
ECM	Extracellular Matrix
ESP	Electrostatic Potential
Fbg	Fibrinogen
FES	Free Energy Surface
FF	Force Field
FITC	Fluorescein isothiocyanate
FM	Funnel Metadynamics
FMDV	Footh-mouth disease Virus
Fn	Fibronectin
FPR	Formyl Peptide Receptor
G4	G-Quadruplex
GA	Genetic algorithms
GPCR	G-Protein Coupled Receptors
HIV	Human Immunodeficiency Virus
HPC	High Performing Computer
HTS	High-troughput Screening
ICL	Intracellular Loops
LBDD	Ligand based drug design
LBVS	Ligand based Virtual Screening
LxA4	Lypoxin A4
MC	Monte Carlo
MD	Molecular Dynamics
MetaD	Metadynamics
MI	myocardial infarction

MIDAS	Metail ion dependent adhesion site
MW	Multiple Walkers
NHE	Nuclease-hypersensitive elements
NMR	Nuclear Magnetic Resonance
PAMP	Pathogen associated molecular patterns
PBC	Periodic Boundary Conditions
PDAC	Pancreatic dual adenocarcinoma
PME	Particle Mesh Ewald
PMF	Potential of Mean Force
PT	Parallel Tempering
PT-WTE	Parallel Tempering in the Well-Tempered Ensemble
QRT-PCR	Quantitative Real-time PCR
RAMD	Replica-averaged Molecular Dynamics
RESP	Restrained Electrostatic Potential
RMSD	Root-mean-square deviation of atomic positions
RvD2	Resolvin D2
SAA	Serum Amyloid A
SAR	Structure-activity relationship
SBDD	Structure based drug design
SBVS	Structure based Virtual Screening
SDL	Specificity determining loop
siRNA	Small-interfering RNA
SPM	Specialized pro-resolving mediators
TGF-β	Transforming Growth Factor-β
TM	Transmembrane
vdW	Van der Waals
Vn	Vitronectin

VS	Virtual Screeninh
VZV	Varicella Zoster Virus
WT-MetaD	Well-Tempered Metadynamics
WTE	Well-Tempered Ensemble

Bibliography

(1) Hughes, J. P.; Rees, S. S.; Kalindjian, S. B.; Philpott, K. L. Principles of Early Drug Discovery. *Br. J. Pharmacol.* **2011**, *162*, 1239–1249.

(2) Paul, S. M.; Mytelka, D. S.; Dunwiddie, C. T.; Persinger, C. C.; Munos, B. H.; Lindborg, S. R.; Schacht, A. L. How to Improve RD Productivity: The Pharmaceutical Industry's Grand Challenge. *Nat. Rev. Drug Discov.* **2010**, *9*, 203–214.

(3) Shaker, B.; Ahmad, S.; Lee, J.; Jung, C.; Na, D. In Silico Methods and Tools for Drug Discovery. *Comput. Biol. Med.* **2021**, *137*, 104851.

(4) Inglese, J.; Auld, D. S. High Throughput Screening (HTS) Techniques: Applications in Chemical Biology. *Wiley Encycl. Chem. Biol.* **2008**, 1–15.

(5) MacArron, R.; Banks, M. N.; Bojanic, D.; Burns, D. J.; Cirovic, D. A.; Garyantes, T.; Green, D. V. S.; Hertzberg, R. P.; Janzen, W. P.; Paslay, J. W.; Schopfer, U.; Sittampalam, G. S. Impact of High-Throughput Screening in Biomedical Research. *Nat. Rev. Drug Discov.* **2011**, *10*, 188–195.

(6) Lionta, E.; Spyrou, G.; Vassilatis, D.; Cournia, Z. Structure-Based Virtual Screening for Drug Discovery: Principles, Applications and Recent Advances. *Curr. Top. Med. Chem.* **2014**, *14*, 1923–1938.

(7) Macalino, S. J. Y.; Gosu, V.; Hong, S.; Choi, S. Role of Computer-Aided Drug Design in Modern Drug Discovery. *Arch. Pharm. Res.* **2015**, *38*, 1686–1701.

(8) Segall, M. D.; Barber, C. Addressing Toxicity Risk When Designing and Selecting Compounds in Early Drug Discovery. *Drug Discov. Today* **2014**, *19*, 688–693.

(9) Robinson, B. S.; Riccardi, K. A.; Gong, Y. F.; Guo, Q.; Stock, D. A.; Blair, W. S.; Terry, B. J.; Deminie, C. A.; Djang, F.; Colonno, R. J.; Lin, P. F. BMS-232632, a Highly Potent Human Immunodeficiency Virus Protease Inhibitor That Can Be Used in Combination with Other Available Antiretroviral Agents. *Antimicrob. Agents Chemother.* **2000**, *44*, 2093–2099.

(10) Krohn, A.; Redshaw, S.; Ritchie, J. C.; Graves, B. J.; Hatada, M. H. Novel Binding Mode of Highly Potent HIV-Proteinase Inhibitors Incorporating the (R)-Hydroxyethylamine Isostere. **1991**, *34*, 3340–3342.

(11) Kempf, D. J.; Marsh, K. C.; Denissen, J. F.; Mcdonald, E.; Flentge, C. A.; Green, B. E.; Fino, L.; Park, C. H.; Kong, P.; Wideburg, N. E.; Saldivar, A.; Ruiz, L.; Kati, W. M.; Sham, H. L.; Robins, T.; Stewart, K. D.; Hsu, A.; Plattner, J. J.; Leonard, J. M.; Norbeck, W.; Flentge, C. A.; Green, B. E.; Fino, L.; Parks, C. H.; Kongt, X.; Stewartt, K. D.; Hsu, A. N. N.; Plattner, J. J.; Leonardtt, J. M.; Norbeck, D. W. ABT-538 Is a Potent Inhibitor of Human Immunodeficiency Virus Protease and Has High Oral Bioavailability in Humans. *PNAS* **1995**, *92*, 2484–2488.

(12) Webber, S. E.; Bleckman, T. M.; Attard, J.; Deal, J. G.; Kathardekar, V.; Welsh, K. M.; Webber, S.; Janson, C. A.; Matthews, D. A.; Smith, W. W.; Freer, S. T.; Jordan, S. R.;

Bacquet, R. J.; Howland, E. F.; Booth, C. L. J.; Ward, R. W.; Hermann, S. M.; White, J.; Morse, C. A.; Hilliard, J. A.; Bartlett, C. A. Design of Thymidylate Synthase Inhibitors Using Protein Crystal Structures: The Synthesis and Biological Evaluation of a Novel Class of 5-Substituted Quinazolinones. *J. Med. Chem.* **1993**, *36*, 733–746.

(13) Rutenber, E. E.; Stroud, R. M. Binding of the Anticancer Drug ZD1694 to E. Coli Thymidylate Synthase: Assessing Specificity and Affinity. *Structure* **1996**, *4*, 1317–1324.

(14) Cushman, HS. Cheung, EF. Sabo, M. Design of Potent Competitive Inhibitors of Angiotensin-Converting Enzyme. *Biochemistry* **1977**, *16*, 5484–5941.

(15) Jiang, Z.; Zhou, Y. Using Bioinformatics for Drug Target Identification from the Genome. *Am. J. PharmacoGenomics* **2005**, *5*, 387–396.

(16) LeeSitao, J.; Zhang, W. Ab Initio Protein Structure Prediction. In *From Protein Structure to Function with Bioinformatics*; 2009; pp 3–25.

(17) Pinzi, L.; Rastelli, G. Molecular Docking: Shifting Paradigms in Drug Discovery. *Int. J. Mol. Sci.* **2019**, *20*.

(18) Hollingsworth, S. A.; Dror, R. O. Molecular Dynamics Simulation for All. *Neuron* **2018**, *99*, 1129–1143.

(19) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. A Geometric Approach to Macromolecule-Ligand Interactions. **1982**.

(20) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Docking and Scoring in Virtual Screening for Drug Discovery: Methods and Applications. *Nature Reviews Drug Discovery*. 2004.

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

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

(23) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J. Comput. Chem.* **2012**, *32*, 174–182.

(24) Dominguez, C.; Boelens, R.; Bonvin, A. M. J. J. HADDOCK: A Protein-Protein Docking Approach Based on Biochemical or Biophysical Information. *J. Am. Chem. Soc.* **2003**, *125*, 1731–1737.

(25) Huang, S. Y.; Zou, X. Advances and Challenges in Protein-Ligand Docking. *Int. J. Mol. Sci.* **2010**, *11*, 3016–3034.

(26) Brooijmans, N.; Kuntz, I. D. Molecular Recognition and Docking Algorithms. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 335–373.

(27) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A Fast Flexible Docking Method Using an Incremental Construction Algorithm. *J. Mol. Biol.* **1996**, *261*, 470–489.

(28) Leach, A. R.; Kuntz, I. D. Conformational Analysis of Flexible Ligands in Macromolecular Receptor Sites. *J. Comput. Chem.* **1992**, *13*, 730–748.

(29) Hastings, W. K. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika* **1970**.

(30) Oshiro, C. M.; Kuntz, I. D.; Dixon, J. S. Flexible Ligand Docking Using a Genetic Algorithm. *J. Comput. Aided. Mol. Des.* **1995**, *9*, 113–130.

(31) Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R. Principles of Docking: An Overview of Search Algorithms and a Guide to Scoring Functions. *Proteins: Structure, Function and Genetics*. 2002.

(32) Taylor, J. S.; Burnett, R. M. DARWIN: A Program for Docking Flexible Molecules. *Proteins Struct. Funct. Genet.* **2000**, *41*, 173–191.

(33) Norel, R.; Lin, S. L.; Wolfson, H. J.; Nussinov, R. Shape Complementarity at Protein–Protein Interfaces. *Biopolymers*. 1994, pp 933–940.

(34) Norel, R.; Petrey, D.; Wolfson, H. J.; Nussinov, R. Examination of Shape Complementarity in Docking of Unbound Proteins. *Proteins Struct. Funct. Genet.* **1999**, *36*, 307–317.

(35) Goodford, P. J. A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules. *J. Med. Chem.* **1985**, *28*, 849–857.

(36) Carlson, H. Protein Flexibility Is an Important Component of Structure-Based Drug Discovery. *Curr. Pharm. Des.* **2005**, *8*, 1571–1578.

(37) Teodoro, M. L.; Kavraki, L. E. *Conformational Flexibility Models for the Receptor in Structure Based Drug Design.*; 2003.

(38) Knegtel, R. M. A.; Kuntz, I. D.; Oshiro, C. M. Molecular Docking to Ensembles of Protein Structures. *J. Mol. Biol.* **1997**.

(39) Österberg, F.; Morris, G. M.; Sanner, M. F.; Olson, A. J.; Goodsell, D. S. Automated Docking to Multiple Target Structures: Incorporation of Protein Mobility and Structural Water Heterogeneity in Autodock. *Proteins Struct. Funct. Genet.* **2002**.

(40) Carlson, H. A.; Masukawa, K. M.; McCammon, J. A. Method for Including the Dynamic Fluctuations of a Protein in Computer-Aided Drug Design. *J. Phys. Chem. A* **1999**.

(41) Böhm, H. J. LUDI: Rule-Based Automatic Design of New Substituents for Enzyme Inhibitor Leads. *J. Comput. Aided. Mol. Des.* **1992**.

(42) Charifson, P. S.; Corkery, J. J.; Murcko, M. A.; Walters, W. P. Consensus Scoring: A Method for Obtaining Improved Hit Rates from Docking Databases of Three-Dimensional Structures into Proteins. *J. Med. Chem.* **1999**.

(43) Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V; Mee, R. P. 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.

(44) Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; Kaus, J. W.; Cerutti, D. S.; Krilov, G.; Jorgensen, W. L.; Abel, R.; Friesner, R. A. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J. Chem. Theory Comput.* **2016**, *12*, 281–296.

(45) Cosconati, S.; Forli, S.; Perryman, A. L.; Harris, R.; Goodsell, D. S.; Olson, A. J. Virtual Screening with AutoDock: Theory and Practice. *Expert Opin. Drug Discov.* **2010**, *5*, 597–607.

(46) Irwin, J. J.; Shoichet, B. K. ZINC - A Free Database of Commercially Available Compounds for Virtual Screening. *J. Chem. Inf. Model.* **2005**, *45*, 177–182.

(47) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. *Advanced Drug Delivery Reviews*. 2012.

(48) Kalliokoski, T.; Salo, H. S.; Lahtela-Kakkonen, M.; Poso, A. The Effect of Ligand-Based Tautomer and Protomer Prediction on Structure-Based Virtual Screening. *J. Chem. Inf. Model.* **2009**, *49*, 2742–2748.

(49) Chang, M. W.; Belew, R. K.; Carroll, K. S.; Olson, A. J.; Goodsell, D. S. Empirical Entropie Contributions in Computational Docking: Evaluation in APS Reductase Complexes. *J. Comput. Chem.* **2008**.

(50) Hopkins, A. L.; Groom, C. R.; Alex, A. Ligand Efficiency: A Useful Metric for Lead Selection. *Drug Discov. Today* **2004**, *9*, 430–431.

(51) Verlet, L. Computer "Experiments" on Classical Fluids. Thermodynamical Properties of Lennard-Jones Molecules. *J. Phys. D. Appl. Phys.* **1976**, *9*, 183–195.

(52) Frenkel, D.; Smit, B. *Understanding Molecular Simulation 2nd Edition*; 2001.

(53) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089.

(54) Rosenbaum, D. M.; Zhang, C.; Lyons, J. A.; Holl, R.; Aragao, D.; Arlow, D. H.; Rasmussen, S. G. F.; Choi, H. J.; Devree, B. T.; Sunahara, R. K.; Chae, P. S.; Gellman, S. H.; Dror, R. O.; Shaw, D. E.; Weis, W. I.; Caffrey, M.; Gmeiner, P.; Kobilka, B. K. Structure and Function of an Irreversible Agonist-B2 Adrenoceptor Complex. *Nature* **2011**, *469*, 236–242.

(55) Dror, R. O.; Green, H. F.; Valant, C.; Borhani, D. W.; Valcourt, J. R.; Pan, A. C.; Arlow, D. H.; Canals, M.; Lane, J. R.; Rahmani, R.; Baell, J. B.; Sexton, P. M.; Christopoulos, A.; Shaw, D. E. Structural Basis for Modulation of a G-Protein-Coupled Receptor by Allosteric Drugs. *Nature* **2013**, *503*, 295–299.

(56) Laio, A.; Parrinello, M. (CUL-ID:1574682) Escaping Free-Energy Minima. *Proc. Natl.*

Acad. Sci. U. S. A. **2002**.

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

(58) Bussi, G.; Gervasio, F. L.; Laio, A.; Parrinello, M. Free-Energy Landscape for b Hairpin Folding from Combined Parallel Tempering and Metadynamics. **2008**, *128*, 13435–13441.

(59) Bonomi, M.; Parrinello, M. Enhanced Sampling in the Well-Tempered Ensemble. *Phys. Rev. Lett.* **2010**, *104*, 1–4.

(60) Deighan, M.; Bonomi, M.; Pfaendtner, J. Efficient Simulation of Explicitly Solvated Proteins in the Well- Tempered Ensemble. *J. Chem. Theory Comput.* **2012**, *8*, 2189–2192.

(61) Barducci, A.; Bonomi, M.; Parrinello, M. Metadynamics. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1*, 826–843.

(62) Limongelli, V.; Bonomi, M.; Parrinello, M. Funnel Metadynamics as Accurate Binding Free-Energy Method. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 6358–6363.

(63) Raniolo, S.; Limongelli, V. Ligand Binding Free-Energy Calculations with Funnel Metadynamics. *Nat. Protoc.* **2020**, *15*, 2837–2866.

(64) Roux, B.; Weare, J. On the Statistical Equivalence of Restrained-Ensemble Simulations with the Maximum Entropy Method. *J. Chem. Phys.* **2013**, *138*.

(65) Cavalli, A.; Camilloni, C.; Vendruscolo, M. Molecular Dynamics Simulations with Replica-Averaged Structural Restraints Generate Structural Ensembles According to the Maximum Entropy Principle. *J. Chem. Phys.* **2013**, *138*, 094112.

(66) Bonomi, M.; Barducci, A.; Parrinello, M. Reconstructing the Equilibrium Boltzmann Distribution from Well-Tempered Metadynamics. *J. Comput. Chem.* **2009**, *30*, 1615–1621.

(67) Tiwary, P.; Parrinello, M. A Time-Independent Free Energy Estimator for Metadynamics. **2014**, *119*, 736–742.

(68) Prakash, M. K.; Barducci, A.; Parrinello, M. Replica Temperatures for Uniform Exchange and Efficient Roundtrip Times in Explicit Solvent Parallel Tempering Simulations. *J. Chem. Theory Comput.* **2011**, *7*, 2025–2027.

(69) Bacolla, A.; Wells, R. D. Non-B DNA Conformations as Determinants of Mutagenesis and Human Disease. *Mol. Carcinog.* **2009**, *48*, 273–285.

(70) Zhao, J.; Bacolla, A.; Wang, G.; Vasquez, K. M. Non-B DNA Structure-Induced Genetic Instability and Evolution. *Cell. Mol. Life Sci.* **2010**, *67*, 43–62.

(71) Amato, J.; Iaccarino, N.; Randazzo, A.; Novellino, E.; Pagano, B. Noncanonical DNA Secondary Structures as Drug Targets: The Prospect of the i-Motif. *ChemMedChem* **2014**.

(72) Balasubramanian, S.; Hurley, L. H.; Neidle, S. Targeting G-Quadruplexes in Gene Promoters: A Novel Anticancer Strategy? *Nat. Rev. Drug Discov.* **2011**, *10*, 261–275.

(73) Bochman, M. L.; Paeschke, K.; Zakian, V. A. DNA Secondary Structures: Stability and Function of G-Quadruplex Structures. *Nat. Rev. Genet.* **2012**, *13*, 770–780.

(74) Caterino, M.; D'Aria, F.; Kustov, A. V.; Belykh, D. V.; Khudyaeva, I. S.; Starseva, O. M.; Berezin, D. B.; Pylina, Y. I.; Usacheva, T.; Amato, J.; Giancola, C. Selective Binding of a Bioactive Porphyrin-Based Photosensitizer to the G-Quadruplex from the KRAS Oncogene Promoter. *Int. J. Biol. Macromol.* **2020**, *145*, 244–251.

(75) Cogoi, S.; Quadrifoglio, F.; Xodo, L. E. G-Rich Oligonucleotide Inhibits the Binding of a Nuclear Protein to the Ki-Ras Promoter and Strongly Reduces Cell Growth in Human Carcinoma Pancreatic Cells. *Biochemistry* **2004**, *1859*, 663–674.

(76) Cogoi, S.; Xodo, L. E. G-Quadruplex Formation within the Promoter of the KRAS Proto-Oncogene and Its Effect on Transcription. *Nucleic Acids Res.* **2006**, *34*, 2536–2549.

(77) Dai, J.; Chen, D.; Jones, R. A.; Hurley, L. H.; Yang, D. NMR Solution Structure of the Major G-Quadruplex Structure Formed in the Human BCL2 Promoter Region. *Nucleic Acids Res.* **2006**, *128*, 1096–1098.

(78) Fernando, H.; Reszka, A. P.; Huppert, J.; Ladame, S.; Rankin, S.; Venkitaraman, A. R.; Neidle, S.; Balasubramanian, S. A Conserved Quadruplex Motif Located in a Transcription Activation Site of the Human C-Kit Oncogene. *Biochemistry* **2006**, *42*, 7854–7860.

(79) Cogoi, S.; Zorzet, S.; Rapozzi, V.; Géci, I.; Pedersen, E. B.; Xodo, L. E. MAZ-Binding G4-Decoy with Locked Nucleic Acid and Twisted Intercalating Nucleic Acid Modifications Suppresses KRAS in Pancreatic Cancer Cells and Delays Tumor Growth in Mice. *Nucleic Acids Res.* **2013**, *41*, 4049–4064.

(80) Miglietta, G.; Cogoi, S.; Marinello, J.; Capranico, G.; Tikhomirov, A. S.; Shchekotikhin, A.; Xodo, L. E. RNA G-Quadruplexes in Kirsten Ras (KRAS) Oncogene as Targets for Small Molecules Inhibiting Translation. *J. Med. Chem.* **2017**, *60*, 9448–9461.

(81) Kang, R.; Xie, Y.; Zhang, Q.; Hou, W.; Jiang, Q.; Zhu, S.; Liu, J.; Zeng, D.; Wang, H.; Bartlett, D. L.; Billiar, T. R.; Zeh, H. J.; Lotze, M. T.; Tang, D. Intracellular HMGB1 as a Novel Tumor Suppressor of Pancreatic Cancer. *Cell Res.* **2017**, *27*, 916–932.

(82) Cox, A. D.; Fesik, S. W.; Kimmelman, A. C.; Luo, J.; Der, C. J. Drugging the Undruggable RAS: Mission Possible? *Nat. Rev. Drug Discov.* **2014**, *13*, 828–851.

(83) Haigis, K. M. KRAS Alleles: The Devil Is in the Detail. *Trends in Cancer* **2017**, *3*, 686–697.

(84) Hobbs, G. A.; Der, C. J.; Rossman, K. L. RAS Isoforms and Mutations in Cancer at a Glance. *J. Cell Sci.* **2016**, *129*, 1750–1759.

(85) McCormick, F. Targeting KRAS Directly. *Annu. Rev. Cancer Biol.* **2018**, *2*, 81–90.

(86) McCormick, F. K-Ras Protein as a Drug Target. *J. Mol. Med.* **2016**, *94*, 253–258.

(87) Janes, M. R.; Zhang, J.; Li, L. S.; Hansen, R.; Peters, U.; Guo, X.; Chen, Y.; Babbar, A.; Firdaus, S. J.; Darjania, L.; Feng, J.; Chen, J. H.; Li, S.; Li, S.; Long, Y. O.; Thach, C.; Liu, Y.; Zarieh, A.; Ely, T.; Kucharski, J. M.; Kessler, L. V.; Wu, T.; Yu, K.; Wang, Y.; Yao,

Y.; Deng, X.; Zarrinkar, P. P.; Brehmer, D.; Dhanak, D.; Lorenzi, M. V.; Hu-Lowe, D.; Patricelli, M. P.; Ren, P.; Liu, Y. Targeting KRAS Mutant Cancers with a Covalent G12C-Specific Inhibitor. *Cell* **2018**, *172*, 578–589.

(88) Lito, P.; Solomon, M.; Li, L. S.; Hansen, R.; Rosen, N. Cancer Therapeutics: Allele-Specific Inhibitors Inactivate Mutant KRAS G12C by a Trapping Mechanism. *Science (80-.).* **2016**, *351*, 604–608.

(89) Alzeer, J.; Vummidi, B. R.; Roth, P. J. C.; Luedtke, N. W. Guanidinium-Modified Phthalocyanines as High-Affinity G-Quadruplex Fluorescent Probes and Transcriptional Regulators. *Angew. Chemie - Int. Ed.* **2009**, *48*, 9362–9365.

(90) Boschi, E.; Davis, S.; Taylor, S.; Butterworth, A.; Chirayath, L. A.; Purohit, V.; Siegel, L. K.; Buenaventura, J.; Sheriff, A. H.; Jin, R.; Sheardy, R.; Yatsunyk, L. A.; Azam, M. Interaction of a Cationic Porphyrin and Its Metal Derivatives with G-Quadruplex DNA. *J. Phys. Chem. B* **2016**, *120*, 12807–12819.

(91) Calabrese, D. R.; Zlotkowski, K.; Alden, S.; Hewitt, W. M.; Connelly, C. M.; Wilson, R. M.; Gaikwad, S.; Chen, L.; Guha, R.; Thomas, C. J.; Mock, B. A.; Schneekloth, J. S. Characterization of Clinically Used Oral Antiseptics as Quadruplex-Binding Ligands. *Nucleic Acids Res.* **2018**, *46*, 2722–2732.

(92) Carvalho, J.; Pereira, E.; Marquevielle, J.; Campello, M. P. C.; Mergny, J. L.; Paulo, A.; Salgado, G. F.; Queiroz, J. A.; Cruz, C. Fluorescent Light-up Acridine Orange Derivatives Bind and Stabilize KRAS-22RT G-Quadruplex. *Biochimie* **2018**.

(93) Lavrado, J.; Brito, H.; Borralho, P. M.; Ohnmacht, S. A.; Kim, N. S.; Leitão, C.; Pisco, S.; Gunaratnam, M.; Rodrigues, C. M. P.; Moreira, R.; Neidle, S.; Paulo, A. KRAS Oncogene Repression in Colon Cancer Cell Lines by G-Quadruplex Binding Indolo[3,2-c]Quinolines. *Sci. Rep.* **2015**, *5*, 9696.

(94) Lavrado, J.; Borralho, P. M.; Ohnmacht, S. A.; Castro, R. E.; Rodrigues, C. M. P.; Moreira, R.; Dos Santos, D. J. V. A.; Neidle, S.; Paulo, A. Synthesis, G-Quadruplex Stabilisation, Docking Studies, and Effect on Cancer Cells of Indolo[3,2-b]Quinolines with One, Two, or Three Basic Side Chains. *ChemMedChem* **2013**, *8*, 1648–1661.

(95) Monchaud, D.; Teulade-Fichou, M. P. A Hitchhiker's Guide to G-Quadruplex Ligands. *Org. Biomol. Chem.* **2008**, *6*, 627–636.

(96) Pattanayak, R.; Barua, A.; Das, A.; Chatterjee, T.; Pathak, A.; Choudhury, P.; Sen, S.; Saha, P.; Bhattacharyya, M. Porphyrins to Restrict Progression of Pancreatic Cancer by Stabilizing KRAS G-Quadruplex: In Silico, in Vitro and in Vivo Validation of Anticancer Strategy. *Eur. J. Pharm. Sci.* **2018**, *125*, 39–53.

(97) D'Aria, F.; D'Amore, V. M.; Di Leva, F. S.; Amato, J.; Caterino, M.; Russomanno, P.; Salerno, S.; Barresi, E.; De Leo, M.; Marini, A. M.; Taliani, S.; Da Settimo, F.; Salgado, G. F.; Pompili, L.; Zizza, P.; Shirasawa, S.; Novellino, E.; Biroccio, A.; Marinelli, L.; Giancola, C. Targeting the KRAS Oncogene: Synthesis, Physicochemical and Biological Evaluation of Novel G-Quadruplex DNA Binders. *Eur. J. Pharm. Sci.* **2020**, *149*, 105337.

(98) Kerkour, A.; Marquevielle, J.; Ivashchenko, S.; Yatsunyk, L. A.; Mergny, J. L.; Salgado, G. F. High-Resolution Three-Dimensional NMR Structure of the KRAS Proto-Oncogene Promoter Reveals Key Features of a G-Quadruplex Involved in Transcriptional Regulation. *J. Biol. Chem.* **2017**.

(99) Castillo-González, D.; Mergny, J. L.; De Rache, A.; Pérez-Machado, G.; Cabrera-Pérez, M. A.; Nicolotti, O.; Introcaso, A.; Mangiatordi, G. F.; Guédin, A.; Bourdoncle, A.; Garrigues, T.; Pallardó, F.; Cordeiro, M. N. D. S.; Paz-Y-Miño, C.; Tejera, E.; Borges, F.; Cruz-Monteagudo, M. Harmonization of QSAR Best Practices and Molecular Docking Provides an Efficient Virtual Screening Tool for Discovering New G-Quadruplex Ligands. *J. Chem. Inf. Model.* **2015**.

(100) Di Leva, F. S.; Novellino, E.; Cavalli, A.; Parrinello, M.; Limongelli, V. Mechanistic Insight into Ligand Binding to G-Quadruplex DNA. *Nucleic Acids Res.* **2014**.

(101) Zhang, D. H.; Fujimoto, T.; Saxena, S.; Yu, H. Q.; Miyoshi, D.; Sugimoto, N. Monomorphic RNA G-Quadruplex and Polymorphic DNA G-Quadruplex Structures Responding to Cellular Environmental Factors. *Biochemistry* **2010**, *49*, 4554–4563.

(102) Dalla Via, L.; Gia, O.; Marciani Magno, S.; Da Settimo, A.; Marini, A. M.; Primofiore, G.; Da Settimo, F.; Salerno, S. Synthesis, in Vitro Antiproliferative Activity and DNA-Interaction of Benzimidazoquinazoline Derivatives as Potential Anti-Tumor Agents. *Farmaco* **2001**, *56*, 159–167.

(103) Di Leva, F. S.; Zizza, P.; Cingolani, C.; D'Angelo, C.; Pagano, B.; Amato, J.; Salvati, E.; Sissi, C.; Pinato, O.; Marinelli, L.; Cavalli, A.; Cosconati, S.; Novellino, E.; Randazzo, A.; Biroccio, A. Exploring the Chemical Space of G-Quadruplex Binders: Discovery of a Novel Chemotype Targeting the Human Telomeric Sequence. *J. Med. Chem.* **2013**.

(104) CarotiI, P.; Ceccotti, C.; Da Settimo, A.; Palla, F.; Primofiore, G. ChemInform Abstract: Synthesis of 5,11-Dihydro-5-Oxopyrido(2',3':4,5)Pyrimido(1,2-a)Benzimidazoles as Potential Antitumor Agents. *ChemInform* **1987**, *18*, 44.

(105) Giancola, C.; Pagano, B. Energetics of Ligand Binding to G-Quadruplexes. *Top. Curr. Chem.* **2013**, *330*, 211–242.

(106) Pagano, B.; Cosconati, S.; Gabelica, V.; Petraccone, L.; De Tito, S.; Marinelli, L.; La Pietra, V.; Saverio di Leva, F.; Lauri, I.; Trotta, R.; Novellino, E.; Giancola, C.; Randazzo, A. State-of-the-Art Methodologies for the Discovery and Characterization of DNA G-Quadruplex Binders. *Curr. Pharm. Des.* **2012**, *18*, 1888–1899.

(107) Siravegna, G.; Mussolin, B.; Buscarino, M.; Corti, G.; Cassingena, A.; Crisafulli, G.; Ponzetti, A.; Cremolini, C.; Amatu, A.; Lauricella, C.; Lamba, S.; Hobor, S.; Avallone, A.; Valtorta, E.; Rospo, G.; Medico, E.; Motta, V.; Antoniotti, C.; Tatangelo, F.; Bellosillo, B.; Veronese, S.; Budillon, A.; Montagut, C.; Racca, P.; Marsoni, S.; Falcone, A.; Corcoran, R. B.; Di Nicolantonio, F.; Loupakis, F.; Siena, S.; Sartore-Bianchi, A.; Bardelli, A. Clonal Evolution and Resistance to EGFR Blockade in the Blood of Colorectal Cancer Patients. *Nature Medicine*. 2015.

(108) Cogoi, S.; Xodo, L. E. G4 DNA in Ras Genes and Its Potential in Cancer Therapy. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*. 2016.

(109) Shelley, J. C.; Cholleti, A.; Frye, L. L.; Greenwood, J. R.; Timlin, M. R.; Uchimaya, M. Epik: A Software Program for PKa Prediction and Protonation State Generation for Drug-like Molecules. *J. Comput. Aided. Mol. Des.* **2007**, *21*, 681–691.

(110) Greenwood, J. R.; Calkins, D.; Sullivan, A. P.; Shelley, J. C. Towards the Comprehensive, Rapid, and Accurate Prediction of the Favorable Tautomeric States of Drug-like Molecules in Aqueous Solution. *J. Comput. Aided. Mol. Des.* **2010**, *24*, 591–604.

(111) Madhavi Sastry, G.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and Ligand Preparation: Parameters, Protocols, and Influence on Virtual Screening Enrichments. *J. Comput. Aided. Mol. Des.* **2013**, *27*, 221–234.

(112) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**.

(113) Ivani, I.; Dans, P. D.; Noy, A.; Pérez, A.; Faustino, I.; Hospital, A.; Walther, J.; Andrio, P.; Goñi, R.; Balaceanu, A.; Portella, G.; Battistini, F.; Gelpí, J. L.; González, C.; Vendruscolo, M.; Laughton, C. A.; Harris, S. A.; Case, D. A.; Orozco, M. Parmbsc1: A Refined Force Field for DNA Simulations. *Nat. Methods* **2015**, *13*, 55–58.

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

(115) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J. Phys. Chem.* **1993**, *97*, 10269–10280.

(116) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas; Foresman, J. B.; Ortiz, J. V; Cioslowski, J.; Fox, D. J. Gaussian 09, Revision B.01. *Gaussian 09, Revision B.01, Gaussian, Inc., Wallingford CT*. 2009.

(117) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graph. Model.* **2006**, *25*, 247–260.

(118) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–1802.

(119) Schiöth, H. B.; Lagerström, M. C. Structural Diversity of g Proteincoupled Receptors and Significance for Drug Discovery. *Nat. Rev. Drug Discov.* **2008**, *7*, 339–357.

(120) Fredriksson, R.; Lagerström, M. C.; Lundin, L. G.; Schiöth, H. B. The G-Protein-Coupled Receptors in the Human Genome Form Five Main Families. Phylogenetic Analysis, Paralogon Groups, and Fingerprints. *Mol. Pharmacol.* **2003**, *63*, 1256–1272.

(121) Katritch, V.; Cherezov, V.; Stevens, R. C. Structure-Function of the G Protein-Coupled Receptor Superfamily. *Annu. Rev. Pharmacol. Toxicol.* **2013**, *53*, 531–556.

(122) Chan, H. C. S.; Li, Y.; Dahoun, T.; Vogel, H.; Yuan, S. New Binding Sites, New Opportunities for GPCR Drug Discovery. *Trends Biochem. Sci.* **2019**, *44*, 312–330.

(123) Rosenbaum, D. M.; Rasmussen, S. G. F.; Kobilka, B. K. The Structure and Function of G-Protein-Coupled Receptors. *Nature* **2009**, *459*, 356–363.

(124) Gonzalez, D. R.; Treuer, A.; Sun, Q.; Stamler, J. S.; Hare, J. M. G Protein–Coupled Receptor Signaling Through b-Arrestin–Dependent Mechanisms. *J. Cardiovasc. Pharmacol.* **2009**, *54*, 188–195.

(125) Wootten, D.; Christopoulos, A.; Marti-Solano, M.; Babu, M. M.; Sexton, P. M. Mechanisms of Signalling and Biased Agonism in G Protein-Coupled Receptors. *Nat. Rev. Mol. Cell Biol.* **2018**, *19*, 638–653.

(126) Congreve, M.; de Graaf, C.; Swain, N. A.; Tate, C. G. Impact of GPCR Structures on Drug Discovery. *Cell* **2020**, *181*, 81–91.

(127) Bylund, D. B.; Eikenberg, D. C.; Hieble, J. P.; Langer, S. Z.; Lefkowitz, R. J.; Minneman, K. P.; Molinoff, P. B.; Ruffolo, R. R.; Trendelenburg, U. IV. International Union of Pharmacology Nomenclature of Adrenoceptors. *Pharmacol. Rev.* **1994**, *46*, 121–136.

(128) Schiffmann, E.; Corcoran, B. A.; Wahl, S. M. N Formylmethionyl Peptides as Chemoattractants for Leucocytes. *Proc. Natl. Acad. Sci. U. S. A.* **1975**, *72*, 1059–1062.

(129) He, H. Q.; Ye, R. D. The Formyl Peptide Receptors: Diversity of Ligands and Mechanism for Recognition. *Molecules* **2017**, *22*.

(130) Perretti, M.; Chiang, N.; La, M.; Fierro, I. M.; Marullo, S.; Getting, S. J.; Solito, E.; Serhan, C. N. Endogenous Lipid- and Peptide-Derived Anti-Inflammatory Pathways Generated with Glucocorticoid and Aspirin Treatment Activate the Lipoxin A4 Receptor. *Nat. Med.* **2002**, *8*, 1296–1302.

(131) Zhang, S.; Gong, H.; Ge, Y.; Ye, R. D. Biased Allosteric Modulation of Formyl Peptide Receptor 2 Leads to Distinct Receptor Conformational States for Pro- and Anti-Inflammatory Signaling. *Pharmacol. Res.* **2020**, *161*, 105117.

(132) Perretti, M.; Godson, C. Formyl Peptide Receptor Type 2 Agonists to Kick-Start Resolution Pharmacology. *Br. J. Pharmacol.* **2020**, *177*, 4595–4600.

(133) Cooray, S. N.; Gobbetti, T.; Montero-Melendez, T.; McArthur, S.; Thompson, D.; Clark, A. J. L.; Flower, R. J.; Perretti, M. Ligand-Specific Conformational Change of the G-Protein-Coupled Receptor ALX/FPR2 Determines Proresolving Functional

Responses. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 18232–18237.

(134) Corminboeuf, O.; Leroy, X. FPR2/ALXR Agonists and the Resolution of Inflammation. *J. Med. Chem.* **2015**, *58*, 537–559.

(135) N. Serhan, C.; Krishnamoorthy, S.; Recchiuti, A.; Chiang, N. *Novel Anti-Inflammatory-Pro-Resolving Mediators and Their Receptors*; 2012; Vol. 11.

(136) Ammendola, R.; Parisi, M.; Esposito, G.; Cattaneo, F. Pro-Resolving FPR2 Agonists Regulate Nadph Oxidase-Dependent Phosphorylation of HSP27, OSR1, and MARCKS and Activation of the Respective Upstream Kinases. *Antioxidants* **2021**, *10*, 1–22.

(137) Inoue, A.; Raimondi, F.; Kadji, F. M. N.; Singh, G.; Kishi, T.; Uwamizu, A.; Ono, Y.; Shinjo, Y.; Ishida, S.; Arang, N.; Kawakami, K.; Gutkind, J. S.; Aoki, J.; Russell, R. B. Illuminating G-Protein-Coupling Selectivity of GPCRs. *Cell* **2019**, *177*, 1933-1947.e25.

(138) Mottola, G.; Chatterjee, A.; Wu, B.; Chen, M.; Conte, M. S. Aspirin-Triggered Resolvin D1 Attenuates PDGF-Induced Vascular Smooth Muscle Cell Migration via the Cyclic Adenosine Monophosphate/Protein Kinase A (CAMP/PKA) Pathway. *PLoS One* **2017**, *12*, 1–20.

(139) Ernst, S.; Lange, C.; Wilbers, A.; Goebeler, V.; Gerke, V.; Rescher, U. An Annexin 1 N-Terminal Peptide Activates Leukocytes by Triggering Different Members of the Formyl Peptide Receptor Family. *J. Immunol.* **2004**, *172*, 7669–7676.

(140) Dalli, J.; Consalvo, A. P.; Ray, V.; Di Filippo, C.; D'Amico, M.; Mehta, N.; Perretti, M. Proresolving and Tissue-Protective Actions of Annexin A1–Based Cleavage-Resistant Peptides Are Mediated by Formyl Peptide Receptor 2/Lipoxin A 4 Receptor . *J. Immunol.* **2013**, *190*, 6478–6487.

(141) Seo, J. K.; Choi, S. Y.; Kim, Y.; Baek, S. H.; Kim, K. T.; Chae, C. B.; Lambeth, J. D.; Suh, P. G.; Ryu, S. H. A Peptide with Unique Receptor Specificity: Stimulation of Phosphoinositide Hydrolysis and Induction of Superoxide Generation in Human Neutrophils. *J. Immunol.* **1997**, *158*, 1895 LP – 1901.

(142) Le, Y.; Gong, W.; Li, B.; Dunlop, N. M.; Shen, W.; Su, S. B.; Ye, R. D.; Wang, J. M. Utilization of Two Seven-Transmembrane, G Protein-Coupled Receptors, Formyl Peptide Receptor-Like 1 and Formyl Peptide Receptor, by the Synthetic Hexapeptide WKYMVm for Human Phagocyte Activation. *J. Immunol.* **1999**, *163*, 6777 LP – 6784.

(143) Tsai, Y.-F.; Yang, S.-C.; Hwang, T.-L. Formyl Peptide Receptor Modulators: A Patent Review and Potential Applications for Inflammatory Diseases. *Expert Opin. Ther. Pat.* **2016**, *26*, 1139–1156.

(144) Maciuszek, M.; Cacace, A.; Brennan, E.; Godson, C.; Chapman, T. M. Recent Advances in the Design and Development of Formyl Peptide Receptor 2 (FPR2/ALX) Agonists as pro-Resolving Agents with Diverse Therapeutic Potential. *Eur. J. Med. Chem.* **2021**, *213*, 113167.

(145) Bürli, R. W.; Xu, H.; Zou, X.; Muller, K.; Golden, J.; Frohn, M.; Adlam, M.; Plant, M. H.; Wong, M.; McElvain, M.; Regal, K.; Viswanadhan, V. N.; Tagari, P.; Hungate, R.

Potent HFPRL1 (ALXR) Agonists as Potential Anti-Inflammatory Agents. *Bioorganic Med. Chem. Lett.* **2006**, *16*, 3713–3718.

(146) García, R. A.; Ito, B. R.; Lupisella, J. A.; Carson, N. A.; Hsu, M. Y.; Fernando, G.; Heroux, M.; Bouvier, M.; Dierks, E.; Kick, E. K.; Gordon, D. A.; Chen, J.; Mintier, G.; Carrier, M.; St-Onge, S.; Shah, H.; Towne, J.; Bucardo, M. S.; Ma, X.; Ryan, C. S.; Wurtz, N. R.; Ostrowski, J.; Villarreal, F. J. Preservation of Post-Infarction Cardiac Structure and Function via Long-Term Oral Formyl Peptide Receptor Agonist Treatment. *JACC Basic to Transl. Sci.* **2019**, *4*, 905–920.

(147) Frohn, M.; Xu, H.; Zou, X.; Chang, C.; McElvaine, M.; Plant, M. H.; Wong, M.; Tagari, P.; Hungate, R.; Bürli, R. W. New "chemical Probes" to Examine the Role of the HFPRL1 (or ALXR) Receptor in Inflammation. *Bioorganic Med. Chem. Lett.* **2007**, *17*, 6633–6637.

(148) Corminboeuf, O.; Cren, S. 1-(p-Tolyl)Cyclopropyl Substituted Bridged Spiro[2.4]Heptane Derivatives as ALX Receptor Agonists. PCT Int. Appl. WO2013/171687, 2013.

(149) Bur, D.; Corminboeuf, O.; Cren, S.; Grisostomi, C.; Leroy, X.; Richard-Bikdstein, S. Fluorinated Aminotriazole Derivatives. PCT Int. Appl. WO2010/143116, 2010, 2010.

(150) Bena, S.; Brancaleone, V.; Wang, J. M.; Perretti, M.; Flower, R. J. Annexin A1 Interaction with the FPR2/ALX Receptor: Identification of Distinct Domains and Downstream Associated Signaling. *J. Biol. Chem.* **2012**, *287*, 24690–24697.

(151) Asahina, Y.; Wurtz, N. R.; Arakawa, K.; Carson, N.; Fujii, K.; Fukuchi, K.; Garcia, R.; Hsu, M. Y.; Ishiyama, J.; Ito, B.; Kick, E.; Lupisella, J.; Matsushima, S.; Ohata, K.; Ostrowski, J.; Saito, Y.; Tsuda, K.; Villarreal, F.; Yamada, H.; Yamaoka, T.; Wexler, R.; Gordon, D.; Kohno, Y. Discovery of BMS-986235/LAR-1219: A Potent Formyl Peptide Receptor 2 (FPR2) Selective Agonist for the Prevention of Heart Failure. *J. Med. Chem.* **2020**, *63*, 9003–9019.

(152) García, R. A.; Lupisella, J. A.; Ito, B. R.; Hsu, M. Y.; Fernando, G.; Carson, N. L.; Allocco, J. J.; Ryan, C. S.; Zhang, R.; Wang, Z.; Heroux, M.; Carrier, M.; St.-Onge, S.; Bouvier, M.; Dudhgaonkar, S.; Nagar, J.; Bustamante-Pozo, M. M.; Garate-Carrillo, A.; Chen, J.; Ma, X.; Search, D. J.; Dierks, E. A.; Kick, E. K.; Wexler, R. R.; Gordon, D. A.; Ostrowski, J.; Wurtz, N. R.; Villarreal, F. Selective FPR2 Agonism Promotes a Proresolution Macrophage Phenotype and Improves Cardiac Structure-Function Post Myocardial Infarction. *JACC Basic to Transl. Sci.* **2021**, *6*, 676–689.

(153) Chen, T.; Xiong, M.; Zong, X.; Ge, Y.; Zhang, H.; Wang, M.; Won Han, G.; Yi, C.; Ma, L.; Ye, R. D.; Xu, Y.; Zhao, Q.; Wu, B. Structural Basis of Ligand Binding Modes at the Human Formyl Peptide Receptor 2. *Nat. Commun.* **2020**, *11*, 1–9.

(154) Zhuang, Y.; Liu, H.; Edward Zhou, X.; Kumar Verma, R.; de Waal, P. W.; Jang, W.; Xu, T. H.; Wang, L.; Meng, X.; Zhao, G.; Kang, Y.; Melcher, K.; Fan, H.; Lambert, N. A.; Eric Xu, H.; Zhang, C. Structure of Formylpeptide Receptor 2-Gi Complex Reveals Insights into Ligand Recognition and Signaling. *Nat. Commun.* **2020**, *11*, 1–12.

(155) Venkatakrishnan, A. J.; Deupi, X.; Lebon, G.; Tate, C. G.; Schertler, G. F.; Madan Babu,

M. Molecular Signatures of G-Protein-Coupled Receptors. *Nature* **2013**, *494*, 185–194.

(156) Dror, R. O.; Pan, A. C.; Arlow, D. H.; Borhani, D. W.; Maragakis, P.; Shan, Y.; Xu, H.; Shaw, D. E. Pathway and Mechanism of Drug Binding to G-Protein-Coupled Receptors. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 13118–13123.

(157) Maestro, Schrödinger, LLC, New York, N. Schrödinger Release 2020-4. 2020.

(158) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *J. Chem. Theory Comput.* **2015**.

(159) Frisch G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratch, D. J., M. J. . T. Gaussian 16, Rev. B.01. *Gaussian, Inc., Wallingford, CT* **2016**.

(160) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation. *Comput. Phys. Commun.* **1995**, *91*, 43–56.

(161) Hess, B., Bekker, H., Berendsen, H.J.C. & Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. Journal of Computational Chemistry. *J. Comput. Chem.* **1997**.

(162) Berendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W. F.; Dinola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, No. 18, 1463–1472.

(163) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.

(164) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.

(165) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New Feathers for an Old Bird. *Comput. Phys. Commun.* **2014**, *185*, 604–613.

(166) Bonomi, M.; Bussi, G.; Camilloni, C.; Tribello, G. A.; Banáš, P.; Barducci, A.; Bernetti, M.; Bolhuis, P. G.; Bottaro, S.; Branduardi, D.; Capelli, R.; Carloni, P.; Ceriotti, M.; Cesari, A.; Chen, H.; Chen, W.; Colizzi, F.; De, S.; De La Pierre, M.; Donadio, D.; Drobot, V.; Ensing, B.; Ferguson, A. L.; Filizola, M.; Fraser, J. S.; Fu, H.; Gasparotto, P.; Gervasio, F. L.; Giberti, F.; Gil-Ley, A.; Giorgino, T.; Heller, G. T.; Hocky, G. M.; Iannuzzi, M.; Invernizzi, M.; Jelfs, K. E.; Jussupow, A.; Kirilin, E.; Laio, A.; Limongelli, V.; Lindorff-Larsen, K.; Löhr, T.; Marinelli, F.; Martin-Samos, L.; Masetti, M.; Meyer, R.; Michaelides, A.; Molteni, C.; Morishita, T.; Nava, M.; Paissoni, C.; Papaleo, E.; Parrinello, M.; Pfaendtner, J.; Piaggi, P.; Piccini, G. M.; Pietropaolo, A.; Pietrucci, F.; Pipolo, S.; Provasi, D.; Quigley, D.; Raiteri, P.; Raniolo, S.; Rydzewski, J.; Salvalaglio, M.; Sosso, G. C.; Spiwok, V.; Šponer, J.; Swenson, D. W. H.; Tiwary, P.; Valsson, O.; Vendruscolo, M.; Voth, G. A.; White, A. Promoting Transparency and Reproducibility in Enhanced Molecular Simulations. *Nat. Methods* **2019**, *16*, 670–673.

(167) Avraamides, C. J.; Garmy-Susini, B.; Varner, J. A. Integrins in Angiogenesis and

Lymphangiogenesis. *Nat. Rev. Cancer* **2008**, *8*, 604–617.

(168) Hamidi, H.; Ivaska, J. Every Step of the Way: Integrins in Cancer Progression and Metastasis. *Nat. Rev. Cancer* **2018**, 1–16.

(169) Hussein, H. A. M.; Walker, L. R.; Abdel-Raouf, U. M.; Desouky, S. A.; Montasser, A. K. M.; Akula, S. M. Beyond RGD: Virus Interactions with Integrins. *Arch. Virol.* **2015**, *160*, 2669–2681.

(170) Campadelli-Fiume, G.; Collins-Mcmillen, D.; Gianni, T.; Yurochko, A. D. Integrins as Herpesvirus Receptors and Mediators of the Host Signalosome. *Annu. Rev. Virol.* **2016**, *3*, 215–236.

(171) Diefenbach, R. J.; Miranda-Saksena, M.; Douglas, M. W.; Cunningham, A. L. Transport and Egress of Herpes Simplex Virus in Neurons. *Rev. Med. Virol.* **2008**, *18*, 35–51.

(172) Jackson, T.; Sheppard, D.; Denyer, M.; Blakemore, W.; King, A. M. Q. The Epithelial Integrin Avβ6 Is a Receptor for Foot-and-Mouth Disease Virus. *J. Virol.* **2000**, *74*, 4949–4956.

(173) Burman, A.; Clark, S.; Abrescia, N. G. A.; Fry, E. E.; Stuart, D. I.; Jackson, T. Specificity of the VP1 GH Loop of Foot-and-Mouth Disease Virus for Av Integrins. *J. Virol.* **2006**, *80*, 9798–9810.

(174) O'Donnell, V.; Pacheco, J. M.; Gregg, D.; Baxt, B. Analysis of Foot-and-Mouth Disease Virus Integrin Receptor Expression in Tissues from Naïve and Infected Cattle. *J. Comp. Pathol.* **2009**, *141*, 98–112.

(175) Kotecha, A.; Wang, Q.; Dong, X.; Ilca, S. L.; Ondiviela, M.; Zihe, R.; Seago, J.; Charleston, B.; Fry, E. E.; Abrescia, N. G. A.; Springer, T. A.; Huiskonen, J. T.; Stuart, D. I. Rules of Engagement between Avβ6 Integrin and Foot-and-Mouth Disease Virus. *Nat. Commun.* **2017**, *8*, 15408.

(176) Krummenacher, C.; Supekar, V. M.; Whitbeck, J. C.; Lazear, E.; Connolly, S. A.; Eisenberg, R. J.; Cohen, G. H.; Wiley, D. C.; Carfí, A. Structure of Unliganded HSV GD Reveals a Mechanism for Receptor-Mediated Activation of Virus Entry. *EMBO J.* **2005**, *24*, 4144–4153.

(177) Gianni, T.; Amasio, M.; Campadelli-Fiume, G. Herpes Simplex Virus GD Forms Distinct Complexes with Fusion Executors GB and GH/GL in Part through the C-Terminal Profusion Domain. *J. Biol. Chem.* **2009**, *284*, 17370–17382.

(178) Carfí, A.; Willis, S. H.; Whitbeck, J. C.; Krummenacher, C.; Cohen, G. H.; Eisenberg, R. J.; Wiley, D. C. Herpes Simplex Virus Glycoprotein D Bound to the Human Receptor HveA. *Mol. Cell* **2001**, *8*, 169–179.

(179) Atanasiu, D.; Saw, W. T.; Cohen, G. H.; Eisenberg, R. J. Cascade of Events Governing Cell-Cell Fusion Induced by Herpes Simplex Virus Glycoproteins GD, GH/GL, and GB. *J. Virol.* **2010**, *84*, 12292–12299.

(180) Gianni, T.; Massaro, R.; Campadelli-Fiume, G. Dissociation of HSV GL from GH by

Avβ6- or Avβ8-Integrin Promotes GH Activation and Virus Entry. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, E3901–E3910.

(181) Gianni, T.; Salvioli, S.; Chesnokova, L. S.; Hutt-Fletcher, L. M.; Campadelli-Fiume, G. Avβ6- and Avβ8-Integrins Serve As Interchangeable Receptors for HSV GH/GL to Promote Endocytosis and Activation of Membrane Fusion. *PLoS Pathog.* **2013**, *9*, 1–14.

(182) Maltsev, O. V.; Marelli, U. K.; Kapp, T. G.; Di Leva, F. S.; Di Maro, S.; Nieberler, M.; Reuning, U.; Schwaiger, M.; Novellino, E.; Marinelli, L.; Kessler, H. Stable Peptides Instead of Stapled Peptides: Highly Potent Avβ6-Selective Integrin Ligands. *Angew. Chemie - Int. Ed.* **2016**, *55*, 1535–1539.

(183) Di Leva, F. S.; Tomassi, S.; Di Maro, S.; Reichart, F.; Notni, J.; Dangi, A.; Marelli, U. K.; Brancaccio, D.; Merlino, F.; Wester, H. J.; Novellino, E.; Kessler, H.; Marinelli, L. From a Helix to a Small Cycle: Metadynamics-Inspired Avβ6 Integrin Selective Ligands. *Angew. Chemie - Int. Ed.* **2018**, No. 57, 14645–14649.

(184) Reichart, F.; Maltsev, O. V.; Kapp, T. G.; Räder, A. F. B.; Weinmüller, M.; Marelli, U. K.; Notni, J.; Wurzer, A.; Beck, R.; Wester, H. J.; Steiger, K.; Di Maro, S.; Di Leva, F. S.; Marinelli, L.; Nieberler, M.; Reuning, U.; Schwaiger, M.; Kessler, H. Selective Targeting of Integrin Avβ8 by a Highly Active Cyclic Peptide. *J. Med. Chem.* **2019**, *62*, 2024–2037.

(185) Nardelli, F.; Ghitti, M.; Quilici, G.; Gori, A.; Luo, Q.; Berardi, A.; Sacchi, A.; Monieri, M.; Bergamaschi, G.; Bermel, W.; Chen, F.; Corti, A.; Curnis, F.; Musco, G. A Stapled Chromogranin A-Derived Peptide Is a Potent Dual Ligand for Integrins Avβ6 and Avβ8. *Chem. Commun.* **2019**, *55*, 14777–14780.

(186) Quigley, N. G.; Tomassi, S.; Saverio Di Leva, F.; Di Maro, S.; Richter, F.; Steiger, K.; Kossatz, S.; Marinelli, L.; Notni, J. Click-Chemistry (CuAAC) Trimerization of an Avβ6 Integrin Targeting Ga-68-Peptide: Enhanced Contrast for in-Vivo PET Imaging of Human Lung Adenocarcinoma Xenografts. *ChemBioChem* **2020**, *21*, 2836–2843.

(187) Bugatti, K.; Bruno, A.; Arosio, D.; Sartori, A.; Curti, C.; Augustijn, L.; Zanardi, F.; Battistini, L. Shifting Towards AVβ6 Integrin Ligands Using Novel Aminoproline-Based Cyclic Peptidomimetics. *Chem. - A Eur. J.* **2020**, *26*, 13468–13475.

(188) Dechantsreiter, M. A.; Planker, E.; Matha, B.; Lohof, E.; Holzemann, G.; Jonczyk, A.; Goodman, S. L.; Kessler, H. N-Methylated Cyclic RGD Peptides as Highly Active and Selective V3 Integrin. *J. Med. Chem.* **1999**, *42*, 3033–3040.

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

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

(191) Mas-Moruno, C.; Beck, J. G.; Doedens, L.; Frank, A. O.; Marinelli, L.; Cosconati, S.; Novellino, E.; Kessler, H. Increasing Avβ3 Selectivity of the Anti-Angiogenic Drug

Cilengitide by N-Methylation. *Angew. Chemie - Int. Ed.* **2011**, *50*, 9496–9500.

(192) Kapp, T. G.; Di Leva, F. S.; Notni, J.; Räder, A. F. B.; Fottner, M.; Reichart, F.; Reich, D.; Wurzer, A.; Steiger, K.; Novellino, E.; Marelli, U. K.; Wester, H. J.; Marinelli, L.; Kessler, H. N -Methylation of Iso DGR Peptides: Discovery of a Selective A5β1-Integrin Ligand as a Potent Tumor Imaging Agent. *J. Med. Chem.* **2018**.

(193) Ovadia, O.; Greenberg, S.; Chatterjee, J.; Laufer, B.; Opperer, F.; Kessler, H.; Gilon, C.; Hoffman, A. The Effect of Multiple N-Methylation on Intestinal Permeability of Cyclic Hexapeptides. *Mol. Pharm.* **2011**, *8*, 479–487.

(194) Merlino, F.; Tomassi, S.; Yousif, A. M.; Messere, A.; Marinelli, L.; Grieco, P.; Novellino, E.; Cosconati, S.; Di Maro, S. Boosting Fmoc Solid-Phase Peptide Synthesis by Ultrasonication. *Org. Lett.* **2019**, *21*, 6378–6382.

(195) Tomassi, S.; D'Amore, V. M.; Di Leva, F. S.; Vannini, A.; Quilici, G.; Weinmüller, M.; Reichart, F.; Amato, J.; Romano, B.; Izzo, A. A.; Di Maro, S.; Novellino, E.; Musco, G.; Gianni, T.; Kessler, H.; Marinelli, L. Halting the Spread of Herpes Simplex Virus-1: The Discovery of an Effective Dual Avβ6/Avβ8 Integrin Ligand. *J. Med. Chem.* **2021**, *64*, 6972–6984.

(196) Boomsma, W.; Ferkinghoff-Borg, J.; Lindorff-Larsen, K. Combining Experiments and Simulations Using the Maximum Entropy Principle. *PLoS Comput. Biol.* **2014**, *10*, 1–9.

(197) D'Annessa, I.; Di Leva, F. S.; La Teana, A.; Novellino, E.; Limongelli, V.; Di Marino, D. Bioinformatics and Biosimulations as Toolbox for Peptides and Peptidomimetics Design: Where Are We? *Front. Mol. Biosci.* **2020**, *7*, 1–11.

(198) Dong, X.; Zhao, B.; Iacob, R. E.; Zhu, J.; Koksal, A. C.; Lu, C.; Engen, J. R.; Springer, T. A. Force Interacts with Macromolecular Structure in Activation of TGF-β. *Nature* **2017**, *542*, 55–59.

(199) Wang, J.; Su, Y.; Iacob, R. E.; Engen, J. R.; Springer, T. A. General Structural Features That Regulate Integrin Affinity Revealed by Atypical AVβ8. *Nat. Commun.* **2019**, *10*, 1–13.

(200) Marelli, U. K.; Frank, A. O.; Wahl, B.; Pietra, V. La; Novellino, E.; Marinelli, L.; Herdtweck, E.; Groll, M.; Kessler, H. Receptor-Bound Conformation of Cilengitide Better Represented by Its Solution-State Structure than the Solid-State Structure. *Chem. - A Eur. J.* **2014**, *20*, 14201–14206.

(201) Cocchi, F.; Menotti, L.; Mirandola, P.; Lopez, M.; Campadelli-Fiume, G. The Ectodomain of a Novel Member of the Immunoglobulin Subfamily Related to the Poliovirus Receptor Has the Attributes of a Bona Fide Receptor for Herpes Simplex Virus Types 1 and 2 in Human Cells. *J. Virol.* **1998**, *72*, 9992–10002.

(202) Desai, P.; Person, S. Incorporation of the Green Fluorescent Protein into the Herpes Simplex Virus Type 1 Capsid. *J. Virol.* **1998**, *72*, 7563–7568.

(203) Marcocci, M. E.; Amatore, D.; Villa, S.; Casciaro, B.; Aimola, P.; Franci, G.; Grieco, P.; Galdiero, M.; Palamara, A. T.; Mangoni, M. L.; Nencioni, L. The Amphibian Antimicrobial Peptide Temporin b Inhibits in Vitro Herpes Simplex Virus 1 Infection.

Antimicrob. Agents Chemother. **2018**, *62*, 1–13.

(204) Fox, T.; Kollman, P. A. Application of the RESP Methodology in the Parametrization of Organic Solvents. *J. Phys. Chem. B* **1998**, *102*, 8070–8079.

(205) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472.

(206) Taliani, S.; Pugliesi, I.; Barresi, E.; Salerno, S.; Marchand, C.; Agama, K.; Simorini, F.; La Motta, C.; Marini, A. M.; Di Leva, F. S.; Marinelli, L.; Cosconati, S.; Novellino, E.; Pommier, Y.; Di Santo, R.; Da Settimo, F. Phenylpyrazolo[1,5-a]Quinazolin-5(4 H)-One: A Suitable Scaffold for the Development of Noncamptothecin Topoisomerase i (Top1) Inhibitors. *J. Med. Chem.* **2013**, *56*, 7458–7462.

(207) Hambley, T. W. Is Anticancer Drug Development Heading in the Right Direction? *Cancer Res.* **2009**, *69*, 1259–1261.

(208) Bartelink, I. H.; Jones, E. F.; Shahidi-Latham, S. K.; Lee, P. R. E.; Zheng, Y.; Vicini, P.; van 't Veer, L.; Wolf, D.; Iagaru, A.; Kroetz, D. L.; Prideaux, B.; Cilliers, C.; Thurber, G. M.; Wimana, Z.; Gebhart, G. Tumor Drug Penetration Measurements Could Be the Neglected Piece of the Personalized Cancer Treatment Puzzle. *Clin. Pharmacol. Ther.* **2019**, *106*, 148–163.

(209) Mould, D. R.; Hutson, P. R. Critical Considerations in Anticancer Drug Development and Dosing Strategies: The Past, Present, and Future. *J. Clin. Pharmacol.* **2017**, *57*, S116–S128.

(210) Dixit, S.; Novak, T.; Miller, K.; Zhu, Y.; Kenney, M. E.; Broome, A. M. Transferrin Receptor-Targeted Theranostic Gold Nanoparticles for Photosensitizer Delivery in Brain Tumors. *Nanoscale* **2015**, *7*, 1782–1790.

(211) Lee, C.; Hwang, H. S.; Lee, S.; Kim, B.; Kim, J. O.; Oh, K. T.; Lee, E. S.; Choi, H. G.; Youn, Y. S. Rabies Virus-Inspired Silica-Coated Gold Nanorods as a Photothermal Therapeutic Platform for Treating Brain Tumors. *Adv. Mater.* **2017**, *29*.

(212) Ruan, S.; Hu, C.; Tang, X.; Cun, X.; Xiao, W.; Shi, K.; He, Q.; Gao, H. Increased Gold Nanoparticle Retention in Brain Tumors by in Situ Enzyme-Induced Aggregation. *ACS Nano* **2016**, *10*, 10086–10098.

(213) Ying, M.; Zhan, C.; Wang, S.; Yao, B.; Hu, X.; Song, X.; Zhang, M.; Wei, X.; Xiong, Y.; Lu, W. Liposome-Based Systemic Glioma-Targeted Drug Delivery Enabled by All- d Peptides. *ACS Appl. Mater. Interfaces* **2016**, *8*, 29977–29985.

(214) Gu, Z.; Da Silva, C. G.; van der Maaden, K.; Ossendorp, F.; Cruz, L. J. Liposome-Based Drug Delivery Systems in Cancer Immunotherapy. *Pharmaceutics* **2020**, *12*, 1–25.

(215) Ran, D.; Mao, J.; Shen, Q.; Xie, C.; Zhan, C.; Wang, R.; Lu, W. GRP78 Enabled Micelle-Based Glioma Targeted Drug Delivery. *J. Control. Release* **2017**, *255*, 120–131.

(216) Hoppenz, P.; Els-Heindl, S.; Beck-Sickinger, A. G. Peptide-Drug Conjugates and Their Targets in Advanced Cancer Therapies. *Front. Chem.* **2020**, *8*, 1–24.

(217) Sugahara, K. N.; Teesalu, T.; Karmali, P. P.; Kotamraju, V. R.; Agemy, L.; Girard, O.

M.; Hanahan, D.; Mattrey, R. F.; Ruoslahti, E. Tissue-Penetrating Delivery of Compounds and Nanoparticles into Tumors. *Cancer Cell* **2009**, *16*, 510–520.

(218) Sugahara, K. N.; Teesalu, T.; Prakash Karmali, P.; Ramana Kotamraju, V.; Agemy, L.; Greenwald, D. R.; Ruoslahti, E. Coadministration of a Tumor-Penetrating Peptide Enhances the Efficacy of Cancer Drugs. *Science (80-.).* **2010**, *328*, 1031–1035.

(219) Kang, S.; Lee, S.; Park, S. IRGD Peptide as a Tumor-Penetrating Enhancer for Tumor-Targeted Drug Delivery. *Polymers (Basel).* **2020**, *12*.

(220) Yin, H.; Zhang, Q.; Yang, J.; Wang, H.; Xu, J.; Zheng, J. IRGD as a Tumor-Penetrating Peptide for Cancer Therapy (Review). *Mol. Med. Rep.* **2017**, *15*, 2925–2930.

(221) Liu, C.; Yao, S.; Li, X.; Wang, F.; Jiang, Y. IRGD-Mediated Core-Shell Nanoparticles Loading Carmustine and O6-Benzylguanine for Glioma Therapy. *J. Drug Target.* **2017**, *25*, 235–246.

(222) Lu, L.; Zhao, X.; Fu, T.; Li, K.; He, Y.; Luo, Z.; Dai, L.; Zeng, R.; Cai, K. An IRGD-Conjugated Prodrug Micelle with Blood-Brain-Barrier Penetrability for Anti-Glioma Therapy. *Biomaterials* **2020**, *230*.

(223) Shi, X.; Ma, R.; Lu, Y.; Cheng, Y.; Fan, X.; Zou, J.; Zheng, H.; Li, F.; Piao, J. G. IRGD and TGN Co-Modified PAMAM for Multi-Targeted Delivery of ATO to Gliomas. *Biochem. Biophys. Res. Commun.* **2020**, *527*, 117–123.

(224) Dean, A.; Gill, S.; McGregor, M.; Broadbridge, V.; Jarvelainen, H. A.; Price, T. J. 1528P Phase I Trial of the First-in-Class Agent CEND-1 in Combination with Gemcitabine and Nab-Paclitaxel in Patients with Metastatic Pancreatic Cancer. *Ann. Oncol.* **2020**, *31*, S941.

(225) Teoh, C.; Tan, S.; Tran, T. Integrins as Therapeutic Targets for Respiratory Diseases. *Curr. Mol. Med.* **2015**, *15*, 714–734.

(226) Teesalu, T.; Sugahara, K. N.; Kotamraju, V. R.; Ruoslahti, E. C-End Rule Peptides Mediate Neuropilin-1-Dependent Cell, Vascular, and Tissue Penetration. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 16157–16162.

(227) Sugahara, K. N.; Braun, G. B.; De Mendoza, T. H.; Kotamraju, V. R.; French, R. P.; Lowy, A. M.; Teesalu, T.; Ruoslahti, E. Tumor-Penetrating IRGD Peptide Inhibits Metastasis. *Mol. Cancer Ther.* **2015**, *14*, 120–128.

(228) Hurtado de Mendoza, T.; Mose, E. S.; Botta, G. P.; Braun, G. B.; Kotamraju, V. R.; French, R. P.; Suzuki, K.; Miyamura, N.; Teesalu, T.; Ruoslahti, E.; Lowy, A. M.; Sugahara, K. N. Tumor-Penetrating Therapy for B5 Integrin-Rich Pancreas Cancer. *Nat. Commun.* **2021**, *12*, 1–13.

(229) Hood, J. D.; Frausto, R.; Kiosses, W. B.; Schwartz, M. A.; Cheresh, D. A. Differential Av Integrin-Mediated Ras-ERK Signaling during Two Pathways of Angiogenesis. *J. Cell Biol.* **2003**, *162*, 933–943.

(230) Xiong, J. P.; Stehle, T.; Zhang, R.; Joachimiak, A.; Frech, M.; Goodman, S. L.; Arnaout, M. A. Crystal Structure of the Extracellular Segment of Integrin AVβ3 in Complex

with an Arg-Gly-Asp Ligand. *Science (80-.).* **2002**, *296*, 151–155.

(231) Van Agthoven, J. F.; Xiong, J. P.; Alonso, J. L.; Rui, X.; Adair, B. D.; Goodman, S. L.; Arnaout, M. A. Structural Basis for Pure Antagonism of Integrin α Vβ 3 by a High-Affinity Form of Fibronectin. *Nat. Struct. Mol. Biol.* **2014**, *21*, 383–388.

(232) Dong, X.; Hudson, N. E.; Lu, C.; Springer, T. A. Structural Determinants of Integrin β-Subunit Specificity for Latent TGF-β. *Nat. Struct. Mol. Biol.* **2014**, *21*, 1091–1096.

(233) Dong, X.; Hudson, N. E.; Lu, C.; Springer, T. A. Structural Determinants of Integrin β-Subunit Specificity for Latent TGF-β. *Nat. Struct. Mol. Biol.* **2014**, *21*, 1091–1096.

(234) Munger, J. S.; Huang, X.; Kawakatsu, H.; Griffiths, M. J. D.; Dalton, S. L.; Wu, J.; Pittet, J. F.; Kaminski, N.; Garat, C.; Matthay, M. A.; Rifkin, D. B.; Sheppard, D. The Integrin Avβ6 Binds and Activates Latent TGFβ1: A Mechanism for Regulating Pulmonary Inflammation and Fibrosis. *Cell* **1999**.

(235) Conroy, K. P.; Kitto, L. J.; Henderson, N. C. Av Integrins: Key Regulators of Tissue Fibrosis. *Cell Tissue Res.* **2016**, *365*, 511–519.

(236) Day, T. J. F.; Honig, B.; Shaw, D. E. A Hierarchical Approach to All-Atom Protein Loop Prediction - Jacobson - 2004 - Proteins: Structure, Function, and Bioinformatics - Wiley Online Library. *Proteins Struct.* **2004**, *55*, 351–367.

(237) Panteva, M. T.; Giambaşu, G. M.; York, D. M. Comparison of Structural, Thermodynamic, Kinetic and Mass Transport Properties of Mg2+ Ion Models Commonly Used in Biomolecular Simulations. *J. Comput. Chem.* **2015**, *36*, 970–982.

(238) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005**, *26*, 1668–1688.

(239) Kräutler, V.; Van Gunsteren, W. F.; Hünenberger, P. H. A Fast SHAKE Algorithm to Solve Distance Constraint Equations for Small Molecules in Molecular Dynamics Simulations. *J. Comput. Chem.* **2001**, *22*, 501–508.

(240) Loncharich, R. J.; Brooks, B. R.; Pastor, R. W. Langevin Dynamics of Peptides: The Frictional Dependence of Isomerization Rates of N-acetylalanyl-N′-methylamide. *Biopolymers* **1992**, *32*, 523–535.