



**UNIVERSITY OF SALERNO**

**DEPARTMENT OF PHARMACY**



**PhD course in Scienza e Tecnologie per l'Industria Chimica,  
Farmaceutica e Alimentare  
XI course NS (XXV)  
2009-2012**

**New techniques of molecular modelling and structural chemistry  
for the development of bioactive compounds**

**Tutor**

Prof. Giuseppe Bifulco

**PhD student**

Gianluigi Lauro

**Coordinators**

Prof. Nunziatina De Tommasi

Prof. Paolo Ciambelli

## **Preface**

My PhD three years course in Pharmaceutical Sciences at the Department of Pharmaceutical and Biomedical Sciences of Salerno University was started in 2010 under the supervision of Prof. Giuseppe Bifulco.

My research activity was mainly focused on studies of ligand-receptor interactions and structural characterization by computational techniques in order to identify new antitumor molecules potentially utilizable in therapy.

In particular, I was mainly interested into the development of a new computational technique named Inverse Virtual Screening. The application of this approach led to the identification of the targets of interaction of several natural compounds.

Furthermore, to improve my knowledge on computational chemistry, I moved to the The University Pompeu Fabra-PRBB (Parc de Recerca Biomèdica de Barcelona) in 2012 (mid-May until end of October 2012) under the supervision of Dr. Gianni De Fabritiis.

During this period in his laboratory, my research work has included learning High-throughput Molecular Dynamics simulations, taking the advantage of his expertise in the field of molecular simulation applied to relevant systems in a drug design perspective, and his unique hardware and software infrastructure (GPU grid-ACEMD software).

In addition to PhD course activities, I was involved in several other projects, mainly regarding the characterization of ligand-targets interactions of ligands on receptor targets (PXR, PPAR- $\gamma$ , HSP70 1A), in order to elucidate the molecular basis of their activity.

**List of publications related to the scientific activity performed during the three years PhD course in Pharmaceutical Sciences:**

1 Dal Piaz, F.; Cotugno, R.; Lepore, L.; Vassallo, A.; Malafrente, N.; Lauro, G.; Bifulco, G.; Belisario, M. A.; De Tommasi, N. Chemical Proteomics Reveals HSP70 1A as a Target for the Anticancer Diterpene Oridonin. *J. Proteomics*, submitted

2 Festa, C.; Lauro, G.; De Marino, S.; D'Auria, M. V.; Monti, M. C.; Casapullo, A.; D'Amore, C.; Renga, B.; Mencarelli, A.; Petek, S.; Bifulco, G.; Fiorucci, S.; Zampella, A. Plakilactones from the marine sponge *Plakinastrella mamillaris*. Discovery of a new class of marine ligands of peroxisome proliferator-activated receptor  $\gamma$ . *J. Med. Chem.* **2012**, *55*, 8303-8317.

3 Sepe, V.; Ummarino, R.; D'Auria, M. V.; Lauro, G.; Bifulco, G.; D'Amore, C.; Renga, B.; Fiorucci, S.; Zampella, A. Modification in the side chain of solomonsterol A: discovery of cholestan disulfate as a potent pregnane-X-receptor agonist. *Org. Biomol. Chem.* **2012**, *10*, 6350-6362.

4 Lauro, G.; Masullo, M.; Piacente, S.; Riccio, R.; Bifulco, G. Inverse Virtual Screening allows the discovery of the biological activity of natural compounds. *Bioorg. Med. Chem.* **2012**, *20*, 3596-3602.

5 Cheruku, P.; Plaza, A.; Lauro, G.; Keffer, J.; Lloyd, J. R.; Bifulco, G.; Bewley, C. A. Discovery and synthesis of namalide reveals a new anabaenopeptin scaffold and peptidase inhibitor. *J. Med. Chem.* **2012**, *55*, 735-742.

6 Lauro, G.; Romano, A.; Riccio, R.; Bifulco, G. Inverse virtual screening of antitumor targets: pilot study on a small database of natural bioactive compounds. *J. Nat. Prod.* **2011**, *74*, 1401-1407.

## Table of Contents

<b>Abstract</b>	.....	<b>I-III</b>
		<b>Page</b>
<b>Introduction</b>	.....	<b>1-47</b>
<b>Chapter 1</b>	Computational chemistry and drug discovery	<b>3-8</b>
<b>1.1</b>	Drug discovery process and computational chemistry	
<b>1.2</b>	Methodologies employed	<b>9</b>
<b>1.2.1</b>	Molecular Docking	<b>9-10</b>
<b>1.2.1.1</b>	Sampling	<b>10-12</b>
<b>1.2.1.2</b>	Scoring	<b>13-16</b>
<b>1.2.1.3</b>	Autodock: An Overview	<b>16-25</b>
<b>1.2.2</b>	Molecular Dynamics	<b>26-31</b>
<b>1.2.3</b>	Methods for the accurate calculation of the binding affinities	<b>31-32</b>
<b>1.2.3.1</b>	“Pathway” methods	<b>31-32</b>
<b>1.2.3.1.1</b>	Free Energy Perturbation	<b>33-35</b>
<b>1.2.3.1.2</b>	Thermodynamic Integration	<b>35-37</b>
<b>1.2.3.2</b>	“Endpoint” methods	<b>37</b>
<b>1.2.3.2.1</b>	Linear Interaction Energy (LIE)	<b>37-41</b>
<b>1.2.3.2.2</b>	MM-GB(PB)SA	<b>41-47</b>

---

<b>Chapter 2</b>	Inverse Virtual Screening	<b>48</b>
<b>2.1</b>	Inverse Virtual Screening: Introduction	<b>49-52</b>
<b>2.2</b>	Pilot Inverse Virtual Screening study: LIBIOMOL library	<b>53-55</b>
<b>2.2.1</b>	Analysis of Predicted Binding Energies	<b>56</b>
<b>2.2.2</b>	Comparing Standard Ligands	<b>56-58</b>
<b>2.2.3</b>	Introducing Ligand Efficiency (LE)	<b>58-60</b>
<b>2.2.4</b>	Normalization of the matrix	<b>61-67</b>
<b>2.3</b>	Re-evaluation of the biological activity of a small library of natural compounds	<b>68-81</b>
<b>2.4</b>	Discovery of peptidase inhibitory activity of the new-anabaenopeptin cyclopeptide namalide	<b>82-94</b>
<b>Chapter 3</b>	High-Throughput Molecular Dynamics for the accurate calculations of the binding affinities	<b>95</b>
<b>3.1</b>	Molecular Dynamics: from to CPU to GPU architecture	<b>96-97</b>
<b>3.2</b>	Re-ranking of Molecular Docking calculations using the Linear Interaction Energy (LIE) method	<b>97-103</b>
<b>Chapter 4</b>	Further applications of in silico screenings on natural compounds	<b>104</b>
<b>4.1</b>	Discovery of cholestan disulfate as a potent pregnane-X-receptor agonist	<b>105-108</b>
<b>4.1.1</b>	Biological studies	<b>108-113</b>
<b>4.1.2</b>	Molecular modelling studies	<b>113-116</b>
<b>4.1.3</b>	Final remarks	<b>116-118</b>

---

<b>4.2</b>	Plakilactones from the marine sponge Plakinastrella Mamillaris, a new class of marine ligands of peroxisome proliferator-activated receptor $\gamma$	<b>119-121</b>
<b>4.2.1</b>	Biological studies	<b>122-128</b>
<b>4.2.2</b>	Molecular modelling studies	<b>128-135</b>
<b>4.2.3</b>	Final remarks	<b>135-139</b>
<b>4.3</b>	HSP70 1A as a Target for the Anticancer Diterpene Oridonin	<b>140-142</b>
<b>4.3.1</b>	Chemical Proteomics results	<b>142-146</b>
<b>4.3.2</b>	Biological studies	<b>146-152</b>
<b>4.3.3</b>	Molecular modelling studies	<b>152-157</b>
<b>4.3.4</b>	Final remarks	<b>158</b>
<b>Chapter 5</b>	Experimental section	<b>159</b>
<b>5.1</b>	Inverse Virtual Screening	<b>160-163</b>
<b>5.2</b>	High- Throughput Molecular Dynamics for the accurate calculations of the binding affinities	<b>163-164</b>
<b>5.3</b>	Further applications of in silico screenings on natural compounds	<b>165-169</b>
<b>Conclusions</b>	.....	<b>170-176</b>
<b>Appendix A</b>	.....	<b>177</b>
<b>A.1</b>	Cancer: Some Data	<b>178-179</b>
<b>A.2</b>	Cancer Pathways	<b>179</b>
<b>A.2.1</b>	Apoptosis	<b>179-183</b>

<b>A.2.2</b>	Replicative lifespan	<b>183</b>
<b>A.2.3</b>	Proliferative signals	<b>183-184</b>
<b>A.2.4</b>	Cell cycle	<b>185-187</b>
<b>A.2.5</b>	Mobilization of resources	<b>187</b>
<b>References</b>		<b>188-213</b>

## Abstract

Computational chemistry represents today a valid and fast tool for the research of new compounds with potential biological activity. The analysis of ligand-macromolecule interactions and the evaluation of possible “binding modes” have a crucial role for the design and the development of new and more powerful drugs. *In silico* Virtual Screening campaigns of large libraries compounds (fragments or drug-like) on a specific target allow the selection of promising compounds, leading the identification of new scaffolds. The accurate analysis and the comparison of different bioactive compounds clarify the molecular basis of their interaction and the construction of pharmacoforic models.

In parallel, another crucial aspect of pharmacological research is the identification of targets of interaction of bioactive molecules, and this is particularly true for compounds from natural sources. In fact, a wide range of drug tests on a large number of biological targets can represent a useful approach for the study of natural products, but often one of the main problems is their limited availability.

Starting from these assumptions, a new computational method named Inverse Virtual Screening is described in details in this thesis. The different works based on this approach were performed considering panels of targets involved in the cancer events, determining the identification of the specific antitumor activity of the natural compounds investigated.

Inverse Virtual Screening studies were performed by means of molecular docking experiments on different natural compounds, organized in small libraries or as single compounds. Firstly, a mathematical method for the exclusion of false positive and false negative results was proposed applying a normalization of the predicted binding energies (expressed in kcal/mol) obtained from the docking calculations (paragraph 2.2). Then this approach was applied on a library of 10

compounds extracted from natural sources (paragraph 2.3), obtaining a good validation through *in vitro* biological tests. Afterwards, another study was performed on the cyclopeptide namalide. Its biological inhibitory activity and selectivity on Carboxipeptidase A target was in accordance with Inverse Virtual Screening results (paragraph 2.4).

Virtual Screening topic was also inspected analyzing the efficacy of Molecular Dynamics-based methods for the accurate calculations of the binding affinities. This work was conducted on a library of 1588 compounds (44 ligands + 1544 decoys) extracted from the DUD database on trypsin target, using the Linear Interaction Energy (LIE) method by means of extensive Molecular Dynamics simulations. Four different LIE results obtained combining different scaling factors were compared with docking results, evaluating and comparing ROC and enrichment curves for each of the considered methods. Poor results were obtained with LIE, and further analysis with MM-GBSA and MM-PBSA approaches are under investigation.

Moreover, *in silico* screenings were performed for the detailed study of natural compounds whose activities are known a priori. With this procedure, several binding modes were reported for a library of compounds on PXR target, whose activity or inactivity were rationalized comparing their binding poses with that of Solomonsterol A, used as a reference compound on this receptor. The presence/absence of biological activity of another library of compounds extracted from the marine sponge *Plakinastrella Mamillaris* on PPAR- $\gamma$  and for the diterpene oridonin on HSP70 1A are described at a molecular level respectively in paragraphs 4.2 and 4.3 with molecular docking and Molecular Dynamics simulations.

The putative binding modes for the reported molecules was described offering a complete rationalization of docking results, evaluating how ligand target specific

interactions (e.g. hydrophobic, hydrophilic, electrostatic contacts) can influence their biological activity.

# **-INTRODUCTION-**

# -CHAPTER 1-

## *Computational chemistry and drug discovery*

## 1.1 Drug discovery process and computational chemistry

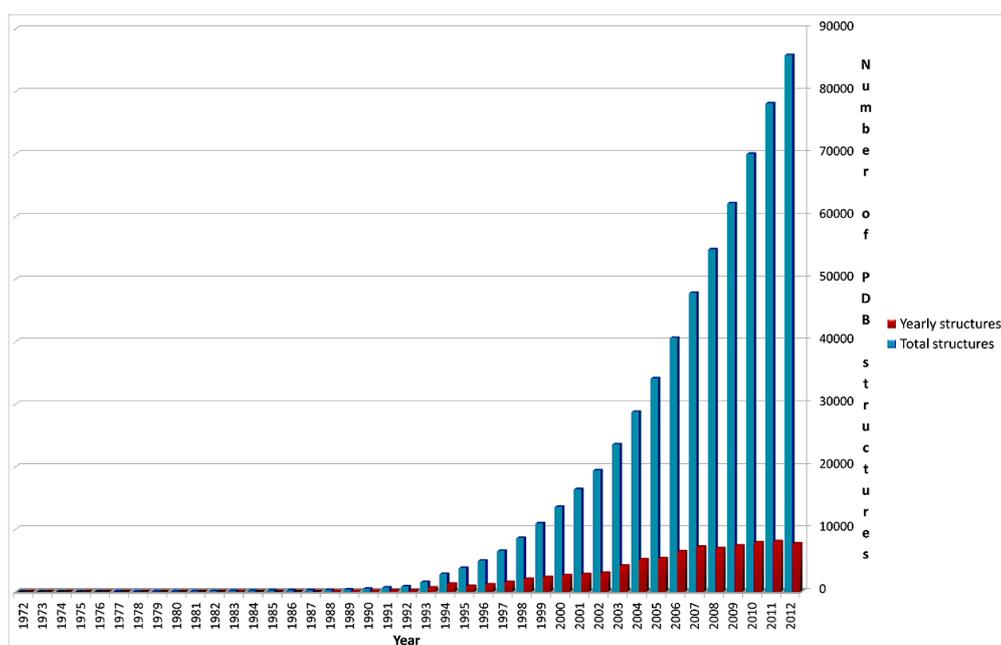
Drug discovery is an expensive process including the identification and validation of a drug target (typically a protein), determination of 3-D structure of that target, discovery of a lead compound that binds to the target and development of the lead into a drug candidate.

Computational chemistry represents today the preferential tool for predicting the putative binding modes and affinities of chemical compounds bound to a target. The difficulties and economic cost of the experimental methods explain the fast development of computational chemistry during the past years. The early development of computational chemistry was dependent on the developments in computational power and techniques. Computing power became widely available in the 1960s and a great number of new numerical models, or algorithms, were constantly produced worldwide. As computer power increased, the mathematical equations were refined in order to model real life more accurately.

Effective use of computational chemistry shortens the development cycle for new drugs and provides pharmaceutical companies a competitive advantage with faster time-to-discovery and reduced costs. Numerous examples of drugs have been discovered and optimized with contributions from computational chemists.<sup>1</sup> The primary goal of computational chemistry in drug discovery is to develop quantitative models that are able to predict activities of compounds quickly and accurately. Computer hardware and software is used to simulate a chemical process or to compute a chemical property. While it is now recognised that high-throughput methods are amply capable of producing greater volumes of data, they do not always increase the productivity and timeliness of a research effort. The problem in laboratories today is not one of acquiring data at each stage of drug discovery, but rather of extracting the useful information from the wealth that is available.

Fortunately, computational power and data-mining techniques are also advancing in step with demand.

Computational methods provide guidance but are by no means able to make perfect predictions. Molecular libraries are screened, and the resulting leads are optimized in a cycle that features design, synthesis and assaying of numerous analogs, and animal studies. Crystal structure determination for complexes of some analogs with the biomolecular target is often possible, which enables “structure-based drug design” (SBDD)<sup>2</sup> and the efficient optimization of leads. SBDD is based on the knowledge of the 3-dimensional structure of the protein and, preferably co-complexed with a ligand as an identifier of the binding site are exploited for the design and the optimization of lead compounds.<sup>3,4,5</sup>



**Figure 1.1** Number of yearly and total PDB structures available on PDB database from 1972 until the end of 2012

As shown in Figure 1.1, a constantly growing number of protein structures have become available each year since 1976. As of February 2013, almost 85000 protein structures have been deposited at the Protein Data Bank (PDB),<sup>6,7</sup> and this strongly demonstrates that structure-based drug design will continue to play a significant role in the field of drug design and discovery.

Briefly, in the SBDD method, starting from the structure of a protein or a nucleic acid (X-ray crystallography, NMR or homology modelling), compounds or fragments of compounds from a database are positioned into a selected region of the structure. These compounds are scored and ranked, and the best compounds are tested with biochemical assays. In a next phase, from the results obtained it is possible to reveal parts of the compound that can be optimized to increase potency. This processes can be re-iterated, and then the optimized compounds usually show marked improvement in binding and/or specificity for the target.

The recognition of the binding site or the active site residues in the target structure is of high importance in SBDD. Basically, it is a small region, a pocket or bumps, where ligand molecules can best fit or bind to activate the receptor and/or target and produce the desirable effect. Since the proteins are capable of undergoing conformational changes, recognizing the accurate binding site residues is difficult;<sup>8</sup> but still there are just a few computational programs, such as Ligsitecsc,<sup>9</sup> Qsite finder<sup>10</sup> and CASTp,<sup>11</sup> that can capably spot out the binding site residues. For example, Qsite finder locates and clusters the favorable binding sites using the interaction energy and Van der Waal's probes, whereas CASTp employs functionally annotated residues for mapping the surface pockets.

In this context, starting from a defined target structure, one of the most popular approaches used is the Virtual Screening (VS) from millions of potential compounds. VS computationally screens large chemical libraries to search for compounds that possess complementarities toward the targets.<sup>12,13</sup> The screening of

compounds in VS is carried out using docking calculations where the compounds are filtered mainly considering their binding energies against the target.<sup>13,14</sup> Because these types of screening techniques are mainly data driven, data accessibility remains highly significant.

Another approach often used is the “ligand-based drug design”, very useful in the absence of an experimental 3D structure.<sup>15,16</sup> Due to the lack of an experimental structure, the known ligand molecules that bind to the drug target are studied and compared to understand their structural and physico-chemical properties that correlate with the desired pharmacological activity.<sup>17</sup> Ligand-based methods may also include natural products or substrate analogues that interact with the target molecule yielding the desired pharmacological effect.<sup>18,19</sup> Here, if sufficient number of reported active compounds (10-40) with diverse activity values are there, one can build a 3D pharmacophore model of these set of compounds by overlapping all of them and finding the common feature among them.

SBDD and LBDD approaches can be used for the de-novo design, a process of creating or building new lead compounds from scratch - the former method being more prevalent than the latter. This process complements VS in hit discovery. The main principle of de novo design is to construct the small-molecule chemical structures that best fit the target space.<sup>20</sup> In receptor-based de novo design high-quality protein structures and their respective binding sites are essential because the hits are designed based on the target structures by placing small fragments in the key interaction sites of the proteins. Receptor-based design can be carried out by two means: linking and growing techniques.

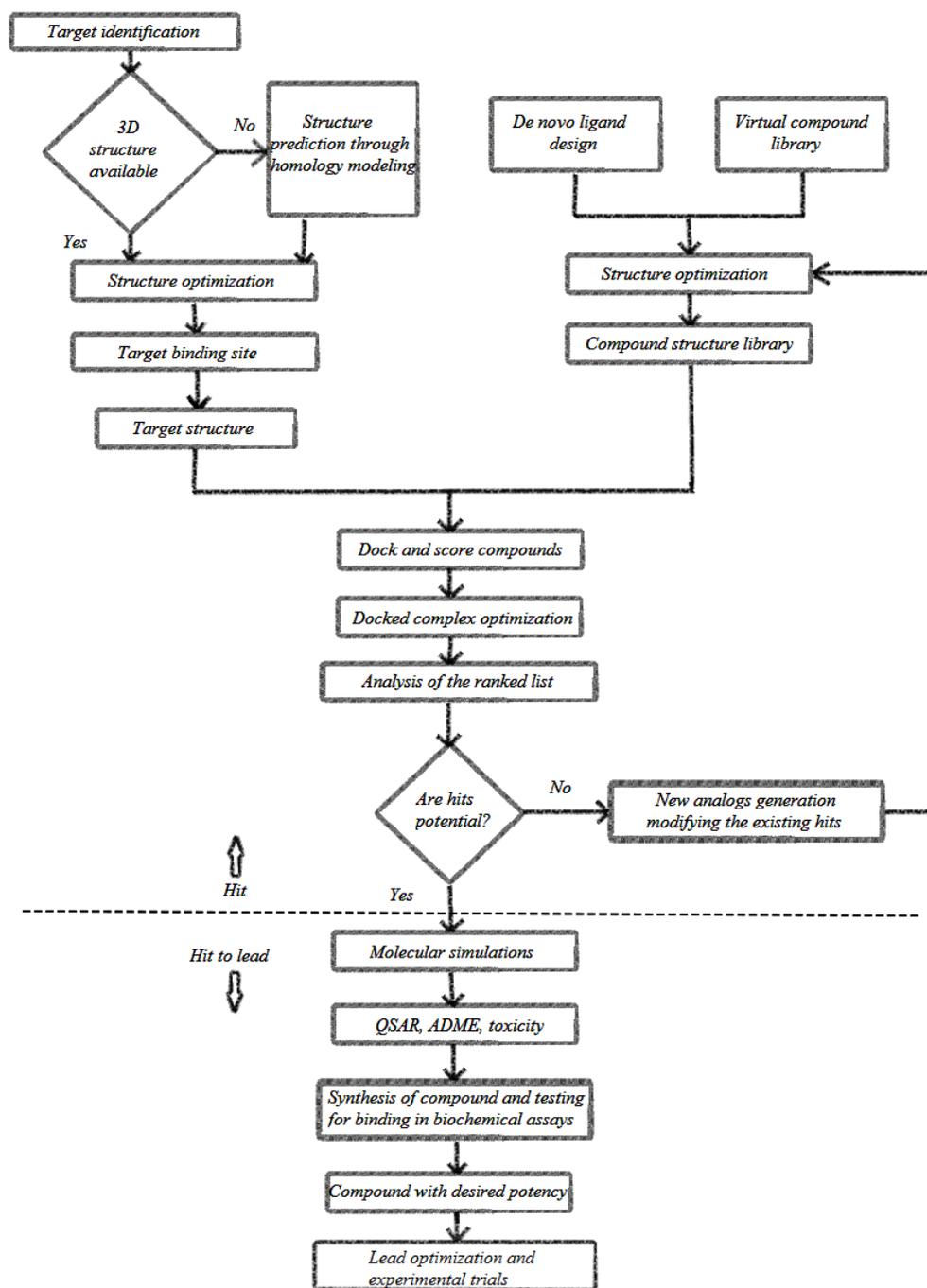


Figure 1.2 Process flow for the "hit identification" and "hit-to-lead" phases.

In the linking process different small fragments from the libraries are added simultaneously to different active site residues of the target.<sup>21</sup> Thus, the small fragments positioned at the binding site link to each other and form a final single compound. This approach is widely preferred because the fragment design strategy is insightful in that most biological targets encompass discrete binding sites for each piece of a ligand.

Whereas, in the growing technique a single small fragment is placed in the active site of the target and this fragment grows well complementarily against the receptor-binding site – thereby resulting in a library of chemical compounds that are more specific to the target. The process flow for the “hit identification” phase and “hit-to-lead” phase is represented in Figure 1.2.

## 1.2 Methodologies employed

A summary of the main methodologies used to realize the project is briefly presented in the next paragraphs.

### 1.2.1 Molecular Docking

One of the more current and fast computational techniques used in the field of the medicinal chemistry is represented by the molecular docking. With this computational tool, based on the protein structures, thousands of possible poses of association are tried and evaluated. Many different docking software have been developed, i.e. GOLD,<sup>22</sup> DOCK,<sup>23</sup> FlexX,<sup>24</sup> Autodock 3,<sup>25</sup> Autodock 4,<sup>26</sup> Autodock 4.2,<sup>27</sup> Autodock-Vina,<sup>28</sup> Glide.<sup>29</sup>

When only the structure of a target and its active or binding site is available, high-throughput docking is primarily used as a hit identification tool. The determination of the binding mode and affinity between the constituent molecules in molecular recognition is crucial to understanding the interaction mechanisms and to designing therapeutic interventions. However, similar calculations are often also used later on during lead optimization, when modifications to known active structures can quickly be tested in computer models before compound synthesis.

Thus, if a particular target structure is known, one can dock a library of different chemical compounds into the same binding site, obtain a scoring value for each pose, and in this way virtually screen for affinity towards the target. High throughput virtual screening (HTVS), where the library may consist of up to  $10^{12}$  (virtual) compounds, is a procedure commonly employed by pharmaceutical companies when starting a new lead discovery process. When desiring more detailed information about a potential ligand–protein complex than can be provided by HTVS, the assumption of the protein being a rigid structure becomes harder to

justify. It is well known that most, if not all, proteins continuously undergo conformational changes when exerting their functions *in vivo*.<sup>30</sup> Specifically, when an agonist binds to a receptor, it is clear that significant conformational changes must take place. In light of this, some attempts have been made to take protein flexibility into account during docking. However, scoring all possible conformational changes is prohibitively expensive in computer time. Docking procedures which permit conformational change, or *flexible docking* procedures, must intelligently select small subset of possible conformational changes for consideration.

In general, there are two aims of docking studies: accurate structural modelling and correct prediction of activity.

Basically, a protein-ligand docking program consists of two essential components, sampling and scoring.

### **1.2.1.1 Sampling**

Sampling refers to the generation of putative ligand binding orientations/conformations near a binding site of a protein and can be further divided into two aspects, ligand sampling and protein flexibility.

Ligand sampling is the most basic element in protein-ligand docking. Given a protein target, the sampling algorithm generates putative ligand orientations/conformations (*i.e.*, poses) around the chosen binding site of the protein.

Treatment of ligand flexibility can be divided into three basic categories:<sup>31</sup> systematic methods (incremental construction, conformational search, databases); random or stochastic methods (Monte Carlo, genetic algorithms, tabu search); and simulation methods (Molecular Dynamics, energy minimization).

Systematic search algorithms generate all possible ligand binding conformations by exploring all degrees of freedom of the ligand.

The most straightforward systematic algorithms are exhaustive search methods, in which flexible-ligand docking is performed by systematically rotating all possible rotatable bonds of the ligand at a given interval. Despite its sampling completeness for ligand conformations, the number of the combinations can be huge with the increase of the rotatable bonds.

In stochastic algorithms, ligand binding orientations and conformations are sampled by making random changes to either a single ligand or a population of ligands at each step in both the conformational space and the translational/rotational space of the ligand, respectively. A newly obtained ligand is evaluated on the basis of a pre-defined probability function. Two popular random approaches are Monte Carlo and genetic algorithms.

For what concerns simulation methods, Molecular Dynamics and energy minimization are the most popular simulation approaches. Molecular Dynamics simulations are often unable to cross high-energy barriers within feasible simulation time periods, and therefore might only accommodate ligands in local minima of the energy surface.

Therefore, an attempt is often made to simulate different parts of a protein–ligand system at different temperatures.<sup>32</sup>

Another strategy for addressing the local minima problem is to perform Molecular Dynamics calculations from different ligand positions. In contrast to Molecular Dynamics, energy minimization methods are rarely used as stand-alone search techniques, as only local energy minima can be reached, but often complement other search methods, including Monte Carlo.

For example, DOCK performs a minimization step after each fragment addition, followed by a final minimization before scoring.

Protein flexibility starts from the assumption that ligand binding commonly induces protein conformational changes, which range from local rearrangements of side-chains to large domain motions. Methods to account for protein flexibility can be grouped into three categories: soft docking, side-chain flexibility, and protein ensemble docking.

Soft docking is the simplest method which considers protein flexibility implicitly. It works by allowing for a small degree of overlap between the ligand and the protein through softening the interatomic van der Waals interactions in docking calculations.

In side-chain flexibility method backbones are kept fixed and side-chain conformations are sampled.

The third type of methods account for protein flexibility by firstly using rigid-body docking to place the ligand into the binding site and then relaxing the protein backbone and side-chain atoms nearby.

Specifically, the initial rigid-body docking allows for atomic clashes between the protein and the placed ligand orientations/conformations in order to consider the protein conformational changes. Then, the formed complexes are relaxed or minimized by Monte Carlo (MC), Molecular Dynamic simulations, or other methods.

In general, the most widely-used type of methods for incorporating protein flexibility utilizes an ensemble of protein structures to represent different possible conformational changes.

The ensemble docking algorithm is not used for generating new protein structures, but instead for selecting the induced-fit structure from a given protein ensemble. Following a similar procedure, Abagyan and colleagues expanded Huang and Zou's algorithm to create ICM's ensemble docking algorithm, referred to as four-dimensional (4D) docking.<sup>33</sup>

### 1.2.1.2 Scoring

Scoring is the prediction of the binding tightness for individual ligand orientations/conformations with a physical or empirical energy function. The top orientation/conformation, namely the one with the lowest energy score, is typically predicted as the binding mode. The scoring function is a key element of a protein-ligand docking algorithm, because it directly determines the accuracy of the algorithm.<sup>34,35,36,37</sup> Speed and accuracy are the two important aspects of a scoring function and an ideal scoring function would be both computationally efficient and reliable. Scoring functions have been developed can be grouped into three basic categories:

- force field
- empirical
- knowledge-based scoring functions.

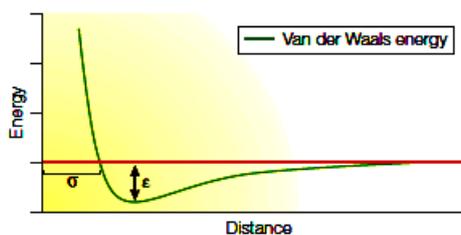
Force field (FF) scoring functions<sup>25,38,39</sup> are based on decomposition of the ligand binding energy into individual interaction terms such as van der Waals (VDW) energies, electrostatic energies, bond stretching/bending/torsional energies, etc., using a set of derived force-field parameters such as AMBER<sup>40</sup> or CHARMM<sup>41,42</sup> force fields. In general, the enthalpic contributions are essentially given by the electrostatic and van der Waals terms, and also considering taking into account the hydrogen bond formation between drug and biological target.

The van der Waals potential energy is often modeled by the Lennard-Jones 12-6 function (Equation 1.1)

$$E_{vdW}(r) = \sum_{j=1}^N \sum_{i=1}^N 4\epsilon \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]$$

**Equation 1.1**

where  $\epsilon$  is the well depth of the potential and  $\sigma$  is the collision diameter of the respective atoms  $i$  and  $j$ . The  $\exp(12)$  is responsible for small-distance repulsion, whereas the  $\exp(6)$  is related to an attractive term which approaches zero as the distance between two atoms increases (Figure 1.3).



**Figure 1.3** Graphical representation of the Lennard-Jones 12-6 function

This Lennard-Jones 12-6 function is also used to describe the hydrogen bond in macromolecule-ligand complex, but is less smooth and angle dependent if compared to the van Der Waals function.

The electrostatic potential energy is represented as the summation of Coulombic interactions, as described in Equation 1.2:

$$E_{coul}(r) = \sum_{i=1}^{N_A} \sum_{j=1}^{N_B} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}$$

**Equation 1.2**

Where  $N$  is the number of atoms in molecules A and B, respectively, and  $q$  is the charge on each atom. The functional form of the internal ligand energy is typically very similar to the ligand-protein interaction energy, and also includes van der Waals contributions and/or electrostatic terms.

One of the major challenges in FF scoring functions is how to account for the solvent effect. The simplest method is to use a distance-dependent dielectric constant ( $r_{ij}$ ) such as the force field scoring function in DOCK<sup>39</sup> (Equation 1.3)

$$E = \sum_i \sum_j \left( \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} + \frac{q_i q_j}{\varepsilon(r_{ij}) r_{ij}} \right)$$

**Equation 1.3**

where  $r_{ij}$  stands for the distance between protein atom  $i$  and ligand atom  $j$ ,  $A_{ij}$  and  $B_{ij}$  are the VDW parameters, and  $q_i$  and  $q_j$  are the atomic charges.  $\varepsilon(r_{ij})$  is usually set to  $4r_{ij}$ , reflecting the screening effect of water on electrostatic interactions. The most rigorous FF methods are to treat water molecules explicitly. However, these methods, together with their simplified approaches such as LIEPROFEC, and OWFEG are computationally expensive.<sup>43</sup> To reduce the computational expense, accelerated methods have been developed while preserving the reasonable accuracy by treating water as a continuum dielectric medium. The Poisson-Boltzmann/surface area (PB/SA) models<sup>44,45,46</sup> and the generalized-Born/surface area (GB/SA) models<sup>47,48,49</sup> are typical examples of such implicit solvent models.

In addition to the challenge on solvent effect, how to accurately account for entropic effect is an even more severe challenge for FF scoring functions. Moreover, whether the free energy of ligand binding can be decomposed into a linear combination of individual interaction terms without calculating the partition function (“ensemble average”) also remains in question, referred to as the nonadditive problem.

The second type of scoring function (empirical scoring functions) works on the sum of a set of weighted empirical energy terms such as VDW energy, electrostatic

energy, hydrogen bonding energy, desolvation term, entropy term, hydrophobicity term, *etc.*(Equation 1.4):

$$\Delta G = \sum_i W_i \cdot \Delta G_i$$

**Equation 1.4**

where  $\{\Delta G_i\}$  represent individual empirical energy terms, and the corresponding coefficients  $\{W_i\}$  are determined by reproducing the binding affinity data of a training set of protein-ligand complexes with known three-dimensional structures, using least squares fitting. GlideScore<sup>50,51</sup>, PLP<sup>52</sup>, SYBYL/F-Score,<sup>24</sup> LigScore,<sup>53</sup> LUDI,<sup>54</sup> SCORE,<sup>55</sup> X-Score,<sup>56</sup> ChemScore,<sup>57</sup> MedusaScore,<sup>58</sup> AIScore,<sup>59</sup> and SFCscore<sup>60</sup> are examples of empirical scoring functions.

In knowledge-based scoring functions protein-ligand complexes are modeled using relatively simple atomic interaction-pair potentials. In essence, it is designed to reproduce experimental structures rather than binding energies. A number of atom-type interactions are defined depending on their molecular environment. Compared to the force field and empirical scoring functions, the knowledge-based scoring functions offer a good balance between accuracy and speed. Namely, because the potentials are extracted from a large number of structures rather than attempting to reproduce the known affinities by fitting, the knowledge-based scoring functions are relatively robust and general. Their pairwise characteristic also enables the scoring process to be as fast as empirical scoring functions.

A technique to improve the performances of scoring functions is clustering-based scoring methods, which incorporate the entropic effects by dividing generated ligand binding modes into different clusters.<sup>61,62,63</sup> The entropic contribution in each cluster is measured by the configurational space covered by the ligand poses or the number of the ligand poses in the cluster.

One restriction in clustering-based scoring methods is that its performance depends on the ligand sampling protocol that is used, *i.e.*, it is docking program-dependent. These methods in combination with ligand conformational sampling using AutoDock have significantly improved binding mode prediction.

### 1.2.1.3 Autodock: An Overview

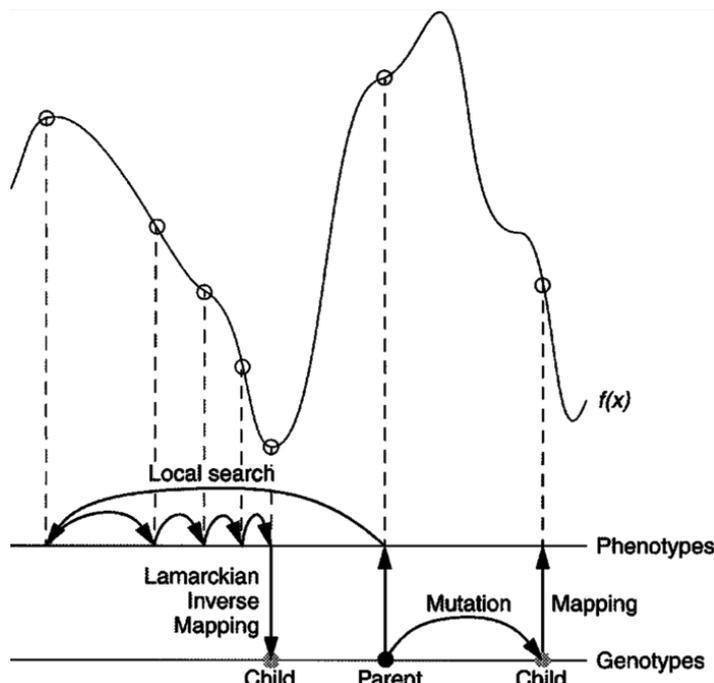
AutoDock currently represents one of the most cited docking softwares,<sup>64</sup> especially in a virtual screening of a compound libraries.<sup>65</sup> For the purposes of this project the software AutoDock 3.0.5,<sup>25</sup> 4,<sup>26</sup> 4.2,<sup>27</sup> and Autodock-Vina<sup>28</sup> have been used.

Basically, the differences between them are related to the speed, macromolecule sidechains flexibility, optimization of the free-energy scoring function based on a linear regression analysis, AMBER force field, larger set of diverse protein-ligand complexes with known inhibition constants; moreover the Lamarckian Genetic Algorithm (LGA) is a big improvement on the Genetic Algorithm, and both genetic methods are much more efficient and robust than SA in the new version of the software.

In AutoDock there are different available search methods, but the Lamarckian Genetic Algorithm (LGA) has been selected for the aim of this study, because it has demonstrated to give the best results compared to the other algorithms.<sup>25</sup>

The vast majority of genetic algorithms mimics the characteristics of Darwinian evolution and applies Mendelian genetics. This is briefly illustrated in Figure 1.4.

This is called the Lamarckian genetic algorithm (LGA), and is an allusion to Jean Batiste de Lamarck's (discredited) assertion that phenotypic characteristics acquired during an individual's lifetime can become heritable traits.<sup>66</sup>



**Figure 1.4** This figure illustrates genotypic and phenotypic search, and contrasts Darwinian and Lamarckian search.<sup>67</sup> The space of the genotypes is represented by the lower horizontal line, and the space of the phenotypes is represented by the upper horizontal line. Genotypes are mapped to phenotypes by a developmental mapping function. The fitness function is  $f(x)$ . The result of applying the genotypic mutation operator to the parent's genotype is shown on the right-hand side of the diagram, and has the corresponding phenotype shown. Local search is shown on the left-hand side. It is normally performed in phenotypic space and employs information about the fitness landscape. Sufficient iterations of the local search arrive at a local minimum, and an inverse mapping function is used to convert from its phenotype to its corresponding genotype. In the case of molecular docking, however, local search is performed by continuously converting from the genotype to the phenotype, so inverse mapping is not required. The genotype of the parent is replaced by the resulting genotype, however, in accordance with Lamarckian principles.

The most important issues arising in hybrids (LGA) of Genetic Algorithm (GA) and the Local Search (LS) revolve around the *developmental mapping*, which transforms genotypic representations into phenotypic ones.

The genotypic space is defined in terms of the genetic operators mutation and crossover in our experiments by which parents of one generation are perturbed to form their children. The phenotypic space is defined directly by the problem,

namely, the energy function being optimized. The local search operator is a useful extension of GA global optimization when there are local “smooth-ness” characteristics (continuity, correlation, etc.) of the fitness function that local search can exploit. In hybrid GA + LS optimizations, the result of the LS is always used to update the fitness associated with an individual in the GA selection algorithm. If, and only if, the developmental mapping function is *invertible*, will the Lamarckian option converting the phenotypic result of LS back into its corresponding genotype become possible. The fitness or energy is usually calculated from the ligand’s coordinates, which together form its phenotype. The developmental mapping simply transforms a molecule’s genotypic state variables into the corresponding set of atomic coordinates. A novel feature of this application of hybrid global-local optimization is that the Solis and Wets LS operator searches through the genotypic space rather than the more typical phenotypic space. This means that the developmental mapping does not need to be inverted. Nonetheless, this molecular variation of the genetic algorithm still qualifies as Lamarckian, because any “environmental adaptations” of the ligand acquired during the local search will be inherited by its offspring. At each generation, it is possible to let a user defined fraction of the population undergo such a local search. The local search frequencies of just 0.06 have found improved efficiency of docking, although a frequency of 1.00 is not significantly more efficient.<sup>67</sup> Both the canonical and a slightly modified version of the Solis and Wets method have been implemented. In canonical Solis and Wets, the same step size would be used for every gene, but we have improved the local search efficiency by allowing the step size to be different for each type of gene: a change of 1 Å in a translation gene could be much more significant than a change of 1° in a rotational or torsional gene. In the Lamarckian genetic algorithm, genotypic mutation plays a somewhat different role than it does in traditional genetic algorithms. Traditionally, mutation plays the role of a local search operator,

allowing small, refining moves that are not efficiently made by crossover and selection alone. With the explicit local search operator, however, this role becomes unnecessary, and is needed only for its role in replacing alleles that might have disappeared through selection. In LGA, mutation can take on a more exploratory role.

The LGA yields a maximum number of 256 potential bioactive conformations: run, whose number can be increased performing more docking calculations. Each conformational solution is the result of a selection. The GA, starting from the input geometry, gives rise to a group of  $n$  conformations or individuals (whose number can be set up) defining for them translational, rotational and torsional variables. By the scoring function, each individual is labeled by the total interaction energy (fitness).

Random pairs of individuals are mated using a process of crossover, in which new individuals inherit geometrical features from their parents leading to new generation of individuals. In addition, some offspring undergo random mutation, in which the translational, rotational and torsional variables are mutated randomly. Selection of the offspring of the current generation occurs based on the individual's fitness: thus the better solutions go on into the next generations, whereas conformations with a low fitness are discarded. This cycle of crossover, mutation to lead new generation is repeated until the better bioactive conformation (run) is given.

The LS performs an energy minimization of the current found conformation. In each generation a fraction of conformations population undergoes the geometry optimization, based on the local search frequency. Rapid energy evaluation is achieved by precalculating atomic affinity potentials (grid maps) for each atom type in the substrate molecule by grid method.<sup>68</sup>

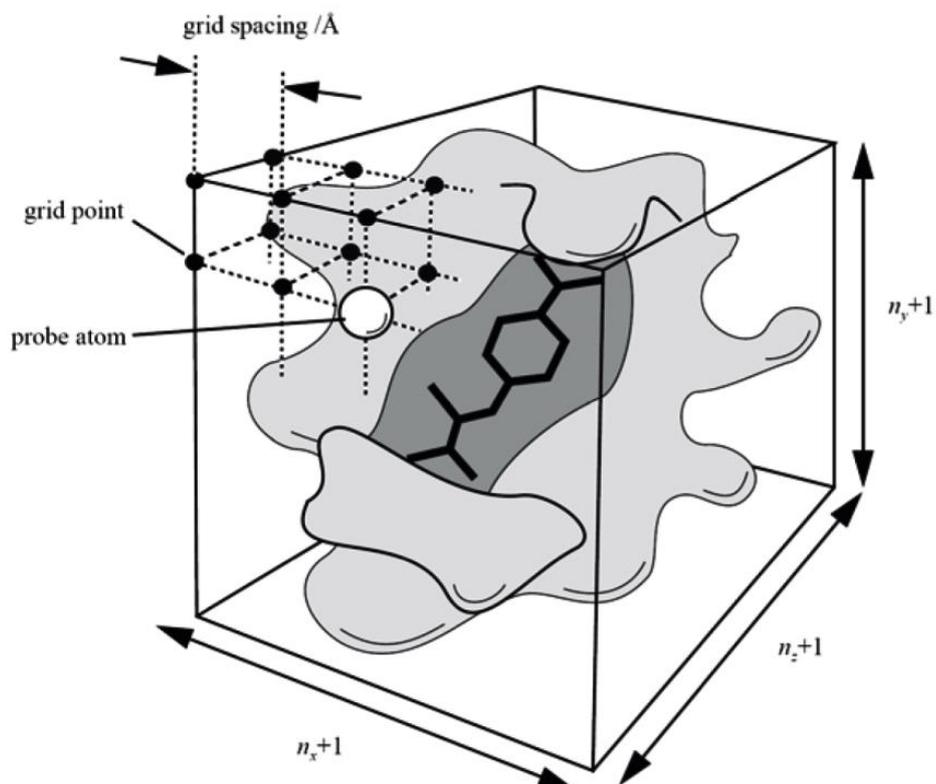
These maps are calculated by AutoGrid. In this procedure the protein is embedded in a three dimensional grid and a probe atom is placed at each grid point (Figure 1.5). The energy of interaction of this single atom with the protein is assigned to the grid point.

An affinity grid is calculated for each type of atom in the substrate, typically carbon, oxygen, nitrogen and hydrogen, as well as a grid of electrostatic potential, either using a point charge of +1 as the probe, or using a Poisson-Boltzmann finite difference method, such as DELPHI.<sup>69</sup> The energetic of a particular substrate configuration is then found by tri-linear interpolation of affinity values of the eight grid points surrounding each of the atoms in the substrate.

The electrostatic interaction is evaluated similarly, by interpolating the values of the electrostatic potential and multiplying by the charge on the atom (the electrostatic term is evaluated separately to allow finer control of the substrate atomic charges).

The time to perform an energy calculation using the grids is proportional only to the number of atoms in the substrate, and is independent of the number of atoms in the protein. An estimated free energy of binding is used to evaluate the docked ligand conformations. This scoring function, based of force field AMBER,<sup>70</sup> comprises terms above described (directional hydrogen bonding, electrostatics, Van der Waals, internal energy) and entropic contribution: desolvation and torsional entropy. The latter describes the loss of entropy upon interaction with macromolecule followed by immobilization in the active site.

The desolvation belongs the displacement of water molecules from the active site upon the binding of ligand to the macromolecular surface and the reorganization of solvent around the complex.



**Figure 1.5** Schematic representation of the grid map.

The scoring function was implemented using the thermodynamic cycle of Wesson and Eisenberg. The function is:

$$\Delta G = \Delta G_{vdw} \sum_{i,j} \left( \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) + \Delta G_{hbond} \sum_{i,j} E(t) \left( \frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right) +$$

$$+ \Delta G_{elec} \sum_{i,j} \frac{q_i q_j}{\epsilon(r_{ij}) r_{ij}} + \Delta G_{tor} N_{tor} + \Delta G_{solv} \sum_{i,j} (S_i V_j + S_i V_i) e^{\left( \frac{-r_{ij}^2}{-2\sigma^2} \right)}$$

**Equation 1.5**

where the five  $\Delta G$  terms on the right hand side are coefficient empirically determined using a linear regression analysis from a set of protein-ligand complexes.<sup>25</sup>

For what concerns AutoDock Vina,<sup>28</sup> this is a open-source program for drug discovery, molecular docking and virtual screening, offering multi-core capability, high performance and enhanced accuracy and ease of use. Vina uses a sophisticated gradient optimization method in its local optimization procedure.

The calculation of the gradient effectively gives the optimization algorithm a “sense of direction” from a single evaluation. In the spectrum of computational approaches to modelling receptor ligand binding molecular dynamics with explicit solvent, Molecular Dynamics and Molecular Mechanics with implicit solvent, molecular docking can be seen as making an increasing trade-off of the representational detail for computational speed.<sup>71</sup> Among the assumptions made by these approaches is the commitment to a particular protonation state of and charge distribution in the molecules that do not change between, for example, their bound and unbound states.

Additionally, docking generally assumes much or all of the receptor rigid, the covalent lengths, and angles constant, while considering a chosen set of covalent bonds freely rotatable (referred to as active rotatable bonds here). Importantly, although Molecular Dynamics directly deals with energies (referred to as force fields in chemistry), docking is ultimately interested in reproducing chemical potentials, which determine the bound conformation preference and the free energy of binding. It is a qualitatively different concept governed not only by the minima in the energy profile but also by the shape of the profile and the temperature.<sup>72</sup> Docking programs generally use a scoring function, which can be seen as an attempt to approximate the standard chemical potentials of the system. When the superficially physics-based terms like the 6–12 van der Waals interactions and

Coulomb energies are used in the scoring function, they need to be significantly empirically weighted, in part, to account for this difference between energies and free energies.<sup>72</sup>

The afore mentioned considerations should make it rather unsurprising when such superficially physics-based scoring functions do not necessarily perform better than the alternatives.

This approach was seen to the scoring function as more of “machine learning” than directly physics-based in its nature. It is ultimately justified by its performance on test problems rather than by theoretical considerations following some, possibly too strong, approximating assumptions

The general functional form of the conformation-dependent part of the scoring function AutoDock Vina is designed to work with is:

$$c = \sum_{i < j} f_{t_i t_j}(r_{ij})$$

**Equation 1.6**

where the summation is over all of the pairs of atoms that can move relative to each other, normally excluding 1–4 interactions, i.e., atoms separated by three consecutive covalent bonds.

Here, each atom  $i$  is assigned a type  $t_i$ , and a symmetric set of interaction functions  $f_{t_i t_j}$  of the interatomic distance  $r_{ij}$  should be defined.

This value can be seen as a sum of intermolecular and intramolecular contributions:

$$c = c_{inter} + c_{intra}$$

**Equation 1.7**

The optimization algorithm attempts to find the global minimum of  $c$  and other low-scoring conformations, which it then ranks.

The predicted free energy of binding is calculated from the intermolecular part of the lowest-scoring conformation, designated as 1:

$$s_1 = g(c_1 - c_{intra1}) = g(c_{inter1})$$

**Equation 1.8**

where the function  $g$  can be an arbitrary strictly increasing smooth possibly nonlinear function.

In the output, other low-scoring conformations are also formally given  $s$  values, but, to preserve the ranking, using  $c_{intra}$  of the best binding mode:

$$s_i = g(c_i - c_{intra1})$$

**Equation 1.9**

For modularity reasons, much of the program does not rely on any particular functional form of  $f_{i-j}$  interactions or  $g$ . Essentially, these functions are passed as a parameter for the rest of the code.

In summary the evaluation of the speed and accuracy of Vina during flexible redocking of the 190 receptor-ligand complexes making up the AutoDock 4 training set showed approximately two orders of magnitude improvement in speed and a simultaneous significantly better accuracy of the binding mode prediction. In addition, Vina can achieve near-ideal speed-up by utilizing multiple CPU cores. However, AutodockVina does not provide very good weight of the energetic contribution derived from the hydrogen bond and electrostatic interactions, especially when the metal ions are presents.

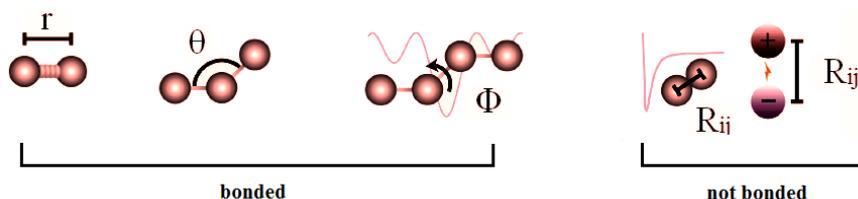
## 1.2.2 Molecular Dynamics

One of the principal tools in the theoretical study of biological molecules is represented by Molecular Dynamics simulations (MD). This computational method calculates the time dependent behavior of a molecular system. MD simulations have provided detailed information on the fluctuations and conformational changes of proteins and nucleic acids. These methods are now routinely used to investigate the structure, dynamics and thermodynamics of biological molecules and their complexes, representing an important tool in the drug discovery process.<sup>73</sup>

Crystallographic studies demonstrate that protein flexibility plays a fundamental role in ligand binding, but they represents long and very expensive methods. As a consequence, computational techniques that can predict protein motions are needed. Unfortunately, the calculations required to describe the absurd quantum-mechanical motions and chemical reactions of large molecular systems are often too complex and computationally intensive for even the best supercomputers. Molecular Dynamics (MD) simulations, first developed in the late 1970s,<sup>74</sup> seek to overcome this limitation by using simple approximations based on Newtonian physics to simulate atomic motions, thus reducing the computational complexity. The forces acting on each of the system atoms are then estimated from an equation like that shown in Equation 1.10<sup>75</sup> and represented in Figure 1.6:

$$E_{total} = \sum_{bonds} K_r (r - r_{eq})^2 + \sum_{angles} K_\theta (\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{V_n}{2} [1 + \cos(n\Phi - \gamma)] \\ + \sum_{i < j} \left[ \frac{A_{ij}}{R_{ij}^{12}} + \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}} \right]$$

**Equation 1.10**



**Figure 1.6** Atomic forces that govern molecular movement can be divided into those caused by interactions between atoms that are chemically bonded to one another and those caused by interactions between atoms that are not bonded

Briefly, these forces arise from interactions between bonded and non-bonded atoms contribute. Chemical bonds and atomic angles are modeled using simple virtual springs, and dihedral angles are modeled using a sinusoidal function that approximates the energy differences between eclipsed and staggered conformations. Non-bonded forces arise due to van der Waals interactions, modeled using the Lennard-Jones 6-12 potential, and charged (electrostatic) interactions, modeled using Coulomb's law.

The energy terms described above are parameterized in order to fit quantum-mechanical calculations and experimental data. The main aim of this parameterization is the identification of the ideal stiffness and lengths of the springs that describe chemical bonding and atomic angles, determining the best partial atomic charges used for calculating electrostatic-interaction energies, identifying the proper van der Waals atomic radii, and so on. Collectively, these parameters are called a 'force field' because they describe the contributions of the various atomic forces that govern Molecular Dynamics. Several force fields are commonly used in molecular dynamics simulations, including AMBER,<sup>75,76</sup> CHARMM,<sup>77</sup> and GROMOS.<sup>78</sup> These differ principally in the way they are parameterized but generally give similar results.

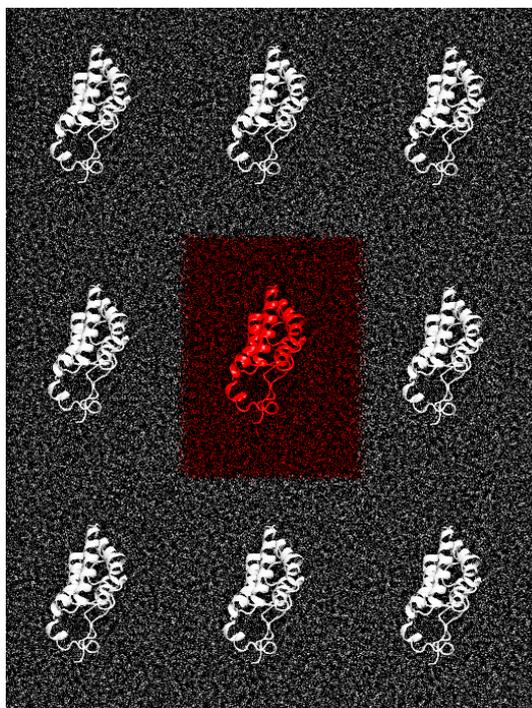
Different parameters determine the better or worse "quality" of a Molecular Dynamics simulation. For example, it is firstly important to include solvent effects.

This can be done at several levels. The simplest treatment is to simply include a dielectric screening constant in the electrostatic term of the potential energy function. In this implicit treatment of the solvent, water molecules are not included in the simulation but an effective dielectric constant is used. Often the effective dielectric constant is taken to be distance dependent. Although this is a crude approximation, it is still much better than using unscreened partial charges. Other implicit solvent models have been developed that range from the relatively simple distance-dependent dielectric constants to models that base the screening on the solvent exposed surface area of the protein. The distance-dependent dielectric coefficient is the simplest way to include solvent screening without including explicit water molecules and it is available in most simulation programs. Recently, several implicit solvent models based on continuum electrostatic theory have been developed.

If water molecules are explicitly included in the simulation, they can provide the electrostatic shielding. In this more detailed treatment of the solvent boundary conditions must be imposed, first, to prevent the water molecules from diffusing away from the protein during the simulation, and second to allow simulation and calculation of macroscopic properties using a limited number of solvent molecules. Several different treatments of the boundary exist, the use of one over another depends strongly on the type of problem the simulation is to address.

Periodic boundary conditions enable a simulation to be performed using a relatively small number of particles in such a way that the particles experience forces as though they were in a bulk solution. The coordinates of the image particles, those found in the surrounding box are related to those in the primary box by simple translations along the three axes (Figure 1.7). The simplest box is the cubic box. Forces on the primary particles are calculated from particles within the

same box as well as in the image box. The cutoff is chosen such that a particle in the primary box does not see its image in the surrounding boxes.



**Figure 1.7** Periodic boundary conditions (PBC). Primary box (red) translated along the three axes (white)

Once the forces acting on each of the system atoms have been calculated, the positions of these atoms are moved according to Newton's laws of motion. The simulation time is then advanced, often by only 1 or 2 quadrillionths of a second, and the process is repeated, typically millions of times. Because so many calculations are required, Molecular Dynamics simulations are typically performed on computer clusters or supercomputers using dozens if not hundreds of processors in parallel (CPUs) or, more recently, moving to the GPU architecture. Many of the most popular simulation software packages, which often bear the same names as

their default force fields (for example AMBER,<sup>79</sup> CHARMM,<sup>77</sup> and NAMD<sup>80,81</sup>), are compatible with multiple processors operating simultaneously.

In the field of the drug discovery, Molecular Dynamics and the insights they offer into protein motion often play important roles. In fact, a single protein conformation tells little about protein dynamics. The static models produced by NMR, X-ray crystallography, and homology modelling provide valuable insights into macromolecular structure, but molecular recognition and drug binding are very dynamic processes. Moreover, it represents the best method for the identification of the sites not immediately obvious from available structures (cryptic sites), or for the allosteric ones.

A link between molecular docking (fast, but with a poor accuracy) and Molecular Dynamics (computationally expensive, but accurate) is a new virtual-screening protocol called the relaxed complex scheme (RCS),<sup>82,83</sup> in which each potential ligand is docked into multiple protein conformations, typically extracted from a Molecular Dynamics simulation. Thus, each ligand is associated not with a single docking score but rather with a whole spectrum of scores. Ligands can be ranked by a number of spectrum characteristics, such as the average score over all receptors. Thus, the RCS effectively accounts for the many receptor conformations sampled by the simulations; it has been used successfully to identify a number of protein inhibitors, including inhibitors of FKBP,<sup>84</sup> HIV integrase.<sup>85</sup>

Moreover, another important application of Molecular Dynamics simulations in the drug discovery is the accurate calculation of the free energy (or binding affinities), widely described in the paragraph 1.2.3.

With constant improvements in both computer power (from CPU to GPU architecture) and algorithm design, the future of computer-aided drug design is promising and Molecular Dynamics simulations are likely to play an increasingly important role in the development of novel pharmacological therapeutics.

### **1.2.3 Methods for the accurate calculation of the binding affinities**

Docking calculations are widely used in high-throughput virtual screening of structurally diverse molecules from available compound libraries/databases against specific targets, but often show many limits, especially in the lead identification stage.

One of the ultimate goals in computer-aided drug design is the accurate prediction of ligand-binding affinities to a macromolecular target. As free energy methods have improved and computational power has continued to grow exponentially, this promise has begun in small part to be fulfilled. Low-throughput computational approaches for the calculation of ligand binding free energies can be divided into “pathway” and “endpoint” methods.<sup>86</sup>

#### **1.2.3.1 “Pathway” methods**

In pathway methods, the system is converted from one state (e.g., the complex) to the other (e.g., the unbound protein/ligand). This can be achieved by introducing a set of finite or infinitesimal “alchemical” changes to the energy function (the Hamiltonian) of the system through free-energy perturbation (FEP) or thermodynamic integration (TI), respectively. In an alchemical transformation, a chemical species is transformed into another via a pathway of nonphysical (alchemical) states. Many physical processes, such as ligand binding or transfer of a molecule from gas to solvent, can be equivalently expressed as a composition of such alchemical transformations. Combined with atomistic Molecular Dynamics (MD) or Monte Carlo (MC) simulations in explicit water solvent models, they represent the most accurate approaches for calculating absolute or relative ligand binding affinities. These methods are applicable with the notable increases of the

computational power, due to the emerging implementation of biomolecular codes on GPU architectures.

### 1.2.3.1.1 Free energy perturbation

Free energy perturbation (FEP) starts from the assumption that internal energy, enthalpy, entropy, and Gibb's free energy all include contributions from the motion of a molecule. Therefore, Molecular Dynamics provides a way to estimate these important thermodynamic parameters. The hypothesis that makes the most sense is that the internal energy,  $\Delta U$ , is the time average of the total energy of the molecule. The total energy of the molecule is the kinetic plus potential energy (Equation 1.11):

$$E = \text{kinetic energy} + \text{potential energy}$$

#### Equation 1.11

The potential energy is just the molecular mechanics steric energy. Molecular Dynamics provides us with the time dependent energy of the molecule; all we need do to get  $\Delta U$  is average the total energy during the trajectory calculation.

Now we turn to the relationship of the steric energy to the Gibb's free energy. In statistical mechanics, we find that the probability of a given state of a system occurring is proportional to the Boltzman weighting factor (Equation 1.12):

$$\text{probability of occurrence} \propto e^{-E/RT}$$

#### Equation 1.12

where  $E$  is the total energy of the system. In other words, states with low total energy are more likely to occur than states with high energy. A state of the system is determined by the conformation and motion of the molecule.

The conformation determines the steric energy and the motions determine the kinetic energy. In perturbation theory, we look at the effect of a small change in the structure of a molecule on its energy. To do the perturbation, the total energy is divided into two parts (Equation 1.13):

$$E = E_0 + E_1$$

**Equation 1.13**

where  $E_0$  is a reference structure and  $E_1$  is a small perturbation from the reference structure.

The perturbation is a small change that we place upon the system, say a small change in bond angle or a small change in the charge on an atom. The corresponding change in free energy of the system caused by the perturbation is given as shown in Equation 1.14:<sup>87</sup>

$$G - G_0 = -RT \ln \langle e^{-E_1/RT} \rangle_0$$

**Equation 1.14**

where  $\langle \rangle_0$  denotes the time average over the motion of the reference structure from a Molecular Dynamics run. The  $e^{-E_1/RT}$  term is the probability of occurrence for the small change in energy caused by the perturbation, from Equation 1.14. The free energy then depends on the time average of the probability of occurrence of the perturbed structure. In other words, if the perturbation produces a small change in energy, that change will contribute to the Gibb's free energy. In our case however, we wish to find the change in free energy for large changes in a molecule. These changes, can contribute to the mutation of a molecule from one state (molecule B) to another (molecule A). First we define a total energy for mutating molecule B to A (Equation 1.15) as

$$E_{\lambda} = \lambda E_A + (1 - \lambda)E_B$$

**Equation 1.15**

where  $E_A$  is the total energy for A and  $E_B$  is that for B, and  $\lambda$  is the coupling parameter. When  $\lambda = 1$  the energy corresponds to molecule A, and when  $\lambda = 0$  the energy corresponds to molecule B.

When  $\lambda$  is at intermediate values, the system is a hypothetical superposition of A and B. It might seem quite strange to have such a combination of two molecules, in fact it is very unphysical; however.

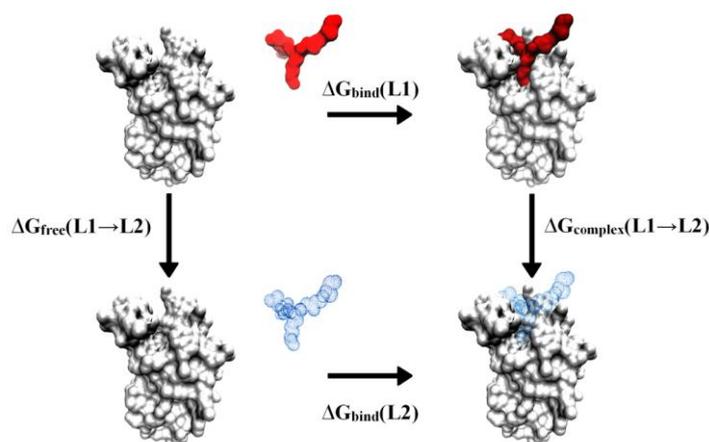
For the complete mutation to take place we vary  $\lambda$  from 0 to 1 over the course of the dynamics run. We divide this full range into short time slices, which are short enough that we can treat the change in each time slice as a perturbation.

Then we apply Equation 1.15 to each time slice and then add up the result for all the time slices. Let the  $\lambda$  value at each time slice be numbered  $\lambda_1, \lambda_2, \lambda_3$ , etc. Then the difference in Equation 1.15 is  $\Delta G(\lambda_i)$  for each time slice,  $i=1, 2, 3, \dots, n$ , for  $n$  total time slices. In words, this simple result means that the change in Gibb's free energy for a perturbation is just the time average of the total perturbation energy.

**1.2.3.1.2 Thermodynamic integration**

Similarly, the thermodynamic integration method (TI) is often used for the “alchemical” computation of *differences* in binding affinities (known as *relative* affinities) among a set of related ligands for the same target protein. In this case, the free energy difference is calculated by defining a thermodynamic path between the states and integrating over enthalpy changes along the path.

These approaches can be used in a thermodynamic cycle, as illustrated in Figure 1.8, often applied in studying the relative strength of ligand-receptor interactions and the relative stability of proteins differing in one or a few amino acids.



**Figure 1.8** Thermodynamic cycle linking the binding of two ligands L1 and L2 to a protein in solution.

Thermodynamic cycle methods were developed because relatively large, complicated changes need to be taken into account when considering the physical phenomena that occur in ligand-receptor binding or the effect of a mutation on protein stability. That is, binding of a drug to a receptor will produce relatively large conformational changes (*i.e.*, the protein will favor a particular set of conformational substates). Binding of a very similar drug to the same site should produce most of the same changes. The thermodynamic cycle is designed to cancel out the large changes that are common to binding of either drug to the receptor.

The horizontal legs describe the experimentally accessible actual binding processes, with free energies  $\Delta G_{\text{bind}}(\text{L1})$  and  $\Delta G_{\text{bind}}(\text{L2})$ . Since the free energy is a state function, the relative binding free energy  $\Delta\Delta G_{\text{bind}}$  is exactly equal to the difference of the free energies in the horizontal or vertical legs:

$$\begin{aligned}\Delta\Delta G_{bind} &= \Delta G_{bind}(L2) - \Delta G_{bind}(L1) \quad \text{Equation 1.16} \\ &= \Delta G_{complex}(L1 \rightarrow L2) - \Delta G_{free}(L1 \rightarrow L2)\end{aligned}$$

#### Equation 1.17

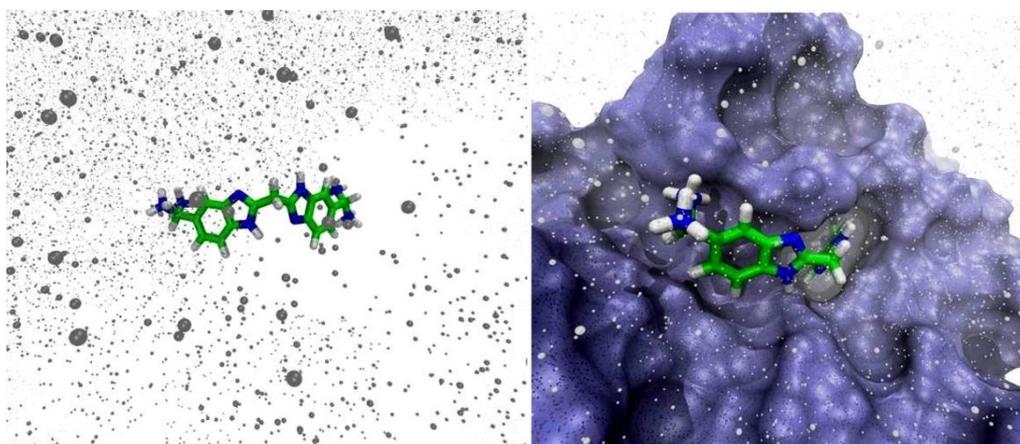
The simulations follow the vertical steps (Equation 1.17) or unphysical processes, by simulations in water solution that gradually change the energy-function of the system from one “endpoint” to the other through a series of intermediate hybrid states. From Figure 1.8, this involves the stepwise “alchemical” transformation of ligand L1 to L2 both in its ‘free’ state (unbound) and in the bound complex, through gradual changes in the forcefield parameters describing the ligand interactions. This leads to the free energy changes  $\Delta G_{free}(L1 \rightarrow L2)$  and  $\Delta G_{complex}(L1 \rightarrow L2)$ , respectively. Averaging over both transformation directions is often used to improve the free-energy estimates, although this is not always the case. These calculations can be accurate, if conducted with the appropriate care.

### 1.2.3.2 “Endpoint” methods

Although we are assisting on a constant improvement in the computational power, much less computationally demanding “endpoint” methods are often successfully applied, such as the molecular mechanics – Poisson Boltzmann (MM-PBSA) and the related molecular mechanics – generalised Born (MM-GBSA) approximation and the linear interaction energy (LIE). All these methods compute binding free energies along the horizontal legs of Figure 1.8, but use only models for the endpoints (bound and unbound states).

### 1.2.3.2.1 LIE (Linear Interaction Energy)

The binding of a ligand to a biological macromolecule can be viewed as a process in which the ligand ( $l$ ) is transferred from one medium, i.e., free in water (f) to another, i.e., the binding site of the water-solvated macromolecular target (b).<sup>88</sup>



**Figure 1.9** Reference state (left) and bound state (right) of a ligand solvated in water

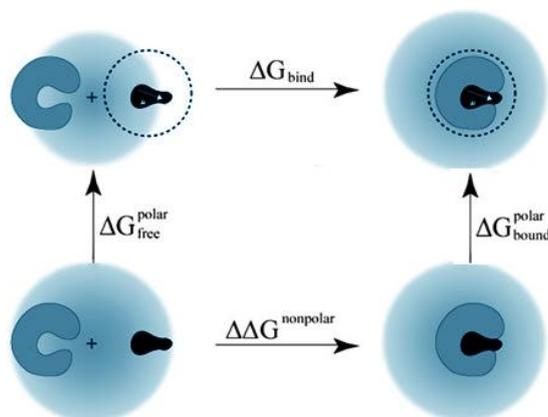
As a consequence, bound state of the ligand and reference state (water-solvated ligand) must be taken into account for a proper description of the total change in free energy associated to the formation of a ligand–receptor molecular complex. This is the correlation behind the LIE method, where the binding free energy is estimated as the free energy of transfer between water and protein environments as:

$$\Delta G_{bind}(l) = \Delta G_{sol}^b(l) - \Delta G_{sol}^f(l)$$

**Equation 1.18**

The main difference with respect to a regular transfer process between two solvents is that the standard state in water (1 M and free rotation) is replaced by restricted translation and rotation in a confined receptor-binding site.

In order to calculate the free energy of binding as a solely function of these two physical, relevant states of the ligand, we can draw a thermodynamic cycle (Figure 1.10), where the upper corners represent these two states (left: free, solvated in water; right: bound to the protein). The two bottom corners will account for two unphysical, intermediate states: a pseudo-ligand without any (intermolecular) electrostatic interactions, in its free (left) or bound (right) state.



**Figure 1.10** The thermodynamic cycle on which is based the estimation of the binding free energies with the linear interaction energy (LIE)

The resolution of such a thermodynamic cycle leads to the following equation:

$$\Delta G_{bind} = (\Delta G_{bound}^{polar} - \Delta G_{free}^{polar}) + \Delta \Delta G_{bind}^{nonpolar} = \Delta \Delta G_{bind}^{polar} + \Delta \Delta G_{bind}^{nonpolar}$$

**Equation 1.19**

where the entropic confinement contributions are hidden in the nonpolar term. Thus, the free energy of binding can be expressed as a sum of the corresponding polar and nonpolar components of the free energy. This is quite convenient, since molecular mechanics force fields analogously split the nonbonded potential energies into electrostatic and nonelectrostatic components. Potential energies ( $U$ ) can be converted into free energies ( $\Delta G$ ) considering that for the polar contribution,

a useful approximation comes from the linear response theory for electrostatic forces,<sup>89,90</sup> which states that the electrostatic part of the solvation free energy is:

$$\Delta G_{sol}^{el} = \frac{1}{2} \{ \langle U_{l-s}^{el} \rangle_{on} + \langle U_{l-s}^{el} \rangle_{off} \}$$

**Equation 1.20**

where the brackets  $\langle \rangle$  indicate thermodynamic averages of the ligand–surrounding (l–s) interaction energies as calculated with standard force-field Molecular Dynamics (or, alternatively, MC or other relevant statistical sampling). The term with the electrostatic interactions turned off in the sampling,  $\langle U_{l-s}^{el} \rangle_{off}$ , corresponds to the average electrostatic energy that would be obtained from the sampled configurations if the interactions instead were turned on (i.e., a “preorganization” term). This term is assumed to be constant or negligible compared to  $\langle U_{l-s}^{el} \rangle_{on}$  (the corresponding energies sampled with the interactions turned on). Thus we will write Equation 1.20 as  $\Delta G_{sol}^{el} = \frac{1}{2} \langle U_{l-s}^{el} \rangle_{on}$ , omitting a possible constant that will be considered below. In applying the linear response approximation to the problem of ligand binding we must also consider the reference state with a dissociated ligand in water. Furthermore, seemingly minor deviations from the exact linear response scaling factor of  $\frac{1}{2}$  have been demonstrated for hydration-free energies that, in fact, are important to take into account in order to improve the accuracy of the method.<sup>91,92</sup> Thus, we will write the expression for the polar component of the free energy in the general form of:

$$\Delta G_{bind}^{polar} = \beta (\langle U_{l-s}^{el} \rangle_b - \langle U_{l-s}^{el} \rangle_f) = \beta \Delta \langle U_{l-s}^{el} \rangle$$

**Equation 1.21**

The other main idea behind the LIE method is to estimate the nonpolar component of the free energy of binding analogously as:

$$\Delta G_{bind}^{nonpolar} = \alpha(\langle U_{l-s}^{vdW} \rangle_b - \langle U_{l-s}^{vdW} \rangle_f) = \alpha\Delta\langle U_{l-s}^{vdW} \rangle + \gamma$$

**Equation 1.22**

where the  $\alpha$  parameter is the empirically-derived nonpolar scaling factor and  $\gamma$  a constant. This was motivated by the observation of linear dependencies of both solvation free energies for nonpolar compounds and  $\langle U_{l-s}^{vdW} \rangle$  on molecular size (which can also be compared to semi-macroscopic approximations such as  $\Delta G_{sol}^{nonpolar} \cong \gamma A + c\Delta\langle U_{l-s}^{vdW} \rangle$ , representing the creation of a cavity and insertion of van der Waals centers into this cavity, where  $\gamma$  is the surface tension,  $A$  the surface area, and  $c$  a scaling factor). However, due to the fact that  $\langle U_{l-s}^{vdW} \rangle$  not only represents “steric” interactions but also is an efficient size measure, Equation 1.22 takes into account all size dependent and constant contributions to the binding free energy, approximating contributions from “cavity creation” confinement effects, and the second term of Equation 1.20.<sup>93</sup> It follows that the full LIE equation, for the estimation of binding affinities based on force-field averaged energies, can be written as:

$$\Delta G_{bind} = \alpha\Delta\langle U_{l-s}^{vdW} \rangle + \beta\Delta\langle U_{l-s}^{el} \rangle + \gamma$$

**Equation 1.23**

It is important to note that with this equation, one can calculate the free energy of binding by averaging the ligand-surrounding potential energies, which are collected only for the two physical states of the ligand involved in the binding process (represented in the upper corners of Figure 1.10): the *free* state (ligand solvated in water  $\langle U_{l-s} \rangle_f$ ) and the *bound* state (ligand in the solvated protein-binding site  $\langle U_{l-s} \rangle_b$ ). This makes a substantial difference compared to other

methods for the estimation of free energies, e.g., in more complicated methods, such as FEP or thermodynamic integration (TI), intermediate unphysical states resulting from mixing of end-point potentials must be explicitly simulated. On the other side, statistical methods such as scoring functions generally only take into account descriptors collected for the bound state, and not the free state, which tends to yield artificial dependencies of binding free energies on ligand size (molecular weight).

### 1.2.3.2.2 MM-GB(PB)SA

In the MM-GB(PB)SA formulation,<sup>94,95,96</sup> the binding free energy of a ligand (L) to a protein (P) to form the complex (PL) is obtained as the following difference (Equation 1.24):

$$\Delta G_{bind} = G(PL) - G(P) - G(L)$$

**Equation 1.24**

The free energy of each of the three molecular systems P, L, and PL is given by the Equation 1.25:

$$G(X) = E_{MM}(X) + G_{solv}(X) - TS(X)$$

**Equation 1.25**

In Equation 1.25,  $E_{MM}$  is the total molecular mechanics energy of molecular system X in the gas phase,  $G_{solv}$  is a correction term (solvation free energy) accounting for the fact that X is surrounded by solvent, and S is the entropy of X. To apply the MM-GB(PB)SA formulation, a representative set of equilibrium conformations for the complex, free protein and free ligand are first obtained by atomistic MD simulations in explicit solvent.

In this post-processing phase, the solvent is discarded and replaced by a dielectric continuum. Changes ( $\Delta$ ) in the individual terms ( $\Delta E_{MM}$ ,  $\Delta G_{solv}$ ,  $-T\Delta S$ ) of Equation 1.25 between the unbound states and the bound (complex) states are calculated, and contribute to the binding free energies according to Equation 1.26.

$E_{MM}$  is the sum of the bonded (internal), and non-bonded electrostatic and van der Waals energies

$$E_{MM} = E_{bonded} + E_{elec} + E_{vw}$$

**Equation 1.26**

These energy contributions are computed from the atomic coordinates of the protein, ligand and complex using the (gas phase) molecular mechanics energy function (or forcefield). The solvation free energy term  $G_{solv}$  contains both polar and non-polar contributions.

The polar contributions are accounted for by the generalized Born, Poisson, or Poisson-Boltzmann model, and the non-polar are assumed proportional to the solvent-accessible surface area (SASA) (Equation 1.27):

$$G_{solv} = G_{PB(GB)} + G_{SASA}$$

**Equation 1.27**

Finally, the entropy  $S$  is decomposed into translational, rotational and vibrational contributions. To improve the accuracy of the computed binding free energies, the various terms of Equation 1.27 are averaged over multiple conformations or MD snapshots (typically a few hundred for the  $E_{MM}$  and  $G_{solv}$  contributions). Depending on the extent of conformational fluctuations in the system under consideration, the convergence into stable values may require relatively long (multi-ns) simulations. The computation of the entropy term, however, requires the extensive minimization of the trajectory conformations for

the protein, ligand and complex to local minima on the potential energy surfaces, followed then by normal mode analysis. The internal energy terms ( $E_{\text{bonded}}$ ) of the protein and complex can be on the order of a few thousand kcal/mol, and can introduce large uncertainties in the computed binding free energies. For what concerns the second contribution in the free energy ( $G_{\text{solv}}(\mathbf{X})$ ), it is important to note that the solvent modifies in a non-trivial manner the intramolecular and intermolecular interactions, an accurate inclusion of solvent effects in biomolecular modelling and simulation is a challenging task. The most rigorous treatment of the solvent in a molecular dynamics simulation is to include explicit molecules in the calculation, with a notable increase in the computational cost. A much less costly approach is to represent the solvent implicitly in the simulation, through the incorporation of additional “potential of mean force” terms in the gas-phase energy function (e.g., Equation 1.28 below). These terms depend only on the atomic coordinates of the solute, and express the solute free energy for a given configuration, after the solvent degrees of freedom have been “integrated out”. Thus, the simulation system has the same number of degrees of freedom as in the gas phase and there is no need for explicit sampling over solvent degrees of freedom. The MM-PB(GB)SA method combine atomistic simulations in explicit solvent for the generation of representative biomolecular conformations with an implicit-solvent estimation of the binding free energies, in a post-processing step.

There are two approaches to generating the necessary ensembles for the bound and unbound state of binding energy calculations: all ensembles can be extracted from a single Molecular Dynamics (MD) or Monte Carlo (MC) trajectory of the bound complex, or trajectories can be generated for each state using separate simulations.

These approaches are called the single trajectory protocol (STP) and multiple trajectory protocol (MTP), respectively, and each approach has distinct advantages

and disadvantages. STP is less computationally expensive than MTP, because only a single trajectory is required to generate all three ensembles. Furthermore, the internal potential terms (e.g., bonds, angles, and dihedrals) cancel exactly in the STP, because the conformations in the bound and unbound ensembles are the same, leading to lower fluctuations and easier convergence in the binding free energy. The STP is appropriate if the receptor and ligand ensembles are comparable in the bound and unbound states. However, the conformations populating the unbound ensembles typically adopt strained configurations when extracted from the bound state ensemble, thereby overstabilizing the binding, compared to the MTP. Conceptually, most implicit solvent models decompose the solvation process into three sequential steps: i) creation of a cavity in solution to accommodate the biomolecule; ii) switching-on dispersion interactions between the biomolecule and surrounding medium, while all atomic charges are set to zero; and iii) switching-on the biomolecular charges. The solvation free energies of steps i) and ii) are normally assumed to be proportional to the SASA of the biomolecule and represent the *non-polar* contributions ( $G_{SASA}$ ) to  $G_{solv}$  in Equation 1.27, although the validity of this approximation has been questioned for step ii). With a positive coefficient of proportionality, an increase in the SASA is associated with an unfavorable increase in solvation free energy, which is partly accounted for by the tendency of non-polar residues to be solvent-excluded. The equation typically used is of the form (Equation 1.28):

$$G_{SASA}(X) = \gamma SASA + \beta$$

**Equation 1.28**

with the  $\gamma$  and  $\beta$  parameter values dependant on the method and solvation model (PBSA or GBSA) used. Meanwhile, step iii) calculates the contribution to solvation free energy due to the charge/electrostatic interactions of the solute with the

surrounding solvent, the *polar* contributions ( $G_{PB(GB)}$ ) to  $G_{\text{solv}}$  in Equation 1.27. In continuum-electrostatics models such as PB and GB, the solute is treated as a low-dielectric cavity embedded in a high dielectric medium. The solute charges are in the simplest and most common approximation centered on the individual atoms. The resulting solvation free energy of a molecule  $X$  is expressed as (Equation 1.29):

$$G_{PB(GB)}(X) = \frac{1}{2} \sum_{i,j \in X} q_i q_j g_{ij}^{PB(GB)}$$

**Equation 1.29**

where the summation is over all the atomic charges  $\{q_i\}$ . The quantity  $g_{ij}^{PB(GB)}$  is determined using the PB model by numerical solution of the Poisson or Poisson-Boltzmann equation (depending on the existence of salt), or using the GB model by an analytical expression with the functional form (Equation 1.30):

$$g_{ij}^{GB} = \left(\frac{1}{\epsilon} - 1\right) \left[ r_{ij}^n + B_{ij} \exp\left(-\frac{r_{ij}^n}{AB_{ij}}\right) \right]^{-1/n}$$

**Equation 1.30**

The parameters  $B_{ij}$  depend on the position (distance from the solute-solvent dielectric boundary) of atoms  $i$  and  $j$ , and the shape of the entire biomolecule;  $\epsilon$  is the solvent dielectric constant, and  $r_{ij}$  is the distance between  $i$  and  $j$ . The constants  $n$  and  $A$  were set to  $n=2$  and  $A=4$  in the original formulation of Still and coworkers.<sup>95</sup> In the PB model, the solute dielectric constant ( $\epsilon_{\text{in}}$ ) affects the computed functions  $g_{ij}^{PB}$  and Equation 1.29. Meanwhile, in the GB model, the solute dielectric constant drops out from the final expression in Equation 1.30, due to the approximations used to arrive at an analytic formula. An  $\epsilon_{\text{in}}$  value other than 1 can still be used; in this case, the first parenthesis on the right-hand side of

Equation 1.30 becomes  $(1/\epsilon - 1/\epsilon_{in})$  and the GB expression yields the free energy of transfer ring the solute from an infinite reference medium with dielectric constant  $\epsilon_{in}$  into solution. Application of Equation 1.29 to a protein:ligand complex (PL) and the dissociated protein (P) and ligand (L) yields the electrostatic (polar) solvation free energy contribution to Equation 1.24 (Equation 1.31):

$$\Delta G_{PB(GB)} = G_{PB(GB)}(PL) - G_{PB(GB)}(P) - G_{PB(GB)}(L)$$

**Equation 1.31**

An advantage of these methods is that they facilitate the decomposition of the total solvation free energy into insightful components. Hou and co-workers evaluated the performance of MM-GBSA and MM-PBSA for predicting binding free energies based on Molecular Dynamics simulations.<sup>97</sup> Their results showed that MM-PBSA performed better in calculating absolute binding free energies compared to MM-GBSA but not necessarily for the relative binding free energies, sufficient for most applications in computational drug design. Interestingly, in a study of the accuracy of continuum solvation models for drug-like molecules, GB methods typically were more stable and gave more accurate results than the widely used PB methods.<sup>98</sup> Recently, Miller et al.<sup>99</sup> implemented the free Python program MMPBSA.py, speeding up the calculation by dividing frames evenly across available processors. The source code is released under the GNU General Public License.

## **-CHAPTER 2-**

### ***Inverse Virtual Screening***

## 2.1 Inverse Virtual Screening: Introduction

The identification of the biological targets of a natural and/or synthetic compound represents a fundamental aim in medicinal chemistry and in the field of natural products. The discovery of the activity of a molecule on a specific target and the careful analysis at a molecular level of the main interactions can be used to rationalize its binding mode. In this context, computational chemistry represents today a valid and fast tool for the scientific research of new compounds with potential pharmacological activity. It is possible to analyze a large number of molecules in very short times, evaluating their binding mode with receptors of pharmacological interest (Virtual Screening<sup>100</sup>). The availability of the crystallographic structures of specific targets permits the evaluation of the binding mode of specific compounds before the synthesis, with a reduction of costs and timing of a drug discovery project. Nevertheless, there are only few examples where the calculations are performed on different targets characterized by their involvement in specific disease processes. This new type of computational approach is known as Inverse Virtual Screening,<sup>101,102,103,104</sup> in which libraries of compounds are tested on panel of targets with the aim of identifying a specific pharmacological activity (panel of antitumour, antiviral, antibacterial targets). This *in silico* tool could allow the identification of new specific biological actions for ligands with a well-known activity and explain in some cases the selectivity or the presence of side effects of a compound.<sup>105,106</sup>

This could be particularly advantageous for the natural products, because the small amounts of compounds obtained from natural sources usually prohibit the performance of pharmacological tests against a large number of receptors. In fact, the chemistry of the natural products has been investigated thoroughly in recent decades, resulting in a better understanding of enzymatic processes and in the development of biosynthetic knowledge and biogenetic theories for a logical

classification connecting a large variety of compounds. Many interesting secondary metabolites are produced by organisms in small quantities. Thus, for the isolation of a few milligrams of pure active metabolites, just enough to conduct preliminary in vitro biological tests, it is often necessary to extract kilograms of an organism. A useful approach for the study of natural products would be to carry out drug tests on a large number of biological targets with a better chance of evaluating their potential activity. Computational methods have been recently shown to be an important complementary tool for the study of the pharmacological activity of natural compounds.<sup>107,108,109,110,111,112</sup>

In this chapter, the molecular docking of several small molecules (organized in libraries or as single compound) against panels of receptor sites in an attempt to find ligands and binding conformations to direct experimental assays on specific biological targets. This approach has been initially applied to the discovery of potential antitumor compounds through the interaction with a number of protein targets involved in cancer, but there are on course other Inverse Virtual Screening studies for the evaluation of the antiviral and antibacterial activities. The panels of targets were built from the Protein Data Bank (PDB), by the selection of proteins involved in different pathologies and if commercially available for subsequent biological tests. The large number of available models for proteins is particularly useful for studying a wide range of molecules with variable biological activity.

It is noteworthy that this approach is also potentially applicable to libraries of synthetic compounds, to accelerate the analysis and to evaluate structure-activity relationships through a virtual method before the experimental study. The Inverse Virtual Screening method is also useful to provide information regarding ligand-protein interactions potentially affecting the physiology of the protein.

Summarizing, this method was applied on:

- a library of natural and marine compounds from LIBIOMOL library<sup>113</sup>

- a second library of natural products (10 compounds) with well-known antiproliferative and antioxidative biological activities<sup>114</sup>
- namalide, a synthetic peptidic compound<sup>115</sup>

In all these cases, using this approach, we were able to identify the targets of interaction of these compounds, confirming these results with biological tests. This highlights the promising usability of this new computational method for the accelerated discovery of the biological activity of natural and synthetic compounds.

Finally, it is important to underline that in each one of the Inverse Virtual Screening studies performed, we used a different number of targets, because the panel was constantly updated. For example, in the first study a panel of 126 targets was used, while in the present state a panel of 210 targets involved in the cancer events is available (Table 2.1). In parallel, we are building other two panels of targets, respectively for the viral (so far, 85 targets) and mycotic infections (so far, 130 targets).

Even if the scope of this thesis is mainly focused on the development of new computational methods for the screening of targets involved in pathological events, it is proper to give an overview of the main events occurring in the cancer disease, that represents the pathology mainly investigated in this type of study. This information are summarized in Appendix A.

protein	PDB code	protein	PDB code	protein	PDB code
14_3_3_epsilon	2BR9	egfr	2J6M	mek4_ancp	3ALN
14_3_3_gamma	2B05	enolase1_site_1	2PSN	mek4_no_ancp	3ALN
14_3_3_sigma	1YWT	enolase1_site_2	2PSN	mek5_pb1_domain	1W10
14_3_3_theta	2BTP	enolase2_site_1	1TE6	metap2	1YW9
14_3_3_zeta	1QJA	enolase2_site_2	1TE6	mk2	3M42
abl2	3HMI	enolase3_site_1	2X6X	mlk1	3DTC
abl	2HYY	enolase3_site_2	2X6X	mmp13	830C
aif	1M6I	EPHa3	3DZQ	mmp3	1HY7
akt1	3MVH	EPHa4	2X9F	mmp8	20Y2
akt2	3D0E	erbB2	1S78	mrp1	2CBZ
alk	2XP2	erbB4	2R4B	msk1	3KN5
alk5	2WOU	erk1	2Z0Q	mthfs	3HY3
ape1	2ISI	erk2	2OJG	mtor	3FAP
ask1	3VW6	fak	3BZ3	mtsp1	3NCL
aurkinA	2W1D	fgf1	1HKH	nek2	2XKF
aurkinB	2VG0	fgfr1	1AGW	nek7	3XQN
bap1	2W15	fgfr2	2PVF	nnos	3JT4
bcl2	2O21	flt3	3QS7	nqo1	2F10
bcl2a1	2VM6	fpps_allo	3N1V	p300	3BIY
bcl6	3LBZ	fpps_ipp	1ZW5	p38	3HEG
bclw	1ZY3	fpps_no_ipp	1ZW5	p53_mut	2X0V
bclx1	1BXL	ftase	1LD8	parp	2JVN
braf	3C4C	fxr	1OSV	pcaf	2RNV
brd2_bd1	2YDW	galectin1	1W6M	pcna	1VYJ
brd2_bd2	3ONI	galectin3	1KJL	pd	3BIK
brd3_bd1	3S91	galectin7	3GAL	pdk1	3NAX
brd3_bd2	3S92	gsk3	3F7Z	peroxiredoxin_1	2RII
brd4_bd1	3MXF	gstm2_2	3GUR	peroxiredoxin_hORF6	1PRX
brd4_bd2	2YEM	gstp1	2A2R	pgm	1YFK
blk	3PIX	hdac1	homology_modelling	pi3k	3ENE
calmodulin	3EWT	hdac2	homology_modelling	pig3	2J8Z
camKIIB	3BHH	hdac3	homology_modelling	pimkin	3JYA
caspace1	2FQQ	hdac4	homology_modelling	pka	3L9L
caspace2	1PYO	hdac6	homology_modelling	pkca	3IW4
caspace3	3EDQ	hdac7	homology_modelling	pkcbII	2I0E
caspace7	1SHL	hdac8b	homology_modelling	pkc_eta	3TXO
caspace8	1QTN	hdac8	homology_modelling	pkc_iota_2	1ZRZ
cathepsinB	1GMY	hgfr_cmet_alt	2WD1	pkc_iota	3A8W
cathepsinG	1AU8	hmt	3HNA	pkc_iota_apo	3A8X
cathepsinK	2R6N	hsc70	3FZH	pkct	2JED
cathepsinL	3HWN	hsp90	2WI6	plk1	3FVH
cbp	2RNY	hspa1a	3XU	pnk	2W30
cdc42_no_alf3	2NGR	hspa11	3GDQ	pop	3DDU
cdk2	2VV9	hspa2	3I33	ppar_g	3AN3
cdk2_alt	2WIH	hspa5	3IUC	pten	1D5R
cdk2_altern	2WFY	hspa6	3FE1	pxr	1M13
cdk4	3G33	ido	2D0T	pyk2	3FZS
cdk5	1H4L	igf	3F5P	raf	3C4C
cdk6	2F2C	irak4	2NRU	ras	3GFT
cdk7	1UA2	jak1	3EYG	ret	2X2K
cdk8	3RGF	jak2	3E64	srpk	1WBP
cdk9	3BLQ	jak3	1YVJ	stat3	1BG1
chk1	2QHN	jmjd3	2XXZ	survivin	1XOX
chk2	2W7X	jmjd3_akg	2XXZ	syk	3FQH
ciap1	3D9U	jnk1	3KVX	tank1	2R8F
ck2	3FL5	jnk2	3NPC	tank2	3KR8
clk1	1Z57	jnk3	2ZDT	tao2	2GCD
clk3	2WU6	kit_kinase	3G0E	tdp1	1RFF
cmet	2WGJ	kras	3GFT	tie2	2O08
CPA	1CBX	lck	3AD4	top1	1K4T
CPU	3D67	lsd1	2EJR	topII	1ZXM
cSRC	3F3V	lyn	2ZVA	topII_atp	1QZR
cxcr4	2K05	mc11	3D7V	tp	1UOU
dapk	3EH9	mdm2	3EQS	ts	3NCL
dhfr	1PD8	mdmx	3EQY	upa	2VIP
diaminoox	3HIG	mek1	3DV3	vegfr1	1FLT
dnmt3a	3A1B	mek1_adp	1S9J	vegfr2	3EWH
dnmt3l	2QRV	mek1_no_adp_no_lig	1S9J	wee1	1X8B
dyrk_1a	3ANQ	mek2_adp	1S9I	xiap	1TFT
e-cadherin	2O72	mek2_no_adp_no_lig	1S9I	zap70	1U59

**Table 2.1** Panel of 210 targets involved in cancer processes built for Inverse Virtual Screening studies

## 2.2 Pilot Inverse Virtual Screening study: LIBIOMOL library

A pilot Inverse Virtual Screening was initially conducted on a library of bioactive compounds (“Library on Bioactive Molecules” - [www.libiomol.unina.it](http://www.libiomol.unina.it)) classified as: (a) molecules with an action on the cytoskeleton; (b) cytotoxic compounds; (c) antitumor agents; (d) antiproliferative substances, and (e) antiangiogenic compounds. When this pilot study was performed 43 molecules were examined, comprising 27 natural compounds, three semisynthetic compounds, and 13 synthetic compounds designed for mimicking selected natural skeletons. This library was screened against a panel of targets selected for their correlations in cancer on the basis of the Inverse Virtual Screening method, allowing the obtainment of a restricted group of promising candidates for subsequent biological tests.

In particular, Autodock-Vina<sup>28</sup> calculations were performed. This software has been shown to produce an increased efficiency in predicting the experimental binding poses and energies, and a two orders of magnitude speed-up compared with Autodock 4.0.<sup>25</sup> It has been designed for parallel computing, representing a particularly suitable tool for this study, for large virtual screening studies in general, and for the investigation of ligands presenting large numbers of active torsion angles, such as naturally occurring compounds.

Docking calculations were performed between 43 molecules (Chart 2.1) previously tested for their potential activity (antitumor, cytotoxic, antiangiogenic, antiproliferative, activity on the cytoskeleton) against a panel of 126 protein targets (Table 2.2). The library used for the calculations included several types of molecules, characterized by similar or different chemical structures.

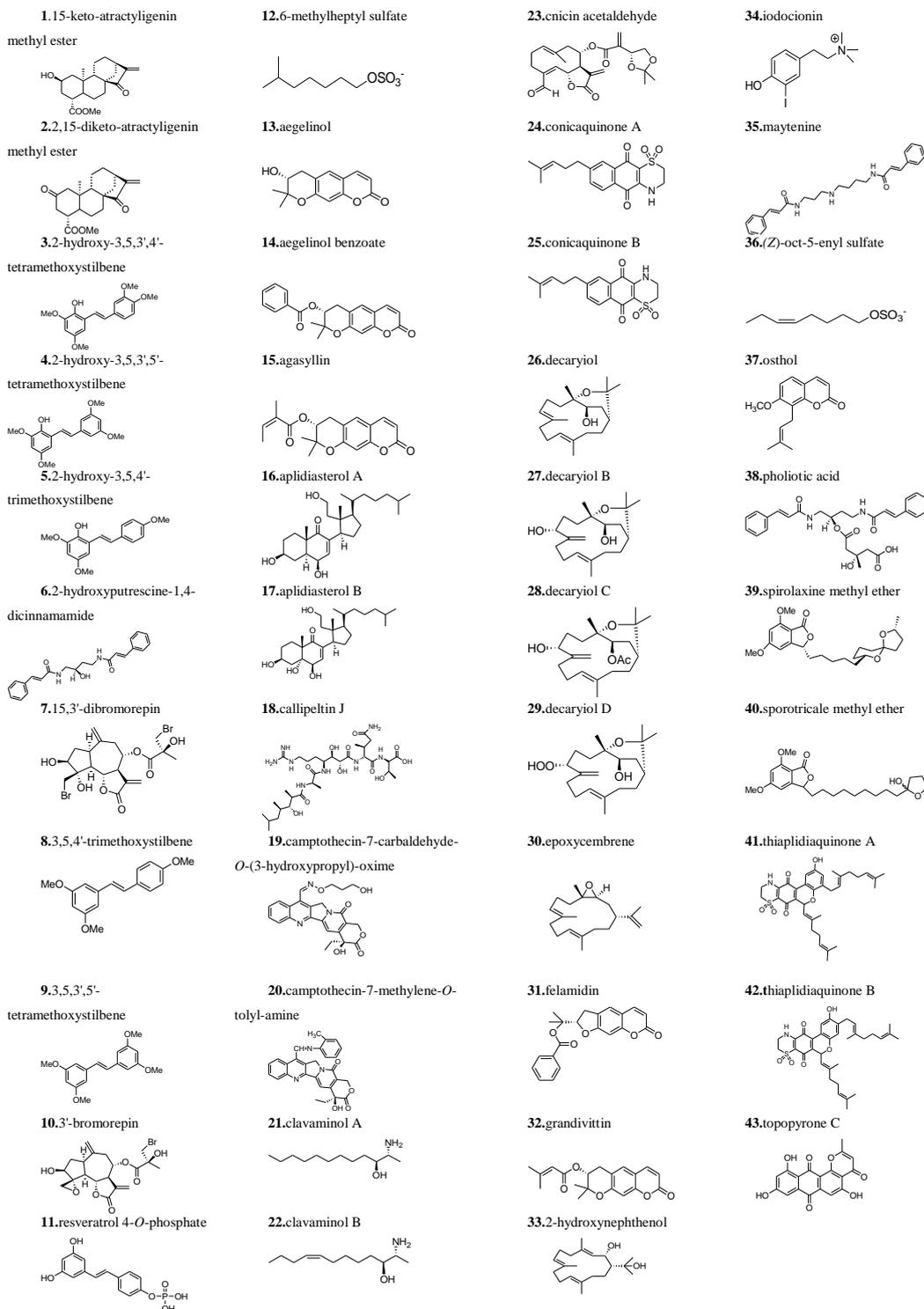


Chart 2.1 Structures of compounds 1-43

protein	PDB code	protein	PDB code	protein	PDB code
abl	2HYY	diaminoox	3HIG	mek1	3DV3
abl2	3HMI	dnmt3a	3A1B	metap2	1YW9
aif	1M6I	dnmt3l	2QRV	mmp3	1HY7
akt	3MVH	e-cadherin	2O72	mmp8	2OY2
apaf1	1Z6T	egfr	2J6M	mmp13	830C
ape1	2ISI	epsilon	2BR9	mrp1	2CBZ
aurkin	2W1H	erbB2	1S78	miTor	3FAP
bap1	2W15	erk1	2ZOQ	nnos	3JT4
bcl2	2O21	erk2	2OJG	nqo1	2F1O
bcl2a1	2VM6	fak	3BZ3	p38	3HEG
bclw	1ZY3	fgf1	1HKN	parp	2JVN
bclxl	1BXL	fgfr1	1AGW	pcaf	2RNW
braf	3C4C	fgfr2	1EV2	pcna	1VYJ
calmodulin	3EWT	fgfr3	3GRW	pd	3BIK
caspase1	2FQQ	ftase	1LD8	pgm	1YFK
caspase2	1PY0	galectin1	1W6M	pi3k	3ENE
caspase3	3EDQ	galectin3	1KJL	pik3	1E7U
caspase7	1SHL	galectin7	3GAL	pimKin	3JYA
caspase8	1QTN	gamma	2B05	plk1	3FWH
cathepsin B	1GMY	gsk3	3F7Z	pop	3DDU
cathepsin G	1AU8	gstp1	2A2R	pten	1D5R
cathepsin K	2R6N	hdac1	homology modelling	pyk2	3FZS
cathepsin L	3HWN	hdac2	homology modelling	raf	3IDP
cbp-p300 <sup>a</sup>	1JJS	hdac3	homology modelling	srpk	1WBP
cd20	3BKY	hdac4	homology modelling	stat1-s1	1BF5
cdk2	2VV9	hdac6	homology modelling	stat1-s2	1BF5
cdk6	2F2C	hdac7	homology modelling	stratifin	1YWT
cdk7	1UA2	hdac8	homology modelling	survivin	2RAW
cdk9	3BLQ	hdac8b	homology modelling	tdp1	1RFF
chk1	2QHN	hmt	3HNA	teta	2BTP
chk2	2W7X	hsp90	2W16	tie2	2O08
ciap1	3D9U	ido	2D07	topI	1K4T
ciap2	3EB5	igf	3F5P	topII	1ZXM
ck2	3FL5	jak1	3EYG	topII_atp	1QZR
clk1	1Z57	jak2	3E64	tp	1UOU
clk3	2WU6	jak3	1YVJ	upa	2VIP
cmet	2WGJ	kit-kinase	3G0E	vegfr1	1FLT
ctl4	1I8L	kras	3GFT	vegfr2	3EWH
cxcr4	2K05	lsd1	2EJR	wee1	1X8B
cyclA-cdk2	2IW9	mcl1	3D7V	xrcc1	2W3O
cyclin A	2WFY	mdm2	3EQS	zeta	1QJA
dhfr	1PD8	mdmx	3EQY		

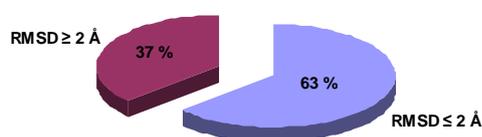
**Table 2.2** The panel of targets used in the pilot Inverse Virtual Screening study

### 2.2.1 Analysis of Predicted Binding Energies

The results of Inverse Virtual Screening were collected in different tables and initially sorted by single ligand vs. target, with the energies expressed in kcal/mol from the highest to the lowest values, in order to explore the possibility of identify ligands with good affinity and selectivity by evaluation of the predicted binding energies.<sup>116</sup> The mere analysis of the binding energies highlighted a restricted group of targets selected with high values of predicted binding energies on a significant number of ligands tested, thus suggesting the use of a re-modulation of the results using a different criterion.

### 2.2.2 Comparing Standard Ligands

Accordingly, docking calculations of crystallized ligands, with a well-known binding mode, were performed in order to obtain a standard energy to be introduced as a filter in the evaluation of the binding energies of the matrix used. In general, in order to assess the efficiency of the docking experiments, the root-mean square deviations (RMSD, expressed in Å) of the docked conformations related to the crystallized ones were calculated. Choosing an exhaustiveness value of 16 for Autodock-Vina calculations and cutoff of 2 Å (an upper limit indicative of a good superimposition), 63% of the correlated structures within this range were identified (Table 2.3 and Figure 2.1).



**Figure 2.1** Distribution of RMSD for the 90 co-crystallized ligands in the PDB files docked with Autodock-Vina software.

abl <sub>lig</sub>	0.917	cathepsinG <sub>lig</sub>	2.157	epsilon <sub>lig</sub>	1.402	kras <sub>lig</sub>	1.527	pimkin <sub>lig</sub>	1.285
abl2 <sub>lig</sub>	0.906	cathepsinK <sub>lig</sub>	1.501	erk1 <sub>lig</sub>	1.408	lsd1 <sub>lig</sub>	2.467	plk1 <sub>lig</sub>	2.670
aif <sub>lig</sub>	1.416	cathepsinL <sub>lig</sub>	1.674	erk2 <sub>lig</sub>	0.458	mcl1 <sub>lig</sub>	9.474	pop <sub>lig</sub>	0.800
AKT <sub>lig</sub>	0.730	cd20 <sub>lig</sub>	3.280	fak <sub>lig</sub>	0.956	mdm2 <sub>lig</sub>	9.524	pten <sub>lig</sub>	0.854
aurkin <sub>lig</sub>	0.729	cdk6 <sub>lig</sub>	1.735	fgfr1 <sub>lig</sub>	8.920	mdmx <sub>lig</sub>	5.776	pyk2 <sub>lig</sub>	0.770
bap1 <sub>lig</sub>	1.619	cdk7 <sub>lig</sub>	1.803	ftase <sub>lig</sub>	0.467	mek1 <sub>lig</sub>	1.715	rafi <sub>lig</sub>	0.170
bcl2 <sub>lig</sub>	0.945	cdk9 <sub>lig</sub>	3.034	galectin3 <sub>lig</sub>	1.299	metap2 <sub>lig</sub>	1.401	srpk <sub>lig</sub>	7.934
bcl2a1 <sub>lig</sub>	2.215	chk1 <sub>lig</sub>	0.951	galectin7 <sub>lig</sub>	1.143	mmp13 <sub>lig</sub>	0.946	stratfin <sub>lig</sub>	2.200
bclw <sub>lig</sub>	2.146	chk2 <sub>lig</sub>	0.350	gamma <sub>lig</sub>	3.655	mmp3 <sub>lig</sub>	2.040	teta <sub>lig</sub>	4.531
bclxl <sub>lig</sub>	7.838	ciap1 <sub>lig</sub>	1.609	gsk3 <sub>lig</sub>	0.805	mmp8 <sub>lig</sub>	0.154	tie2 <sub>lig</sub>	1.341
braf <sub>lig</sub>	0.410	ck2 <sub>lig</sub>	0.837	gstp1 <sub>lig</sub>	1.398	mrp1 <sub>lig</sub>	4.235	top1 <sub>lig</sub>	1.250
calmodulin <sub>lig</sub>	6.137	clk1 <sub>lig</sub>	0.307	hgfr <sub>lig</sub>	0.453	mtor <sub>lig</sub>	0.219	topII <sub>lig</sub>	0.837
caspase1 <sub>lig</sub>	1.688	clk3 <sub>lig</sub>	1.108	hmt <sub>lig</sub>	2.073	nno5 <sub>lig</sub>	1.035	topII_atp <sub>lig</sub>	2.563
caspase2 <sub>lig</sub>	5.858	cmet <sub>lig</sub>	1.025	ido <sub>lig</sub>	0.541	nqo1 <sub>lig</sub>	4.185	tp <sub>lig</sub>	0.693
caspase3 <sub>lig</sub>	4.380	diaminoox <sub>lig</sub>	2.257	igfi <sub>lig</sub>	2.478	p38 <sub>lig</sub>	1.247	upa <sub>lig</sub>	0.919
caspase7 <sub>lig</sub>	2.372	dnmt3a <sub>lig</sub>	1.974	jak1 <sub>lig</sub>	0.331	pcaf <sub>lig</sub>	15.307	vegfr2 <sub>lig</sub>	0.399
caspase8 <sub>lig</sub>	2.270	dnmt3 <sub>lig</sub>	2.729	jak2 <sub>lig</sub>	1.135	pcna <sub>lig</sub>	10.441	wee1 <sub>lig</sub>	0.330
cathepsinB <sub>lig</sub>	1.992	egfr <sub>lig</sub>	1.256	kit_kinase <sub>lig</sub>	1.202	pgm <sub>lig</sub>	2.042	zeta <sub>lig</sub>	2.250

**Table 2.3** Root mean square deviations (RMSD, expressed in Å) calculated between crystallized and docked conformations for the 90 ligands in complex with related targets in the PDB files.

This procedure was initially assessed to 44 targets of the panel for which docked ligands (standards) complied with this requirement. In particular, the efficiency in the binding was evaluated through the ratio between the binding energies of the ligands and the standards, as indicated in Equation 2.1:

$$\delta = \Delta G_{\text{compounds}} / \Delta G_{\text{standards}}$$

#### Equation 2.1

Compounds showing a  $\delta \geq 1$  were selected from the library.

From this analysis, 335 associations between the ligands and the targets on 1892 (43 ligands of the library  $\times$  44 targets investigated) calculations performed showed a  $\delta \geq 1$ , suggesting a consistent number of false positives.<sup>116</sup> Careful analysis of the chemical structures highlighted a connection between high values of  $\delta$  and high molecular weights of Libiomol ligands, especially when these were correlated to low molecular weights of standard ligands.

Examples are the crystallized ligands (standards) for the targets pten (PDB code = 1D5R; standard C<sub>4</sub>H<sub>6</sub>O<sub>6</sub>, mw = 150.09 g/mol), tp (PDB code = 1UOU; standard C<sub>9</sub>H<sub>11</sub>ClN<sub>4</sub>O<sub>2</sub>, mw = 242.66 g/mol), clk1 (PDB code = 1Z57; standard C<sub>11</sub>H<sub>11</sub>N<sub>5</sub>O<sub>2</sub>, mw = 245.24 g/mol).

This trend could be explained by considering the importance of molecular size in the predicted docking energy; in particular an increase of this parameter may influence the amount of van der Waals interactions representing an important factor for the calculations with docking software.<sup>25</sup>

### **2.2.3 Introducing Ligand Efficiency (LE)**

For the above considerations, subsequent analysis were performed considering the “ligand efficiency” of all the molecular structures. Ligand efficiency, a term that has recently attracted the attention of researchers involved in the drug discovery field, is generally defined as the binding energy of a ligand normalized by its size.

Successful drug discovery involves the optimization of many variables, such as compound potency, selectivity, cellular activity, solubility, metabolic stability, bioavailability, and acceptable toxicity.

Recently, the concept of ligand efficiency as a measure for lead selection was suggested. Ligand efficiency reduces the number of variables by combining potency with molecular weight and polar surface area, and it represents a particularly suitable parameter in the field of the drug discovery.

This parameter is useful for effective and efficient drug discovery, and might provide the basis for a mathematically robust optimization of the drug discovery process.<sup>117</sup>

The ligand efficiency depends on the size of the ligand, as smaller ligands have a higher efficiency than the larger ligands.

One of the reasons behind this principle is the reduction in the area accessible to the ligand increasing the size of the ligand. These findings have important implications in the screening of libraries of compounds.

Ligand efficiency is calculated using the Equation 2.2:

$$LE = \Delta G / N$$

**Equation 2.2**

where N is the number of non-hydrogen atoms.<sup>117</sup>

On this basis, we calculated the ligand efficiency of our database of compounds and for the standard ligands, and the results are reported as ratios between the values obtained for the ligands and the standards, as indicated in Equation 2.3:

$$\delta_{LE} = LE_{compound} / LE_{standard}$$

**Equation 2.3**

For each receptor considered, selected compounds were those complying with the following conditions (Equation 2.4):

$$a) \delta_{LE} \geq 1; (b) \delta_{LE} \geq M + 3\sigma$$

**Equation 2.4**

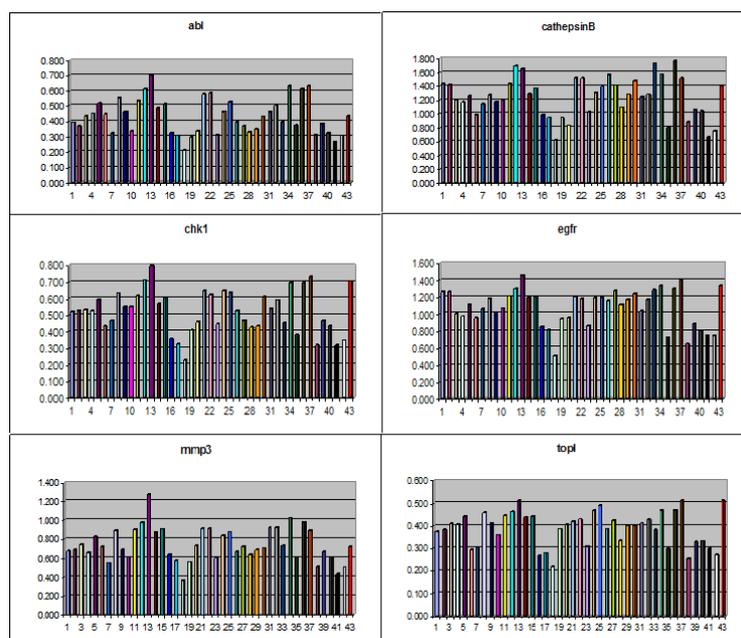
where M is the average value of  $\delta_{LE}$  for all the compounds and  $\sigma$  is the standard deviation (a value of  $3\sigma$  was used in order to select the best results from this type of screening).

Histograms associated to each target were drawn to assess the overall behavior of the compounds analyzed for ligand efficiency.<sup>116</sup>

The trends observed for each target proved to be very similar, as shown in histograms related to a sample of 6 targets (Figure 2.2). In a matrix of 43 compounds and 44 targets, the molecules with a low molecular weight showed the best values ( $\delta_{LE} \geq 1$ ): 6-methylheptyl sulfate<sup>118</sup> and aegelinol<sup>119,120,121</sup> in docking with 18 targets, 2-hydroxynephtenol<sup>122</sup> in docking with eight targets, iodocionin<sup>123</sup> in docking with 17 targets, and (Z)-oct-5-enyl sulfate<sup>118</sup> and osthol<sup>124,125</sup> in docking

with 17 and 14 targets, respectively. The common targets for these compounds were *abl2*, *akt*, *bap1*, cathepsin B, cathepsin K, *cdk6*, *egfr*, *mtor*, and *pyk2*.

The most important feature of this analysis is that small molecules with a better ligand efficiency than standard ligands were selected systematically by the screening procedure.



**Figure 2.2** Histograms of  $\delta_{LE}$  values for six target samples. On the x axis the id of the compound and on the y axis the  $\delta_{LE}$  are reported. Also considering different y axis scales, it is clear the same distribution of  $\delta_{LE}$  for the compounds investigated.

Ligand efficiency is very important to establish limits in the building of new structures adapted on an active protein site, but the evaluation of this parameter was not considered a useful method of screening in this study.

## 2.2.4 Normalization of the Matrix

To overcome the lack of selectivity of the molecules chosen for the library to the panel of receptors, the binding energy (kcal/mol) data were organized in a matrix of 43 structures vs. the 126 targets of antitumor panel.

The aim was to exclude the false positives through a mathematical filter aimed at eliminating systematic errors associated with molecules and targets in their interaction.

To obtain the normalization of binding energy values in the matrix, Equation 2.5 was applied:

$$V = V_0 / [(M_L + M_R) / 2]$$

### Equation 2.5

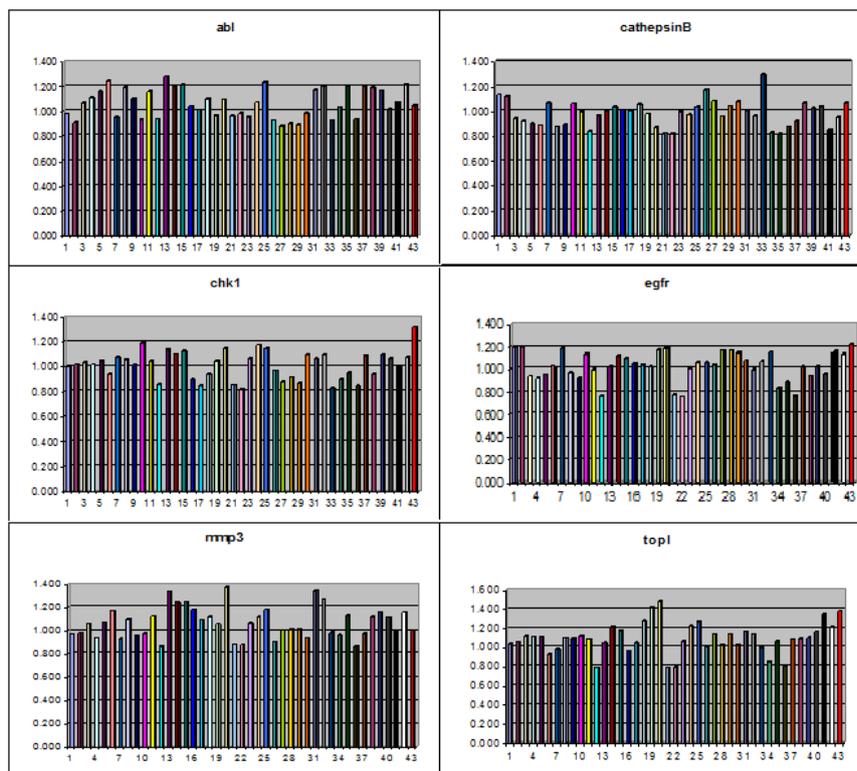
In this convention,  $V$  is the new value associated to each compound,  $V_0$  is the value of binding energy obtained from the docking calculation,  $M_L$  is the average binding energy of each ligand (on different targets), and  $M_R$  is the average binding energy associated to each target (on the various ligands). Every single value in the matrix representing a single ligand vs. a specific receptor was accordingly normalized taking simultaneously into account the influence of the two specific averages contained in Equation 2.5.<sup>116</sup>

In this case, the values obtained led to the creation of histograms showing a different trend on the individual compounds against every single target, as shown in Figure 2.3 for the same sample of targets shown previously in Figure 2.2.

The molecules were selected through calculation of the standard deviation from the average of matrix ( $M$ ); in particular the molecules were chosen up to the value of  $M + 3\sigma$  to classify compounds with the best interactions. Selected results are reported in Table 2.5.

targets↓	1	2	3	M <sub>R</sub> <sup>a</sup>
abl	-7.5	-7.0	-7.9	-8.2
abl2	-6.2	-6.1	-8.2	-7.5
aif	-9.4	-9.3	-8.1	-8.6
akt	-8.6	-8.2	-7.1	-8.0
ape1	-6.6	-6.7	-5.7	-6.6
aurkin	-7.4	-7.0	-8.2	-8.2
bap1	-6.0	-6.0	-6.7	-7.0
bcl2	-7.9	-7.9	-6.9	-7.5
bclx1	-8.2	-8.0	-6.6	-7.4
braf	-8.6	-8.8	-7.6	-8.5
calmodulin	-6.6	-6.6	-5.7	-6.8
caspase1	-5.7	-5.5	-4.6	-5.3
caspase2	-5.2	-5.2	-4.9	-5.0
caspase3	-7.2	-7.4	-6.4	-7.1
caspase7	-8.4	-8.4	-6.8	-7.7
caspase8	-6.9	-6.9	-6.3	-6.6
cathepsin B	-8.0	-7.9	-6.4	-6.9
cathepsin G	-7.7	-7.7	-6.5	-6.8
cathepsin K	-6.3	-6.5	-5.8	-6.1
cathepsin L	-6.4	-6.1	-5.9	-6.3
cdk2	-8.5	-8.6	-7.8	-8.1
cdk6	-7.9	-8.6	-6.9	-7.8
cdk7	-8.9	-9.0	-7.8	-8.3
cdk9	-8.3	-8.4	-6.1	-7.6
chk1	-7.3	-7.4	-7.2	-7.3
chk2	-6.2	-6.3	-8.7	-8.0
ciap1	-5.9	-5.7	-5.6	-5.9
ck2	-7.8	-7.9	-7.7	-8.1
clk1	-7.6	-8.0	-7.8	-8.1
clk3	-8.7	-9.1	-7.4	-8.5
cmet	-6.4	-6.5	-7.4	-7.6
cyclin A	-7.5	-7.5	-5.5	-6.4
dhfr	-8.5	-8.5	-7.4	-7.9
M <sub>L</sub> <sup>b</sup>	-7.2	-7.2	-6.6	

**Table 2.4** Values of Binding Energies for Three Sample Ligands and 33 Sample Targets  
(<sup>a</sup>Average of values for targets. <sup>b</sup>Average of the values for ligands).



**Figure 2.3** Histograms related to the same six target samples reported in Figure 2.2 after normalization of predicted binding energies

compounds	targets	V
19	top1	1.418
20	chk2	1.446
	lsd1	1.450
	mTor	1.690
	top1	1.483
28	caspase2	1.409
31	abl2	1.398
41	mTor	1.438
43	wee1	1.429

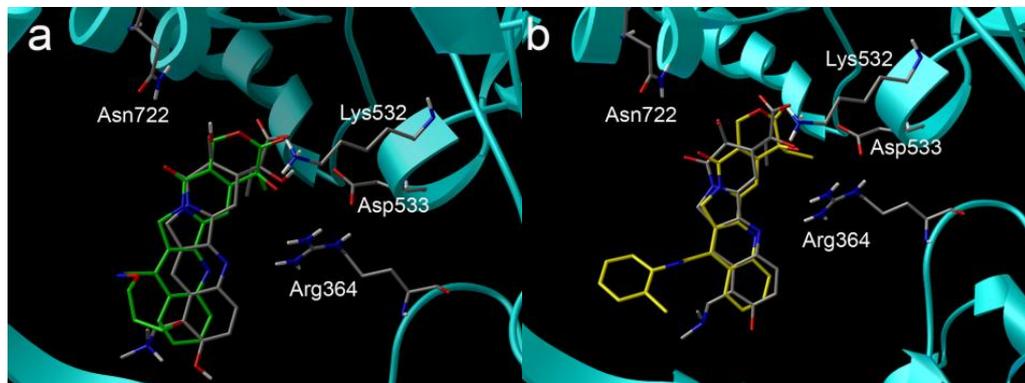
**Table 2.5** Selected compounds after normalization of the matrix

Analysis of the tables confirmed the validity of the method. In particular, as reported in the literature, both camptothecin-7-carbaldehyde-*O*-3-hydroxypropyl-oxime (**19**)<sup>126</sup> (Figure 2.4 a) and camptothecin-7-methylene-*O*-tolyl-amine (**20**)<sup>127</sup> (Figure 2.4 b) are semisynthetic derivatives of the naturally occurring

camptothecin, which in biological assays show an action on topoisomerase I.<sup>128</sup> They stimulate topoisomerase I-mediated DNA cleavage and the persistence of the cleavable complex; these compounds were evaluated for their cytotoxicity against the H460 human non-small lung carcinoma cell line, using topotecan as a reference compound ( $IC_{50} = 0.40 \mu\text{M}$ ; topotecan  $1.38 \mu\text{M}$ ).<sup>129</sup>

Topotecan and these compounds are very well overlapped in the pocket-receptor, establishing many common interactions. From Figure 2.4 the accuracy of Autodock-Vina calculations on these two compounds on the panel of receptors in identifying the target of choice was clear.

The two derivatives of camptothecin establish van der Waals interactions with the same residues of the pocket occupied by topotecan (Asn722, Lys 532, Asp533, Arg364), with  $K_d = 2.65 \times 10^{-9} \text{ M}$  (**19**) and  $K_d = 2.96 \times 10^{-10} \text{ M}$  (**20**). As a reference, the same Vina calculation has been performed on topotecan used as model and the result was  $-10.2 \text{ kcal/mol}$  ( $K_d = 3.34 \times 10^{-8} \text{ M}$ ).



**Figure 2.4** (a) Superimposition of topotecan (colored by atom type: O red, N blue, C grey)/camptothecin-7-carbaldehyde-*O*-3-hydroxypropyl-oxime (**19**) (colored by atom type: O red, N blue, C green) in docking with topoisomerase I (PDB code = 1K4T). (b) Superimposition of topotecan (colored by atom type: O red, N blue, C grey)/camptothecin-7-methylene-*O*-tolyl-amine (**20**) (colored by atom type: O red, N blue, C yellow) in docking with topoisomerase I.<sup>130</sup>

Two other positive results were obtained in the calculated interactions, between the natural cytotoxic thiaplidiaquinone A<sup>131</sup> (**41**) with the receptor mTor (PDB code = 3FAP) and the cytotoxic topopyrone C<sup>132,133</sup> (**43**) with the receptor weel kinase (PDB code = 1X8B; Figure 2.5 a).<sup>134</sup>

Topopyrone C is very well superimposed on a crystallized inhibitor reported in the literature on weel kinase (9-hydroxy-4-phenylpyrrolo[3,4-*c*]-carbazole-1,3(2*H*,6*H*)-dione). It establishes van der Waals interactions with Ile305, Val313, Lys323, Ala326, Phe433, and Gly382, like the crystallized ligand, with a  $K_d = 1.35 \times 10^{-9}$  M, when compared with a  $K_d$  for the standard ligand of  $5.81 \times 10^{-10}$  M.

Moreover, topopyrone C is a synthetic compound evaluated for its cytotoxicity against the H460 cell line, using topotecan as a reference compound ( $IC_{50} = 29.50 \mu\text{M}$ ; topotecan  $1.38 \mu\text{M}$ ). It induces the same sequence selectivity of topoisomerase I-mediated DNA cleavage shown by camptothecin derivatives.

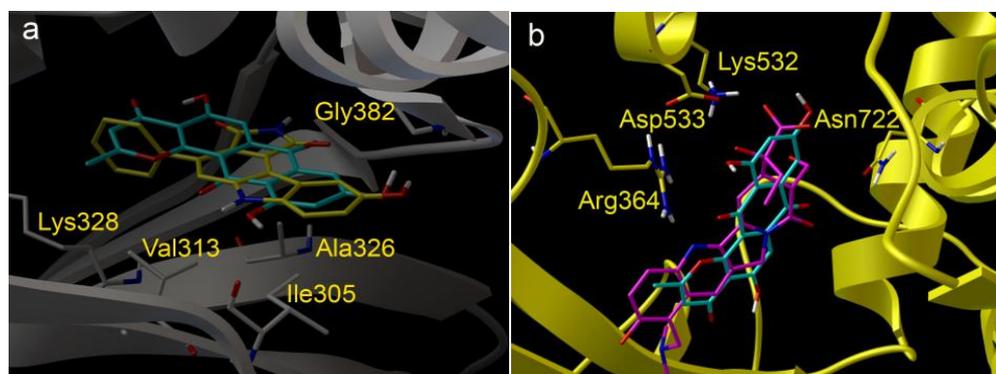
As reported in the literature on pharmacological assays, the interaction with topoisomerase I has been found.

Topotecan and topopyrone C are very well overlapped in the pocket receptor (Figure 2.5 b); topopyrone C establishes van der Waals interactions with the same residues of the pocket occupied by topotecan (Asn722, Lys532, Asp533, and Arg364) with  $K_d = 2.65 \times 10^{-9}$  M, while for topotecan the value is  $K_d = 3.34 \times 10^{-8}$  M. The high *V* value related to the interaction of **43** with TopI (*V* = 1.379) confirms these observations.

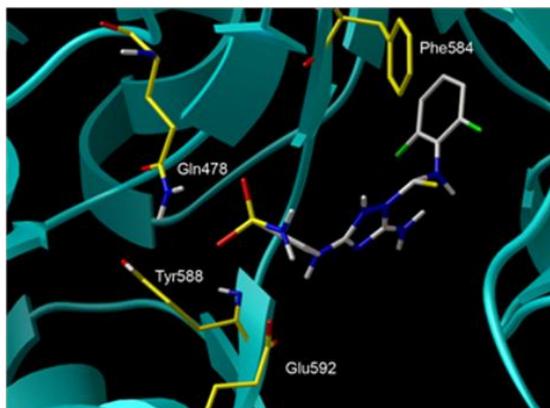
To obtain confirmation of the proposed method, the table organized in a matrix has been integrated with the Autodock-Vina results on two standard known molecules as ligands of the targets abl2 (PDB code = 3HMI) and FTase (PDB code = 1LD8). Both ligands (5-amino-3-{[4-(aminosulfonyl)phenyl]amino}-*N*-(2,6-difluorophenyl)-1*H*-1,2,4-triazole-1-carbothioamide for abl2 and (2*S*)-19,20,21,22-tetrahydro-19-oxo-5*H*-18,20-ethano-12,14-etheno-6,10-metheno-18*H*-

benz [d] imidazo [4,3-*k*] [1,6,9,12] oxatriaza-yclooctadecosine-9-carbonitrile for FTase) are as crystals in the corresponding PDB. Docking calculations of the two compounds were performed on all the target members of the panel. The aim was to verify that the two known ligands would show significant *V* values when interacting with their specific target.

The candidate targets for these molecules were selected through calculation of the standard deviation from the average of matrix (*M*) choosing *V* values up to  $M + 3\sigma$  to classify the targets with the best interactions.



**Figure 2.5** (a) Superimposition of 9-hydroxy-4-phenylpyrrolo[3,4-*c*]-carbazole-1,3(2*H*,6*H*)-dione (colored by atom type: O red, N blue, H white, C yellow)/topopyrone C (colored by atom type: O red, N blue, H white, C cyan) in docking with wee1-kinase (PDB code = 1X8B). (b) Superimposition of topotecan (colored by atom type: O red, N blue, C violet)/topopyrone C (colored by atom type: O red, N blue, H white, C cyan) in docking with topoisomerase I (PDB code = 1K4T).



**Figure 2.6** Interactions of the crystallized ligand of abl2 (colored by atom type: F green, O red, N blue, H and C grey) with nNos.

For the crystallized ligand of abl2, two receptors were selected, namely, its receptor abl2 and nNos (PDB code = 3JT4). The structures obtained from Vina calculations for the abl2 ligand were very well superimposed to the crystallized compound, finding the same interactions with the receptor reported in [www.pdb.org](http://www.pdb.org) (Ligand Explorer).

On the other hand, the good interaction of the same ligand with the receptor nNos (Figure 2.6) is due to common interactions with its known ligand (*N*-5-[(3-(ethylsulfanyl)propanimidoyl)-l-ornithine]).<sup>135</sup> The molecule establishes van der Waals interactions with Phe584, Glu592, and Tyr588, and one H-bond with Gln478, as reported in the literature for the crystallized ligand, and is very well accommodated in the pocket of the receptor. Also, the structure obtained from Vina calculations for FTase ligand<sup>136</sup> is well superimposed to the crystallized compound for this receptor, displaying analogous interactions.

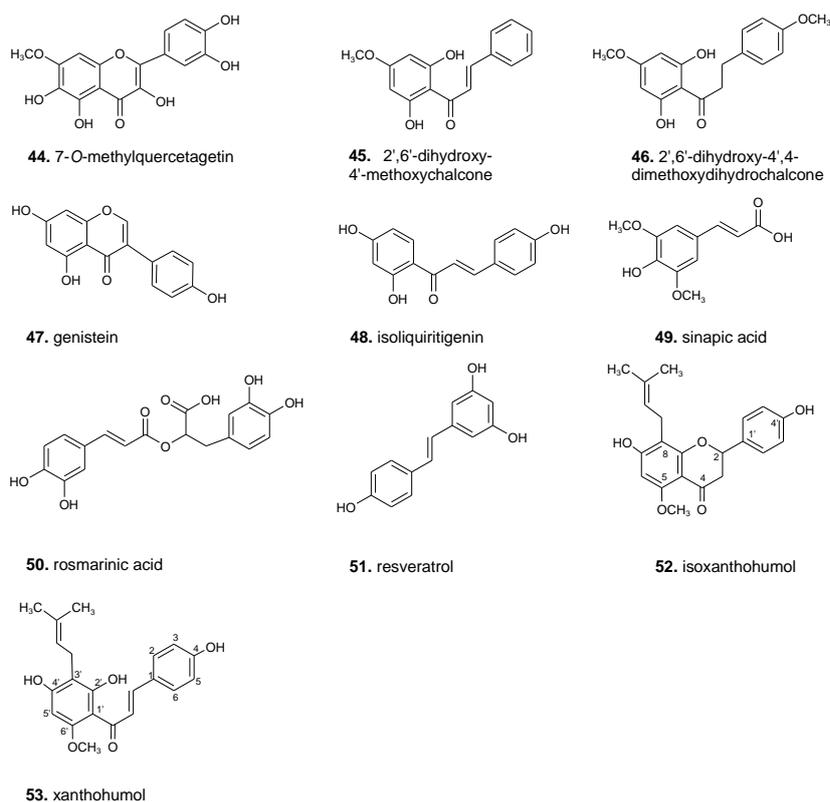
The data discussed above are useful to confirm the validity of the proposed computational method, and, besides the interactions with experimentally known targets, apparently discordant results are justifiable through careful analysis of the

observed interactions and can open the way for the discovery or further inspections of the targets of interaction of bioactive compounds.

## 2.3 Re-evaluation of the biological activity of a small library of natural compounds

Starting from the encouraging results on the first pilot Inverse Virtual Screening study, we proposed an application of the Inverse Virtual Screening on a small set of phenolic natural compounds tested on a panel enlarged to 163 targets involved in the cancer processes (Table 2.6).

The small library of molecules used in this second study consisted of 10 compounds with a certain variability of the scaffolds (Chart 2.2) and included compounds extracted from various plants and widely examined for their different pharmacological actions.



**Chart 2.2** The library of 10 natural compounds used for the screening.

protein	PDB code	protein	PDB code	protein	PDB code
abl	2HYY	epsilon (14-3-3)	2BR9	mmp13	830C
abl2	3HMI	erbB2	1S78	mrp1	2CBZ
aif	1M6I	erbB4	2R4B	msk1	3KN5
akt1	3MVH	erk1	2ZOQ	mthfs	3HY3
akt2	3D0E	erk2	2OJG	mTor	3FAP
alk	2XP2	fak	3BZ3	mtsp1	3NCL
alk5	2WOU	fgf1	1HKN	nek2	2XKF
ape1	2ISI	fgfr1	1AGW	nek7	2WQN
aurkin	2W1H	fgfr2	2PVF	nnos	3JT4
aurkin B	2VGO	ftase	1LD8	nqo1	2F1O
bap1	2V15	galectin1	1W6M	p300	3BIY
bcl2 <sup>a</sup>	2O21	galectin3	1KJL	p38	3HEG
bcl2a1	2VM6	galectin7	3GAL	p53_mut	2X0V
bcl6	3LBZ	gamma (14-3-3)	2B05	parp <sup>a</sup>	2JVN
bclw <sup>a</sup>	1ZY3	gsk3	3F7Z	pcaf <sup>b</sup>	2RNW
bclxl <sup>a</sup>	1BXL	gstt2_2	3GUR	pcna	1VYJ
braf	3C4C	gstp1	2A2R	pd	3BIK
btik	3PIX	hdac1	homology modelling	pdk1	3NAX
calmodulin	3EWT	hdac2	homology modelling	peroxiredoxin hORF6	1PRX
camKIIIB	3BHH	hdac3	homology modelling	pgm	1YFK
caspace1	2FQQ	hdac4	homology modelling	pi3k	3ENE
caspace2	1PYO	hdac6	homology modelling	pig3	2J8Z
caspace3	3EDQ	hdac7	homology modelling	pimKin	3JYA
caspace7	1SHL	hdac8	homology modelling	pka	3L9L
caspace8	1QTN	hdac8b	homology modelling	pkca	3IW4
cathepsin B	1GMY	hgfr (c-MET alternat)	2WD1	pkcθ	2JED
cathepsin G	1AU8	hmt	3HNA	plk1	3FVH
cathepsin K	2R6N	hsc70	3FZH	pnk	2W3O
cathepsin L	3HWN	hsp90	2WI6	pop	3DDU
cbp <sup>a</sup>	2RNY	hspa1a1	3JXU	pten	1D5R
cdk2	2VV9	hspa1l	3GDQ	pyk2	3FZS
cdk6	2F2C	hspa2	3I33	raf	3IDP
cdk7	1UA2	hspa5	3IUC	ret	2X2K
cdk9	3BLQ	hspa6	3FE1	srpk	1WBP
chk1	2QHN	ido	2D0T	stratifin (14-3-3-sigma)	1YWT
chk2	2W7X	igf	3F5P	survivin	1XOX
ciap1	3D9U	irak4	2NRU	syk	3FQH
ck2	3FL5	jak1	3EYG	tank1	2RF5
clk1	1Z57	jak2	3.00E+64	tdp1	1RFF
clk3	2WU6	jak3	1YVJ	theta (14-3-3)	2BTP
cmet	2WGJ	jmjd3	2XXZ	tie2	2O08
CPA	1CBX	jmjd3_akg	2XXZ	topI	1K4T
CPU	3D67	kit-kinase	3G0E	topII	1ZXM
cSRC	3F3V	kras	3GFT	topII_atp	1QZR
cxcr4 <sup>a</sup>	2K05	lck	3AD4	tp	1UOU
cyclin A (cdk2_altern)	2WFY	lsd1	2EJR	ts	1HVY
dapk	3EH9	lyn	2ZVA	upa	2VIP
dhfr	1PD8	mc11	3D7V	vegfr1	1FLT
diamineox	3HIG	mdm2	3EQS	vegfr2	3EWH
dnmt3a	3A1B	mdmx	3EQY	wee1	1X8B
dnmt3l	2QRV	mek1	3DV3	xiap	1TFT
e-cadherin	2O72	metap2	1YW9	zap70	1U59
egfr	2J6M	mk2	3M42	zeta (14-3-3)	1QJA
EPHa3	3DZQ	mmp3	1HY7		
EPHb4	2X9F	mmp8	2OY2		

Table 2.6 The panel of 163 targets used

The biological properties of these compounds were recognized and their involvement in metabolic disorders,<sup>137</sup> oxidative damage events,<sup>138</sup> atherosclerosis<sup>139</sup> and also in cancer prevention<sup>140,141,142</sup> was largely demonstrated. The bioactivity of several of these compounds was noteworthy and in some cases the specific targets of interactions were already known. Moreover, the complexity of the pharmacological activity of a given compound could be often explained considering its capacity of interacting with more than one target.

The chemopreventive activity of most of the compounds featuring these scaffolds was largely investigated and often proven by testing them on several cancer cell lines.<sup>143,144</sup> This capacity was mainly attributed to their well-known antioxidant properties.<sup>145</sup> However, many different actions on several targets involved more directly in the progression of this pathology were also shown.<sup>146,147,148</sup>

In particular, compounds **45**, **46**, **48** and **53** belong to the class of chalcones, for which antioxidant, chemopreventive, anticancer, anti-inflammatory, antifungal and antibacterial activities are reported.<sup>149</sup>

A number of chalcone derivatives have also been found to inhibit several important enzymes in cellular systems, including xanthine oxidase, aldose reductase, heme oxygenase, protein tyrosine kinase, quinine reductase and tyrosinase.<sup>141</sup> Among chalcone derivatives, the prenylated chalcone xanthohumol (**53**) and its derivative isoxanthohumol (**52**) occurring in the cones of *Humulus lupulus*, have attracted a lot of attention because of their biological activities, among which a broad-spectrum antiinfective activity against several microorganisms.<sup>150</sup> Xanthohumol (**53**) has been shown to inhibit the initiation, promotion, and progression stages of carcinogenesis, hence behaving as a potential broad-spectrum chemopreventive agent.<sup>151</sup>

Recently, it was demonstrated that xanthohumol (**53**) decreases the viability of the T98G human malignant glioblastoma cell line.

Apoptosis induced by xanthohumol is associated with activation of caspase-3, caspase-9, and PARP cleavage and is mediated by the mitochondrial pathway, as exemplified by mitochondrial depolarization, cytochrome c release, and downregulation of the antiapoptotic Bcl-2 protein. Moreover, xanthohumol induces intracellular reactive oxygen species (ROS), which provides a specific environment that results in MAPK-induced cell death.<sup>152</sup>

The small library of selected antioxidative phenolic compounds comprises also sinapic acid (**49**), a cinnamic acid derivative, with anti-anxiety properties,<sup>153</sup> rosmarinic acid (**50**) exerting antiinflammatory, antimutagen, antibacterial and antiviral activities<sup>154</sup> and resveratrol (**51**), the well known natural phytoalexin found in considerable amounts in the skin of grapes. In the last years resveratrol (**51**) has received a lot of attention because of its biological activities, as antimutagenic, antiviral, antiinflammatory, and cancer preventing. In particular it is believed that because of its antioxidant properties, resveratrol is responsible for the reduced risk of cardiovascular disease associated with a moderate consumption of red wine.<sup>155,156</sup> Moreover the isoflavone genistein (**47**) showed a topoisomerase<sup>149,157</sup> and tyrosine kinase<sup>158</sup> activity.

By considering the wide spectrum of activity of these compounds, a re-evaluation of their biological properties may further clarify their modulatory activity in the cancer events. On the other hand, the phases of extraction and purification imply small quantities of compounds from natural sources and make complex the performance of biological tests on more than one target. In this context, Inverse Virtual Screening represents a useful tool for the re-evaluation and/or the identification of the specific interactions of the library of compounds here considered.

Using the Inverse Virtual Screening approach we screened the library of 10 compounds on a panel of 163 targets involved in the cancer progression and collected the results in a matrix. Then we normalized the predicted binding energies, using the Equation 2.5, as previously described:

$$V = V_0 / [(M_L + M_R) / 2]$$

**Equation 2.5**

where  $V$  is the normalized value associated to each compound,  $V_0$  is the value of predicted binding energy obtained from the docking calculation (kcal/mol),  $M_L$  is the average binding energy of each ligand (on different targets, kcal/mol), and  $M_R$  is the average binding energy associated to each target (on the various ligands, kcal/mol) (Table 2.7).<sup>159</sup>

It is noteworthy that  $V$  is an absolute number. This mathematical manipulation causes the loss of the original significance of the binding energy as a value for the prediction of activity of a given compound. The normalized values can be used to generate a ranking in which the best values represent a promising interaction between a compound and a target from the panel. Moreover, taking into account the average trends in the Equation 2.5 the selection of false positive results can be avoided,<sup>113</sup> reminding that the main aim of this study is the identification of the targets interacting with a compound.

It is important to underline that the correspondence between the predicted and the calculated binding energies is much more difficult with respect to a classical Virtual Screening, in which only one target is studied, and mainly for two reasons. First, the comparison of the results for several targets even if normalized reduces but does not completely eliminate the problem of the variability of the interacting binding sites. In the second place, Autodock-Vina is a very fast and accurate software for the docking calculations, but in some cases a sensible deviation from the experimental results could be observed in the prediction of the binding

energies. This could depend in a variable way also by the number of active rotatable bonds of the investigated compounds.<sup>28</sup>

The selection of the best results was so conducted sorting and analyzing the normalized results of the screening from the best to the worse value. We observed that the best two normalized results highlighted the correlation between isoxanthohumol (**52**) with PKC- $\alpha$ <sup>160</sup> (Protein Kinase C  $\alpha$ , V value=1.286, position nr.1 in the final ranking on 1630 total calculations) and xanthohumol (**53**) with PDK1<sup>161</sup> (phosphoinositide-dependent kinase 1, V value=1.264, position nr.2 in the final ranking).

Moreover, the presence in the panel of the PKC- $\theta$ <sup>162</sup> isoform, characterized by a binding site related to PKC- $\alpha$ ,<sup>163</sup> prompted us to explore the behavior of isoxanthohumol (**52**) with this target. Interestingly, the normalized result (V=1.213) is at the significant position number 14. On the other hand the normalized results for **53** indicated a poor value of V for PDK1 (V=0.982, position number 921).

	44	45	46	47	48	49	50	51	52	53	M <sub>R</sub>
abl	-8.4	-8.5	-9	-9.8	-9.1	-6.7	-8.9	-9.1	-7.6	-9	-8.6
abl2	-8.3	-8.2	-7.7	-8.3	-8.8	-7	-9.8	-8.8	-6.9	-9.1	-8.3
aif	-8.7	-8.6	-8.5	-8.4	-8.7	-6.7	-8.7	-8.2	-9.1	-9.6	-8.5
akt1	-8.5	-8	-7.8	-7.8	-8.2	-6.3	-8.3	-7.4	-8.7	-8.3	-7.9
akt2	-8.1	-8.1	-7.6	-8.2	-8	-6.4	-8.7	-7.5	-8.6	-7.7	-7.9
alk	-7.4	-6.7	-6.2	-7.7	-7.1	-5.6	-7.2	-6.6	-7.7	-7.5	-7
alk5	-8.1	-7.3	-7.7	-8.3	-8.2	-6.4	-8.8	-7.8	-8.5	-8	-7.9
ape1	-6.9	-6.4	-6.1	-7.3	-6.9	-5.5	-6.8	-6.8	-7.1	-6.6	-6.6
aurkin	-9	-8	-8	-8.7	-8.4	-6.7	-8.7	-8.5	-9.7	-8.5	-8.4
aurkinB	-7.7	-7.2	-7.1	-7.8	-7.7	-5.7	-7.1	-7	-7.8	-7.3	-7.2
Other targets											
M <sub>L</sub>	-7.7	-7.2	-7.1	-7.6	-7.5	-5.9	-7.8	-7.1	-7.8	-7.6	

**Table 2.7** Predicted binding energies (V<sub>0</sub>, kcal/mol), M<sub>R</sub> and M<sub>L</sub> average values for a sample of 10 ligands on 30 targets for the calculations of the V normalized values.

For what concerns **53**, while the normalized value and the related position in the ranking showed a good predicted activity for PDK1, poor results are observable for PKC- $\alpha$  and PKC- $\theta$  (positions 892 and 563, respectively). In Table 2.8 the values of V and the positions in the final ranking for **52** and **53** with PDK1, PKC- $\alpha$  and PKC- $\theta$  are reported.

These targets identified from the screening play a fundamental role in the progression of the tumour events,<sup>164,165,166</sup> and new inhibitors are searched for the development of new anti-cancer agents. For these reasons, in the next phase we validated these preliminary observations with in vitro biological tests (Table 2.9).

	PDK1	PKC- $\alpha$	PKC- $\theta$
<b>52</b>	0.982 ( <b>921</b> )	1.286 ( <b>1</b> )	1.213 ( <b>14</b> )
<b>53</b>	1.264 ( <b>2</b> )	0.986 ( <b>892</b> )	1.038 ( <b>563</b> )

**Table 2.8** V normalized values for **52** and **53** on PDK1, PKC- $\alpha$  and PKC- $\theta$  targets. On parenthesis are shown the relative positions in the final ranking on the 1630 total calculations.

	PDK1	PKC- $\alpha$	PKC- $\theta$
<b>52</b>	58.7 $\mu$ M	45.3 $\mu$ M	31.6 $\mu$ M
<b>53</b>	6.6 $\mu$ M	> 100 $\mu$ M	> 100 $\mu$ M

**Table 2.9** IC<sub>50</sub> values for **52** and **53** on PDK1, PKC- $\alpha$  and PKC- $\theta$  targets.

In order to validate the efficacy of the screening also negative controls were considered in this phase (**52** on PDK1, **53** on PKC- $\alpha$  and PKC- $\theta$ ). Regarding the most promising results, we observed that these compounds confirmed our predictions, showing an inhibitory activity in the  $\mu$ M range. In more detail, xanthohumol (**53**) show a best activity and selectivity on PDK1 (6.6  $\mu$ M). Moreover, IC<sub>50</sub> value calculated for **53** on PKC- $\alpha$  and  $\theta$  isoforms show no inhibitory activity at concentrations as high as  $10^{-4}$  M and this is in perfect

agreement with the normalized results from the screening. Isoxanthohumol (**52**) is active (less than the previous result) and not selective on PKC- $\alpha$  and  $\theta$  confirming the selection of the related good positioning from the final ranking. Surprisingly a moderate activity of **52** on PDK1 was highlighted, in spite of the low  $V$  value (0.982) and of the low position in the ranking (921). It is important to observe that the best  $V$  value found is 1.286, and this means that in this case 921 interactions out of 1630 total calculation are found within a restricted difference of  $V$  values of 0.304.

Firstly, we wondered whether this lack of sensibility of the approach to avoid a false negative result was caused by the normalization of the affinities. In this context, the range from the best to the worse value for the normalized binding energies and for the predicted binding energies before the normalization (Range $V$  and RangeBE, respectively, Equation 2.6) was considered:

$$\text{Range}V = 1.286 - 0.546 = 0.74; \text{Range}BE = -10.7 \text{ kcal/mol} - (-4.2 \text{ kcal/mol}) = -6.5 \text{ kcal/mol}$$

**Equation 2.6**

For the interaction of **52** on PDK1, we calculated the deviations of the  $V$  value and of the predicted binding energy from the best ones (Equation 2.7).

$$\text{Dev}V_{52\text{-PDK1}} = V_{\text{best}} - V_{52\text{-PDK1}}; \text{Dev}BE_{52\text{-PDK1}} = BE_{\text{best}} - BE_{52\text{-PDK1}}$$

**Equation 2.7**

Where  $V_{\text{best}}$  and  $BE_{\text{best}}$  are the best values of  $V$  and predicted binding energies for the 1630 total calculations, and  $V_{52\text{-PDK1}}$  and  $BE_{52\text{-PDK1}}$  are the  $V$  and the predicted binding energies values for the interaction of **52** with PDK1.

Then we divided the these two deviations for the two ranges for the two types of calculation (Equation 2.8).

$$\text{Ratio}V_{52-PDK1} = \frac{\text{Dev}V_{52-PDK1}}{\text{Range}V}; \text{Ratio}BE_{52-PDK1} = \frac{\text{Dev}BE_{52-PDK1}}{\text{Range}BE}$$

**Equation 2.8**

These parameters indicated the nearness of the two values for the interaction of **52** with PDK1 to the best ones in the two rankings. We found two very similar values ( $\text{Ratio}V_{52-PDK1} = 0.41$ ;  $\text{Ratio}BE_{52-PDK1} = 0.40$ ). We can conclude that this unexpected result strongly depends from the original value of predicted binding energy.

Moreover, it is important to note that the evaluation of the binding energies before the normalization do not allow the selection of this false negative. Furthermore, we can observe a low value of  $\text{Dev}BE_{52-PDK1}$  that corresponds in an overcrowding of many values better than  $BE_{52-PDK1}$  in a restricted range, as we have initially observed for the normalized values.

This means that low variations of the values of predicted binding energies may cause large variations in the two final rankings. For these reasons, in an Inverse Virtual Screening study the choice of the parameters that can affect the estimation of the calculated affinities (i.e the protein preparation, the exhaustiveness values and the grid boxes for Autodock-Vina) is of primary importance.

Accordingly, as we have previously demonstrated that in other case,<sup>113,115</sup> the normalization could be a useful tool to avoid false positive and negative results.

In order to give a further validation the method, we performed a similar study using another library of 10 compounds able to bind 10 targets in the panel of 163 targets with high efficacy. We chose these compounds considering the availability of the crystallographic structures of the complexes with the partner receptors.

Also in this case we obtained a matrix of 1630 calculations and then we normalized them. In Table 2.10 are listed the V values, the positions in the ranking, the experimental and predicted values of  $IC_{50}$  for the 10 compounds with their specific targets.

Reference ligands	V (positions in the ranking)	Exper IC <sub>50</sub> *	Predicted affinities
ABL_lig	1.421 (4)	170.0 nM	0.35 nM
ALK5_lig	1.145 (168)	72.0 nM	65 nM
CK2_lig	1.446 (3)	52.0 nM	5.2 nM
CLK3_lig	1.062 (451)	29.2 nM	129 nM
JAK1_lig	1.097 (304)	1.6 nM	497 nM
PDK1_lig	1.585 (1)	1.0 nM	0.0085 nM
PKC- $\alpha$ _lig	1.230 (38)	2.1 nM	6.2 nM
PKC- $\theta$ _lig	1.341 (8)	na	1.35 nM
RAF_lig	1.495 (2)	1.6 nM	0.055 nM
TP_lig	1.115 (242)	20.0 nM	1160 nM

**Table 2.10** V values, positions in the ranking, the experimental IC<sub>50</sub> and predicted affinities for the 10 reference compounds with their specific targets; \*Values extracted from the papers related to the PDB codes listed in Table 2.6

For the listed compounds Autodock-Vina found the crystallographic poses with high accuracy (RMSD < 2 Å), and so the predicted binding energies could be related to the experimental ones. Also in this case, using the Inverse Virtual Screening approach we can highlight that four right correlations between ligands and targets are found in the first four positions, a fifth is identified in the first eight positions, and a sixth in the first thirtheight.

The remaining four correct correlations are found far from the high ranking positions, so resulting as false negative mainly for their lower values of predicted binding energies.

The availability of the ligands crystallized in the binding sites of the targets PDK1, PKC- $\alpha$ , and PKC- $\theta$  and active in the nM range allowed a more precise comparison with the V values calculated for the compounds **52** and **53** emerging from the screening. We built a third matrix for the 163 targets interacting with **52**, **53**, ( $\mu$ M range of activity on the targets identified) **PDK1\_lig**, **PKC- $\alpha$ \_lig**, and **PKC- $\theta$ \_lig** (nM range of activity on their related targets).

In Table 2.11 are reported the V values for **52** and **53** on the three targets compared with the V values of the reference compounds.

	<b>52</b>	<b>53</b>	<b>Reference compounds</b>
<b>PDK1</b>	0.898	1.154	1.529
<b>PKC-<math>\alpha</math></b>	1.147	0.878	1.235
<b>PKC-<math>\theta</math></b>	1.050	0.897	1.291

**Table 2.11** V values for **52**, **53**, **PDK1\_lig**, **PKC- $\alpha$ \_lig** and **PKC- $\theta$ \_lig** on PDK1, PKC- $\alpha$  and PKC- $\theta$  targets.

In this way a more precise and accurate correspondence between the V values and the experimental IC<sub>50</sub> was found. Above all, these new scale of values also justified the moderate activity of the compounds **52** and **53** on the three kinases.

In summary, while the normalization method could reduce the possibility of selection of false positive and negative results, the comparison with the the V values of reference compounds could give qualitative indications for the prediction of the range of activity for a given set of compounds. In particular, for compounds **44-53**, five predictions of activity on six actually fitted with the biological tests.

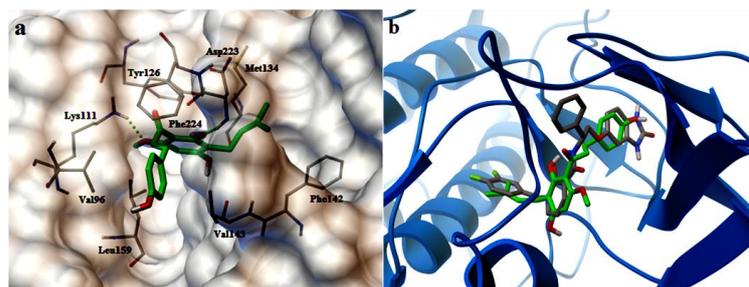
Finally, an accurate analysis of the main interactions of the compounds selected from the screening in the binding sites of the different targets allows an explanation of the different activity and selectivity of the different compounds on the panel of targets used. The best pose from the docking calculations for each compound on the three specific targets was considered in this phase.

Regarding the xanthohumol (**53**) case we observed a very good occupancy of the binding site of PDK1. As illustrated in Figure 2.7 a, **53** is able to establish a set of hydrophobic and electrostatic interactions in the binding pocket composed by Val96, Tyr126, Met134, Phe142, Val143, Leu159, and Phe224. It also involved in two Hydrogen bonds with the nitrogen in the side chain of Lys111 (with the oxygen of the -OCH<sub>3</sub> in 6' position) and with the nitrogen in the backbone of the Asp223 (with the -OH in 2' position). Moreover, the isopentenyl portion of the compound is accommodated in an internal hydrophobic pocket of PDK1. A good

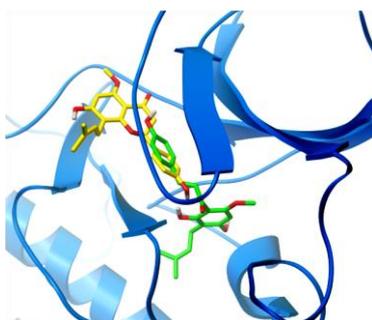
superposition between **53** and 1-(3,4-difluorobenzyl)-2-oxo-N-[(1R)-2-[(2-oxo-2,3-dihydro-1H-benzimidazol-5-yl)oxy]-1-phenylethyl]-1,2-dihydropyridine-3-carboxamide crystallized in the binding site of PDK1<sup>161</sup> could be considered as a further confirm (Figure 2.7 b). The two phenyl portions of **53** are overlapped to two aromatic portions of the crystallized ligand, respecting their distances and orientations. For what concerns the interactions between isoxanthohumol (**52**) and PDK1, we can primarily observe a most external occupancy of the binding site than the xanthohumol (**53**) and an inversion of the aromatic ring presenting the isopentenyl part, that in this case is oriented on the external face of the binding site. The different accommodation of the two compounds and of their pharmacophoric portions in a more deep position of the PDK1 binding site could explain the different activities on this target (Figure 2.8). The binding pocket of PDK1 involved in the interactions with **52** is composed by Leu88, Gly89, Ala109, Val143, Leu159, Glu166, Leu212, Thr222 and Phe224. A Hydrogen bond is established between the oxygen of the carbonyl in position 4 of **52** and the nitrogen in the backbone of Ala162 (Figure 2.9a). The analysis of the docked structures of **52** with the PKC ( $\alpha$  and  $\theta$  isoforms) revealed the accommodation of the compound in the binding sites mainly through hydrophobic interactions. In fact, in the case of the  $\alpha$  isoform the compound interacts with the following residues: Leu345, Phe350, Val353, Ala366, Thr401, Met417, Tyr419, Val420, Ala480 and Asp481. A hydrogen bond is also observable between the nitrogen in the backbone of Val420 and the oxygen in position 4 of the carbonyl of **52** (Figure 2.9 b). On the other hand, the best binding pose in the binding site of PKC- $\theta$  isoform show the involvement of a pocket of residues composed by Leu386, Phe391, Val394, Ala407, Thr442, Met458, Ala521, Asp522 and Phe664. Also in this case, a hydrogen bond is established between the nitrogen in the backbone of the Leu461 and the 4-carboxylic oxygen of **52** (Figure 2.9c).

We considered this theoretical hypothesis of binding modes reliable for two reasons. In the first place, the good result from the biological tests could be considered as an experimental proof of the binding of **52** and **53** in these specific sites of action. Moreover, these compounds show a simple chemical structure mainly characterized by a small number of aromatic rings and of active rotatable bonds.

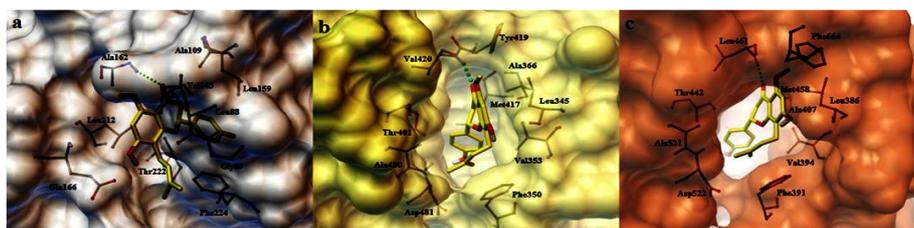
In these conditions Autodock-Vina software shows a high accuracy of the prediction of the experimental crystallographic poses, with low RMSD values compared to the predicted ones. For these reasons these results could be considered as reasonable models of binding.



**Figure 2.7** a) **53** (coloured by atom types: C green, H grey, O red) in docking with PDK1. Hydrogen bonds are displayed with green spheres; b) Superimposition between **53** (coloured by atom types: C green, H grey, O red) and 1-(3,4-difluorobenzyl)-2-oxo-N-[(1R)-2-[(2-oxo-2,3-dihydro-1H-benzimidazol-5-yl)oxy]-1-phenylethyl]-1,2-dihydropyridine-3-carboxamide (coloured by atom types: C dark grey, H grey, O red, N blue) with PDK1.



**Figure 2.8** Superimposition between **52** (coloured by atom types: C yellow, H grey, O red) and **53** (coloured by atom types: C green, H grey, O red) in the binding site of PDK1.



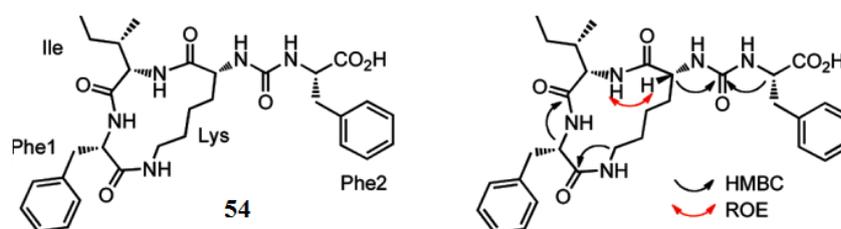
**Figure 2.9** **52** (coloured by atom types: C green, H grey, O red) in docking with: a) PDK1; b) PKC- $\alpha$ ; c) PKC- $\theta$ . Hydrogen bonds are displayed with green spheres.

## 2.4 Discovery of peptidase inhibitory activity of the new-anabaenopeptin cyclopeptide namalide.

The diversity of natural products coming from marine invertebrates and possessing interesting biological activities has been well-documented, yet discovery of new compounds and their associated modes of action continues apace.<sup>167</sup> To mention only a few examples, marine natural products that can inhibit biological targets linked to cancer,<sup>168,169,170,171</sup> inflammation,<sup>172</sup> and bacterial cell division<sup>173</sup> have been reported. Here the identification and Fmoc-based solid phase synthesis of a new ureido-containing cyclic peptide, namalide (**54**), and its potent inhibitory activity on carboxypeptidase A (CPA) is described. Namalide was isolated from the same collection of the marine sponge *Siliquariaspongia mirabilis* that provided the anti-HIV lipopeptides mirabamides A–D,<sup>174</sup> the antitumor polyketide mirabalin,<sup>175</sup> and the known antifungals aurantosides A and B.<sup>176</sup> Namalide represents a new anabaenopeptin-type scaffold possessing a 13-membered macrolactam core.

Previously, the discovery of three separate classes of natural products coming from the aqueous extract of a single collection of the marine sponge *S. mirabilis* was reported.<sup>174,175,176</sup> Mirabamides, mirabalin, and aurantosides were isolated from the n-BuOH fraction of the aqueous extract after fractionation on a Sephadex LH-20 column eluting with MeOH, followed by HPLC purification. During these studies, a separate group of Sephadex fractions containing what appeared to be a novel compound on the basis of its molecular weight (ESI-MS) was detected. In addition, these fractions showed strong inhibition of the enzyme CPA. Active fractions were combined and purified by C12 RP-HPLC to give just 0.4 mg of the active compound (~90% pure), a new natural product named namalide (**54**, Figure 2.10). Combining HR-ESI-MS and 2D NMR experiments (HSQC, HMBC, HOHAHA, ROESY and DQF-COSY) a structural characterization of **54** was

determined. In particular, the molecular formula requires **54** to be a cyclic peptide and ROESY experiments indicated that **54** comprised a three-residue macrocycle N-linked to an exocyclic Phe residue via a ureido bridge.

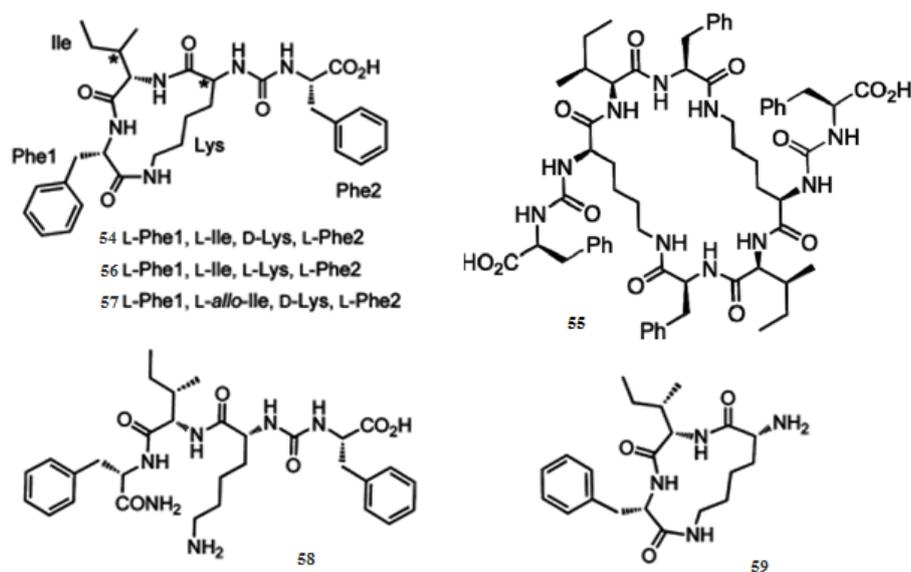


**Figure 2.10** Structure of namalide (**54**) (left) and key HMBC and ROE correlations (right).

The absolute configurations of the alpha carbons for both Phe residues and Ile were readily established as L from LC-MS analyses of the L- and D-FDLA (1-fluoro-2,4-dinitrophenyl-5-L/ D-leucinamide) derivatives<sup>177</sup> of an acid hydrolysate of **54**, in conjunction with comparison of retention times with authentic standards. However, because of our limited material or the limitations of the method, this analysis did not allow us to assign confidently the configuration of the Lys residue and C-3 of Ile. It is worth noting that although all anabaenopeptin-type compounds derived from cyanobacteria contain D-Lys exclusively,<sup>178,179</sup> peptides reported to contain L-Lys rather than D have been isolated from marine sponges.<sup>180,181,182,183</sup> To address these unknowns and provide additional compound for biological screening, a synthetic route to provide namalide and its stereoisomers was developed.

To establish configurations at C-3 of Ile and C-2 of Lys, two additional namalide analogs containing L-Ile/L-Lys and L-allo-Ile/D-Lys (**56**, **57**, Figure 2.11) were synthesized. Comparisons of the NMR spectra and RP-HPLC retention times of the three synthetic peptides **54**, **56**, and **57** with that of the natural product

showed that the isomer containing D-Lys and L-Ile corresponded to the natural material.



**Figure 2.11** Synthetic namalide analogs.

Members of the anabaenopeptin<sup>184</sup> family of cyclic peptides are known to inhibit carboxypeptidases and other proteases.<sup>185,186,187,188</sup> Although a high resolution structure of an anabaenopeptin-type peptide in complex with any carboxypeptidase has not been determined, selectivity profiles of natural products and synthetic analogs provide good evidence that the C-terminal ureido amino acid confers specificity toward different proteases.<sup>188,189</sup> Peptides **54**, **55**, and **56–59** (Figure 2.11) were evaluated as inhibitors of bovine pancreas CPA using N-(4-methoxyphenylazofornyl)-phenylalanine as a colorimetric substrate in a 96-well plate format. The results are summarized in Table 2.12.

The synthetic version of natural namalide (**54**) containing D-Lys and L-Ile was the most potent inhibitor of CPA with an IC<sub>50</sub> value of 250 ± 30 nM. Peptide **56**, the corresponding L-Lys analog, was inactive at concentrations as high as 30 μM,

and analog **57** bearing L-allo-Ile/D-Lys appeared to be inactive, although its insolubility yielded poor assay results. The linear version of namalide, **58**, showed an 18-fold reduction in activity relative to **54** with an IC<sub>50</sub> value of 4.5 μM. In contrast cyclic tripeptide **59**, which lacks the C-terminal exocyclic ureido Phe, and the namalide dimer **55** were inactive.

For comparison, compounds **54**, **56**, **58**, and **59** were tested against carboxypeptidase U (CPU), also known as activated thrombin-activable fibrinolysis inhibitor (TAFIa), an enzyme known to recognize C-terminal basic amino acids in its S1 pocket, 25 and α-chymotrypsin, a serine protease that recognizes aliphatic and aromatic amino acids.<sup>190</sup>

None of these peptides inhibited either enzyme at concentrations as high as 60 μg/mL. Together, these results demonstrate the importance of the lysine configuration for CPA inhibition in this new tricyclic peptide scaffold. The lack of inhibitory activity of the des-ureido-Phe analog **59** further suggests that namalides may inhibit through a similar mode of binding as the larger anabaenopeptin-type compounds.

Peptide	IC <sub>50</sub> (μM)
<b>54</b>	0.25 ± 0.03
<b>55</b>	na <sup>b</sup>
<b>56</b>	na <sup>b</sup>
<b>57</b>	nt <sup>c</sup>
<b>58</b>	4.5 ± 0.9
<b>59</b>	na <sup>b</sup>

<sup>a</sup>Tested in duplicate; <sup>b</sup>not active at 30 μM;

<sup>c</sup>insoluble

**Table 2.12** Inhibition of carboxypeptidase A.

Because namalide represents a new natural product scaffold, we were interested in applying our recently described Inverse Virtual Screening in silico approach<sup>113</sup> using Autodock Vina software<sup>28</sup> to assist in identifying other possible new targets

of namalide. For docking calculations, a 3-D model of namalide was prepared by performing combined Monte Carlo conformational searches and Molecular Dynamics simulations, followed by energy and geometry minimization of the obtained structure. When this study was performed, the minimized model was used against a panel of 159 receptors involved in cancer processes and whose coordinates were extracted from the Protein Data Bank (Table 2.13).

For normalization of the results, a prebuilt matrix containing the affinity values of a library of 22 natural compounds (used as “blanks”) with structural and molecular weight properties similar to those of **54** was employed.

In particular, we calculated an average value of binding energy for each target receptor on all of the compounds in the matrix and then normalized the affinity values of 1 using the Equation 2.9:

$$V = V_0/V_R$$

**Equation 2.9**

where V is the normalized value of binding energy,  $V_0$  is the value of binding energy before the normalization, and  $V_R$  is the average value of binding energy for each targets. We were gratified to find that of the 159 proteins screened, CPA was identified as the third best hit on the basis of normalized binding energy (Table 2.14).

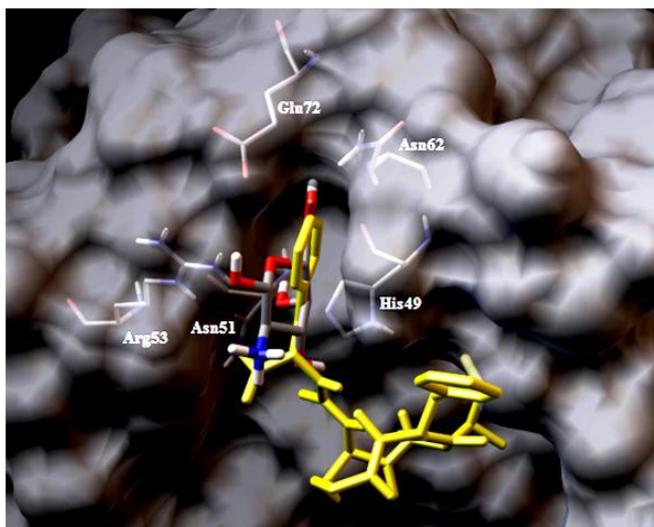
The top two hits corresponded to galectin 7 and calmodulin (CaM). Inspection of the docked models showed **54** to be located in the galactose binding site of galectin 7, anchored through the exocyclic Phe residue (Figure 2.12), and in an extensive hydrophobic channel in CaM, making only partial contacts with this site (Figure 2.13). Because these interactions were not optimal relative to the known ligands, we did not pursue them further.

protein	PDB code	protein	PDB code	protein	PDB code
abl	2HYY	egfr	2J6M	mk2	3M42
abl2	3HMI	EPHa3	3DZQ	mmp3	1HY7
aif	1M6I	EPHb4	2X9F	mmp8	2OY2
akt	3D0E	epsilon	2BR9	mmp13	830C
akt1	3MVH	erbB2	1S78	nrp1	2CBZ
alk	2XP2	erbB4	2R4B	msk1	3KN5
alk5	2WOU	erk1	2ZOQ	mthfs	3HY3
ape1	2ISI	erk2	2OJG	mTor	3FAP
aurkin	2W1H	fak	3BZ3	mtsp1	3NCL
aurkin B	2VGO	fgf1	1HKN	nek2	2XKF
bap1	2W15	fgfr1	1AGW	nek7	2WQN
bcl2 <sup>a</sup>	2O21	fgfr2	1EV2	nmos	3JT4
bcl2a1	2VM6	ftase	1LD8	nqo1	2F1O
bcl6	3LBZ	galectin1	1W6M	p38	3HEG
bclw <sup>a</sup>	1ZY3	galectin3	1KJL	p53_mut	2X0V
bclxl <sup>a</sup>	1BXL	galectin7	3GAL	parp <sup>a</sup>	2JVN
braf	3C4C	gamma	2B05	pcaf <sup>a</sup>	2RNW
btk	3PIX	gsk3	3F7Z	pcna	1VYJ
calmodulin	3EWT	gstm2_2	3GUR	pd	3BIK
camKIIb	3BHH	gstp1	2A2R	pdk1	3NAX
caspace1	2FQQ	hdac1	homology modelling	peroxiredoxin	1PRX
caspace2	1PY0	hdac2	homology modelling	pgm	1YFK
caspace3	3EDQ	hdac3	homology modelling	pi3k	3ENE
caspace7	1SHL	hdac4	homology modelling	pimKin	3JYA
caspace8	1QTN	hdac6	homology modelling	pka	3L9L
cathepsin B	1GMY	hdac7	homology modelling	plk1	3FWH
cathepsin G	1AU8	hdac8	homology modelling	pop	3DDU
cathepsin K	2R6N	hdac8b	homology modelling	pten	1D5R
cathepsin L	3HWN	hgr	2WD1	pyk2	3FZS
cbp <sup>a</sup>	2RNY	hmt	3HNA	raf	3IDP
cd20	3BKY	hsc70	3FZH	ret	2X2K
cdk2	2VV9	hsp90	2W16	srpk	1WBP
cdk6	2F2C	hspa1a1	3JXU	stratifin	1YWT
cdk7	1UA2	hspa11	3GDQ	survivin	2RAW
cdk9	3BLQ	hspa2	3I33	syk	3FQH
chk1	2QHN	hspa5	3IUC	tank1	2RF5
chk2	2W7X	hspa6	3FE1	tank2	3KR8
ciap1	3D9U	ido	2D07	tdp1	1RFF
ck2	3FL5	igf	3F5P	teta	2BTP
clk1	1Z57	irak4	2NRU	tie2	2O08
clk3	2WU6	jak1	3EYG	top1	1K4T
cmet	2WGJ	jak2	3.00E+64	topII	1ZXM
CPA	1CBX	jak3	1YVJ	topII_atp	1QZR
CPU	3D67	kit-kinase	3G0E	tp	1UOU
cSRC	3F3V	kras	3GFT	ts	1HVY
cxcr4 <sup>a</sup>	2K05	lck	3AD4	upa	2VIP
cyclin A	2Wfy	lsd1	2EJR	vegfr1	1FLT
dapk	3EH9	lyn	2ZVA	vegfr2	3EWH
dhfr	1PD8	mcl1	3D7V	wee1	1X8B
diamineox	3HIG	mdm2	3EQS	xiap	1TFT
dnmt3a	3A1B	mdmx	3EQY	xrcc1	2W3O
dnmt3l	2QRV	mek1	3DV3	zap70	1U59
e-cadherin	2O72	metap2	1YW9	zeta	1QJA

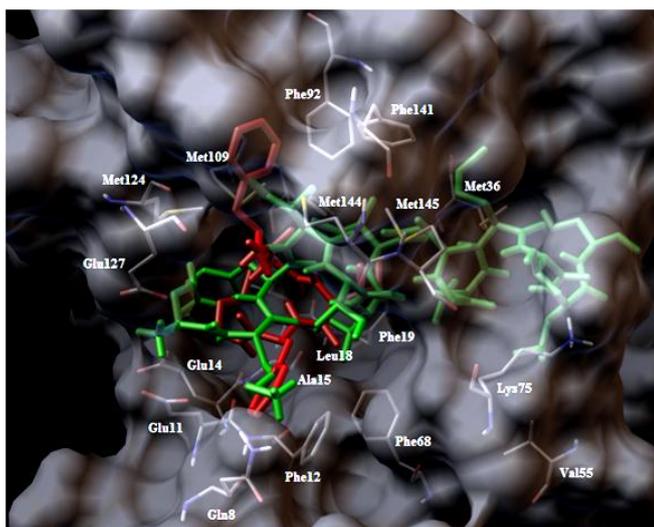
Table 2.13 The panel of 159 targets used in the study on namalide

Position in the ranking	Targets	V values	Position in the ranking	Targets	V values	Position in the ranking	Targets	V values
1	calmodulin	1.267	54	pd	1.047	107	peroxiredoxin	0.954
2	galectin7	1.221	55	EPHa3	1.045	108	tie2	0.950
3	CPA	1.218	56	galectin3	1.045	109	aurkin	0.948
4	bclw	1.193	57	cathepsinK	1.044	110	ck2	0.948
5	lsd1	1.190	58	hdac3	1.044	111	clk3	0.946
6	bcl6	1.182	59	erk2	1.037	112	egfr	0.942
7	bcl2a1	1.166	60	mrp1	1.037	113	top1	0.941
8	stratifin	1.159	61	nqo1	1.035	114	xrcc1	0.932
9	mcl1	1.153	62	plk1	1.032	115	mmp3	0.924
10	caspase7	1.150	63	mthfs	1.032	116	cyclinA	0.923
11	vegfr1	1.149	64	ape1	1.029	117	pi3k	0.921
12	bcll	1.149	65	caspase2	1.023	118	cathepsinB	0.921
13	mtor	1.137	66	fgf1	1.021	119	tank1	0.920
14	survivin	1.133	67	pcaf	1.018	120	cdk9	0.919
15	tdp1	1.132	68	lck	1.016	121	ts	0.917
16	caspase3	1.125	69	cSRC	1.011	122	hspa5	0.908
17	mdmx	1.118	70	hdac6	1.011	123	alk	0.908
18	hsc70	1.117	71	parp	1.009	124	cdk2	0.907
19	e-cadherin	1.116	72	nek7	1.008	125	pten	0.907
20	ido	1.113	73	tank2	1.008	126	galectin1	0.897
21	srpk	1.110	74	hspa1a	1.007	127	vegfr2	0.895
22	caspase1	1.110	75	hdac8b	1.005	128	clk1	0.894
23	kras	1.107	76	AKT	1.005	129	top1_atp	0.892
24	p38	1.095	77	dnmt3a	1.004	130	msk1	0.888
25	nmos	1.092	78	caspase8	1.003	131	zap70	0.886
26	xiap	1.090	79	topII	1.002	132	pgm	0.879
27	ftase	1.088	80	bap1	1.002	133	jak3	0.878
28	mtsp1	1.088	81	hdac2	0.999	134	gsk3	0.876
29	erbB2	1.087	82	zeta	0.999	135	jak2	0.875
30	cd20	1.081	83	bcl2	0.995	136	dhfr	0.874
31	gstm2_2	1.080	84	hspa6	0.994	137	EPHb4	0.870
32	tp	1.078	85	pyk2	0.993	138	ret	0.867
33	cathepsinL	1.076	86	cathepsinG	0.989	139	erk1	0.867
34	pop	1.076	87	p53_mut	0.988	140	fak	0.855
35	mdm2	1.075	88	epsilon	0.984	141	pimkin	0.845
36	aif	1.072	89	upa	0.983	142	camKIIB	0.844
37	lyn	1.070	90	cmet	0.983	143	hspa2	0.844
38	syk	1.069	91	raf	0.982	144	mmp8	0.842
39	ciap1	1.067	92	dnmt3l	0.982	145	fgfr1	0.839
40	hdac8	1.066	93	mk2	0.981	146	chk1	0.823
41	hdac7	1.066	94	pcna	0.975	147	irak4	0.809
42	hsp90	1.066	95	mmp13	0.973	148	cdk6	0.806
43	aurkinB	1.065	96	btik	0.973	149	CPU	0.800
44	teta	1.064	97	pdk1	0.970	150	wee1	0.794
45	hspa11	1.060	98	cxcr4	0.969	151	abl2	0.786
46	igf	1.059	99	diaminoox	0.968	152	dapk	0.772
47	akt1	1.056	100	cbp	0.968	153	jak1	0.767
48	gamma	1.054	101	metap2	0.965	154	erbB4	0.766
49	abl	1.053	102	nek2	0.964	155	braf	0.751
50	hdac4	1.053	103	gstp1	0.964	156	alk5	0.732
51	hdac1	1.051	104	hmt	0.963	157	chk2	0.695
52	fgfr2	1.050	105	cdk7	0.962	158	pka	0.624
53	kit_kinase	1.048	106	mek1	0.955	159	hgfr	0.582

Table 2.14 Normalized V values related to compound 54

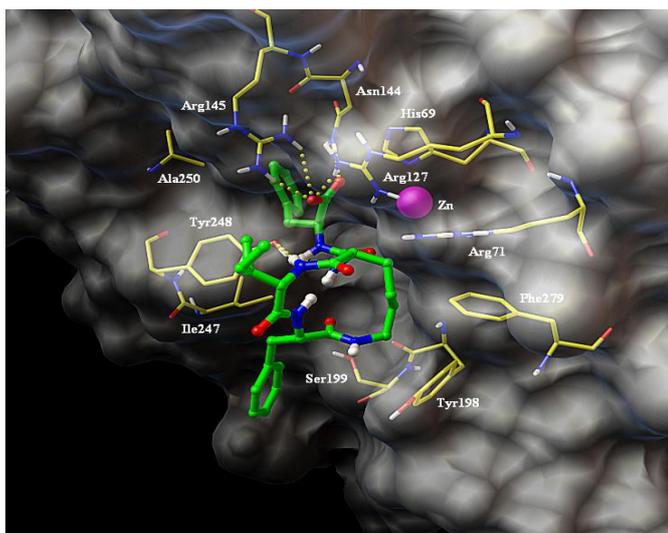


**Figure 2.12** Superposition between **54** (yellow) and galactosamine (colored by atom type: C grey, O red, N blue and H light grey) in the Galectin7 binding site of sugars.



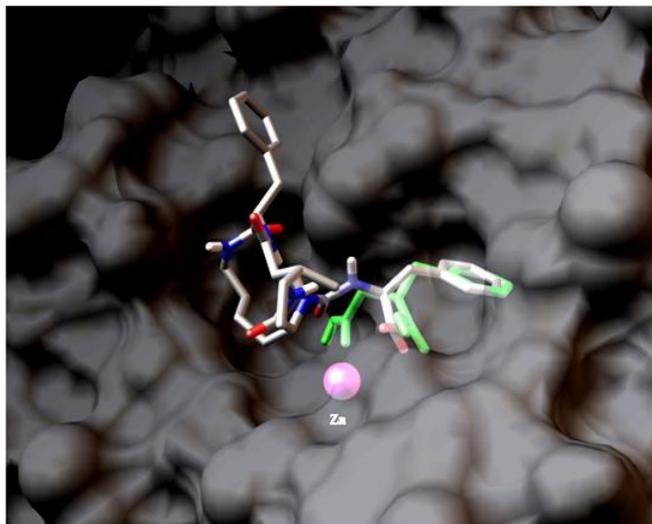
**Figure 2.13** Superposition between **54** (red) and a crystallized peptide (green) in the CaM binding domain.

As shown in Figure 2.14, a refined molecular docking model of **54** to CPA using Autodock 4.2<sup>27</sup> software places the C-terminal carboxylate in close proximity to the  $Zn^{2+}$  atom and within hydrogen bonding distance with the guanidine groups of Arg 127 and Arg 145, key residues in the active site and specificity pocket of CPA<sup>190,191</sup> as well as the side chain of Asn 144. Additional interactions are seen between the amide of Ile and the hydroxyl of Tyr 248. Similarly, the ureido group is involved in electrostatic interactions with Arg 127 and Tyr 248. Despite its reduced size relative to the pentapeptide core of anabaenopeptins, both electrostatic and aliphatic interactions are observed between CPA and residues in the cyclic portion of **54** in the docked model. In particular, the side chain of D-Lys is positioned close to Phe 279, and the amide of Ile is hydrogen-bonded to the -OH of Tyr 248.



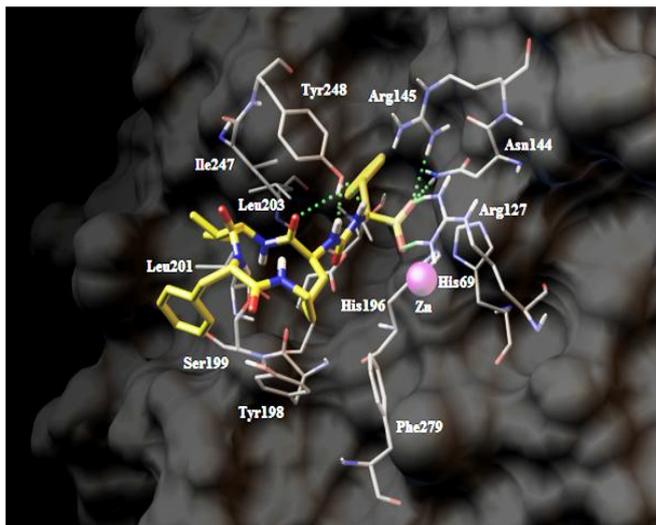
**Figure 2.14** Docked model of **54** to CPA. Protein is shown as a gray surface,  $Zn^{2+}$  atom is shown as a magenta sphere, and residues in contact with **54** are shown as yellow (C atoms), blue (N atoms), red (O atoms), and gray (H atoms) sticks; namalide (**54**) is rendered as green (C atoms), blue (N atoms), red (O atoms), and gray (H atoms) sticks and balls.

Consistent with other models and the biological activity, the C-terminal Phe is directed toward the hydrophobic specificity pocket formed in part by Ile247, Tyr 248, and Ala 250, and is well superimposed with the similar moieties in the chemical structure of the crystallized inhibitor L-benzylsuccinate (Figure 2.15).

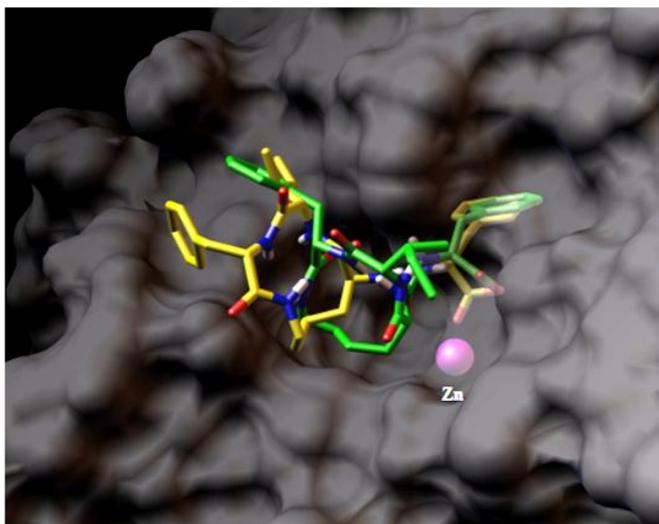


**Figure 2.15** Superposition between **54** (colored by atom type: C grey, O red, N blue and H light grey) and the crystallized inhibitor L-benzylsuccinate (green).  $\text{Zn}^{2+}$  is as a magenta sphere.

To investigate the specificity of namalide for CPA, we performed molecular docking of **56**, the L-Lys-containing analog, to CPA and of **54** to CPU using the same protocol as above. In the lowest energy model of **56** bound to CPA, interactions between the exocyclic Phe and its carboxylate group with CPA are preserved. However, the inverted configuration of Lys results in a flipping of the ring that moves the ureido group far from the  $\text{Zn}^{2+}$  atom and reduced interactions between the ring amino acids and CPA (Figure 2.16 and Figure 2.17).



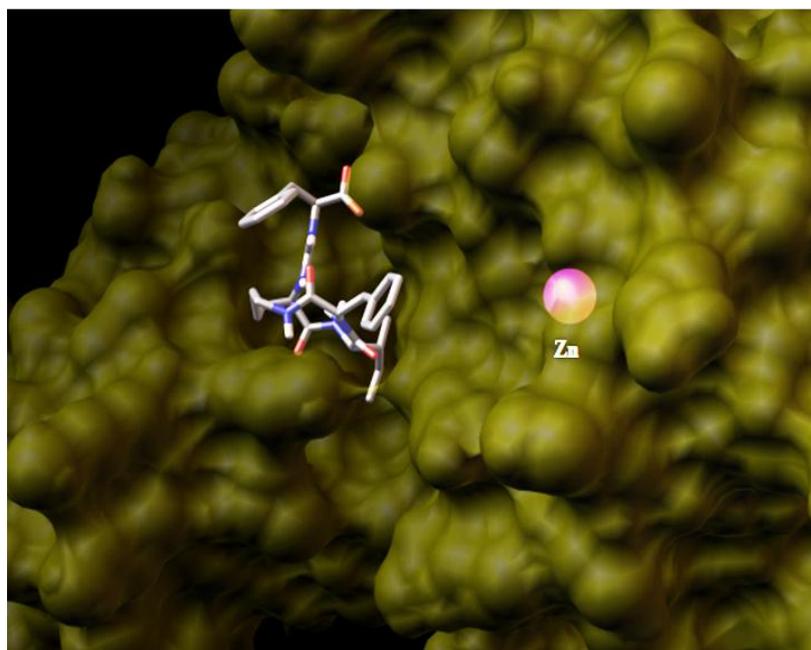
**Figure 2.16** Docked model of **56** to CPA. Protein is shown as a grey surface,  $\text{Zn}^{2+}$  atom as a magenta sphere, and residues in contact with **56** as grey (C atoms), blue (N atoms), red (O atoms), light grey (H atoms) sticks; **56** is rendered as yellow (C atoms), blue (N atoms), red (O atoms), light grey (H atoms) sticks and balls. Hydrogen bonds are displayed as green spheres.



**Figure 2.17** Superposition between **54** (colored by atom type: C green, O red, N blue and H light grey) and **56** (colored by atom type: C yellow, O red, N blue and H light grey).  $\text{Zn}^{2+}$  is as a magenta sphere.

Docking of **54** to CPU failed to give any reasonable models with **54** bound to or near the active site. For the above considerations, the possibility of Namalide to interact with Carboxypeptidase U is restricted on the external face of this binding cavity and then far from the Zinc ion.

However, it is important to observe that in this case the Autodock-Vina value of binding energy is of -6.5 kcal/mol, with a correspondent V value of 0.800 and a position nr.149 on the final classification of the 159 targets. The binding poses found in which the molecule is able to interact with the metal ion present positive values of energy, due to steric hindrance effects (Figure 2.18).



**Figure 2.18** **54** (colored by atom types: C grey, N blue, H cyan, O red) in docking with the CPU. Zinc is coloured in violet.

Both CPU and carboxypeptidase B are exopeptidases that preferentially cleave basic C-terminal residues, recognizing Arg and Lys as opposed to the aromatic

residues preferred by CPA. The lack of inhibitory activity toward CPU lends further support that this scaffold binds peptidases in a similar fashion to the larger anabaenopeptins.<sup>178,186</sup>

In summary, structure elucidation and synthesis of the natural product establish the absolute configuration of all amino acids as L, with the exception of D-Lys. In contrast with other anabaenopeptins, the macrocycle comprises just three rather than the typical five amino acids, leading to a 13-membered macrolactam ring versus the usual 19-membered one. The presence of this strained ring likely accounts for the side product of a namalide dimer formed during the on-resin cyclization employed in the synthesis of **54**. In keeping with specificity patterns described for a few peptides of this class, the carboxypeptidase inhibitory activity depends on the presence of D-Lys, and the exocyclic amino acid appears to dictate specificity. Because namalide inhibits CPA with potencies comparable to the more common hexapeptides, it will be interesting to evaluate other designed namalide analogs against CPA and related hydrolases.

## **-CHAPTER 3-**

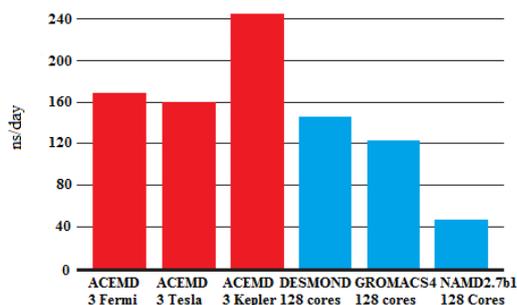
***High-throughput Molecular Dynamics for the  
accurate calculations of the binding affinities***

### **3.1 Molecular Dynamics: from CPU to GPU architecture**

The constant and notable improvement of the computational power determined the advance of the computational chemistry as a basic tool in the drug discovery. Remarkable results in less and less times can be obtained, providing potential indications for the subsequent experimental phases and allowing a strong reduction of the costs of a synthetic chemistry project. As mentioned in paragraph 1.2.2, Molecular Dynamics (MD) simulations are often required in order to clarify molecular mechanism of interaction of a ligand to a biological target. Contemporary Molecular Dynamics simulations are able to access microseconds<sup>192</sup> for all-atom systems. This impressive increase in speed if compared with time-scales of the previous years is accounted for in part by increasing algorithmic sophistication, but is mainly due to the parallelization of code to run on multiprocessor supercomputers.<sup>81,193</sup> Although these parallel MD codes represent a new frontier in the field of the drug discovery, they must be run on expensive, and high performance computing (HPC) resources. As a consequence, MD studies have focused on obtaining and analyzing a small number of trajectories that are long enough to completely sample the process of interest. The future of this approach could be represented by the development of novel,<sup>194,195</sup> specialized hardware and new MD protocols.<sup>196,197,198</sup> In particular, modern graphics processing units (GPUs) have recently been shown to be highly capable at MD simulations. This is a profound qualitative change, mainly because the cost of performing long simulations is now low. This represents a shift from MD as an expensive activity to a low cost one that can be performed on cheap commodity hardware.

In this context, another challenge was the development of the MD softwares able to exploit GPUs in parallel. Recently, ACEMD<sup>194</sup> software was implemented, allowing microsecond long trajectories on workstation hardware. It reads CHARMM/NAMD and AMBER input files with a simple and powerful

configuration interface, and represents the computational engine behind one of the largest distributed computing project worldwide GPUGRID.net nowadays summing thousands of GPUs. The performance of ACEMD is summarized in Figure 3.1.<sup>199</sup>



**Figure 3.1.** Comparative performance on DHFR (dihydrofolate reductase, solvated in water, 23558 atoms) system for ACEMD, DESMOND, GROMACS4, NAMD2.7b1. Fermi is a GTX580 GPU. Kepler is a GTX680 GPU. Tesla is a M2090. ECC off. CUDA4.2 and ACEMD ver 2400. Periodic boundary conditions, 9 Å cutoff, PME long range electrostatic 64×64×64, hydrogen mass repartitioning, rigid bonds, Langevin thermostat, time step 4 fs. Time step 2 fs NAMD and AMBER, 2.5 DESMOND and 4 fs GROMACS. AMBER 8 Å cutoff.

### 3.2 Re-ranking of molecular docking calculations using the Linear Interaction Energy (LIE) method

In the field of drug discovery, molecular docking methodology represents a particularly fast and suitable tool in the search of new potential active scaffolds related to a specific target. In particular, the main interest is to obtain in fast times ranking of compounds of potential pharmaceutical interest (Virtual Screening), whose biological activities would be in agreement with their calculated binding free energies toward a given protein. On the other hand, the affinity characterizes

the strength of the specific recognition of a ligand by a protein and plays a noteworthy role in correlating the structure and function of proteins.

From a general point of view, molecular docking methods are in point of fact limited by a poor accuracy in the prediction of the binding free energies, and this is mainly due to the simplified energy models. Scoring functions are often built and parameterized in order to have the maximum correspondence with the experimental affinities values, but actually are not derived from a well-defined physical model. For example, most of the scoring functions usually have poor treatment of the electrostatic and solvation contributions that limit the achievement of truthful results. Numerous docking methods are capable of producing near native docking poses with a good accuracy, but the development of a consistent scoring function still remains a challenge.

In particular, a low computational cost is evermore required, in order to obtain reliable results in short times for the applicability of the docking to large set of compounds. This aspect is of fundamental relevance in a Virtual Screening study, in which an inaccurate scoring function can produce a final ranking with a large number of false positives and more importantly false negatives, that will be excluded from the next phase of validation by experimental biological tests.

An improvement is required to assign better energy scores to select high affinity ligands from a large set of decoy compounds (enrichment), trying to correlate at the same time scores with measured binding affinities of known ligands.

Another important point is that the scoring functions, derived from empirical equations, are often system dependent, and different methods perform better on different systems. Results are sometimes improved when compared from multiple scoring functions, and this because a ligand that simultaneously ranks high using different scoring functions is more likely to bind strongly.<sup>200,201,202</sup> Thus the

development of a reliable and accurate scoring function has been the focus of many ongoing studies.

On the other hand, Molecular Dynamics simulations clarifies protein-ligand binding processes, but can be used also for the accurate estimations of the binding affinities (Paragraph 1.2.3). Obviously, the reliability of the results of these simulations is strictly related to their accuracy, and in particular the use of the explicit solvent models is often required.

Considering the huge and constant progress of the computational power, even more time-dispending computational techniques are advancing.

The development of ACEMD,<sup>194</sup> a software able to exploit the accelerated graphical processing units (GPUs) on a distributed computing network (GPUGRID) notably contributed to the performance of high-throughput Molecular Dynamics, providing detailed descriptions of a single ligand-protein binding, simulating this process in the order of the microseconds.

For the topic of interest of this study, these huge timescales can be exploited not only on one defined system, but it is possible to split the Molecular Dynamics simulations to a consistent number of different systems.

During the period mid May-November 2012 I was involved, under the supervision of Dr. Gianni De Fabritiis (University Pompeu Fabra, Barcelona), in a project regarding a Virtual Screening study using Molecular Dynamics simulations on 1588 systems involving different ligands and decoys from the DUD database, previously docked to the trypsin receptor.

The schedule was based on the investigation and the comparison of the efficacy of molecular docking, linear interaction energy (LIE),<sup>88</sup> and molecular mechanics Poisson-Boltzmann (MM-PBSA)<sup>94</sup> and Generalized-Born (MM-GBSA)<sup>95</sup> solvent accessible surface area methods.

All these studies are ongoing, and here only partial results are shown.

The main aim of this study was to give a more careful view of the applicability of these different accurate methods for a better estimation of the predicted binding affinities, comparing these results with those deriving from the docking calculations.

In particular, we performed for each of these systems (complex and ligand alone)  $3 \times 10$  ns simulations in explicit solvent, obtaining a total time of simulation of  $\sim 100 \mu\text{s}$  ( $\sim 50 \mu\text{s}$  for the complexes,  $\sim 25000$  atoms for each solvated system;  $\sim 50 \mu\text{s}$  for the ligands alone systems,  $\sim 3500$  atoms for each solvated system).

As known in the Linear Interaction Energy theory, the binding energy is calculated taking into account the polar and nonpolar contributions of potential energy, and averaging the ligand-surrounding potential energies, collected only for the two physical states of the ligand involved in the binding process, that are the free state and the bound state (Equation 1.23).

Moreover, into this equation three scaling factors ( $\alpha, \beta, \gamma$ ) empirically derived are used to fit the experimental affinity values. In literature several combinations of these scaling factors have been reported, and each of these is related and optimized to a specific system. In this study, four different combinations of different scaling factors were considered, in order to verify how much LIE is able to predict binding affinities in a dependent way from these scaling factors. Scaling factors considered were:

a)  $\alpha=0.476, \beta=0.165, \gamma=0.000^{203}$

b)  $\alpha=0.236, \beta=0.146, \gamma=0.010^{203}$

c)  $\alpha=0.418, \beta = 0.087, \gamma = 0.000^{204}$

d)  $\alpha=0.224, \beta = 0.085, \gamma = 0.000^{205}$

These initial parameters were chosen considering the different related influence of  $\alpha$  and  $\beta$  (and then the weight of polar and non polar parts) on the final calculations of the binding energy. For example, the a) and c) cases differ for the different weight of  $\beta$ ; b) and d) cases show lower  $\alpha$  values with respect to a) and c), and moreover differ from each other for the different weight of  $\beta$ .

The results were analyzed using enrichment and Receiver Operating Characteristic (ROC) curves.

In the first case, the ratio between the true positive percentage rate (ligands) and the percentages of the ranked database is considered. It represents a useful tool in a Virtual Screening study, because it indicates on a certain percentage of the whole database how many true positive compounds are identified (enrichment factor). The most interesting enrichment factors considered are at 1, 2, 5 and 10% (EF<sub>1</sub>, EF<sub>2</sub>, EF<sub>5</sub>, EF<sub>10</sub>) of the ranked database.

Figure 3.2, Figure 3.3, Figure 3.4 respectively show the enrichment curves considering the 100% (whole), 10% and 5% of the ranked database, while enrichment factors are summarized in Table 3.1.

These data show from a general point of view slightly better results for the four LIE methods versus docking until the 2% of the ranked database, but over this point docking curve is significantly able to overtake LIE curves.

% of ranked database	EF LIE ( $\alpha=0.476$ ; $\beta=0.165$ , $\gamma=0.000$ )	EF LIE ( $\alpha=0.236$ ; $\beta=0.146$ , $\gamma=0.010$ )	EF LIE ( $\alpha=0.418$ ; $\beta=0.087$ , $\gamma=0.000$ )	EF LIE ( $\alpha=0.224$ ; $\beta=0.085$ , $\gamma=0.000$ )	EF docking (GOLD)
1	11.364	13.636	11.364	11.364	9.091
2	7.955	7.955	7.955	9.091	7.955
5	6.364	5.909	5.455	6.364	9.091
10	4.318	4.318	3.636	4.091	7.500

**Table 3.1** Enrichment factors (EF) calculated for the different methods analyzed at 1, 2, 5, and 10 % of the ranked database.

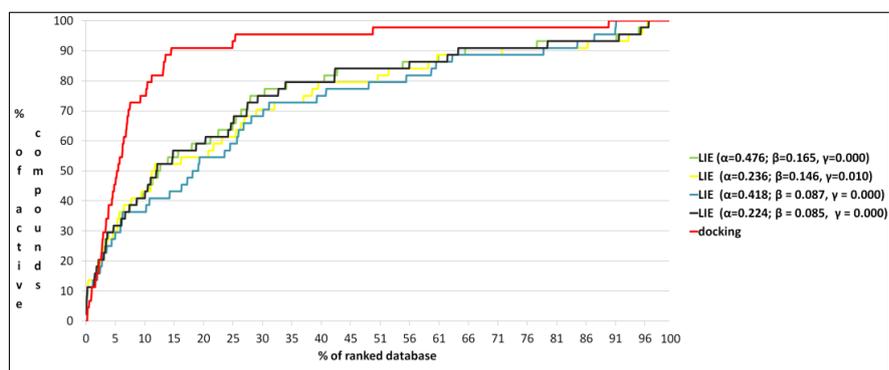


Figure 3.2 Enrichment curves for docking and LIE calculations related the whole ranked database

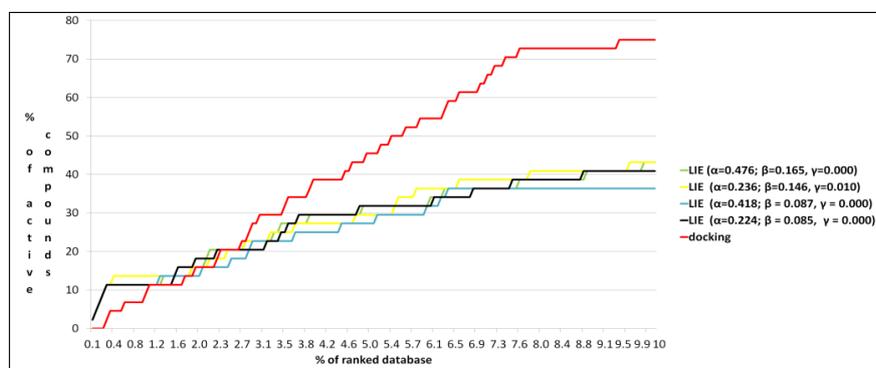


Figure 3.3 Enrichment curves for docking and LIE calculations related the first 10% of the ranked database

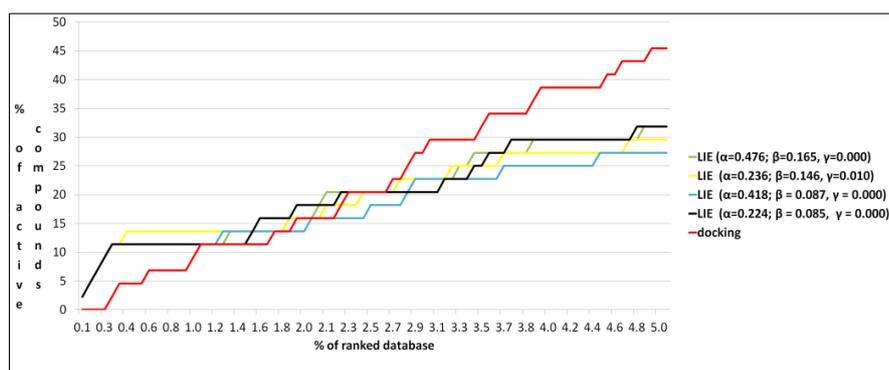
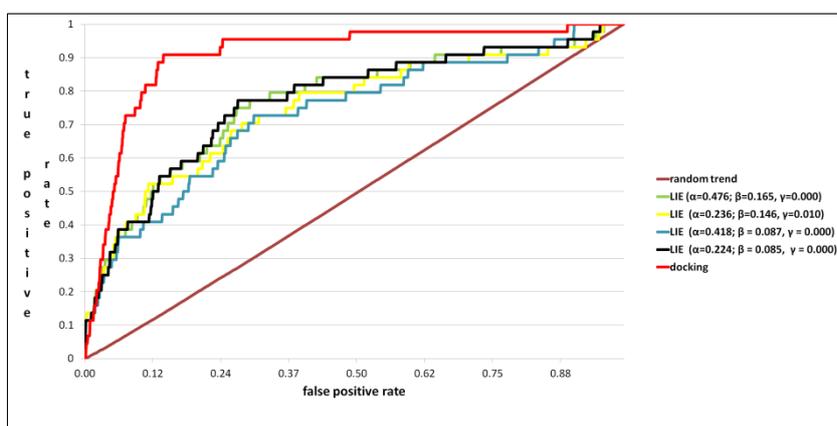


Figure 3.4 Enrichment curves for docking and LIE calculations related the first 5% of the ranked database

All these observations are confirmed by the ROC curve, where the ratio between the true positive (ligands) rate and the false positive (decoys) rate is taken into account (Figure 3.5).



**Figure 3.5** ROC curves related to molecular docking and LIE calculations

In general, the data presented highlight the poor results obtained by the four LIE methods used, especially considering the fastness of the molecular docking calculations. These data seem to be interesting concerning the reliability of the LIE calculations. On the other hand, remaining in the field of the LIE, further investigations are in progress regarding the dependence of results from the scaling factors used. In parallel, we are considering other MD-based methods for the re-ranking of the docking calculations.

For these reasons, other studies are under evaluation, specifically concerning:

- Use of ligand specific LIE scaling factors<sup>206</sup>
- MM-PB(GB)SA methods.

## **-CHAPTER 4-**

### ***Further applications of in silico screenings on natural compounds***

#### **4.1 Discovery of cholestan disulfate as a potent pregnane-X-receptor agonist**

PXR (pregnane-X-receptor) is a master gene regulating the activity of a variety of genes involved in xeno- and endo-biotic metabolism in the liver and gastrointestinal tract.<sup>207-210</sup>

Once activated, PXR heterodimerizes with the retinoid-X-receptor (RXR), binds to regulatory DNA sequences in the promoter of responsive genes and modulating their transcription. PXR is a master gene orchestrating the expression/function of a number of genes involved in the detoxification/excretion of endo- and xeno-biotics thus preventing toxic accumulation of metabolites within cells. In addition, PXR is recognized as an important regulatory factor in modulation of important effector functions in the immune system through inhibition of proinflammatory transcription factor NF- $\kappa$ B in epithelial and immune cells.<sup>211,212</sup>

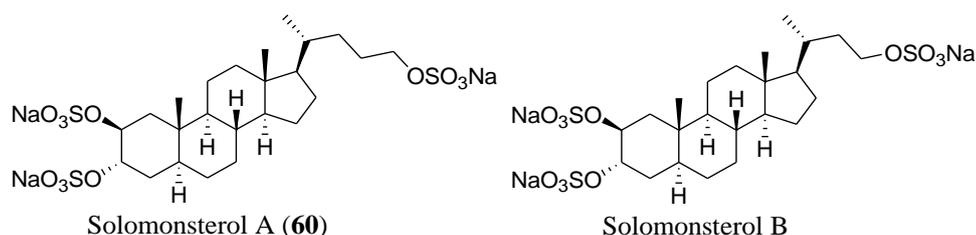
A role for PXR in the pathogenesis of inflammatory bowel disease (IBD) is increasingly supported by genetic and pharmacological evidence. Thus, gene expression analysis of colon tissues from ulcerative colitis and patients with Crohn's disease have revealed a significant reduction in the expression/function of PXR and its target genes compared with normal intestinal samples.<sup>213</sup> Moreover rifaximin, a human PXR activator, is in clinical trials for treatment of IBD and has demonstrated efficacy in Crohn's disease and active ulcerative colitis.<sup>214,215</sup> It is speculated that by activating PXR, rifaximin could contribute to the maintenance of the intestinal barrier integrity by regulating the metabolism of xenobiotics and increasing the expression and activity of PXR and PXR-regulated genes.<sup>216</sup>

Therefore PXR represents an important pharmacological target and the discovery of potent and selective PXR agonists holds potential in the discovery of new drugs for the treatment of human disorders characterized by dysregulation of innate immunity.

Recently research group of Professor Angela Zampella reported the biochemical decodification of several steroids of marine origin as ligands of two nuclear receptors, FXR and PXR.<sup>217-223</sup> Among these, two sulfated steroids, solomonsterols A (**1**) and B (Chart 4.1) isolated from the sponge *Theonella swinhoei*,<sup>224</sup> were proved to be potent inducers of PXR transactivation in human hepatocyte cell line (HepG2 cells) stimulating the expression of CYP3A4 and MDR1 (Cytochrome P450 3A4 and multidrug resistance 1), two well characterized PXR responsive genes in the same cell line (Chart 4.1).<sup>225</sup>

In addition, through a deep pharmacological investigation on transgenic mice expressing the human PXR, we have demonstrated that solomonsterol A (**60**) effectively protects against development of clinical signs and symptoms of colitis, reduces the generation of TNF- $\alpha$  and enhances the expression of TGF- $\beta$  and IL-10, two potent counter-regulatory cytokines in IBD, via inhibition of NF- $\kappa$ B activation in a PXR dependent mechanism.<sup>226</sup>

All these data pointed towards the identification of solomonsterol A (**60**) as a new lead in the treatment of IBD. One of the possible limitation to its use in clinical settings is that, when administered *per os*, solomonsterol A (**60**) could undergo absorption from the GIT before reaching the colon causing severe systemic side effects resulting from the activation of PXR in the liver.



**Chart 4.1** Chemical structures of solomonsterol A and B

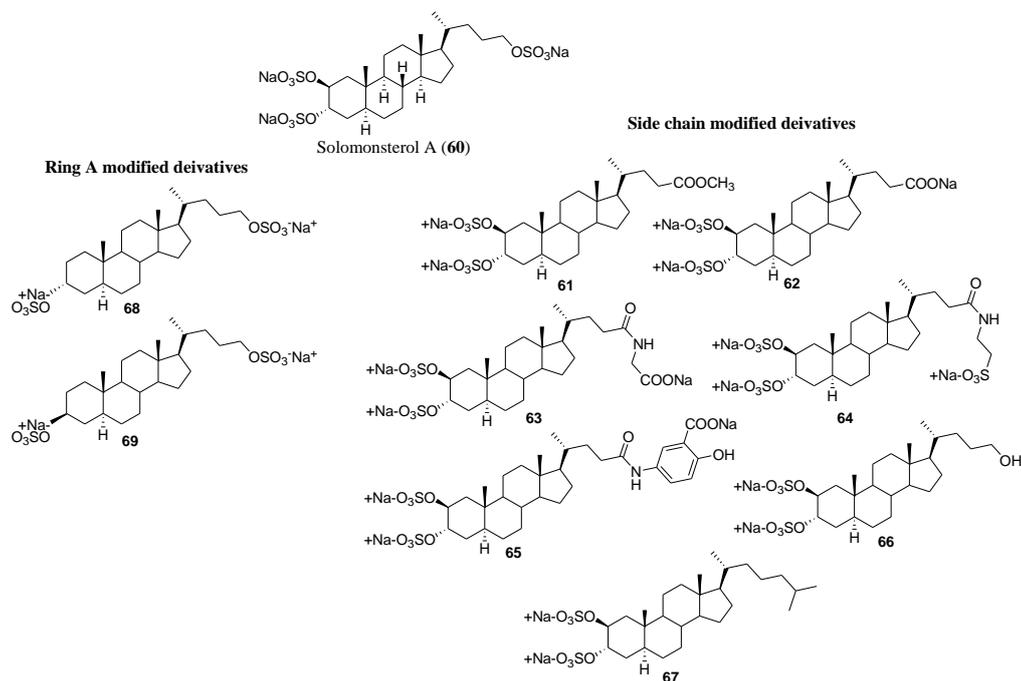
One of the best approaches used for colon specific drug delivery is based on the formation of a prodrug through chemical modification of the drug structure, usually by the conjugation with a suitable carriers, such as amino acids, sugar, glucuronic acid, dextrans or polysaccharides.

Since the luxuriant microflora presents in the colon, the prodrugs undergo enzymatic biotransformation in the colon thus releasing the active drug molecule.

Another challenging task is the design of dual-drugs able to release in the colon two molecules acting in a synergic manner. For example the possible eventual chemical linkage of solomonsterol A (**60**) to 5-ASA, one of the oldest anti-inflammatory agents in use for the treatment of IBD, could produce a dual-drug with enhanced potency. Upon enzymatic hydrolysis in the colon, this kind of molecule could release solomonsterol A and 5-ASA, potent agonists of PXR and PPAR- $\gamma$ ,<sup>227</sup> respectively, two nuclear receptors playing key roles in colon inflammation diseases.

When synthesizing prodrugs, the first step is the introduction of a functional group on the drug molecule suitable of conjugation with a selected carrier (*e.g.*, an hydroxyl group that could enter into a glycosidic linkage with various sugars, or alternatively a carboxyl group to form ester *e/o* amide conjugates with cyclodextrin, amino acids etc).

Inspection of chemical structure of solomonsterol A (**60**) revealed that the presence of three sulfate groups hampered any further derivatization *e/o* conjugation. In order to introduce a function group suitable for further derivatization, several solomonsterol A derivatives with a modified side chain but preserving the steroidal tetracyclic nucleus (compounds **61-67** in Chart 4.2) were prepared. Another opportunity was to speculate the pharmacoforic role played by ring A, preparing derivatives **68-69** with a sulfate group at C3 in  $\alpha$  and  $\beta$  orientation, respectively.

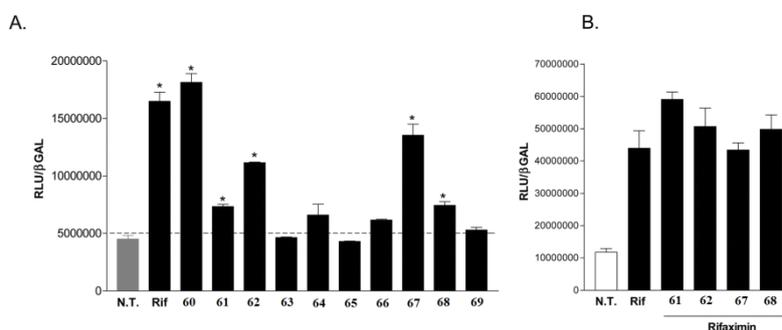


**Chart 4.2** Modified solomonsterol A derivatives

The small library of derivatives obtained (Chart 4.2) was subjected to pharmacological evaluation and docking analysis. This study allowed the discovery of a synthetic solomonsterol analogue, 2 $\beta$ ,3 $\alpha$ -cholestan disulfate (ColdiSolf, **67**), as simplified new potent PXR agonist.

#### 4.1.1 Biological studies

To investigate whether these compounds act on PXR and eventually PXR regulated genes, we have carried out a luciferase reporter assay on human hepatocyte cell line (HepG2 cells) transiently transfected with pSG5-PXR, pSG5-RXR, pCMV- $\beta$ galactosidase, and p(CYP3A4)-TK-Luc vectors (Figure 4.1). Cells were then stimulated with rifaximin, a well known PXR agonist, and with compounds **62-65** at the concentration of 10  $\mu$ M each.

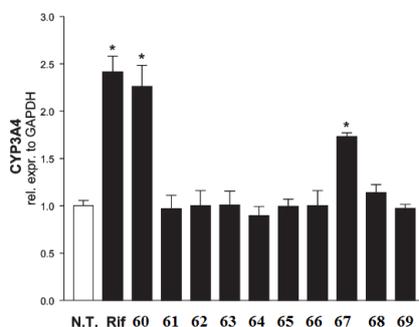


**Figure 4.1** Luciferase reporter assay. HepG2 cells were transiently transfected with pSG5-PXR, pSG5-RXR, pCMV-βgalactosidase and p(CYP3A4)-TK-Luc vectors and then stimulated with (A) 10 μM rifaximin or compounds **60–69** for 18 h, or (B) 10 μM rifaximin alone or in combination with 50 μM of compounds **61, 62, 67** and **68**. N.T., not treated. Rif, rifaximin. \*P < 0.05 versus cells left untreated. Data are mean ± SE of three determinations.

As shown in Figure 4.1 A, beside the closely structural resemblance with solomonsterol A (**60**), only carboxylate (**62**) showed a slight activity in transactivating PXR. Besides at first sight this behaviour should be ascribable to a scarce bioavailability, the scarce activity also for the methyl ester **61** and the complete loss of activity for C-24 alcohol **66**, obtained through LiBH<sub>4</sub> reduction of **61** (75% yield), pointed towards unfavourable pharmacodynamic features. Indeed, although compounds **62–65** (COOH, taurina, gly, asa) possess a negative charge on their side chains, most likely they are unable to form polar interactions with Lys210 or alternatively with other polar amino acids of PXR LBD. As previously reported, PXR presents a large ligand binding cavity<sup>228,229,230</sup> allowing the accommodation of different kind of molecules and the possible binding modes are characterized an adequate balance between hydrogen bond and Van der Waals interactions established between a small molecule and the receptor.

Therefore the lack of a polar interaction should be get over by increasing the contribution of the hydrophobic interactions on the side chain and the derivative **67**, 2 $\beta$ ,3 $\alpha$ -cholestan disulfate (ColdiSolf), has been prepared.

As shown in Figure 4.1 a, compound **67** with its hydrophobic side chain is able to transactivate PXR with a potency comparable with the parent solomonsterol A (**60**). Although compound **68** induces a slight PXR transactivation, the lack of sulfate group at C-2 as well as the inversion of configuration at C-3 are responsible for a general loss in the agonistic activity towards PXR (Figure 4.1 a). To investigate whether these compounds could act as potential antagonists of PXR we have carried out a transactivation experiment in HepG2 cells stimulated with rifaximin (10  $\mu$ M) and compounds **61**, **62**, **67** and **68** at a concentration of 50  $\mu$ M each. As shown in Figure 4.1 b, all compounds failed to reverse the induction of luciferase caused by rifaximin, indicating that none of these solomonsterol A (**60**) derivatives is a PXR antagonist.

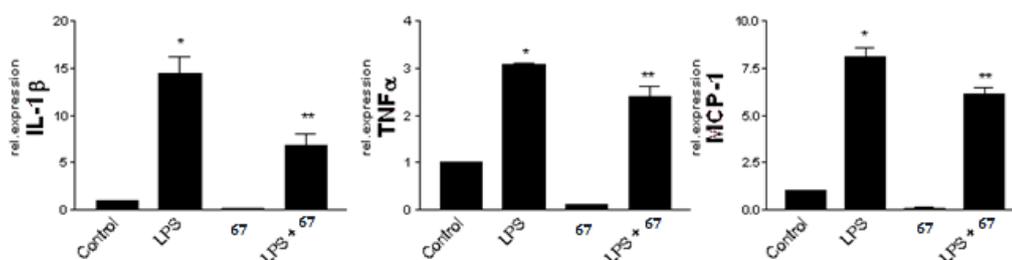


**Figure 4.2** Real-time PCR of CYP3A4 carried out on cDNA isolated from HepG2 not stimulated or primed with 10  $\mu$ M rifaximin, and compounds **60–69**. N.T., not treated. Rif, rifaximin. \*P < 0.05 versus N.T. cells.

To further examine the activity of compound **67** as a PXR activator and further clarify the behavior of compounds **61**, **62** and **68**, we have tested the effects of all members of our series on the expression of CYP3A4, a canonical PXR target gene

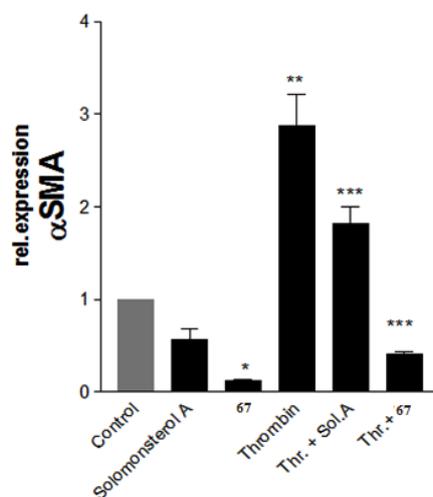
(Figure 4.2). Despite compounds **61**, **62** and **68** causing a slight transactivation of PXR, they failed to modulate the expression of CYP3A4 at the concentration of 10  $\mu$ M. In contrast, confirming data shown in Figure 4.1, compound **67** effectively increased the expression of CYP3A4 (Figure 4.2) in HepG2 cells, with a magnitude similar to that of rifaximin and solomonsterol A (**60**). While these data do not exclude that compound **62** could also stimulate CYP3A4 expression in this system, the need for higher concentrations to display a full PXR agonistic activity precluded its further development. To further investigate whether compound **67** displays a full PXR agonistic activity, we then evaluated the effect of **67** in regulating immune response using THP1 cells, a human macrophage/ monocytic cell line, challenged with lipopolysaccharide (LPS), a potent agonist of Toll like receptor (TLR)-4. Previous studies<sup>231</sup> have shown that activation of PXR in this setting attenuates immune response triggered by LPS, a toll-like receptor 4 ligand, and key modulator of innate immunity.

Results shown in Figure 4.3, demonstrate that compound **67** effectively attenuates induction of IL-1 $\beta$ , TNF $\alpha$  and MCP-1 induced by LPS.



**Figure 4.3** Effect of compound **67** on cytokine release induced by LPS in THP1 cells.  $3 \times 10^6$  THP-were starved for 24 h and then pre-treated with 10  $\mu$ M of compound **67** for 3 h and then stimulated for 18 h with LPS 1  $\mu$ g/ml. Cytokine expression was assessed by RT-PCR. Data shown are m mean  $\pm$  SE of 9 assays from three different sets of experiments. \*P<0.05 versus control cells; \*\* P<0.05 versus LPS alone.

Because the above mentioned data indicate that compound **67** effectively modulates immune response in human monocytes, additional experiments were carried out to investigate the effect of this compound in another model of inflammation-driven activation, using hepatic stellate cells (HSCs). HSCs are a liver-resident cell population that proliferate in response to liver injury. In response to immune activation, HSCs undergo a complex phenotype rearrangement characterized by resetting expression of nuclear receptors, including PXR, and acquisition of an activated, myofibroblast-like phenotype whose main characteristic is the ability to express  $\alpha$ -smooth muscle actin ( $\alpha$ SMA). HSCs are recognized as the main source of extracellular matrix production in the fibrotic liver. Previous studies have shown that, along with other nuclear receptors, PXR ligands reverse this phenotype and reduce  $\alpha$ -SMA production.<sup>232,233,234</sup>



**Figure 4.4** Hepatic stellate cells (HSC) were starved for 24 h and then stimulated with thrombin, 10 U/ml, in the presence of solomonsterol A or compound **67**, 10  $\mu$ M each.  $\alpha$ -SMA expression was assessed by RT-PCR. Data shown are mean  $\pm$  of three experiments. \*  $P < 0.05$  versus control cells; \*\*  $P < 0.05$  thrombin versus control cells; \*\*\*  $P < 0.05$  versus thrombin alone.

For this purpose HSCs were exposed to thrombin, a proteinase activated receptor (PAR)-1 agonist alone or in combination with compound **67**. Previous studies have shown that thrombin drives HSCs trans-differentiation and its inhibition reverses HSCs from an activated to a quiescent phenotype.<sup>235</sup> Results shown in Figure 4.4 demonstrate that, similarly to solomonsterol A (**60**), not only does compound **67** effectively reduce basal expression of  $\alpha$ SMA, but it also attenuates HSCs trans-differentiation (i.e. induction of  $\alpha$ SMA expression) triggered by thrombin.

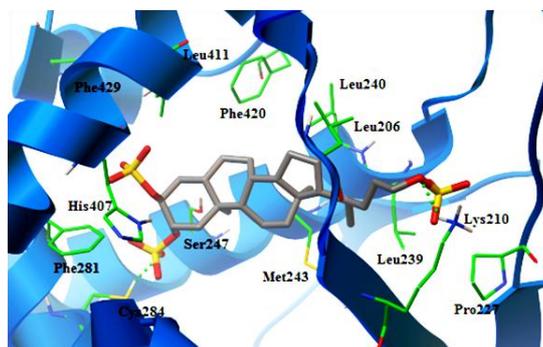
#### 4.1.2 Molecular modelling studies

In order to clarify the different activities of the here described compounds at a molecular level we performed docking calculations, using the Autodock 4.2 software,<sup>27</sup> we examined the positioning of all the compounds in the binding site of PXR,<sup>236</sup> and in particular we analyzed the crucial interactions with the Ser247, His407 and finally with the Lys210.

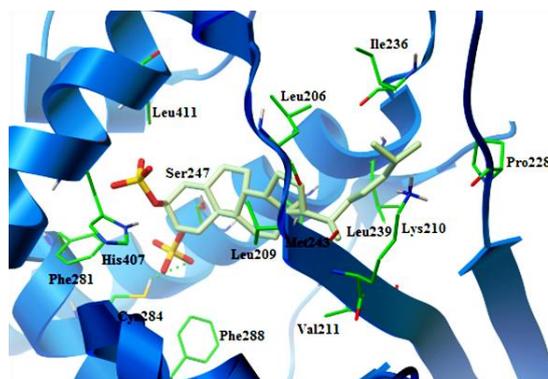
These residues, in particular the Ser247, are involved in hydrogen bonds with various PXR's ligands.<sup>229,236,237</sup>

The three sulfate groups of the agonist solomonsterol A (**60**) act as key points of interactions with these amino acids, and contribute to accommodate the steroid nucleus in a mostly hydrophobic part of the binding site of PXR. The compound establishes hydrogen bonds (Figure 4.5) with the Cys284 (2-*O*-sulfate) and with the Lys210 (24-*O*-sulfate) and electrostatic interactions with the Ser247 (2-*O*-sulfate) and His407 (3-*O*-sulfate).<sup>226</sup>

ColdiSolf (**67**), presenting the C8 aliphatic side chain of cholesterol, is well superimposed with the binding pose of **60**, and is able to interact with the



**Figure 4.5** Solomonsterol A (**60**) (coloured by atom types: C grey, O red, S yellow) in docking with PXR-LBD (residues are coloured by atom type: C green, H light grey, O red, N blue). Hydrogen bonds are displayed with green spheres.



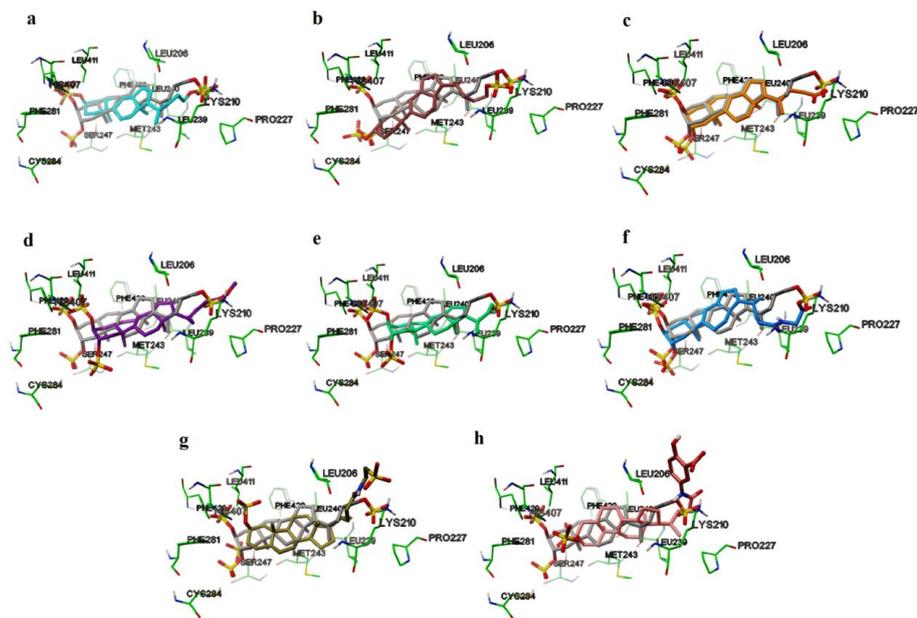
**Figure 4.6** ColdiSolf (**67**) (coloured by atom types: C light green, O red, S yellow) in docking with PXR-LBD (residues are coloured by atom type: C green, H light grey, O red, N blue). Hydrogen bonds are displayed with green spheres.

Ser247, Cys284 and the His407 through its two sulfate groups in the ring A (Figure 4.6). Moreover, **67** establishes hydrophobic interactions with almost all the residues observed for solomonsterol A (**60**) (Leu209, Val211, Pro228, Leu239, Met243, Phe281, Phe288, Leu411). The presence of a hydrophobic chain allows to gain two more Van der Waals interactions (with the Leu209 and Val211) that may counter the loss of electrostatic interaction observed for the sulfate group at C24 of parent

solomonsterol A. Nevertheless, the weaker nature of this Van der Waals interaction could explain the decrease of the activity of coldiSolf (**67**) on PXR (difference of predicted binding energies **60-67**=1.05 kcal/mol).

On the other hand, the absence of the sulfate group at C-2 in the steroid nucleus causes the observed lack of activity, due to an inability to interact simultaneously with the three key point of contact previously described.

For example, compounds **68** and **69** are able to interact with the Lys210 but they fail to respect the key interactions involving the internal part of the binding site. Regarding compound **62**, its tetracyclic nucleus is well superimposed with **60**, but its shorter side chain causes a poor interaction with the nitrogen of Lys210. The two oxygens of its terminal carboxylic part are not well overlapped with the oxygens of the 24-*O*-sulfate of the **60**, and the different arrangement of the side chain causes also a loss of two Van der Waals interactions with the Leu239 and Pro227. In other cases a displacement of the compound to the solvent part is observable. For example, the rings A of compounds **61** and **66** are in the place occupied by the ring B of **60** and, as a consequence, the 2-*O*-sulfate and/or 3-*O*-sulfate are in a less deep position. Compounds **63**, **64** and **65** present a longer and more functionalized side chain compared with the previous derivatives, but also in this case the steroid nucleus are placed toward the external part of the binding site of PXR (**63**, **65**). Moreover, compound **64** is unable to bind in the above described fashion and accommodates in a reverse orientation (a flipping of ~ 180° along the major axis of the steroid nucleus) of their steroid nucleus. The overall result is an inverted disposition of all the chemical groups (sulfates/methyl groups, and side chain) in the binding pocket of PXR and then a different pattern of interactions. (Figure 4.7)



**Figure 4.7** Superimposition between **60** (coloured by atom types: C grey, O red, S yellow) and: a) **68** (coloured by atom types: C sky-blue, O red, S yellow); b) **69** (coloured by atom types: C brown, O red, S yellow); c) **62** (coloured by atom types: C orange, O red, S yellow); d) **61** (coloured by atom types: C purple, O red, S yellow); e) **66** (coloured by atom types: C turquoise green, O red, S yellow); f) **63** (coloured by atom types: C dodger blue, O red, S yellow); g) **64** (coloured by atom types: C dark green, O red, S yellow); h) **65** (coloured by atom types: C pink, O red, S yellow) in PXR-LBD (residues are coloured by atom type: C green, H light grey, O red, N blue).

### 4.1.3 Final remarks

In summary, coldiSolf (**67**) has been identified as a new PXR agonist. The ability of this compound to function as a PXR agonist was first demonstrated in transactivation assay using HepG2 cells transiently transfected with a PXR vector. The results of these experiments demonstrate that compound **67**, and less effectively compounds **61**, **62**, **68**, efficiently transactivates human PXR with a relative

potency that was very similar to that of rifaximin a well know PXR agonist. The functionality of the interaction of compound **67** with PXR was further investigated in two different cell models. Using THP1 cells, a mielo-monocytic cell line, we have provided evidence that compound **67** attenuates cytokine generation induced by LPS. Moreover, in agreement with transactivation experiments, the PCR data demonstrate that compound **67** increases the expression of CYP3A4, a well characterized PXR responsive gene, in liver cells. The functionality of the interaction of compound **67** with PXR was further investigated in two different cell models. Using THP1 cells, a monocytic cell line, we have provided evidence that compound **67** attenuates cytokine generation induced by LPS. Because previous studies have provided robust evidence that PXR activation by rifaximin and solomonsterol A<sup>226</sup> exerts anti-inflammatory activity in rodent models of colitis by attenuating inflammation driven-immune dysfunction and cytokine accumulation in inflamed tissues, the present results extend on the role of PXR ligands in regulating immune function, and pave the way for the use of compound **67** in preclinical models of inflammation. Despite the fact that we have not investigated the mechanism mediating inhibition of cytokines by compound **67**, we and others have provided evidence that PXR agonists inhibit NF- $\kappa$ B activation.<sup>226,238</sup> Extending on the role of PXR as an endogenous braking signal for inflammation, we have then examined whether compound **67** would have been effective in reducing collagen production by HSCs, a myofibroblast-like cell line. HSCs acquire an activated phenotype in response to liver injury and release collagen and express  $\alpha$ SMA in response to toxic and immunological stimuli in a variety of liver disorders. Previous studies have provided evidence that nuclear receptors, including FXR, SHP and PPAR $\gamma$ , modulate a collagen release by HSCs and might function as important therapeutic targets for treating liver fibrosis.<sup>239</sup> PXR ligation attenuates liver fibrosis and HSCs activation.<sup>240</sup> Here we have shown that solomonsterol A

(**60**) and compound **67** reduce  $\alpha$ SMA accumulation triggered by thrombin. There is substantial evidence to support the notion that PXR activators are anti-fibrogenic in human liver myofibroblasts in vitro<sup>241</sup> and in in vivo animal models of liver fibrosis. The role of the PXR in regulating HSCs activation has been unequivocally established using mice with a disrupted PXR gene. The mechanism throughout which PXR regulates production of extracellular matrix proteins is likely associated with a function for the PXR that is not related with its recognized function as a regulator of genes associated with endobiotic and xenobiotic clearance, and might be linked to inhibition of intracellular signaling including NF- $\kappa$ B, or involved in the regulation of transdifferentiation of these fat-storing cells as demonstrated for other nuclear receptors including PPAR $\gamma$  and farnesoid-X-receptor, FXR.<sup>242</sup>

Because of its simplified structure and efficacy in attenuating immune activation in macrophages and HSCs, compound **67** is a suitable candidate for further development in preclinical models of inflammatory diseases.

## **4.2 Plakilactones from the marine sponge *Plakinastrella mamillaris*, a new class of marine ligands of peroxisome proliferator-activated receptor $\gamma$**

The peroxisome proliferator-activated receptors (PPARs) are ligand activated transcription factors belonging to the nuclear receptor superfamily. Three distinct receptor subtypes, PPAR $\alpha$ , PPAR $\gamma$  and PPAR $\beta$ ( $\delta$ ), have been identified. While the PPAR subtypes share a high level of sequence and structural homology, each subtype has distinct physiological functions and exhibits a unique tissue expression pattern. PPAR $\gamma$ , the most widely investigated PPAR subtype, is predominately expressed in the adipose tissue with lower levels in heart, colon, kidney, spleen, intestine, skeletal muscle, liver and macrophages. PPAR $\gamma$  is generally recognized as a pivotal transcription factor in the regulation of adipocyte gene expression and differentiation. In addition, PPAR $\gamma$  has been shown to be an important regulator of target genes involved in glucose and lipid metabolism and is the mainstay of therapy for type 2 diabetes.<sup>243,244</sup> Furthermore, PPAR $\gamma$  transrepresses the expression of genes involved in inflammatory responses,<sup>245</sup> and suppression of the inflammatory response by PPAR $\gamma$  agonists is closely linked to the anti-diabetic and anti-atherosclerotic effects of this receptor. Thus, PPAR $\gamma$  agonists have been found effective in the treatment of several inflammatory and degenerative disorders including cancer, atherosclerosis, rheumatoid arthritis and inflammatory bowel disease.

The ligand binding domain (LBD) of PPAR $\gamma$  allows the accommodation of a large variety of structurally different chemicals, including many food-derived substances such as polyunsaturated fatty acids, flavonoids, terpenoids and polyphenols.<sup>246</sup> Despite the extraordinary chemical diversity exhibited by marine natural products, only two reports have identified marine natural products as putative PPAR $\gamma$  agonists. Thus, the screening of 2688 extracts from marine

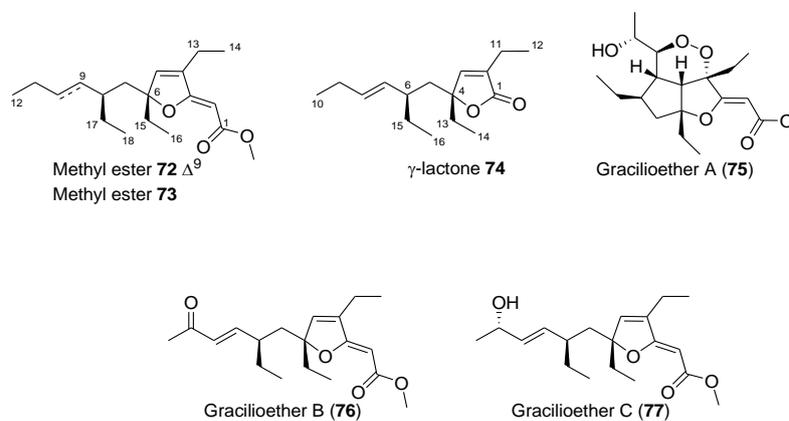
organisms led to the identification of psammaphin A as the first PPAR $\gamma$  agonist from a marine sponge.<sup>247</sup> Similarly, from the screening of a library of 90 bioactive marine extracts for their ability to stimulate PPAR $\alpha$  and PPAR $\gamma$  transcriptional activity, sargaquinoic acid and sargahydroquinoic acid were identified as novel PPAR $\alpha/\gamma$  dual agonists from *Sargassum yezoense*.<sup>248</sup>

As part of our search for human nuclear receptor modulators from marine organisms,<sup>217,218,219,220,221,222,223,224,226</sup> a library of oxygenated polyketides from the sponge *Plakinastrella mamillaris*, collected at Fiji Islands, was investigated as PPAR $\gamma$  modulators. Marine sponges of the genera *Plakortis* and *Plakinastrella* are known to produce a great variety of oxygenated polyketides, formed by the combination of acetyl-, propionyl-, and/or butyryl-CoA units. They include plakortolide, plakinic acid, plakortic acid, plakortone, or plakortide families.<sup>249</sup> Several activities have been ascribed to the members of this class, including anti-proliferative,<sup>250</sup> antifungal, anti-inflammatory or activation of cardiac SR-Ca<sup>2+</sup>-pumping ATPase.

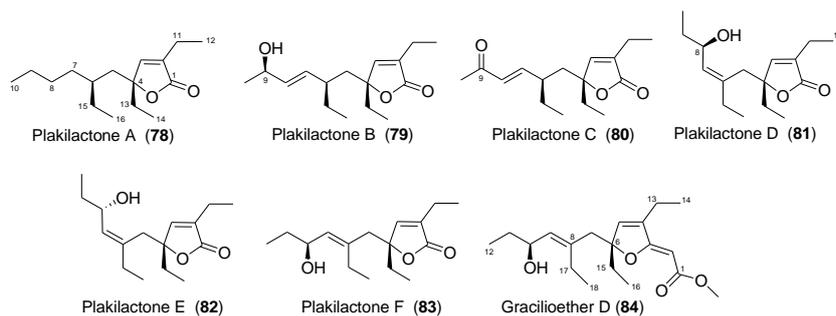
Interestingly, compounds containing the 1,2 dioxolane system such as plakortin, plakortides and gracilioether A exhibit potent antiprotozoan activity against *Plasmodium falciparum* and *Leishmania major*.<sup>251-259</sup> A specimen of *P. mamillaris* Kirkpatrick, 1900 (Homoscleromorpha) was collected at the Fiji Islands. The lyophilized sponge was extracted with MeOH, and the combined extracts were fractionated according to the Kupchan partitioning procedure.<sup>260</sup>

The major components of the hexane extract were proved to be the previously reported methyl esters **72**<sup>261</sup> and **73**,<sup>262</sup> the  $\gamma$ -lactone **74**<sup>261</sup> (Figure 4.8) and the new  $\gamma$ -lactone **78** (Figure 4.9), which we named plakilactone A. A careful analysis of the chloroformic extract afforded several more polar derivatives such as gracilioethers A-C (**75-77**), previously reported from the marine sponge *Agelas gracilis* (Figure 4.8),<sup>259</sup> five new non-peroxy plakortin derivatives, plakilactones B-

F (**79-83**), featuring the  $\alpha,\beta$ -unsaturated  $\gamma$  lactone moiety, and gracilioether D (**84**) showing the  $\alpha,\beta,\gamma,\delta$ -unsaturated methyl ester moiety (Figure 4.9).



**Figure 4.8.** Known polyketides from *Plakinastrella mamillaris*.

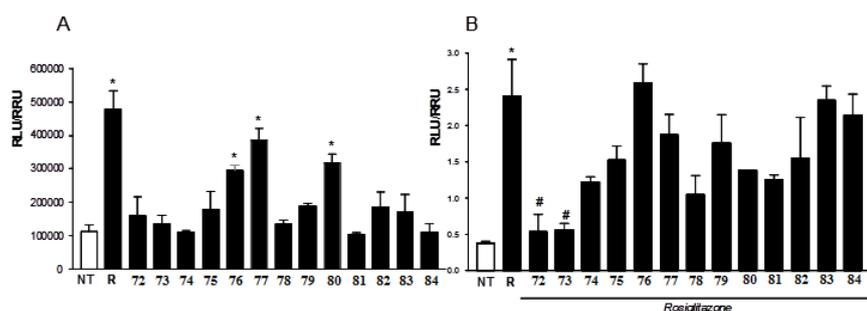


**Figure 4.9.** New compounds from *Plakinastrella mamillaris*.

The structural characterizations of these compounds were determined by detailed analysis of HRESIMS and NMR experiments ( $^{13}\text{C}$  and 2D NMR data, namely COSY, HSQC, HMBC and ROESY ).

## 4.2.1 Biological studies

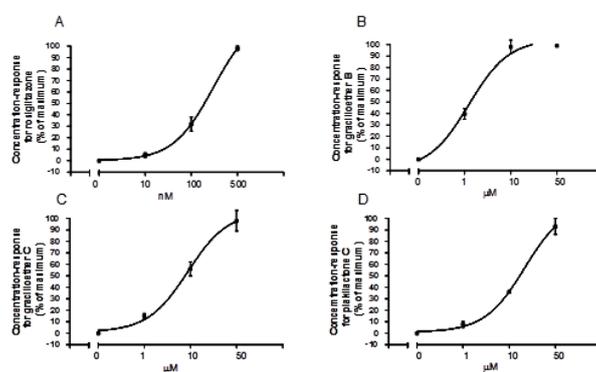
PPAR $\gamma$  can be activated by a number of natural lipid metabolites, including oxidized fatty acids, several cyclooxygenase (COX) and lipoxygenase (LOX) metabolites and 15-deoxy- $\Delta$ 12,14- prostaglandin J<sub>2</sub> (15d-PGJ<sub>2</sub>), the first endogenous ligand identified. These molecules act as covalent ligands possessing a common core moiety, an  $\alpha,\beta$ -unsaturated ketone, able to form a covalent bond with a cysteine residue in the PPAR $\gamma$ -LBD through a Michael addition.<sup>263,264,265</sup> Thus, the presence of a Michael acceptor, an  $\alpha,\beta$ -unsaturated ketone or alternatively an  $\alpha,\beta,\gamma,\delta$ -unsaturated methyl ester moiety, in all polyketides isolated from *Plakinastrella* prompted us to investigate their capability to transactivate PPAR $\gamma$ .



**Figure 4.10** PPAR $\gamma$  transactivation assay. HepG2 cells were transiently transfected with a chimeric receptor expressing plasmid pSG5GAL4-PPAR $\gamma$ LBD and with the reporter vector p(UAS)5<sub>x</sub>TK-Luc. 24 h post transfection cells were stimulated for 18 h with (A) 100 nM rosiglitazone (R) and compounds **72-84**, 10  $\mu$ M. (B) 100 nM Rosiglitazone (R) alone or in combination with compounds **72-84**, 50  $\mu$ M. Data are the mean  $\pm$  S.E. of three experiments. \*P<0.05 versus not treated cells (NT). #P<0.05 versus rosiglitazone stimulated cells.

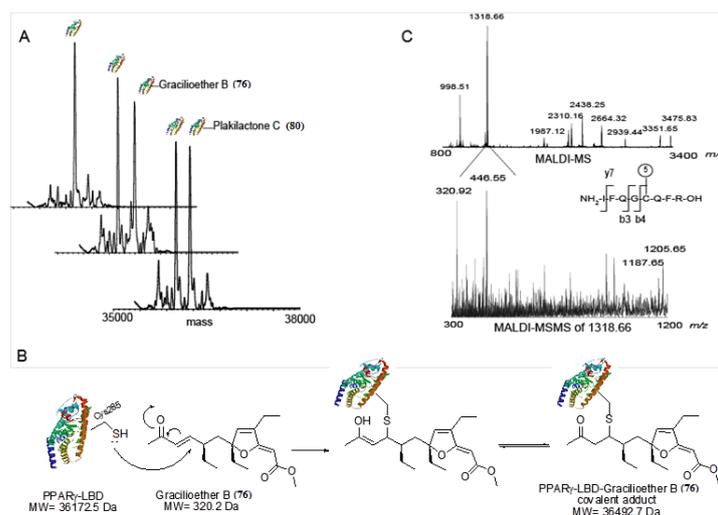
As shown in the Figure 4.10, several members of this series effectively transactivated PPAR $\gamma$  with compounds **76**, **77** and **80** being the most potent agonists. In addition, when incubated in presence of the synthetic ligand

rosiglitazone, methyl esters **72** and **73** attenuated the transactivation induced by this agent, thus acting as PPAR $\gamma$  antagonists. Analysis of the concentration/response curves for transactivation of PPAR $\gamma$  in response to rosiglitazone and to gracilioether B (**76**), gracilioether C (**77**) and plakilactone C (**80**) demonstrates that the marine compounds activate PPAR $\gamma$  in a dose-dependent manner with a relative EC<sub>50</sub> of  $\approx$  5, 10 and 2  $\mu$ M for compounds **76**, **77** and **80**, respectively (Figure 4.11). We then moved to a detailed analysis of the interaction mechanism at a molecular level, in order to assess the binding mode of these agents within PPAR $\gamma$ -LBD. First, a liquid chromatography-ESI MS (LC-ESI-MS)<sup>266,267,268</sup> approach was applied to detect the potential formation of PPAR $\gamma$ -LBD covalent complexes with gracilioether B (**76**), gracilioether C (**77**), plakilactone C (**80**), the methyl ester **72**, the  $\gamma$ -lactone **74** and plakilactone B (**79**), in physiologically relevant conditions.



**Figure 4.11** HepG2 cells were transiently transfected with pSG5GAL4-PPAR $\gamma$ LBD and p(UAS)5 $\times$ TK-Luc. 24 h Post transfection cells were treated with increasing concentrations of (A) rosiglitazone, (B) gracilioether B (**76**), (C) gracilioether C (**77**), (D) plakilactone C (**80**) and cell extracts subsequently assayed for luciferase activity. Data are the mean  $\pm$  S.E. of three experiments.

After incubation with PPAR $\gamma$ -LBD, a time-course analysis of the reaction mixtures was performed for each ligand, and the chromatograms revealed the presence of stable covalent complexes solely in presence of gracilioether B (**76**) and plakilactone C (**80**). Two species were detected in the LC-MS runs of gracilioether B (**76**) and plakilactone C (**80**), (Figure 4.12 panel A), that were identified, on the basis of their MW, as the unmodified PPAR $\gamma$ -LBD (MW of 36172.7 $\pm$ 0.3 Da) and as the 1:1 PPAR $\gamma$ -LBD/gracilioether B (**76**) or plakilactone C (**80**) covalent adducts, the last ones giving mass increments of 320 Da and 264 Da, respectively, compared to the free protein (Figure 4.12 panel A). These mass differences supported the hypothesis of a Michael addition between the natural compounds and PPAR $\gamma$ -LBD (Figure 4.12 panel C).

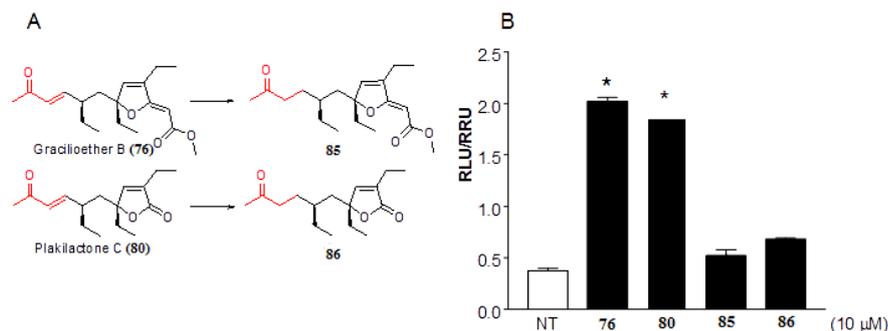


**Figure 4.12** Panel A: deconvoluted spectra of PPAR $\gamma$ -LBD alone (back), upon gracilioether B (**76**) (middle) and plakilactone C (**80**) incubation (front). Panel B: mechanism of the covalent modification of PPAR $\gamma$ -LBD by Michael addition on the  $\alpha,\beta$ -unsaturated ketone moiety in the side chain of **76** (or **80**). Panel C: MALDI-MS spectrum of PPAR $\gamma$ -LBD/gracilioether B (**76**) complex tryptic digestion and MALDI-MS/MS analysis of the ion at  $m/z$  1318.66.

Then, we moved to the identification of the punctual site of the covalent modification on PPAR $\gamma$ -LBD by gracilioether B (**76**) through a combination of classical protein chemistry protocols, MALDI-MS and MS/MS techniques. The chromatographic fraction containing the covalent complex was subjected to an extensive proteolysis with trypsin and analyzed by MALDI-MS. As reported in Figure 4.12 panel C and, the MALDI spectrum led us to identify a peak at  $m/z$  1318.66, corresponding to the peptide 281-288 (NH<sub>2</sub>-IFQGCQFR-COOH) containing the Cys285 and increased of 320.2 Da in its MW. Finally, MALDI-MS/MS analysis confirmed the correct peptide identification through the formation of the daughter ions at  $m/z$  390.92, 446.55, and 1205.65, attributed to b3, b4 and y7 fragmentations, respectively. Thus, Cys285 was unequivocally determined as the protein residue involved in the covalent binding with gracilioether B (**76**).

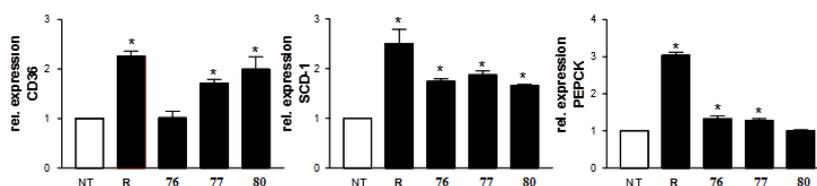
On the basis of structural considerations, the  $\alpha,\beta$ -unsaturated carbonyl moiety exclusively present in the side chain of gracilioether B (**76**) and plakilactone C (**80**) should act as specific Michael acceptor (Figure 4.12 panel B). To demonstrate this hypothesis, gracilioether B (**76**) and plakilactone C (**80**) were subjected to mild hydrogenation of the side chain double bond, giving the derivatives **85** and **86**, respectively (Figure 4.13, panel A).

Derivatives **85** and **86** were separately incubated with PPAR $\gamma$ -LBD, submitted to LC-ESI-MS analysis (see before) and, as expected, were unable to form covalent adducts with the PPAR $\gamma$ -LBD (covalent complex PPAR $\gamma$ -LBD/Agonists not found by LC-ESI-MS experiments) and, importantly, to transactivate PPAR $\gamma$  (Figure 4.13, panel B). All these data unequivocally revealed that the Cys285 targets the  $\alpha,\beta$ -unsaturated carbonyl moiety on the side chain of gracilioether B (**76**) and plakilactone C (**80**), moreover confirming gracilioether C (**77**), methyl esters **72** and **73** as non covalent PPAR $\gamma$  ligands.



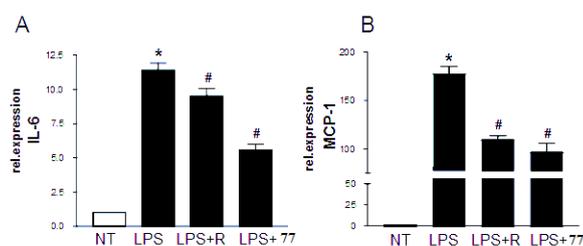
**Figure 4.13** A) Mild hydrogenation of gracilioether C and plakilactone C ( $H_2/Pt/C$ , 5 min, r.t). B) Luciferase reporter assay performed in HepG2 cells transiently transfected with a chimeric receptor expression plasmid pSG5GAL4-PPAR $\gamma$ LBD and with the reporter vector p(UAS)5 $\times$ TK-Luc. 24 h post transfection cells were stimulated for 18 h with 10  $\mu$ M compounds **76**, **80**, **85** and **86**. Data are the mean  $\pm$  S.E. of three experiments. \*P,0.05 versus not treated cells (NT).

We next examined whether gracilioethers B (**76**) and C (**77**) and plakilactone C (**80**) regulate the expression of genes that are known target of PPAR $\gamma$  in HepG2, a human hepatocarcinoma cell line, and in THP-1, a human monocytic leukemia cell line. For this purpose, HepG2 cells were exposed to rosiglitazone or gracilioethers B (**76**) and C (**77**) and plakilactone C (**80**) and the relative mRNA expression of SCD-1 (stearoyl-CoA desaturase-1), CD36 (cluster of differentiation 36) and PEPCCK (phosphoenolpyruvate carboxykinase)<sup>269,270</sup> were assessed by quantitative RT-PCR. As shown in Figure 4.14, gracilioether B (**76**), gracilioether C (**77**) and plakilactone C (**80**) exhibited a pattern of pharmacological activities full compatible with their ability to bind and transactivate PPAR $\gamma$ . All these agents increased the expression of SCD-1. Furthermore, gracilioether C (**77**) and plakilactone C (**80**) increased the expression of CD36, and gracilioethers B (**76**) and C (**77**) the expression of PEPCCK.



**Figure 4.14.** Serum starved HepG2 cells were stimulated for 18 h with 100 nM of rosiglitazone (R) or gracilioethers B (**76**), C (**77**) and plakilactone C (**80**), 10  $\mu$ M. Total RNA was extracted to perform Real-Time PCR of (A) CD36, (B) SCD-1 and (C) PEPCK. Values were normalized relatively to GAPDH mRNA and are expressed relatively to content of these genes in untreated cells, which are arbitrarily set to 1. Analysis was carried out in triplicate and the experiment was repeated twice. \* $P < 0.05$  versus not treated cells.

Finally, we measured whether gracilioether C (**77**), the non-covalent PPAR $\gamma$  agonist, effectively modulated PPAR-regulated genes in macrophages, a prototypical target of this nuclear receptor. As shown in Figure 4.15, pre-treating of THP-1 cells with rosiglitazone or gracilioether C (**77**) counter-regulated the induction of both the pro-inflammatory cytokine IL-6 (interleukine-6) and the MCP-1 (monocyte chemotactic protein-1) chemokine caused by LPS administration.



**Figure 4.15** Serum starved THP-1 cells were pre-treated for 3 h with 100 nM rosiglitazone (R) or gracilioether C (**77**), 10  $\mu$ M, before the administration of LPS (1  $\mu$ g/ml) for 18 h. Total RNA was extracted to analyze the relative mRNA expression of (A) IL-6 and (B) MCP-1 by Real-Time PCR. Values were normalized with respect to GAPDH mRNA and are expressed with respect to those of the untreated cells, which were arbitrarily set to 1. The analysis was carried out in triplicate and the experiment was repeated twice. \* $P < 0.05$  versus not treated cells. # $P < 0.05$  versus LPS treated cells.

## 4.2.2 Molecular modelling studies

In order to rationalize the binding mode in PPAR $\gamma$  of the afore mentioned compounds, docking calculations and Molecular Dynamics simulations were performed using Autodock-Vina<sup>28</sup> and Macromodel 8.5 software packages, respectively, taking into account that gracilioether B (**76**) and plakilactone C (**80**) are covalent ligands.

In this context, it should be mentioned that the activation of PPAR $\gamma$  by a covalent ligand depends also by its ability to establish, apart the covalent bond, additional weak interactions in the LBD.<sup>263,264,271</sup>

Indeed, Waku et al.<sup>265</sup> proposed a model, defined “dock and lock”, in which the first step (docking step) involves several non-covalent interactions of the putative ligand in the LBD, whereas in the second step (locking step) the covalent binding to the Cys285 is observed. As the apo and the locked form of this protein showed remarkable structural differences, in our docking studies on the covalent ligands **76** and **80**, two crystallographic structures of PPAR $\gamma$  were used: the apo-form (PDB code: 2ZK0) and a covalent complexed form with nitro-233 (PDB code: 2ZK5),<sup>265</sup> that was removed before the docking of our compounds.

The docking poses of gracilioether B (**76**), gracilioether C (**77**), and plakilactone C (**80**), the most active components of the library, were compared with the antagonists, methyl esters **72** and **73**, and with the inactive compounds plakilactone B (**79**) and gracilioether D (**84**).

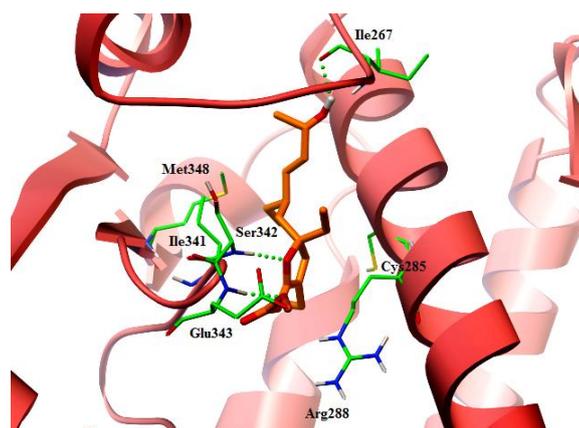
Within this approach, we confirmed that the active covalent ligands are docked in the apo form in poses compatible with the positioning of their reactive moieties around the Cys285, showing at the same time several interactions with key amino acid residues in the LBD. In the locked form of the receptor, we firstly tried to analyze the formation of the covalent bond using the recent introduced covalent docking methodologies.<sup>27</sup> Unfortunately, using the flexible side chain method, a

restricted space in the binding site of PPAR $\gamma$  was observed, thus determining steric clashes after the formation of the covalent bond.

On the other hand, using the covalent grid based approach, poses compatible with the covalent bond were found but with unfavorable predicted binding energies.

For these reasons, we concluded that putative models of the covalent ligands could be better detected by using Molecular Dynamics simulations, analyzing their motions in the LBD of PPAR $\gamma$ .

Results from these experiments allowed us to identify some significant poses in which the distances between the reactive part of these molecules and the sulfur atom on the Cys285 were compatible with the formation of the C-S covalent bond.



**Figure 4.16** Gracilioether C (**77**), (coloured by atom types: C orange, O red, -OH hydrogen light grey) in docking with PPAR $\gamma$ -LBD of the apo form. Residues are coloured by atom type: C green, H light grey, O red, N blu, S yellow.

For what concerns non-covalent ligands, the docking model in the apo form shows that the agonist gracilioether C (**77**) entails a set of weak interactions, with the -OH group in the side chain determining a favorable accommodation in the LBD of PPAR $\gamma$ . In particular, in the apo form, **77** establishes Van der Waals

interactions with the Ile341 and the Met348, and polar interactions with the Arg288.

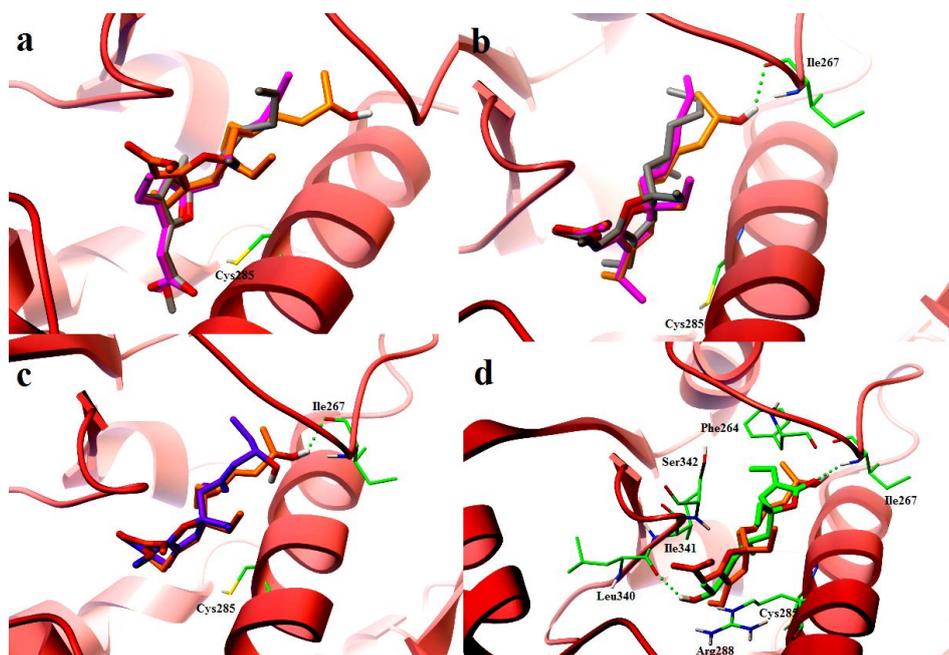
Three hydrogen bonds stabilize the compound in the LBD, between the ether oxygen of the cycle and the -NH of the Ser342, between the terminal oxygen of the ester moiety and the -NH of the Glu343, and finally between the -OH at position-11 and the CO group of the Ile267 (Figure 4.16).

Methyl esters **72** and **73** differ from gracilioether C (**77**) for the presence of a fully hydrophobic side chain that causes a different orientation in the PPAR $\gamma$ -LBD, with a flip of the ring of  $\sim 180^\circ$  on the major axis of the compounds (Figure 4.17 a) and then with the loss of some essential interactions.

Indeed, we found alternative poses in which the  $\alpha,\beta,\gamma,\delta$ -unsaturated methyl ester moiety of **77**, **72** and **73** are well superimposed (Figure 4.17 b), but also in this case the hydrophobic side chains of **72** and **73** do not allow any polar interaction as observed for **77**, mainly the hydrogen bond with the Ile267.

As demonstrated for 15d-PGJ<sub>2</sub>, this residue plays a fundamental role in the activity of a putative PPAR $\gamma$  covalent agonist. Indeed, even if the covalent binding at the Cys285 is maintained, mutations at this residue abolish the transcriptional activation induced by this endogenous PPAR $\gamma$  agonist.<sup>265</sup>

Similar considerations could be done for gracilioether D (**84**). Also in this case, the superposition with the gracilioether C (**77**) shows a similar orientation of the  $\alpha,\beta,\gamma,\delta$ -unsaturated methyl ester core but a different accommodation of the side chain with the -OH group at C-10 far from the Ile267 (Figure 4.17c).



**Figure 4.17** Superimposition between **77** (coloured by atom types: C orange, O red, -OH hydrogen light grey), and: a) first docking model and b) alternative docking model of **72** (coloured by atom types: C grey, O red), and **73** (coloured by atom types: C purple, O red); c) gracilioether D (**84**) (coloured by atom types: C violet, O red, -OH hydrogen light grey); d) plakilactone B (**79**) (coloured by atom types: C green, O red, -OH hydrogen light grey) in docking with PPAR $\gamma$ -LBD of the apo form. Residues are coloured by atom type: C green, H light grey, O red, N blue, S yellow. Hydrogen bonds are displayed with green spheres.

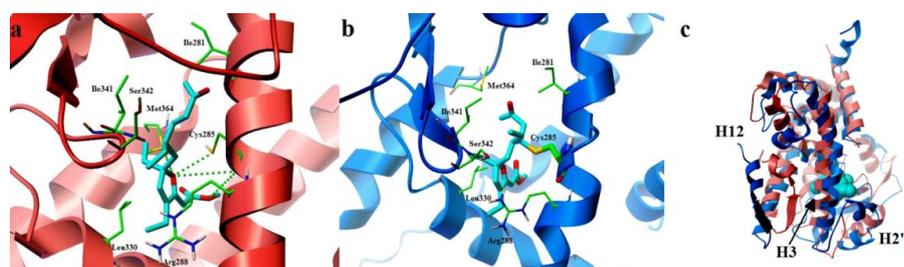
In summary, the inverted configuration at C-10 and the presence of a  $\Delta^8$  double bond with respect to the side chain of gracilioether C (**77**) could explain the loss of the bioactivity for **84**.

Of interest is the case of the inactive plakilactone B (**79**) displaying the same lactone moiety of plakilactone C (**80**) and the side chain of gracilioether C (**77**). Its accommodation in the PPAR $\gamma$ -LBD is inverted (Figure 4.17 d), probably due to the absence of the  $\alpha,\beta,\gamma,\delta$ -unsaturated methyl ester moiety. As a consequence, the

smaller cyclic  $\alpha$ - $\beta$  unsaturated lactone moiety is oriented to form a hydrogen bond with the Ile267, and the -OH in the side chain is able to establish only one hydrogen bond (Leu340) where **77** generated a wide set of polar interactions. In summary, **79** shows some potential interesting points in its chemical structure, but in perfect agreement with absence of activity towards PPAR $\gamma$ , the inability to create a covalent bond and the presence of a small cyclic part cause an unstable placement in the LBD.

Regarding the covalent agonists, for gracilioether B (**76**) a docking pose in the apo form of PPAR $\gamma$  in which the  $\alpha$ - $\beta$  unsaturated ketone in the side chain is oriented towards the Cys285 was found, and this conformation is stabilized by several hydrogen bonds between the  $\alpha,\beta,\gamma,\delta$ -unsaturated methyl ester moiety and the Cys285, showing a strong point of attach of the compound in proximity of this residue. Moreover, **76** establishes further polar interactions with the Ile281 and the Arg288, as well as hydrophobic interactions with the Met264, the Leu330, and the Ile341 (Figure 4.18 a). Therefore, as previously demonstrated,<sup>265</sup> in the “docking step” a putative covalent ligand searches, through a wide range of weak interactions, the best conformation favorable to the covalent bond.

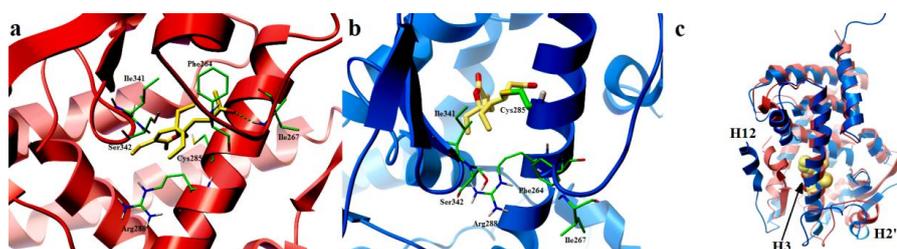
In the locked form, a conformation in which the  $\alpha$ - $\beta$  unsaturated ketone of **76** is closed to the Cys285 (distance between the sulfur of Cys285 and the reactive  $\beta$  carbon of the  $\alpha$ - $\beta$  unsaturated ketone of **76** = 3.439 Å) was found. Starting from this conformation, several fast steps of Molecular Dynamics simulations were applied observing a gradual reduction of this distance. When a pose with a distance between these groups compatible with the C-S covalent bond was found (~1.8 Å), the covalent bond was manually generated and the complex was processed to a further fast step of Molecular Dynamics simulation (Figure 4.18b).



**Figure 4.18** Docking and covalent models of gracilioether B (**76**) (coloured by atom types: C sky blue, O red) in docking with PPAR $\gamma$ -LBD of a) apo form; b) locked form after Molecular Dynamics simulations. Residues are coloured by atom type: C green, H light grey, O red, N blue, S yellow; c) superimposition between the apo form of PPAR $\gamma$  (secondary structure represented in red) and the locked form complexed with gracilioether B (**76**) (secondary structure represented in blue, **76** is in CPK representation and coloured by atom types: C sky blue, O red).

As previously reported, several alterations in the regions following the helix H2' and a considerable rearrangement of the helices H3 and H12 could be observed comparing the apo form and the locked form of the receptor (Figure 4.18c).<sup>265</sup>

Similarly, the side chain  $\alpha$ - $\beta$  unsaturated ketone of plakilactone C (**80**) is prone to function as Michael acceptor and, in Figure 4.19 a, a pose in the apo form in which this moiety is near to the Cys285 is shown. A hydrogen bond was observed between the carbonyl oxygen of the  $\alpha$ , $\beta$ -unsaturated lactone moiety and the -NH of the Ile267 (Figure 4.19 a)<sup>265</sup> as well as further polar interactions are established between the cyclic core and the Arg288 and the His266. In the locked form of PPAR $\gamma$ , the docking results confirmed the nearness of the  $\alpha$ - $\beta$  unsaturated ketone in the side chain of **80** to the Cys285 (distance between the sulfur of Cys285 and the reactive  $\beta$  carbon of the  $\alpha$ - $\beta$  unsaturated ketone of **80** = 3.918 Å).



**Figure 4.19.** Docking and covalent models of plakilactone C (**80**) (coloured by atom types: C light yellow, O red) in docking with PPAR $\gamma$ -LBD of a) apo form; b) locked form after Molecular Dynamics simulations. Residues are coloured by atom type: C green, H light grey, O red, N blue, S yellow. Hydrogen bonds are displayed with green spheres; c) superimposition between the apo form of PPAR $\gamma$  (secondary structure represented in red) and the locked form complexed with plakilactone C (**80**) (secondary structure represented in blue, **80** is in CPK representation and coloured by atom types: C light yellow, O red).

As for **76**, this complex (Figure 4.19 b) was processed through Molecular Dynamics simulations observing also in this case the rearrangement of the helices H3, H12 and H2' (Figure 4.19 c).

### 4.2.3 Final remarks

In summary, the isolation and pharmacological characterization of a family of oxygenated polyketides from the *Plakinastrella mamillaris* sponge was reported. The detailed analysis of pharmacological properties of these agents allowed us to demonstrate that members of this library are robust and selective ligands of the nuclear receptor PPAR $\gamma$ . PPAR $\gamma$  is a well validated pharmacological target. Thiazolidinedione, rosiglitazone and pioglitazone are potent PPAR $\gamma$  agonists and insulin-sensitizers and have been extensively used in the treatment of type 2 diabetes.

Thiazolidinediones induce the transcription of PPAR $\gamma$  responsive genes and control lipid synthesis and storage in the adipose tissue, liver and many other tissues, however their use is associated with side effects including weight gain, fluid retention, and increased risk of heart failure. Rosiglitazone has been withdrawn from the market in 2011, and pioglitazone is contraindicated in patients with New York Heart Association (NYHA) Class III or IV heart failure in USA, while in Europe is contraindicated in patients with any stage of heart failure.<sup>270,272-277</sup>

Despite pioglitazone does not increase the risk of myocardial infarction, and its use associates with a reduction in all-cause mortality, there is an urgent need for development of novel PPAR $\gamma$  agonists or modulators.

Previous studies from our group have led to the demonstration that marine organisms are a rich source of ligands for nuclear receptors. Thus, we have identified ligands for two major targets, the farnesoid-X-receptor (FXR) and the pregnane-X-receptor (PXR).<sup>246</sup> Furthermore marine environment, mainly sponge organisms, has been reported as the source of nuclear receptor antagonists,<sup>278</sup> that are currently being developed for their biomedical potential.

Members of this oxygenated polyketides library showed the ability to activate PPAR $\gamma$  in a transactivation assay in HepG2 cells transfected with a viral vector containing the LBD of the receptor whereas others reverted the effect of the synthetic ligand rosiglitazone, thus acting as PPAR $\gamma$  antagonists. Results from these studies demonstrated that gracilioether B (**76**), gracilioether C (**77**) and plakilactone C (**80**) activate the receptor with EC<sub>50</sub> ranging from 2-9  $\mu$ M, and are therefore 20-90 fold less potent than rosiglitazone whose EC<sub>50</sub> is  $\approx$ 100 nM in this assay.

Despite their reduced potency, the efficacy of gracilioethers B (**76**), gracilioether C (**77**) and plakilactone C (**80**) in term of receptor transactivation was very similar to that of rosiglitazone ( $\approx$ 80%), and in addition, all three compounds effectively triggered the transcription of PPAR $\gamma$ -regulated genes, although with a difference in the relative potency (Figure 4.14). Using HepG2 cells, we have shown that gracilioether C (**77**) and plakilactone (**80**) induce the expression of CD36, a scavenger receptor involved in the hepatic uptake of oxidized lipoprotein, a typical effect of PPAR $\gamma$  ligands in the liver. Interestingly, gracilioether B (**76**) fails to increase the expression of this gene in HepG2.

This finding might be of relevance because increased expression of CD36 caused by PPAR $\gamma$  ligands is thought to mediate lipid accumulation in macrophages and liver. Thus, the fact that gracilioether B (**76**) causes a different pattern of gene expression in comparison to rosiglitazone, might be associated with a specific pharmacological profile *in vivo*. In addition, we also demonstrated that gracilioether B (**76**), gracilioether C (**77**) and plakilactone C (**80**) increase the expression of SCD-1 in liver cells. SCD-1 catalyzes the rate-limiting reaction of monounsaturated fatty acid synthesis and plays an important role in the development of fatty liver. Finally, we demonstrated that gracilioether B (**76**) and gracilioether C (**77**) but not plakilactone C (**80**) up-regulate the expression of

PEPCK. PEPCK is a rate limiting enzyme involved in gluconeogenesis and glyceroneogenesis pathways and its expression is tightly regulated at the transcriptional level by hormones controlling glucose homeostasis with glucagon and glucocorticoids having a strong gluconeogenic action while insulin inhibiting hepatic gluconeogenesis by repressing the expression of this gene.<sup>279</sup>

The expression of PEPCK is positively regulated by different transcription factors and co-activators including hepatic nuclear factor 4 $\alpha$  (HNF4 $\alpha$ ), Forkhead box O1 (Foxo1) and PPAR $\gamma$ , while the PPAR $\gamma$  coactivator 1 alpha (PGC1- $\alpha$ ) has been shown to cooperate in regulating the expression of this gene in the fasting state. Induction of PEPCK in HepG2 cells by rosiglitazone and gracilioethers B (76) and C (77) is therefore of interest. Indeed, since HepG2 cells are grown in a low glucose medium, induction of PEPCK expression drives cell metabolism to glycerogenesis rather than gluconeogenesis and might be involved in development of lipid accumulation in hepatocytes, a common side effect in the rosiglitazone therapy.<sup>279</sup> All together these data suggest the possibility to develop novel PPAR $\gamma$  modulators.

Because gracilioether C (77) activates PPAR $\gamma$  in a non covalent fashion, we have then investigated whether this agent still exerts the same range of effects of rosiglitazone. Using THP-1, a monocytic cell line, we demonstrated that gracilioether C (77) effectively modulates the expression of two inflammatory mediators, IL-6 and MCP-1.

Thus, similarly to rosiglitazone, gracilioether C (77) causes a robust attenuation of the expression of IL-6 and MCP-1 triggered by LPS. Because the inhibition of pro-inflammatory mediators is a common theme in the pharmacology of PPAR $\gamma$  ligands, present data provide a robust evidence that gracilioether C (77) might be a potential agent in the treatment of inflammatory disorders.<sup>280,281</sup>

At molecular level we demonstrated that gracilioether B (**76**) and plakilactone C (**80**) covalently bind to a cysteine residue in the PPAR $\gamma$ -LBD through a Michael addition reaction to the  $\alpha,\beta$ -unsaturated ketone in their side chains. Such findings were also supported by an integrated analysis of docking and Molecular Dynamics simulations.

Besides the Cys285 is conserved in all three PPAR subtypes, gracilioether B (**76**) and plakilactone C (**80**) showed a significant specificity towards PPAR $\gamma$ . This finding indicates, as previously reported for several selective covalent PPAR $\gamma$  ligands,<sup>282,283</sup> that other amino acid residues confer specificity in the recognition process to PPAR $\gamma$ -LBD and moreover points towards the importance of the docking step in which the putative covalent ligand establishes several non-covalent interactions.

Also for gracilioether C (**77**) a detailed docking analysis was performed in order to rationalize the structural requirements for its non covalent interaction in the receptor's LBD and insights were gained to explain its peculiar mode of action. The analysis of the docking poses in comparison with the antagonists methyl esters **72** and **73** and several non active members of this series clarified the chemical requirements for the PPAR $\gamma$  agonistic activity and could be useful for the future de-novo design and for the prediction of the bioactivity of a set of new ligands.

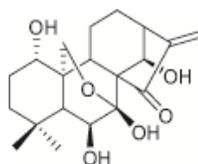
Also the discovery that methyl esters **72** and **73** are PPAR $\gamma$  antagonists<sup>284</sup> that counteract the receptor transactivation caused by rosiglitazone is noteworthy. Because PPAR $\gamma$  antagonists are of pharmacological and therapeutic relevance, we are currently elaborating on these structures to gain further insights on their pharmacological profiles.

In conclusion this study discloses a new class of marine PPAR $\gamma$  ligands structurally unrelated to all synthetic and natural ligands so far reported and

reaffirms the extraordinary chemodiversity and therapeutic potential of marine natural compounds.

### 4.3 HSP70 1A as a Target for the Anticancer Diterpene Oridonin

Kaurane diterpenes are a very important class of natural compounds, identified from numerous medicinal plants, possessing several biological activity such as plant growth regulators, antimicrobial, antiviral, inflammation, and antitumoral activity.<sup>285</sup> Kaurane diterpene oridonin is the principal active metabolite of *Rabdosia rubescens* (Hemsl.) Hara (Donglingcao), Hook. f. (Leigongteng) used in Chinese medicine to treat several diseases. Numerous biological activities such as cytotoxic effect, antiproliferative, anticancer, anti-inflammatory and antibacterial activities have been reported for this compound.<sup>286</sup> Oridonin (**87**) (Figure 4.20) has aroused high interest especially in cancer researchers due to its potential to be developed into tumor chemotherapeutic drug.



**Figure 4.20** Chemical structure of oridonin

The ability to inhibit tumor cell growth either by *in vitro* as well by *in vivo* experimental models has indeed been repeatedly confirmed by many research groups.<sup>287,288</sup> Inhibition of tumor cell growth by oridonin was ascribed to the ability of the drug to affect cell cycle progression and/or to promote cell death by apoptosis and autophagy.<sup>289</sup> Depending on cell type, oridonin has been shown to induce cell cycle arrest in G2/M or G1/S and to modulate the expression/activity of different cell cycle regulatory proteins. Again, depending on the cell type and the experimental conditions used, oridonin has been shown to modulate the expression of proteins implicated in either death receptor-mediated (i.e. FAS, FAS ligand) or

mitochondria-dependent apoptotic pathways (i.e. increased BAX/Bcl2 ratio, caspases). Moreover oridonin was reported to modulate several proteins, such as MAPKs, PI3K/Akt, and NFkB, responsible of death (apoptosis/autophagy) or survival cell fate have been observed in different cell lines.<sup>290</sup> Modulation of some of the aforementioned proteins (i.e. p21 and NFkB)<sup>291</sup> by oridonin was also related to the simultaneous induction of apoptosis and autophagy or changed balance between the two cell death mechanisms.

Despite the large number of proteins which activity and/or expression have been shown to be modulated by oridonin, it has not yet identified which is the primary target of this diterpene. Hence, we have tried to mine target proteins of oridonin by employing a chemical proteomic approach to gain more insight into the molecular mechanisms of oridonin.

Chemical proteomics is a compound-centric approach aimed to describe the mechanism of action of bioactive small molecule at a molecular level; it is a powerful mass spectrometry (MS)-based affinity chromatography approach for identifying proteome-wide small molecule-protein interactions; mapping these interactions on signaling and metabolic pathways, could comprehensively characterize drug targets, profile toxicity of known drugs, and lead to the identification of possible off-target activities.<sup>292</sup>

Recently, there is an increasing interest in application of chemical proteomics experiments to bioactive natural products. Indeed, even if natural products have been the single most productive source of leads for the development of drugs,<sup>293</sup> information concerning mechanism of action at a molecular level of many of them are very poor or completely missed.

Chemical Proteomics revealed the binding and inhibition of oridonin to the multifunctional, stress-inducible molecular chaperone HSP70-1A. This specific

activity was rationalized combining docking experiments and Molecular Dynamics simulations.

However, since it is clearly emerging that many secondary metabolites can act on various targets at the same time,<sup>293</sup> further studies will be performed to validate the possible interaction of oridonin with other proteins identified in chemical proteomics studies (i.e. Peroxiredoxin-1,  $\alpha$ -Enolase, and Cofilin-1).

### 4.3.1 Chemical proteomics results

Chemical proteomics experiments were performed to identify oridonin molecular target(s) responsible for its anticancer activity. In that aim, oridonin was biotinylated, taking care to prevent the reaction of the  $\alpha,\beta$ -unsaturated carbonyl group, essential for the activity of ent-kaurane diterpenes<sup>294,295</sup> and known to contribute to the biological activity of many other natural compounds such as sesquiterpenes<sup>296</sup> and indolizidine alkaloids.<sup>297</sup> Before using biotinylated oridonin in chemical proteomic analysis for target identification, it was mandatory to verify that the chemical modification did not cause a marked loss of oridonin biological activity. To that aim, we compared Jurkat cell proliferation inhibition potential of biotinylated oridonin to that of oridonin.

Chemically modified oridonin and oridonin inhibited cell growth to a comparable extent, as inferred by the  $7.5 \pm 0.5 \mu\text{M}$  and  $5 \pm 0.3 \mu\text{M}$   $\text{IC}_{50}$  values (at 24 h of exposure) measured for biotinylated oridonin and oridonin, respectively. Even if biotin conjugation can produce an alteration of small molecules membrane permeability,<sup>298</sup> this result demonstrated that biotinylated oridonin could be used for identifying oridonin cellular target(s). In characterizing binding partners for a small molecule by chemical proteomics, the major challenge is to identify bona fide interacting partners since very high sensitivity of MS analysis can permit the

identification of almost all proteins, even contaminants present at very low levels in the sample.

Therefore, Jurkat cell lysates were incubated with biotinylated oridonin or with biotin, as a negative control, to discriminate between proteins specifically interacting with the diterpene and unspecific background.<sup>292</sup> Samples were purified by affinity chromatography on a streptavidin resin and the resulting protein mixtures were resolved by SDS-PAGE; each gel line was then cut in 13 pieces, digested with trypsin, and analyzed by mass spectrometry through nanoflow reversed-phase HPLC MS/MS.

Doubly and triply charged peptide species were fragmented, and all the MS/MS spectra were evaluated by a Mascot database search. Chemical proteomic experiments were performed in triplicate to ensure data reproducibility leading to the identification of about 85% of identical proteins in the three experiments.

To establish the proteins specifically captured by biotinylated oridonin, we compared the list of the proteins detected with those in control experiment. This analysis led to the identification of four potential partners of oridonin (Table 4.1). These proteins were always detected in the samples incubated with biotinylated oridonin and never revealed in those from control gels.

The same experiment was also performed using proteins extracted from PC3 cell line (human prostate cancer cell line) leading to the definition of the same putative targets. Among the four putative oridonin target the higher score was achieved for HSP70 1A.

These findings together with the multiple role of HSP70 in the control of cell proliferation, apoptosis and autophagy,<sup>299</sup> prompted us to further characterize the interaction of oridonin with this key target.

Swiss-Prot code	Protein Identified	Mr	Sequence coverage	Matches	Sequences	Mascot Score
HSP71_HUMAN	Heat shock 70 kDa protein 1A/1B	70052	17%	7	7	265
ENOA_HUMAN	Alpha-enolase	47169	36%	6	6	181
PRDX1_HUMAN	Peroxiredoxin-1	22110	15%	5	5	158
COF1_HUMAN	Cofilin-1	18502	28%	4	4	156

**Table 4.1** Proteins identified by chemical proteomic approach as putative oridonin molecular targets

Heat shock protein 70 (HSP70) is a member of a ubiquitously expressed family of molecular chaperones that are involved in protein homeostasis. In its role as a mediator of protein fate, this chaperone has been linked to multiple tasks, including roles in *de novo* protein folding, subcellular trafficking, protein disaggregation, proteasome-mediated degradation, and autophagy. To accomplish its various chaperone functions, HSP70 physically interacts with the exposed hydrophobic residues of polypeptides *via* its C-terminal substrate binding domain (SBD). Hydrolysis of ATP in the adjacent, N-terminal nucleotide-binding domain (NBD) propagates an allosteric change to the SBD, resulting in an approximately 10-fold enhancement in substrate affinity. These findings suggest an important role for the nucleotide state in controlling the interactions of HSP70 with misfolded substrates.

HSP70 has a long-term connection with the cancer problematic: at the initial stages of tumorigenesis, HSP70 can protect cells undergoing transformation from oncogenic stress induced by over-expression of oncogenes, e.g. c-myc.<sup>299</sup> In addition, HSP70 has been shown to suppress cellular senescence, an important anti-tumor mechanism at the early stages of tumorigenesis and also important in the proper response to anti-cancer therapy.

HSP70 over-expression has been routinely associated with poor prognosis in multiple form of cancer and is thought to provide a survival advantage to cancer

cells interacting with multiple component of both caspase-dependent and caspase-independent apoptotic pathway. Moreover, many evidences demonstrated a critical role played by HSP70 in the control of so-called chaperone-mediated autophagy.<sup>300,301</sup>

With the critical roles played by HSP70s in cancer, neurodegenerative disease, viral infection, and injury, one might anticipate that many small molecule modulators for this class of chaperones have been identified, at moment few molecules are known as HSP70 inhibitors.<sup>302,303</sup> Indeed, several compounds were demonstrated to modulate the ATPase activity of HSP70 and/or to affect its affinity towards client proteins or co-chaperones,<sup>306</sup> but in most cases the exact site and mode of binding of these molecules to the chaperone are currently unknown.<sup>304,305</sup> The complete elucidation of the binding site is however not secondary, since it could also determine the selectivity and the efficiency of an inhibitor.<sup>306</sup> Using a mass based-chemical proteomics approach, HSP70 was revealed as a molecular target of the anticancer diterpene oridonin.

This finding was validate by different western blot analyses, also confirming that the diterpene can interact with HSP70 under pseudo-physiological conditions inside Jurkat cells and demonstrating the selectivity of oridonin towards this chaperone. Biochemical and structural studies were performed allowing to elucidate the molecular basis of the inhibition of HSP70 by oridonin, demonstrating that the interaction of the diterpene with the NBD of the chaperone prevents HSP70 interaction with ATP.<sup>313</sup>

Using a molecular docking approach, a model that rationalize the inhibitory activity of oridonin describing the dynamic modification of the orientation of the helixes surrounding the NBD of HSP70 is proposed. Besides, our data demonstrated that oridonin interaction with HSP70 NBD is also stabilized by the formation of a covalent bond involving the thiolic group of Cys267; however, this

covalent binding seems not to be critical for the biological activity of oridonin, as inferred by the observation that the diterpene also inhibits ATPase activity of a prokaryotic HSP70 (DnaK) lacking of this residue.

The identification of NBD as the binding region of oridonin on HSP70 suggests that oridonin could affect both HSC70 and HSP70 activity because the N-terminal region of these two proteins shares about 99% of their structure.<sup>307</sup>

This result is relevant since a study by Power et al. demonstrated that the simultaneous silencing of HSP70 and HSC70 is required to impair HSP90 function, thus enhancing the pro-apoptotic effects of HSP70 inhibition.<sup>306</sup>

Conformation changes occurring on the NBD could also account for the enhanced affinity between HSP70 and its co-chaperone BAG3, observed following oridonin interaction.

### 4.3.2 Biological studies

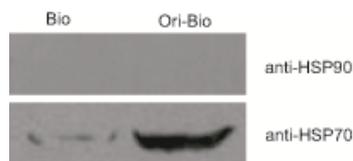
Oridonin/HSP70 interaction was investigated in Jurkat cells, taken as a more physiological model. Cells were exposed to doses (10  $\mu$ M and 20  $\mu$ M) higher than  $IC_{50}$  of biotinylated oridonin for 5 h.

We chose to use high doses to increase the possibility to detect the biotinylated oridonin-HSP70 complex.

On the other hand, the time was shortened to 5 h to prevent extensive cell death. Binding of biotinylated oridonin with HSP70 or HSP90 was monitored *in cell* fishing procedure followed by Western blotting.

Resulting blots (Figure 4.21) showed that also inside cells oridonin efficiently interacts with HSP70 1A, but not with HSP90  $\alpha$ .

The affinity of oridonin towards HSP70 1A was evaluated by a surface plasmon resonance (SPR) based binding assay.

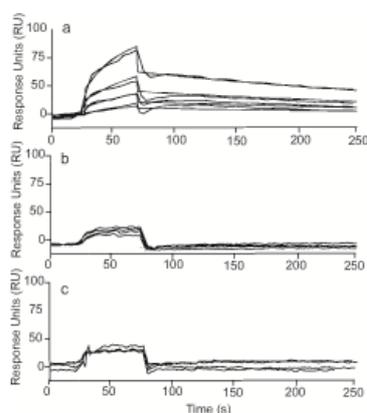


**Figure 4.21** *In cell fishing* procedure. Western blotting analysis of the binding of biotinylated oridonin to HSP70 or HSP90. The same blot, representative of two replicate experiments, was sequentially probed with anti-HSP70 and anti HSP90 antibodies.

The diterpene efficiently interacted with the immobilized protein, as demonstrated by the concentration dependent responses and by the clearly discernible exponential curves, during both the association and dissociation phases (Figure 4.22 a).

Fitting the achieved sensorgram to a single-site bimolecular interaction model ( $A+B=AB$ ), a thermodynamic dissociation constant ( $K_D$ ) of  $26.0 \pm 1.2$  nM and a kinetic dissociation constants ( $k_{off}$ ) of  $2.7 \pm 0.4$  ms<sup>-1</sup> were calculated for the HSP70/oridonin complex.

This result demonstrated high affinity of oridonin towards the chaperone, and the low  $k_{off}$  measured suggested a very high stability of the resulting complex. Since HSP70s are ATP dependent chaperones, SPR analysis was then performed adding 5  $\mu$ M ATP into the running buffer. Remarkably, the presence of an ATP concentration 10-times higher than its reported  $K_D$  (500 nM)<sup>308</sup> almost completely inhibited oridonin binding to the chaperone (Figure 4.22 b).



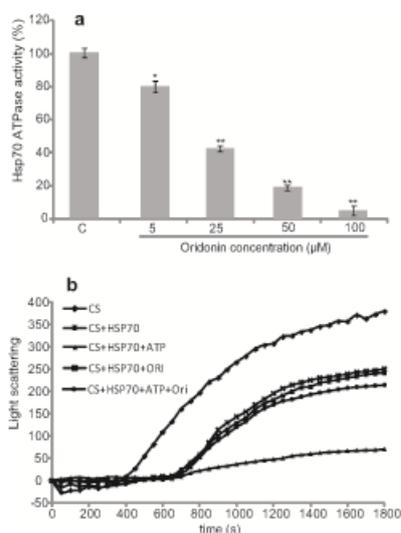
**Figure 4.22** SPR results. Sensorgrams obtained by injecting different concentrations (from 0.020 to 1  $\mu\text{M}$ ) of oridonin on immobilized HSP70 using PBS (A) or 5  $\mu\text{M}$  ATP in PBS as running buffer (B), or on immobilized HSP90 (C)

This result could arise either from a competition of oridonin and ATP for the same binding site or from conformational changes occurring on the HSP70 structure following ATP binding and preventing oridonin interaction. To verify the selectivity of oridonin towards HSP70, the diterpene was also injected on a HSP90  $\alpha$  modified chip; the obtained sensorgrams (Figure 4.22 c) showed no interaction.

To evaluate the effect of oridonin binding on HSP70, ATPase activity was measured in the presence of different amounts of oridonin using the approach described by Chang et al.<sup>309</sup> oridonin was found to inhibit in a dose-dependent manner HSP70 ATPase activity (Figure 4.23 a). The ability of oridonin to affect the chaperone activity of HSP70 was tested monitoring the thermal-induced citrate synthase (CS) aggregation under different experimental conditions (Figure 4.23 b). Upon incubation at elevated temperatures, CS underwent quantitative protein aggregation, but the presence of stoichiometric amounts of HSP70 changed the aggregation kinetics. As expected, the addition of a 4-fold molar excess of ATP further decreased the curve slope, since ATP binding to the N-terminal domain of HSP70 enhances the  $k_{on}$  of its interaction with substrate proteins.<sup>310</sup> When also

oridonin was added, CS aggregation curve became almost comparable to that observed without ATP, thus indicating that oridonin can revert the effect of the triphosphate nucleotide on HSP70 chaperone activity.

Finally, we investigated whether oridonin binding to HSP70 affected the ability of the chaperone to interact with other proteins. In particular, since oridonin inhibits ATPase activity, we focused our attention on those co-chaperones modulating ADP release from HSP70 (Nucleotide Exchanging Factors, NEF), such as the BAG proteins.<sup>311</sup>



**Figure 4.23** Oridonin effect on HSP70 activity. Inhibition of the ATPase activity of HSP70 (DnaK) by different concentration of oridonin (A). Data are the mean of two independent experiments performed in triplicate and were analyzed by *t* student test (Control vs testing compounds): \**p* < 0.05, \*\**p* < 0.005. Aggregation kinetics of CS at 43 °C determined by light scattering (B). The spontaneous aggregation of CS at 43 °C (◆) and the aggregation of CS at 43 °C in the presence of 0.075 µM Hsp70 (■), of 0.075 µM HSP70 and 0.3 µM ATP (▲), of 0.075 µM HSP70 and 0.3 µM oridonin (x), or of 0.075 µM HSP70, 0.3 µM ATP and 0.3 µM oridonin (●) are shown. Kinetics traces reported are the averages of two measurements.

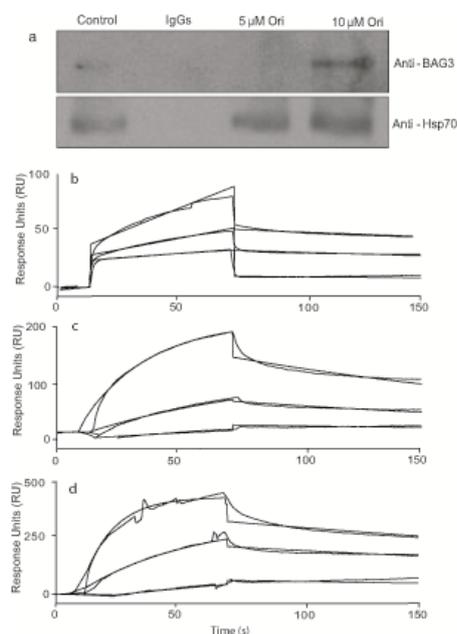
Moreover, the effect of oridonin on chaperone interaction with some client proteins (i.e. EGFR and ERK1/2) was evaluated. Therefore, we performed co-

immunoprecipitation in Jurkat cells exposed to 5  $\mu\text{M}$  and 10  $\mu\text{M}$  oridonin for 5 h. The most interesting result was achieved on the interaction between HSP70 and its co-chaperone BAG3, since we found that oridonin, even though only at the highest concentration tested (10  $\mu\text{M}$ ), promoted HSP70-BAG3 binding (Figure 4.24 a).

To confirm the ability of oridonin to enhance the affinity of HSP70 towards BAG3, some SPR experiments were carried out injecting BAG3 on immobilized HSP70, in the presence of different concentrations of the diterpene. The interaction was firstly tested using PBS as running buffer, allowing measuring a  $K_D$  for the HSP70/BAG3 complex of  $49.8 \pm 1.9$  nM (Figure 4.24 b); when the same analyses were performed using 1  $\mu\text{M}$  or 10  $\mu\text{M}$  oridonin in PBS as running buffers (Figure 4.24 c,d), the  $K_D$  of the resulting complex lowered to  $12.1 \pm 0.7$  nM and  $8.5 \pm 0.5$  nM, respectively, indicating that the interaction of oridonin with HSP70 increased the binding efficiency of the chaperone with its partner.

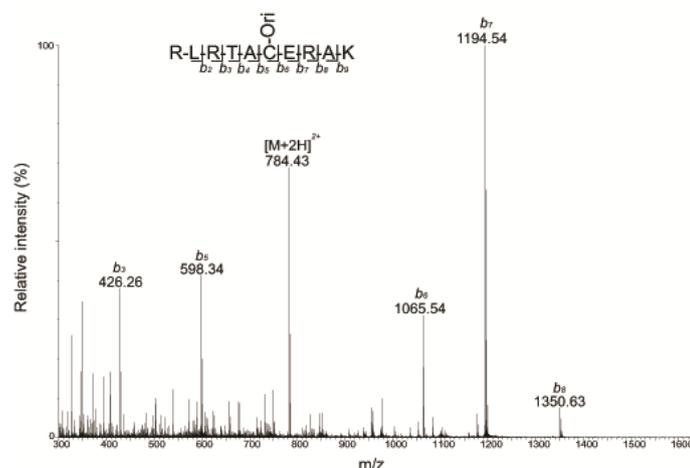
The high stability of the HSP70/oridonin complex observed in SPR analyses and the presence of the  $\alpha,\beta$ -unsaturated carbonyl group in the structure of the diterpene, led us to investigate the possible formation of covalent bond(s) between oridonin and protein thiol group(s), via Michael reaction.

Therefore, HSP70 was incubated with a 2-fold molar excess of oridonin under pseudo-physiological conditions and the resulting complex underwent a classical peptide mapping procedure aimed to identify possible covalently modified peptides.



**Figure 4.24** Effect of oridonin binding on HSP70-BAG3 interaction. Jurkat cells were exposed to 5  $\mu\text{M}$  and 10  $\mu\text{M}$  oridonin or vehicle only (control) for 5 h. Lysates were immunoprecipitated with anti-HSP70 antibody or IgGs and blots probed with anti-BAG3 antibody (A). The blot shown is representative of two replicate experiments. Sensorgrams obtained by injecting different concentrations (from 0.010 to 1  $\mu\text{M}$ ) of BAG3 on immobilized HSP70 using PBS (B), 1  $\mu\text{M}$  oridonin in PBS (C), or 10  $\mu\text{M}$  oridonin in PBS as running buffer (D).

This analysis also revealed the presence of two species displaying a molecular mass of 1693.661 and 1566.835 respectively, assigned to peptides 259-269 and 262-271, both carrying one oridonin linked on the same Cys residue (Cys267). This identification was confirmed by MS/MS spectra (Figure 4.25), and indicated Cys267 as the only covalent binding site of oridonin on HSP70.



**Figure 4.25** MS/MS analysis of oridonin-modified peptide. CID fragmentation spectrum of the double charged ion  $[M+2H]^{2+}$  at  $m/z$  784.43, assigned to HSP70 peptide 259-269 modified by one oridonin linked on the Cys267. The peptide sequence is also reported.

It is noteworthy that the Cys267 is in the ATPase domain of HSP70; it has been experimentally demonstrated that molecules able to covalently bind this residue of the inducible members of the HSP70 family cause a deep reduction of the affinity for ATP, and this because of the proximity of Cys267 to the NBD with a consequent steric inhibition of nucleotide binding.<sup>312</sup>

### 4.3.3 Molecular modelling studies

The binding mode of oridonin in HSP70 was rationalized through docking calculations and Molecular Dynamics simulations, using respectively Autodock 4.0<sup>26</sup> and Macromodel 8.5 (Schrödinger, LLC, New York, 2003) software packages. The activity of oridonin is related to the presence in its chemical structure of the  $\alpha,\beta$ -unsaturated ketone conjugate moiety, that represents a Michael acceptor function to the Cys267 in the binding site of HSP70.

For these reasons, we firstly tried to analyze the formation of the covalent linkage using the recent introduced covalent docking methodologies,<sup>27</sup> using the

crystallographic structure of the N-terminal domain of HSP70, in which the protein is complexed with the ADP, magnesium and inorganic phosphate (PDB code: 3JXU).<sup>313</sup> Unfortunately, using the flexible side chain method we firstly observed a restricted space in the binding site of HSP70. This feature determined steric clashes in the binding site after the formation of the covalent bond required for the next phase of flexible docking.

Moreover, using the covalent grid based approach we did not find poses compatible with the covalent linkage, because also in this case the restricted space did not allow the contact between the reactive part of oridonin and the Cys267 counterpart of HSP70. Moreover, the predicted binding energy values related to these conformations confirmed the poor results, with presence of steric clashes in this binding site.

Therefore, a putative models of oridonin covalently bound to HSP70 was possible with better accuracy using Molecular Dynamics simulations, analyzing its motions in the binding site of this target. In particular, using this method we found some poses of these compounds in which the distances between their reactive parts and the sulfur of the Cys267 were compatible with the C-S linkage experimentally determined. Combining docking calculations and Molecular Dynamics simulations, we were able to build the models of these compounds covalently bound to the target.

It is noteworthy that the Cys267 is in the ATPase domain of HSP70. It has been experimentally demonstrated that molecules able to covalently bind this residue in the HSP72 inducible member of the HSP70 family cause a deep reduction of the affinity for ATP, and this because of the proximity of Cys267 to the NBD with a consequent steric inhibition of nucleotide binding.<sup>312</sup>

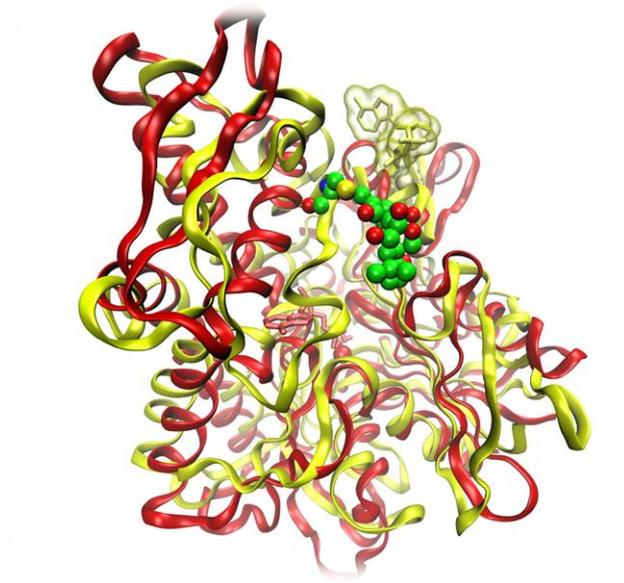
It has also been suggested that nucleotide-dependent conformational changes due to subdomain reorientations represent an intrinsic property of all NBDs, crucial

for their functions.<sup>314,315</sup> Nucleotides bind at the bottom of the deep central cleft at the interface between subdomains IB and IIB, close to the site in which oridonin is able to covalently bind HSP70.

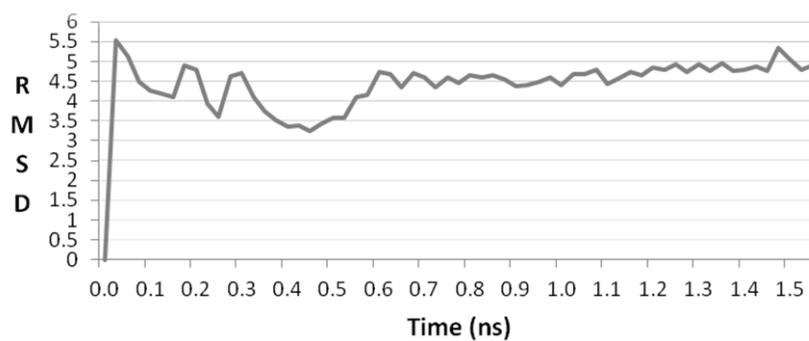
Starting from the docked conformation, we applied several fast steps of Molecular Dynamics simulations using Macromodel 8.5 (Schrödinger, LLC, New York, 2003), observing a gradual reduction of the distance between the  $\alpha$ - $\beta$  unsaturated ketone of oridonin to the sulfur of the Cys267. When we found a pose with a distance between these groups compatible with the covalent linkage C-S ( $\sim 1.8$  Å), we manually generated the covalent linkage. During the simulated approach of oridonin, we noticed a significant conformational change of the protein, and especially in the nucleotide binding site. In particular, as shown in Figure 4.26, a re-orientation of the  $\alpha$ -helix presenting the Cys267 toward the binding site is observable, causing the exit of the nucleotide complex on the outer part of the protein.

The RMSD plot, showing considerable variations during this step, confirmed these observations (Figure 4.27). Moreover, a Molecular Dynamics simulation of this generated complex was performed for 10 ns, and the results highlighted the stabilization of this inhibited conformation of HSP70 induced by the covalent linkage of oridonin, able in this bound state to establish interactions with several polar residues on different helices in the nucleotide binding site (Figure 4.28).

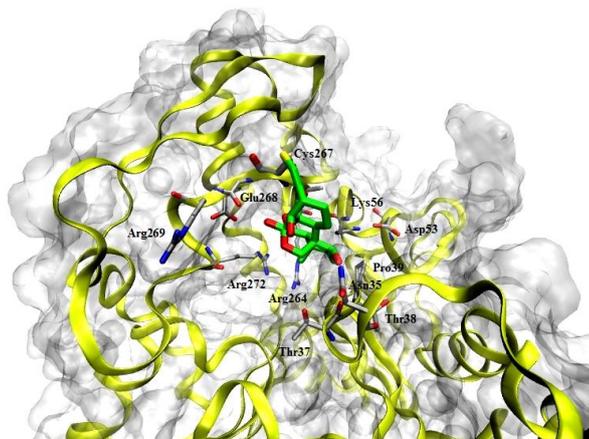
In agreement with this hypothesis, in this case the RMSD plot showed a more stable trend (Figure 4.29).



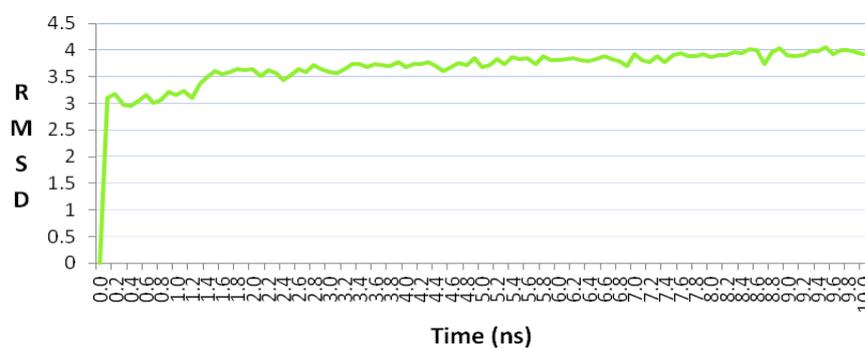
**Figure 4.26** Superposition between the crystal structure of HSP70 (PDB code: 3JXU, secondary structure of the protein, ADP, magnesium and inorganic phosphate as stick and balls colored in red) and the covalent model of oridonin (represented as CPK and colored by atom types: C green, O red) with HSP70 after Molecular Dynamics simulations (secondary structure of the protein, ADP, magnesium and inorganic phosphate as stick and balls colored in yellow).



**Figure 4.27** RMSD plot related to the simulated approach of oridonin to the identified binding site of HSP70.

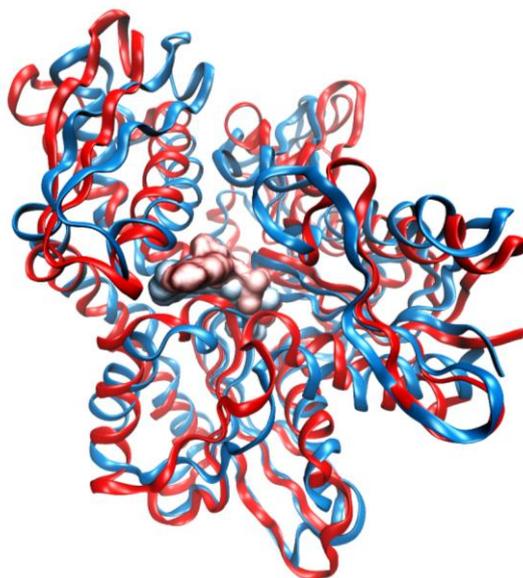


**Figure 4.28** Covalent model of oridonin (colored by atom types: C green, O red) bound to the Cys267 of HSP70 (secondary structure colored in yellow, molecular surface colored in light grey, residues in contact with oridonin represented as stick and balls, colored by atom types: C gray, O red, N blue).

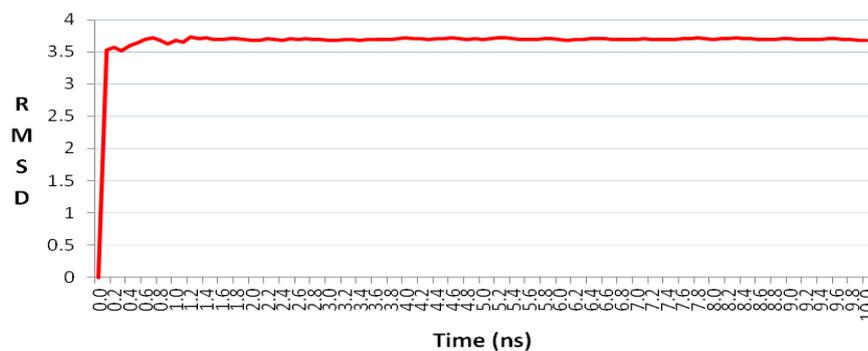


**Figure 4.29** RMSD plot related to the Molecular Dynamics simulation of HSP70-oridonin complex.

In order to exclude the possibility of an artifact from the computational methodology used, a “blank” Molecular Dynamics simulation of 10 ns was performed starting from the crystal structure of HSP70 without oridonin. As expected, the protein preserved its conformation, without movements of the  $\alpha$ -helix involved in the binding of oridonin, and also the nucleotide complex did not show significant movements in the binding site (Figure 4.30 and Figure 4.31).



**Figure 4.30** Superposition between the crystal structure of HSP70 (secondary structure of the protein, molecular surface of ADP, magnesium and inorganic phosphate colored in red) and the final structure obtained after a “blank” Molecular Dynamics simulation of 10 ns (secondary structure of the protein, molecular surface of ADP, magnesium and inorganic phosphate colored in blue).



**Figure 4.31** RMSD plot related to the Molecular Dynamics simulation of HSP70 not complexed with oridonin.

#### **4.3.4 Final remarks**

In summary, the model here proposed rationalizes the inhibitory activity of the covalent binder oridonin describing the dynamic modification of the orientation of the helixes surrounding the NBD of HSP70.

Our observations revealed a mechanism of action of oridonin consistent with the multiple biological activities described for this diterpenes. Oridonin, inhibiting HSP70 activity can affect the activity of many client proteins of this chaperone, simultaneously affecting several molecular pathways. Moreover, our data could also account for the effects of oridonin treatment on intracellular level of HSP70 reported elsewhere.<sup>316</sup> Shedding light on the molecular basis of the biological activity of oridonin, our findings may be relevant for possible therapeutic applications of oridonin, such as its use in combination and the design of new therapeutic approaches. Even if efficacy of HSPs inhibitors as single-agent therapy is still limited for pharmacological and pharmaco-dynamic reasons, they have been reported to enhance cytotoxicity of various antitumor agents<sup>317</sup> or the efficiency of HSP90 inhibitors.<sup>318</sup> Indeed, recent preclinical and clinical studies exploring the effect of a combination of HSPs inhibitors with other anticancer agents in cancer therapy, demonstrated that in most cases they produce additives or synergic effects.<sup>319,320,321</sup> On the other hand, discovering small molecules ATP competitive inhibitors of the NBD of HSP70 has proved extremely challenging providing an effective tool to study the biology of this protein;<sup>304</sup> oridonin could therefore represent a promising new tool to advance basic investigations on the varied activities of HSP70.

## **-CHAPTER 5-**

### ***Experimental section***

## 5.1 Inverse Virtual Screening

### Pilot Inverse Virtual Screening study: LIBIOMOL library

**Chemical structure preparation.** The library of compounds was downloaded from the website [www.libiomol.unina.it](http://www.libiomol.unina.it). Chemical structures were built through MacroModel Maestro software Version 8.5. Optimization of the three-dimensional structures was applied with Monte Carlo Conformational Search and Molecular Dynamics simulations. Molecular mechanics/dynamics calculations were performed on a Pentium IV 2008 using the Macromodel 9.0 software package and the OPLS force field. The Monte-Carlo multiple minimum (MCMM) method (5000 steps) was used first in order to allow a full exploration of the conformational space. Molecular Dynamics simulations were performed at a temperature of 600K. A constant dielectric term mimicking the presence of water was used in the calculations to reduce artifacts derived from the absence of the solvent.

**Inverse Virtual Screening.** Protein targets, known to be involved in tumor processes, were prepared by a search of crystallized structures on the Protein Data Bank database. Water molecules were removed and polar hydrogens were added with AutodockTools 1.4.5.

Molecular docking calculations were performed using Autodock-Vina software. The grids focused on receptors were built using as reference the binding mode of crystallized ligands in PDB files. For the docking studies, we used an exhaustiveness of 16. For all the investigated compounds, all open-chain bonds were treated as active torsional bonds. Autodock-Vina results were analyzed with Autodock Tools 1.4.5.

## **Re-evaluation of the biological activity of a small library of natural compounds**

**Chemical structure preparation.** We built and processed the chemical structures of the library of compounds with Macromodel 8.5 (Schrödinger, LLC, New York, 2003). Molecular mechanics/dynamics calculations were performed on a 4 × AMD Opteron SixCore 2.4 GHz using Macromodel 8.5 and the OPLS force field. To allow a full exploration of the conformational space, we used the Monte Carlo multiple minimum (MCMM) method (5000 steps). Molecular Dynamics simulations were performed at a temperature of 600 K and with a simulation time of 10 ns. A constant dielectric term, mimicking the presence of the solvent, was used in the calculations to reduce artifacts. To identify a possible three-dimensional starting model of each compounds for the subsequent docking calculations, we applied an optimization (Conjugate Gradient, 0.05 Å convergence threshold) of the structures.

**Inverse Virtual Screening.** We built the panel of protein targets by a search of crystallized structures in the Protein Data Bank database ([www.rcsb.org](http://www.rcsb.org)). Water molecules were removed, and polar hydrogens were added with AutodockTools 1.4.5. Molecular docking calculations were performed using Autodock-Vina<sup>28</sup> software and then normalized. In the configuration file we specified only the exhaustiveness value to 64 and the coordinate values for the targets, focusing the grids on the sites of presumable pharmacological interest. When it was possible, we used as reference the binding mode of crystallized ligands in PDB files. For all the investigated compounds, all open-chain bonds were treated as active torsional bonds. Autodock-Vina results were analyzed with Autodock Tools 1.4.5.

## **Discovery of peptidase inhibitory activity of the new-anabaenopeptin cyclopeptide namalide**

**Chemical structure preparation.** The chemical structures of **54** and **56** were built and processed with Macromodel 8.5 (Schroedinger, New York, 2003). Molecular mechanics/dynamics calculations were performed on a quad-core Intel Xeon 3.4 GHz using Macromodel 8.5 and the OPLS force field. The Monte Carlo multiple minimum (MCMM) method (5000 steps) was used first to allow a full exploration of the conformational space. Molecular dynamics simulations were performed at 600 K and with a simulation time of 10 ns. A constant dielectric term, mimicking the presence of the solvent, was used in the calculations to reduce artifacts. Finally, an optimization (Conjugate Gradient, 0.05 Å convergence threshold) of the structures was applied for the identification of a possible 3-D starting models of **54** and **56** for the subsequent steps of docking calculations.

**Inverse Virtual Screening.** The panel of protein targets was prepared by a search of crystallized structures in the Protein Data Bank. Water molecules were removed, and polar hydrogens were added with AutodockTools 1.4.5. Molecular docking calculations were performed using Autodock-Vina software using an exhaustiveness of 64. The selected grids focused on presumed sites of pharmacological interest on the basis of the binding modes of crystallized ligands in the PDB files wherever possible. A more accurate analysis of the interaction between **54** and CPA was conducted with the Autodock 4.2 software package. To have an accurate weight of the electrostatics, we derived the partial charge of Zn = 1.136 and of the amino acids involved in the catalytic center by DFT calculations at the B3LYP level by the 6-31G(d) basis set and ChelpG<sup>322</sup> method for population analysis (Gaussian 03 Package software<sup>323</sup>). Ten calculations consisting of 256 runs were performed, obtaining 2560 structures (256 × 10). The Lamarckian genetic

algorithm was used for dockings. An initial population of 450 randomly placed individuals, a maximum number of  $10.0 \times 10^6$  energy evaluations, and a maximum number of  $8.0 \times 10^6$  generations was taken into account. A mutation rate of 0.02 and a crossover rate of 0.8 were used. Results differing by  $<3.0 \text{ \AA}$  in positional root-mean-square deviation (rmsd) were clustered together. For all of the investigated compounds, all open-chain bonds were treated as active torsional bonds. Autodock Vina results were analyzed with Autodock Tools 1.4.5.

## 5.2 High-Throughput Molecular Dynamics for the accurate calculations of the binding affinities

**Compounds data set.** In order to investigate the performance of LIE, we used the trypsin receptor and data set from the directory of useful decoys (DUD, <http://dud.docking.org/>).

The trypsin data set is composed of 49 ligands and 1664 decoys. Since the presence for the same ligand or decoys of more IDs, mainly characterized by different protonation states, we provided a final library of unique ligands (44 ligands and 1544 decoys).

**Molecular Dynamics.** Regarding the Molecular Dynamics simulations, all the ligands were parameterized using the Antechamber software (AmberTools 1.5), assuming the compounds already with charges (.mol2 files), adding GAFF atom types, and then all the ligand-specific parameters not included in the GAFF forcefield were guessed with the *parmchk* command.

Compounds-protein systems and the compounds systems were solvated and neutralized with *sleap* software (AmberTools 1.5). In particular, all the systems were solvated (in a cube built considering the more external part of the system and

distant from this point 10 Å) with TIP3P water molecules and neutralized, and then minimized and equilibrated before the MD runs.

After the minimization step, the systems were equilibrated in two phases. The first phase was conducted under NVT conditions at 300 K atm for 0.1 ns, with a time step of 4 fs. Then a second equilibration phase was done under NPT conditions at 300 K and 1 atm for 2ns, and a time step of 4 fs. During this run,  $C\alpha$  atoms of the complex and of the ligand alone were restrained with a  $1 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$  harmonic potential to prevent the system reorienting.

Then all the systems were submitted to the production runs:  $3 \times 10$  ns MD simulation runs were performed for each complex and ligand alone related to the same system, under NVT conditions, at 300 K, using a time step of 4 fs. For all the simulations, a cutoff of 9 Å and particle-mesh Ewald long range electrostatic were set.

The minimization phases were performed on standard GPU desktop server, while the production runs (~ 9500 simulations) were performed on the distributed computing project GPUGRID, using ACEMD software.

Energy extraction was performed using NAMD energy plugin, version 1.4, and for each simulation an average of the energies related to each frame was considered. Finally, the average of the three simulations was calculated and used for the LIE calculations.

## 5.3 Further applications of in silico screenings on natural compounds

### Discovery of cholestan disulfate as a potent pregnane-X-receptor agonist

**Chemical structure preparation.** Prior to docking calculations, we built and processed the chemical structures of the compounds with Macromodel 8.5 (Schrödinger, LLC, New York, 2003). Molecular Mechanics/Dynamics calculations were performed on a 4 × AMD Opteron SixCore 2.4 GHz using Macromodel 8.5 and the OPLS force field. The Monte Carlo multiple minimum (MCM) method (5000 steps) was used first in order to allow a full exploration of the conformational space. Molecular Dynamics simulations were performed at a temperature of 600 K and with a simulation time of 10 ns. A constant dielectric term, mimicking the presence of the solvent, was used in the calculations to reduce artifacts. Finally, we applied an optimization (Conjugate Gradient, 0.05 Å convergence threshold) of the structures for the identification of a possible three-dimensional starting models for the subsequent steps of docking calculations.

**Docking calculations.** Docking calculations were performed using Autodock 4.2 software. A grid box size of 90×108×96 with spacing of 0.375 Å between the grid points and centered at 14.282 (x), 74.983 (y), and 0.974 (z) was used for the PXR receptor. We performed 10 calculations consisting of 256 runs, obtaining 2560 structures (256 × 10), using the Lamarckian genetic algorithm for dockings. An initial population of 450 randomly placed individuals, a maximum number of  $5.0 \times 10^6$  energy evaluations, and a maximum number of  $6.0 \times 10^6$  generations were taken into account. A mutation rate of 0.02 and a crossover rate of 0.8 were used. We clustered together results differing by less than 3.0 Å in positional root-mean-square deviation (rmsd). For all the investigated compounds, all open-chain

bonds were treated as active torsional bonds. Docking results were analyzed with Autodock Tools 1.4.5.

**Plakilactones from the marine sponge *Plakinastrella mamillaris*, a new class of marine ligands of peroxisome proliferator-activated receptor  $\gamma$**

**Chemical structure preparation.** The chemical structures of the compounds were built and processed with Macromodel 8.5 (Schrödinger, LLC, New York, 2003). Molecular mechanics/dynamics calculations were performed on a 4 × AMD Opteron SixCore 2.4 GHz using Macromodel 8.5 and the OPLS force field. The Monte Carlo multiple minimum (MCMM) method (5000 steps) was used to allow a full exploration of the conformational space. Molecular Dynamics simulations were performed at 600 K and with a simulation time of 10 ns. A constant dielectric term, mimicking the presence of the solvent, was used to reduce artifacts. Finally, the optimization (Conjugate Gradient, 0.05 Å convergence threshold) of the structures was applied to identify the three-dimensional starting models for the subsequent steps of docking calculations.

**Docking calculations.** Docking calculations were performed using the Autodock-Vina software. In the configuration files of the two crystallographic structures of PPAR $\gamma$  we specified only the exhaustiveness value to 64 and the coordinate values for the targets, focusing the grid on the site of presumable pharmacological interest. In particular a grid box size of 22 × 22 × 30 and centered at 17.464 (x), 64.919 (y), and 19.625 (z) was used for the PPAR $\gamma$  receptor (apo form), and of 20 × 20 × 20 and centered at 17.654 (x), 64.696 (y), and 11.136 (z) (locked form), with spacing of 1.0 Å between the grid points.

For all the investigated compounds, all open-chain bonds were treated as active torsional bonds. Docking results were analyzed with Autodock Tools 1.4.5.

**Molecular Dynamics simulations.** Molecular Dynamics simulations steps were performed for the compounds **76** and **80** in order to observe their contacts with the reactive counterpart of PPAR $\gamma$ , using Macromodel 8.5. In each step the distances between the sulfur of the Cys285 and the reactive  $\beta$  carbon of the  $\alpha$ - $\beta$  unsaturated ketone parts of **76** and **80** were constrained in order to reduce the distance at 0.5 Å. The obtained structures were reprocessed until a distance and an orientation of the reactive moieties compatible with the C-S bond were found. In all these steps, for the equilibration phase, an equilibration time of 10 ps was considered. Molecular Dynamics simulations were performed at 300 K, with a time step of 4.0 fs and a simulation time of 0.1 ns. Then the covalent bond between the reactive points was manually built, and the two covalent complexes (**76** or **80** with PPAR $\gamma$ ) were subjected to a further fast Molecular Dynamics simulation in order to eliminate eventual steric clashes. In all the simulations, a constant dielectric term, mimicking the presence of the solvent, was used in to reduce artifacts.

### **HSP70 1A as a Target for the Anticancer Diterpene Oridonin**

**Chemical structure preparation.** We built and processed the chemical structure of oridonin (**87**) with Macromodel 8.5 (Schrödinger, LLC, New York, 2003). Molecular mechanics/dynamics calculations were performed on a 4 × AMD Opteron 16 Core 2.3 GHz using Macromodel 8.5 and the OPLS force field. The Monte Carlo multiple minimum (MCMM) method (5000 steps) was used first in order to allow a full exploration of the conformational space. Molecular Dynamics simulations were performed at a temperature of 600 K and with a simulation time of 10 ns. A constant dielectric term, mimicking the presence of the solvent, was used in the calculations to reduce artifacts. Finally, we applied an optimization (Conjugate Gradient, 0.05 Å convergence threshold) of the structures for the

identification of a possible three-dimensional starting models for the subsequent steps of docking calculations.

**Docking calculations.** We performed covalent docking calculations using the Autodock 4.0 software using the recently introduced grid based approach.<sup>27</sup> In the configuration file of the crystallographic structure of HSP70 (PDB code: 3JXU) we specified a grid box size of  $48 \times 46 \times 48$  with spacing of  $0.375 \text{ \AA}$  and centered at  $-9.765(x)$ ,  $-29.49(y)$ , and  $20.277(z)$ .

We performed 3 calculations consisting of 256 runs, obtaining 768 structures ( $256 \times 3$ ), using the Lamarckian genetic algorithm for dockings. An initial population of 450 randomly placed individuals, a maximum number of  $5.0 \times 10^6$  energy evaluations, and a maximum number of  $6.0 \times 10^6$  generations were taken into account. A mutation rate of 0.02 and a crossover rate of 0.8 were used. We clustered together results differing by less than  $3.0 \text{ \AA}$  in positional root-mean-square deviation (rmsd).

For the investigated compound, all open-chain bonds were treated as active torsional bonds. Docking results were analyzed with Autodock Tools 1.4.5.

**Molecular Dynamics simulations.** Molecular Dynamics simulations steps were performed for the oridonin in order to observe their contacts with the reactive counterpart of HSP70, using Macromodel 8.5. We constrained the distances between the sulfur of Cys267 and the reactive  $\beta$  carbon of the  $\alpha$ - $\beta$  unsaturated ketone parts of oridonin in order to reduce the distance of from the sulfur in the side chain of Cys267. The obtained structures were reprocessed until we found a distance and an orientation of the reactive moieties compatible with the C-S linkage. In all these steps, for the equilibration phase an equilibration time of 50 ps was considered. Molecular Dynamics simulations were performed at a temperature of 300 K, with a time step of 2.0 fs and a simulation time of 0.5 ns. Then we manually built the covalent linkage between the reactive points, and then the

covalent complex was submitted to a further Molecular Dynamics simulation of 10 ns after an equilibration phase of 1 ns. Also in this case, a time step of 2.0 fs was considered. In parallel we performed also a Molecular Dynamics simulation of 10 ns on the crystal structure of HSP70 complexed with ADP, magnesium and inorganic phosphate, using the same parameters previously described. In all the simulations, a constant dielectric term, mimicking the presence of the solvent, was used in the calculations to reduce artifacts.

# **-CONCLUSIONS-**

Classical computational studies are oriented towards the evaluation of the predicted affinities on one defined target protein, in order to perform subsequent biological assays for the evaluation of the biological activity. This represents the main concept on which is based the Virtual Screening *in silico* protocol. Thus, the main risk is to confine these studies to a single target of a pathological process.

In this thesis, the concept of Inverse Virtual Screening has been introduced and examined. The application of this method may facilitate the prediction of the activity of secondary metabolites from natural or synthetic sources on known different receptors involved in the development of a pathology (i.e. cancer; viral, bacterial, mycotic infections).

However, the most important evidence arising from the analysis of the results was the impossibility of consider the mere predicted binding energies as parameter of selection of the best results, and this was mainly due to the variability of the binding sites of the different targets. To overcome this problem, a normalization of the matrix that collects all values of predicted binding energies from the calculations (Equation 2.5, Paragraph 2.2.4) was applied. The different results pointed out in this thesis demonstrate that through docking calculations on large panels of ligands and targets, a screening of energies after normalization is possible in order to select the best interactions. From these selections, experimental tests could be started on a restricted number of proteins. Thus, Inverse Virtual Screening may be considered as a new accurate tool to facilitate the drug discovery process.

From a general point of view, Inverse Virtual Screening represents a new computational tool for the identification of targets of pharmacological interest rather than a method for the precise prediction of the range of activity for one or more compounds. Indeed, the normalization of the predicted binding energies considering the average trends of the ligands (on the targets) and of the targets (on

the ligands) is useful to identify significant results by a large set of data from heterogeneous sources.

On the other hand, this approach makes possible the discovery of new molecular scaffolds on a specific receptor as lead compounds and suggested a possible improvement of their potency and selectivity focused on a precise and defined biological context. In particular, precise modifications in the pharmacophore models of the compounds identified could increase their inhibition on the targets selected from the screening.

As shown in the first pilot attempt (Paragraph 2.2), applying this mathematical normalization, compounds **19** and **20** on Topoisomerase I target were selected. These data represented encouraging results regarding the reliability of this approach. In fact, in the library of 43 molecules used, compounds **19** and **20** corresponded in point of fact to internal “standard” compounds, because their bioactivity on Topoisomerase I was yet known.

Proceeding on this research line, the Inverse Virtual Screening method was implemented using another library of 10 compounds extracted from plants against a panel of 163 target involved in the cancer events (Paragraph 2.3). The normalization of the predicted affinities addressed the biological tests on 2 compounds (**52** and **53**) against three targets (PDK1, PKC- $\alpha$  and PKC- $\theta$ ). Also in this case, a good correlation between the Inverse Virtual Screening results and the biological data, corroborated by 5 prediction of activity/inactivity on 6, was observed.

For compounds **52** and **53** identified in this study moderate biological activities were found ( $\mu\text{M}$  range). This is in line with the hypothesis of a modulatory role of these molecules in the cancer events. Moreover, the  $\mu\text{M}$  range of activity is fully compatible with the role of cancer prevention that is achievable by a consistent exposure to non-toxic agent (i.e. food product).

Moreover, in order to explain this moderate activities found for the best results, another Inverse Virtual Screening study was performed considering **52**, **53**, and the ligands crystallized in the binding sites of the targets PDK1, PKC- $\alpha$ , and PKC- $\theta$  (namely **PDK1\_lig**, **PKC- $\alpha$ \_lig**, and **PKC- $\theta$ \_lig**). The biological activity of this “standard” compounds in the nM range on their respective targets allowed a more precise comparison with the V values calculated for the compounds **52** and **53** emerging from this new screening (Paragraph 2.3, Table 2.11 ), because in this case their normalized results are related to a new scale of values.

In parallel, the case of the Inverse Virtual Screening on a single compound was also considered, and specifically for the cyclopeptide namalide (**54**, Paragraph 2.4), featuring a new mini-anabaenopeptin scaffold.

A modification of the way to normalize the docking results was proposed, introducing “blank” compounds for the calculations of the affinities indispensable for the normalization. **54** belongs to the family of anabaenopeptides, known to inhibit carboxipeptidases. Using a panel of 159 targets involved in the cancer, Carboxypeptidase A (CPA) protein was identified as the third best hit in the final ranking, while the isoform Carboxipeptidase U (CPU) was on position 149. These results were found to be in total accordance with the biological results ( $IC_{50} = 250$  nM), that highlighted the nM inhibitory activity and selectivity of **54** against CPA, with a total absence of activity on CPU. The fact that **54** inhibits CPA with potencies comparable to the more common hexapeptides, together with the docking analysis and rationalization of the inactivity of other cyclopeptides synthesized, can open the way to further QSAR investigations in order to design other namalide analogs against CPA and related hydrolases.

Virtual Screening combined with accuracy of the calculated affinities was also examined using methods other than molecular docking. In particular, high-throughput Molecular Dynamics simulations were applied on a library of ligands

and decoys specific for the trypsin receptor. In general, these libraries are open-source available on Internet, and are very useful for comparing the efficiency of different computational methods in distinguishing true positive results from false positive ones. In the specific case, a comparison between the docking results, LIE (Linear Interaction Energy) and MM-GB(PB)SA was taken into account. So far, only preliminary results are available and are related to molecular docking and LIE method, considering for the latter method four results deriving from four different combinations of scaling factors. The efficacy of these two approaches was investigated and graphically summarized using ROC and enrichment curves. In general, LIE show better results than molecular docking only on the 2% of the ranked database, and after this point docking shows enrichment factors particularly better than those of LIE.

MM-GBSA and MM-PBSA evaluation of the binding affinities are ongoing and will be compared with the results obtained until now, in order to offer a more comprehensive view in the field of the calculations of the binding affinities by Molecular Dynamics-based methods.

*In silico* screenings for the detailed study of natural compounds whose activities are known a priori were performed on different targets. In these case studies, the main aim was to offer qualitative considerations regarding the absence/presence and orientations of the pharmacophoric points of these compounds in the binding sites of the targets investigated. The rationalization of the molecular basis of these interactions can be useful for the next design of new molecules for these specific systems.

In more details, a small library of 9 compounds derivated by the modification of the side chain of the natural compound Solomonsterol A, a potent agonist of PXR target, was carefully analyzed by means of molecular docking experiments.

In particular, compound **67** showed a remarkable agonistic activity ( $\mu\text{M}$  range) that was clarified at molecular level, elucidating the fundamental presence of two sulfate group in precise positions on the ring A, and of an hydrophobic side chain derived from the cholesterol. For all the other inactive compounds in the library, the absence of this pharmacophoric points or their shift in different positions were identified as the cause of their total absence of activity on PXR.

Moreover, a set of compounds extracted from the marine sponge *Plakinastrella Mamillaris* was analyzed on the target PPAR $\gamma$ . At molecular level it was demonstrated that gracilioether B (**76**) and plakilactone C (**80**) covalently bind to a cysteine residue in the PPAR $\gamma$ -ligand binding site (LBD) through a Michael addition reaction to the  $\alpha,\beta$ -unsaturated ketone in their side chains.

An integrated analysis of docking and Molecular Dynamics simulations showed, as previously reported for several selective covalent PPAR $\gamma$  ligands, that steps based on non-covalent interactions confer specificity in the recognition process to PPAR $\gamma$ -LBD and then, after the covalent binding, a conformational change in the protein is observable, accounting for the agonist activity observed in the biological tests. For gracilioether C (**77**) a detailed docking analysis laid the basis for the structural requirements for its non covalent interaction in the receptor's LBD. Further insights were gained to explain the antagonist activities of methyl esters **72** and **73** and several non active members of this series.

In a similar way, computational studies based on molecular docking and Molecular Dynamics simulations assisted the experimental evidences from Chemical Proteomics and several biological studies about the inhibitory activity of compound oridonin (**87**) on the target HSP70 1A.

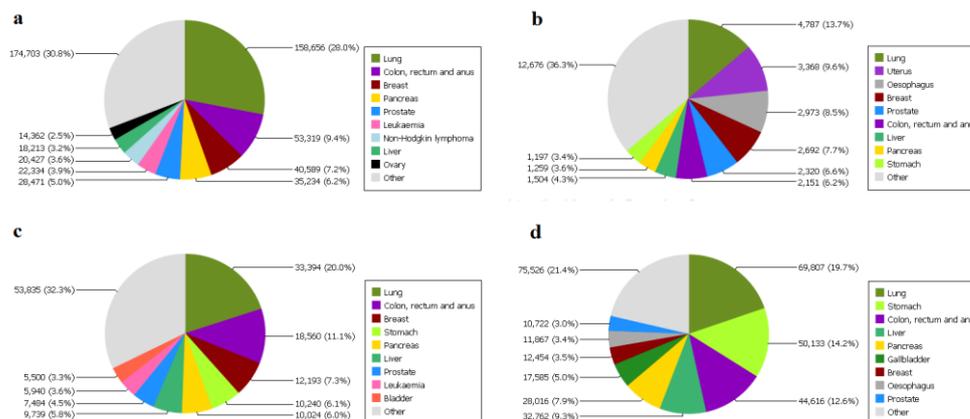
The formation of the covalent bond between its  $\alpha,\beta$ -unsaturated ketone moiety and the reactive Cys267 and the conformational change of the protein induced by

this linkage was shown, demonstrating the re-orientation of the binding site of HSP70.

**-Appendix A-**

## A.1 Cancer: some data

Cancer represents today a global public health burden, and the summary of cancer incidence and mortality patterns in the relative rankings by cancer site reveal differences between industrialized and developing nations. Reliable data on the magnitude of the cancer problem are essential for monitoring the health of the community, assessing the performance of the health care system and allowing authorities to make informed decisions. Several migrant studies have documented that cancer rates in successive generations of migrants shift in the direction of the prevailing rates in the host country, suggesting that the international variations in cancer rates for most cancers largely reflect differences in environmental risk factors (including lifestyle and culture) rather than genetic differences (Figure 5.1).<sup>324</sup>



**Figure 5.1** Distribution of several specific tissue cancer types in: a)usa (2008) b)southafrica (2009) c)Italy (2009) d)japan (2010)

Cancer research has generated a rich and complex body of knowledge, revealing this pathology to be a disease involving dynamic changes in the genome. In particular, mutations that produce oncogenes with dominant gain of function and

tumor suppressor genes with recessive loss of function have been identified through their alteration in human and animal cancer cells. To date, many pathways involving several protein with different locations (extracellular, intracellular and nuclear) and structural features are known.

## **A.2 Cancer pathways**

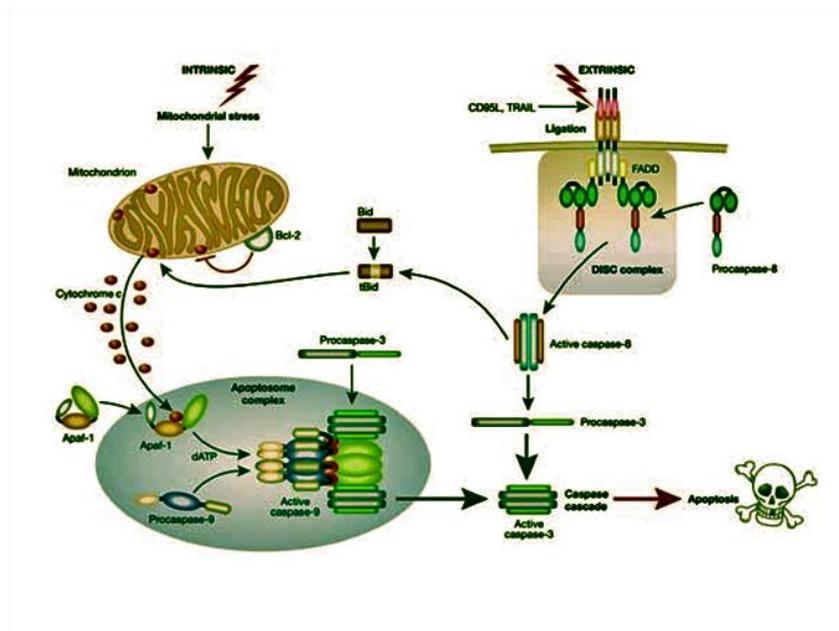
In the next paragraphs an overview of the main pathways in the cancer process are reported.

### **A.2.1 Apoptosis**

Apoptosis has been recognized and accepted as a distinctive and important mode of “programmed” cell death, which involves the genetically determined elimination of cells. Irradiation or drugs used for cancer chemotherapy results in DNA damage in some cells, which can lead to apoptotic death through a *p53*-dependent pathway. The mechanisms of apoptosis are complex and sophisticated, involving an energy-dependent cascade of molecular events. More specifically, caspases are widely expressed in an inactive proenzyme form in most cells and once activated can often activate other pro-caspases, allowing initiation of a protease cascade. Caspases have proteolytic activity and are able to cleave proteins at aspartic acid residues, although different caspases have different specificities involving recognition of neighboring amino acids. Once caspases are initially activated, there seems to be an irreversible commitment towards cell death. To date, ten major caspases have been identified and broadly categorized into initiators (caspase-2,-8,-9,-10), effectors or executioners (caspase-3,-6,-7) and inflammatory caspases (caspase-1,-4,-5).<sup>325</sup> The other caspases that have been identified include caspase-11, which is reported to regulate apoptosis and cytokine maturation during septic shock, caspase-12, which mediates endoplasmic-specific apoptosis and cytotoxicity by amyloid-  $\beta$ , caspase-

13, which is suggested to be a bovine gene, and caspase-14, which is highly expressed in embryonic tissues but not in adult tissues.<sup>326</sup>

It is possible to distinguish two main apoptotic pathways: the extrinsic or death receptor pathway and the intrinsic or mitochondrial pathway. However, there is now evidence that the two pathways are linked and that molecules in one pathway can influence the other (Figure 5.2).



**Figure 5.2** Intrinsic and extrinsic apoptotic pathways

The extrinsic signaling pathways that initiate apoptosis involve transmembrane receptor-mediated interactions. These involve death receptors that are members of the tumor necrosis factor (TNF) receptor gene superfamily. The sequence of events that define the extrinsic phase of apoptosis are best characterized with the FasL/FasR and TNF- $\alpha$ /TNFR1 models. In these models, there is clustering of receptors and binding with the homologous trimeric ligand. Upon ligand binding, cytoplasmic adapter proteins are recruited which exhibit corresponding death

domains that bind with the receptors. The binding of Fas ligand to Fas receptor results in the binding of the adapter protein FADD and the binding of TNF ligand to TNF receptor results in the binding of the adapter protein TRADD with recruitment of FADD and RIP. FADD then associates with procaspase-8 via dimerization of the death effector domain. At this point, a death-inducing signaling complex (DISC) is formed, resulting in the autocatalytic activation of procaspase-8. Once caspase-8 is activated, the execution phase of apoptosis is triggered.

Apoptosis can also occur via Intrinsic pathways. The Intrinsic Apoptosis pathway begins when an injury occurs within the cell. Intrinsic stresses such as Oncogenes, direct DNA damage, Hypoxia, and survival factor deprivation, can activate the Intrinsic Apoptotic pathway. p53 is a sensor of cellular stress and is a critical activator of the intrinsic pathway. The DNA checkpoints proteins, ATM (Ataxia Telangiectasia Mutated protein), and Chk2 (Checkpoints Factor-2) directly phosphorylate and stabilize p53 and inhibit MDM2 (Mouse Double Minute-2 Homolog) mediated ubiquitination of p53. MDM2 binds p53 and mediates the nuclear export. When bound to MDM2, p53 can no longer function as an activator of transcription. p53 initiates Apoptosis by transcriptionally activating proapoptotic Bcl2 family members and repressing antiapoptotic Bcl2 proteins and CIAPs. Other p53 targets include BAX, Noxa, PUMA (p53-Upregulated Modulator of Apoptosis) and BID. p53 also transactivates other genes that may contribute to Apoptosis including PTEN (Phosphatase and Tensin Homolog Deleted On Chromosome-10), APAF1, Perp, p53AIP1 (p53-regulated Apoptosis-Inducing Protein-1), and genes that lead to increases in ROS (Reactive Oxygen Species). These ROS lead to generalized oxidative damage to all Mitochondrial components. Damage to Mitochondrial DNA disrupts Mitochondrial oxidative phosphorylation, contributing to a number of Human diseases.<sup>327</sup> Other proteins released from Damaged Mitochondria, SMAC (Second Mitochondria-Derived Activator of

Caspase)/ Diablo, Arts and Omi/HTRA2 (High Temperature Requirement Protein-A2), counteract the effect of IAPs (Inhibitor of Apoptosis Proteins), which normally bind and prevent activation of Caspase3. The interaction between Bcl family members, IAPs, SMAC and Omi/HTRA2 is central to the intrinsic Apoptosis pathway. Another nuclease, EndoG (Endonuclease-G), is specifically activated by Apoptotic stimuli and is able to induce nucleosomal fragmentation of DNA independently of Caspase and DFF (DNA-Fragmentation Factor)/ CAD (Caspase-Activated DNase). EndoG is a mitochondrion-specific nuclease that translocates to the nucleus and cleaves chromatin DNA during Apoptosis. Another protein, AIF (Apoptosis Inducing Factor) has also been attributed a role in Apoptosis, becoming active upon translocation from mitochondria to nuclei, where it initiates chromatin condensation and large-scale DNA fragmentation.<sup>328</sup> Gain and loss of function models of genes in the core Apoptotic pathway indicate that the violation of cellular homeostasis can be a primary pathogenic event that results in disease. Evidence indicates that insufficient Apoptosis can manifest as Cancer or Autoimmunity, while accelerated cell death is evident in Acute and Chronic Degenerative diseases, Immunodeficiency, and Infertility.<sup>329</sup> One important way of blocking apoptosis in cancer cells is by inactivating the p53 tumour-suppressor pathway. Levels of this transcription factor increase in response to a wide range of genotoxic insults, and many of its target genes encode pro-apoptotic proteins (for simplicity, only BAX is shown).<sup>330</sup> p53 is targeted for degradation in the proteasome by MDM2, an oncogenic ubiquitin ligase<sup>331</sup>. In turn, MDM2 is inhibited by ARF (also known as p14 in humans and p19 in mice).<sup>332</sup>

As well as blocking pro-apoptotic pathways, cancer cells can upregulate anti-apoptotic pathways. The growth-factor-mediated activation of phosphatidylinositol 3-kinase (PI3K) is particularly important in this regard.

### **A.2.2 Replicative lifespan**

One of the hallmarks of cancer is unlimited replicative potential. In human cells, telomere maintenance serves as a generational clock that counts cell divisions and regulates cell lifespan.<sup>333,334</sup> Cancer cells have at least two ways of maintaining their telomeres. The more common mechanism, which is active in 85-90% of human cancer cells, is to switch on TERT, the protein component of telomerase (a ribonucleoprotein enzyme that synthesizes telomeres and maintains telomeric ends). A small proportion of cancer cells use a yet undefined pathway that is known as alternative lengthening of telomeres (ALT) and allows telomere maintenance in the absence of active telomerase. As indicated, although telomeres have a crucial function in specifying cell lifespan, other pathways contribute to the replicative potential of a tumour cell.

### **A.2.3 Proliferative signals**

Extracellular signals have important functions in the homeostatic regulation of cell growth and proliferation. Many of these signals impinge on a linear series of signalling molecules that link the cell surface to intracellular machinery that effects cell growth and division. Activating mutations in many of these genes are oncogenic and serve to liberate cancer cells from these normal homeostatic mechanisms, allowing self-sufficient proliferation. Proliferative signalling pathways often begin with the activation of a receptor tyrosine kinase (RTK) by a growth factor, although activation of some G-protein-coupled receptors (GPCRs) can also activate certain branches of this signalling pathway.

Depending on the proteins that are subsequently recruited by the activated receptor, several downstream signalling pathways might be activated. It would be impossible to illustrate all of these permutations here, but two important pathways are shown.

The first of these is the RAS pathway. The binding of a growth factor to an RTK recruits and activates the adaptor proteins GRB2 and SHC which, in turn, enable SOS to activate the small GTP-binding protein RAS.

This association activates a cascade of serine/threonine kinases (such as RAF and MEK), culminating in the activation of a mitogen-activated protein kinase (MAPK, usually an extracellular signal regulated kinase ERK).

MAPKs are translocated to the nucleus where, by phosphorylating transcription factors such as FOS and JUN, they modulate the expression of a wide range of genes that are involved in cell growth and survival. RAS has a number of effectors other than MAPK (not all shown),<sup>335</sup> and many of these might also impinge on the neoplastic phenotype.

A second branch downstream of growth factors that is often deregulated in cancer is the phosphatidylinositol 3-kinase (PI3K) pathway, which activates the serine/threonine kinase AKT.<sup>336</sup> Inactivation of PTEN, a lipid phosphatase, also results in activation of this pathway, and inherited loss of PTEN confers susceptibility to many types of cancer.<sup>337</sup>

RAS activation can also lead to activation of PI3K, indicating that these pathways are interconnected at several levels.

#### **A.2.4 Cell cycle**

Several tumour-suppressor proteins monitor intrinsic and extrinsic signals and integrate these inputs to decide whether the cell should remain in a quiescent state or enter into the cycle of active growth and division. In addition to these decisions, which are made in the G1 phase of the cell cycle, checkpoints that operate in other phases of the cycle exert quality control by determining that certain steps have been properly completed before subsequent steps are undertaken.

Two tumour-suppressor pathways that block progression through the cell cycle - the retinoblastoma (RB) pathway and the p53 pathway - are inactivated in most, if not all, cancer cells.<sup>331,338</sup> RB represses the transcriptional activation of genes controlled by E2Fs, a family of transcription factors that regulate the expression of several genes involved in cell-cycle progression.<sup>339</sup> The cyclin-D-CDK4/6 (cyclin-dependent kinase 4/6) complexes and cyclin-E-CDK2 complexes regulate RB activity by phosphorylation; hyperphosphorylation of RB renders it unable to suppress the activity of the E2Fs. These cyclin-D-CDK4/6 complexes are, in turn, antagonized by a series of CDK inhibitors, known as INKs. The two best studied of these, INK4A (also known as p16) and INK4B (also known as p15), have prominent functions in blocking RB phosphorylation in many cell types.

As well as activating apoptosis (see apoptosis box, p53 blocks progression through the cell cycle by activating the transcription of another CDK inhibitor, WAF1 (also known as p21). WAF1 inhibits the activity of a number of cyclin-CDK complexes, among them cyclin-E-CDK2.

Normal cells seem to have a hard-wired safety mechanism that defends the cell against illegitimate activation of oncogenic stimuli. One of the best-characterized is the response of cells to persistent activation of the RAS pathway, which activates both the RB and the p53 pathways, thereby leading to cell-cycle arrest.<sup>340,341</sup> Oncogenic activation of RAS therefore does not lead to tumorigenicity unless this safety catch is removed by inactivation of the p53 and RB pathways.

Another mechanism for putting the brakes on the cell cycle, at least in some cell types, is the transforming growth factor- $\beta$  (TGF- $\beta$ ) pathway.<sup>342</sup> Members of the TGF- $\beta$  family activate the intrinsic serine/threonine kinase activity of their receptors (TGF- $\beta$ Rs), leading to the phosphorylation and nuclear translocation of a group of transcription factors called SMADs. Depending on the presence of other co-activator or corepressor molecules, SMADs can regulate the transcription of a

wide range of molecules that are involved in cell-cycle regulation. For example, they can upregulate the expression of the INK4B CDK inhibitor and can repress the expression of the *MYC* oncogene. However, there is a darker side to TGF- $\beta$  activation as, under some circumstances, it can lead to invasion and metastasis.

Progression through the cell cycle is activated by many oncogenic signalling pathways, including those that activate RAS, which indirectly induces cyclin D expression. The WNT-Frizzled signalling pathway also stimulates cell-cycle progression and is dysregulated in many colorectal cancers.<sup>343</sup> Activation of the Frizzled receptor leads to the inhibition of glycogen synthase kinase-3 $\beta$  (GSK3 $\beta$ ); this regulates the assembly of a ubiquitin ligase complex that normally keeps levels of free cytoplasmic b-catenin low. When GSK3 $\beta$  is inhibited, free b-catenin accumulates and gains access to the nucleus, where it acts as a co-activator for a transcription factor called TCF. The TCF-b-catenin complex induces the transcription of several important regulators of the cell cycle, such as cyclin D and *MYC*, but evidence is also accumulating that some of the targets of the TCF-b-catenin complex are important for driving metastasis.

### **A.2.5 Mobilization of resources**

Many of the intracellular and extracellular changes that are found in cancer cells lead to changes in gene expression and protein metabolism. Activation of these pathways is necessary to mobilize the cellular resources that are necessary for the cancer-cell phenotype. In particular, some of these changes lead to expression of inappropriate gene programs and activation of metabolic programs that confer specific advantages to a continuously dividing cancer cell. Among the alterations that are included in this category are those that affect ribosome biosynthesis, expression of differentiation-associated antigens, enzymes involved in nutrient metabolism and enzymes that regulate oxidative potential.

For example, as well as blocking apoptotic pathways, the PI3K-AKT-PTEN pathway might be involved in regulating cell size by activating biosynthetic pathways.<sup>344</sup> Protein phosphatase 2A (PP2A) and the serine/threonine kinase TOR are also thought to be involved in this biosynthetic route; both of these can activate ribosomal S6 kinase (RSK), an important regulator of ribosome assembly. PP2A also activates eukaryotic initiation factor 4E (eIF4E), a protein that is involved in ribosome biogenesis. The identification of gene mutation or loss of each of these molecules in human cancers indicates that these biosynthetic pathways have important functions in cancer pathogenesis.

## References

---

- 1 Jorgensen, W. L. *Science* **2004**, *303*, 1813–1818.
- 2 Kalyaanamoorthy, S.; Chen, Y. P. *Drug Discov. Today* **2011**, *16*, 831-839.
- 3 Schneider, G.; Böhm, H. J. *Drug Discov. Today* **2002**, *7*, 64-70.
- 4 Kontoyianni, M.; McClellan, L. M.; Sokol, G. S. *J. Med. Chem.* **2004**, *47*, 558-565.
- 5 Mohan, V.; Gibbs, A. C.; Cummings, M. D.; Jaeger, E. P.; DesJarlais, R. L. *Curr. Pharm. Des.* **2005**, *11*, 323-333.
- 6 Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235-242.
- 7 RCSB Protein Data Bank. <http://www.pdb.org>.
- 8 Goodnow, J. R. A. *Drug Discov. Today: Technol.* **2006**, *3*, 367–375.
- 9 Huang, B.; Schroeder, M. *BMC Struct. Biol.* **2006**, *6*, 19-30.
- 10 Laurie, A.; Jackson, R. *Bioinformatics* **2005**, *21*, 1908–1916.
- 11 Binkowski, T.; Naghibzadeh, S.; Liang, J. *Nucleic Acids Res.* **2003**, *31*, 3352–3355.
- 12 Keseru, G. M.; Makara, G. M. *Nat. Rev. Drug Discov.* **2009**, *8*, 203–212.
- 13 Shoichet, B. K. *Nature* **2004**, *432*, 862–865.
- 14 Phatak, S. S.; Stephan, C. C.; and Cavasotto, C. N. *Exp. Opin. Drug Discov.* **2009**, *4*, 947-959.
- 15 Mason, J. S.; Good, A. C.; Martin, E. J. *Curr. Pharm. Des.* **2001**, *7*, 567–597.
- 16 Loew, G. H.; Villar, H. O.; Alkorta, I. *Pharm. Res.* **1993**, *10*, 475–486.
- 17 Kurogi, Y.; Guner, O. F. *Curr. Med. Chem.* **2001**, *8*, 1035–1055.
- 18 Koehn, F. E.; Carter, G. T. *Drug. Discov.* **2005**, *4*, 206–220.
- 19 Kuntz, I. D. *Science* **1992**, *257*, 1078–1082.

- 
- 20 Thomas, G. *Computer-aided drug design. In Medicinal Chemistry: An Introduction (2nd Edn.)*, Thomas, G., ed., Wiley **2007**.
- 21 Keseru, G. M.; Makara, G. M. *Drug Discov. Today* **2006**, *11*, 741–748.
- 22 Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. *J. Mol. Biol.* **1997**, *267*, 727–748.
- 23 Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. *J. Mol. Biol.* **1982**, *161*, 269–288.
- 24 Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. *J. Mol. Biol.* **1996**, *261*, 470–489.
- 25 Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. *J. Comput. Chem.* **1998**, *19*, 1639-1662.
- 26 Huey, R.; Morris, G. M.; Olson A. J.; Goodsell; D. S. *J. Comput. Chem.* **2007**, *28*, 1145-1152.
- 27 Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. *J. Comput. Chem.* **2009**, *30*, 2785–2791.
- 28 Trott, O.; Olson, A. J. *J. Comput. Chem.* **2010**, *31*, 455-461.
- 29 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. *J. Med. Chem.* **2004**, *47*, 1739–1749.
- 30 Gerstein, M.; Krebs, W. *Nucleic Acids Res.* **1998**, *26*, 4280-4290.
- 31 Brooijmans, N.; Kuntz, I. D. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 335–373.
- 32 Di Nola, A.; Berendsen, H. J. C.; Roccatano, D. *Proteins* **1994**, *19*, 174–182.
- 33 Abagyan, R. A.; Totrov, M. M.; and Kuznetsov, D. A. *J. Comput. Chem.* **1994**, *15*, 488-506.
- 34 Gohlke, H.; Klebe, G. *Curr. Opin. Struct. Biol.* **2001**, *11*, 231-235.

- 
- 35 Schulz-Gasch, T.; Stahl, M. *Drug Discov. Today: Tech.* **2004**, *1*, 231-239.
- 36 Jain, A. N. *Curr. Protein Pept. Sci.* **2006**, *7*, 407-420.
- 37 Rajamani, R.; Good, A. C. *Curr. Opin. Drug. Discov. Devel.* **2007**, *10*, 308-315.
- 38 Huang, N.; Kalyanaraman, C.; Irwin, J. J.; Jacobson, M. P. *J. Chem. Inf. Model.* **2006**, *46*, 243-253.
- 39 Meng, E. C.; Shoichet, B. K.; Kuntz, I. D. *J. Comput. Chem.* **1992**, *13*, 505-524.
- 40 Weiner, P. K.; Kollman, P. A. *J. Comput. Chem.* **1981**, *2*, 287-303.
- 41 Nilsson, L.; Karplus, M. *J. Comput. Chem.* **1986**, *7*, 591-616.
- 42 Brooks, B. R.; Brucoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187-217.
- 43 Wang, W.; Donini, O.; Reyes, C. M.; Kollman, P. A. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 211-243.
- 44 Wang, J.; Morin, P.; Wang, W.; Kollman, P. A. *J. Am. Chem. Soc.* **2001**, *123*, 5221-5230.
- 45 Kuhn, B.; Gerber, P.; Schulz-Gasch, T.; Stahl, M. *J. Med. Chem.* **2005**, *48*, 4040-4048.
- 46 Kuhn, B.; Kollman, P. A. *J. Med. Chem.* **2000**, *43*, 3786-3791.
- 47 Zou, X.; Sun, Y.; Kuntz, I. D. *J. Am. Chem. Soc.* **1999**, *121*, 8033-8043.
- 48 Liu, H.-Y.; Zou, X. *J. Phys. Chem. B* **2006**, *110*, 9304-9313.
- 49 Lyne, P. D.; Lamb, M. L.; Saeh, J. C. *J. Med. Chem.* **2006**, *49*, 4805-4808.
- 50 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. *J. Med. Chem.* **2004**, *47*, 1739-1749.
- 51 Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. *J. Med. Chem.* **2004**, *47*, 1750-1759.

- 
- 52 Gehlhaar, D. K.; Verkhivker, G. M.; Rejto, P. A.; Sherman, C. J.; Fogel, D. B.; Freer, S. T. *Chem. Biol.* **1995**, *2*, 317-324.
- 53 Krammer, A.; Kirchhoff, P. D.; Jiang, X.; Venkatachalam, C. M.; Waldman, M. *J. Mol. Graph. Model.* **2005**, *23*, 395-407.
- 54 Böhm, H. J. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 243-256.
- 55 Wang, R.; Liu, L.; Lai, L.; Tang, Y. *J. Mol. Model.* **1998**, *4*, 379-394.
- 56 Wang, R.; Lai, L.; Wang, S. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 11-26.
- 57 Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 425-445.
- 58 Yin, S.; Biedermannova, L.; Vondrasek, J.; Dokholyan, N. V. *J. Chem. Inf. Model.* **2008**, *48*, 1656-1662.
- 59 Raub, S.; Steffen, A.; Kämper, A.; Marian, C. M. *J. Chem. Inf. Model.* **2008**, *48*, 1492-1510.
- 60 Sottriffer, C. A.; Sanschagrin, P.; Matter, H.; Klebe, G. *Proteins: Struct. Funct. Bioinf.* **2008**, *73*, 395-419.
- 61 Ruvinsky, A. M. *J. Comput.-Aided Mol. Des.* **2007**, *21*, 361-370.
- 62 Chang, M. W.; Belew, R. K.; Carroll, K. S.; Olson, A. J.; Goodsell, D. S. *J. Comput. Chem.* **2008**, *29*, 1753-1761.
- 63 Lee, J.; Seok, C. *Proteins: Struct. Funct. Bioinf.* **2008**, *70*, 1074-1083.
- 64 Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. *Proteins: Struct. Funct. Bioinf.* **2006**, *65*, 15-26.
- 65 Park, H.; Lee, J.; Lee, S. *Proteins: Struct. Funct. Bioinf.* **2006**, *65*, 549-554.
- 66 Lamarck, J. B. *Zoological Philosophy*, Macmillan, London, 1914.
- 67 Hart, W. E.; Kammeyer, T. E.; Belew, R. K. *In Foundations of Genetic Algorithms III*, D. Whitley and M. Vose, Eds., Morgan Kaufman, San Francisco, CA, **1994**.
- 68 Goodford, P. J. *J. Med. Chem.* **1985**, *28*, 849-857.

- 
- 69 a) Sharp, K.; Fine, R.; Honig, B. *Science* **1987**, *236*, 1460-1463; b) Allison, S. A.; Bacquet, R. J.; McCammon, J. *Biopolymers* **1988**, *27*, 251-269.
- 70 Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comput. Chem.* **1986**, *7*, 230-252.
- 71 Gilson, M. K.; Zhou, H.-X. *Annu. Rev. Biophys. Biomole Struct.* **2007**, *36*, 21-42.
- 72 a) Gilson, M.; Given, J.; Bush, B.; McCammon, J. *Biophys. J.* **1997**, *72*, 1047-1069; b) Chang, C.-E. A.; Chen, W.; Gilson, M. K. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 1534-1539.
- 73 Durrant, J. D.; McCammon, J. A. *BMC Biology* **2011**, *9*, 71-80.
- 74 McCammon, J. A.; Gelin, B. R.; Karplus, M. *Nature* **1977**, *267*, 585-590.
- 75 Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179-5197.
- 76 Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157-1174.
- 77 Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187-217.
- 78 Christen, M.; Hünenberger, P. H.; Bakowies, D.; Baron, R.; Bürgi, R.; Geerke, D. P.; Heinz, T. N.; Kastenholz, M. A.; Kräutler, V.; Oostenbrink, C.; Peter, C.; Trzesniak, D.; van Gunsteren, W. F. *J. Comput. Chem.* **2005**, *26*, 1719-1751.
- 79 Case, D. A.; Cheatham, T. E. 3rd; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M. Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668-1688.

- 
- 80 Kale, L.; Skeel, R.; Bhandarkar, M.; Brunner, R.; Gursoy, A.; Krawetz, N.; Phillips, J.; Shinozaki, A.; Varadarajan, K.; Schulten, K. *J. Comput. Phys.* **1999**, *151*, 283-312.
- 81 Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781-1802.
- 82 Amaro, R. E.; Baron, R.; McCammon, J. A. *J. Comput. Aided Mol. Des.* **2008**, *22*, 693-705.
- 83 Lin, J. H.; Perryman, A. L.; Schames, J. R.; McCammon, J. A. *Biopolymers* **2003**, *68*, 47-62.
- 84 Lin, J. H.; Perryman, A. L.; Schames, J. R.; McCammon, J. A. *J. Am. Chem. Soc.* **2002**, *124*, 5632-5633.
- 85 Schames, J. R.; Henchman, R. H.; Siegel, J. S.; Sotriffer, C. A.; Ni, H.; McCammon, J. A. *J. Med. Chem.* **2004**, *47*, 1879-1881.
- 86 Deng, Y.; Roux, B. *J. Phys. Chem.* **2009**, *113*, 2234-2246.
- 87 Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420-1426.
- 88 Guitiérrez-de-Teràn, H.; Åqvist, J. *Computational Drug Discovery and Design, Methods in Molecular Biology*, vol. 819, Riccardo Baron, ed., **2012**.
- 89 Lee, F. S.; Chu, Z. T.; Bolger, M. B.; Warshel, A. *Prot. Eng.* **1992**, *5*, 215-228.
- 90 Marcus, R. A. *Phys. Chem.* **1964**, *15*, 155-196.
- 91 Almlof, M.; Carlsson, J.; Åqvist, J. *J. Chem. Theor. Comput.* **2007**, *3*, 2162-2175.
- 92 Åqvist, J.; Hansson, T. *J. Phys. Chem.* **1996**, *100*, 9512-9521.
- 93 Almlof, M.; Åqvist, J.; Smalas, A. O.; Brandsdal, B. O. *Biophys. J.* **2006**, *90*, 433-442.

- 
- 94 Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. *Acc. Chem. Res.* **2000**, *33*, 889-897.
- 95 Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127-6129.
- 96 Gohlke, H.; Kiel, C.; Case, D. A. *J. Mol. Biol.* **2003**, *330*, 891-913.
- 97 Hou, T.; Wang, J.; Li, Y.; Wang, W. *J. Comput. Chem.* **2011**, *32*, 866-877.
- 98 Kongsted, J.; Söderhjelm, P.; Ryde, U. *J. Comput. Aided Mol. Des.* **2009**, *23*, 395-409.
- 99 Miller, B. R.; Dwight McGee Jr., T.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. *J. Chem. Theory Comput.* **2012**, *8*, 3314-3321.
- 100 Rester, U. *Curr. Opin. Drug Discovery Devel.* **2008**, *11*, 559-568.
- 101 Chen, Y. Z.; Zhi, D. G. *Proteins: Struct. Funct. Bioinf.* **2001**, *43*, 217-226.
- 102 Hui-Fang, L.; Qing, S.; Jian, Z.; Wei, F. *J. Mol. Graph. Mod.* **2010**, *29*, 326-330.
- 103 Li, H.; Gao, Z.; Kang, L.; Zhang, H.; Yang, K.; Yu, K.; Luo, X.; Zhu, W.; Chen, K.; Shen, J.; Wang, X.; Jiang, H. *Nucl. Acids Res.* **2006**, *34*, 219-224.
- 104 Zahler, S.; Tietze, S.; Totzke, F.; Kubbutat, M.; Meijer, L.; Vollmar, A. M.; Apostolakis, J. *Chem. Biol.* **2007**, *14*, 1207-1214.
- 105 Keiser, M. J.; Setola, V.; Irwin, J. J.; Laggner, C.; Abbas, A. I.; Hufeisen, S. J.; Jensen, N. H.; Kuijper, M. B.; Matos, R. C.; Tran, T. B.; Whaley, R.; Glennon, R. A.; Hert, J.; Thomas, K. L. H.; Edwards, D. D.; Shoichet, B. K.; Roth, B. L. *Nature* **2009**, *462*, 175-181.
- 106 Mestres, J.; Seifert, S. A.; Oprea, T. I. *Clin. Pharmacol. Ther.* **2011**, *90*, 662-665.

- 
- 107 Gavernet, L.; Talevi, A.; Castro, E. A.; Bruno-Blanch, E. *QSAR Comb. Sci.* **2008**, *27*, 1120-1129.
- 108 Rollinger, J. M.; Steindl, T. M.; Schuster, D.; Kirchmair, J.; Anrain, K.; Ellmerer, E. P.; Langer, T.; Stuppner, H.; Wutzler, P.; Schmidtke, M. *J. Med. Chem.* **2008**, *51*, 842-851.
- 109 Schneider, G. *Nat. Rev. Drug Discov.* **2010**, *9*, 273-276.
- 110 Rollinger, J. M. *Phytochem. Lett.* **2009**, *2*, 53-58.
- 111 Paul, N.; Kellenberger, E.; Bret, G.; Müller, P.; Rognan, D. *Proteins: Structure, Function Bioinf.* **2004**, *54*, 671-680.
- 112 Lamb, M. L.; Burdick, K. W.; Toba, S.; Young, M. M.; Skillman, A. G.; Zou, X.; Arnold, J. R.; Kuntz, I. D. *Proteins: Structure, Function Genet.* **2001**, *42*, 296-318.
- 113 Lauro, G.; Romano, A.; Riccio, R.; Bifulco, G. *J. Nat. Prod.* **2011**, *74*, 1401-1407.
- 114 Lauro, G.; Masullo, M.; Piacente, S.; Riccio, R.; Bifulco, G. *Bioorg. Med. Chem.* **2012**, *20*, 3596-3602.
- 115 Cheruku, P.; Plaza, A.; Lauro, G.; Keffer, J.; Lloyd, J. R.; Bifulco, G.; Bewley, C. A. *J. Med. Chem.* **2012**, *55*, 735-742.
- 116 Supporting Information of Ref.113
- 117 Abad-Zapatero, C.; Metz, J. T. *Drug Discov. Today* **2005**, *10*, 464-469.
- 118 Aiello, A.; Carbonelli, S.; Esposito, G.; Fattorusso, E.; Iuvone, T.; Menna, M. *J. Nat. Prod.* **2000**, *63*, 517-519.
- 119 Erdelmeier, C. A. J.; Sticher, O. *Planta Med.* **1985**, *51*, 407-409.
- 120 Basile, A.; Sorbo, S.; Spadaro, V.; Bruno, M.; Faraone, N.; Rosselli, S. *Molecules* **2009**, *14*, 939-952.
- 121 Kim, H. H.; Bang, S. S.; Choi, J. S.; Han, H.; Kim, I. H. *Cancer Lett.* **2005**, *223*, 191-201.

- 
- 122 Tursch, B.; Braekman, J. C.; Dalozze, D. *Bull. Soc. Chim. Belg.* **1975**, *84*, 767-774.
- 123 Aiello, A.; Fattorusso, E.; Imperatore, C.; Menna, M.; Müller, W. E. G. *Mar. Drugs* **2010**, *8*, 258-291.
- 124 El-Razek, M. H. A.; Ohta, S.; Ahmed, A. A.; Hirata, T. *Phytochemistry* **2001**, *57*, 1201-1203.
- 125 Rosselli, S.; Maggio, A.; Bellone, G.; Formisano, C.; Basile, A.; Cicala, C.; Alfieri, A.; Mascolo, N.; Bruno, M. *Planta Med.* **2007**, *73*, 116-120.
- 126 Dallavalle, S.; Ferrari, A.; Biasotti, B.; Merlini, L.; Penco, S.; Gallo, G.; Marzi, M.; Tinti, M. O.; Martinelli, R.; Pisano, C.; Carminati, P.; Carenini, N.; Beretta, G.; Perego, P.; De Cesare, M.; Pratesi, G.; Zunino, F. *J. Med. Chem.* **2001**, *44*, 3264-3274.
- 127 Dallavalle, S.; Merlini, L.; Morini, G.; Musso, L.; Penco, S.; Beretta, G.; Tinelli, S.; Zunino, F. *Eur. J. Med. Chem.* **2004**, *39*, 507-513.
- 128 Burke, T. G.; Bom, D. *Ann. NY Acad. Sci.* **2006**, *922*, 36-45.
- 129 Kollmannsberger, C.; Mross, K.; Jakob, A.; Kanz, L.; Bokemeyer, C. *Oncology* **1999**, *56*, 1-12.
- 130 Staker, B. L.; Hjerrild, K.; Feese, M. D.; Behnke, C. A.; Burgin, A. B.; Stewart, L. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 15387-15392.
- 131 Aiello, A.; Fattorusso, E.; Luciano, P.; Macho, A.; Menna, M.; Muñoz, E. *J. Med. Chem.* **2005**, *48*, 3410-3416.
- 132 Gattinoni, S.; Merlini, L.; Dallavalle, S. *Tetrahedron Lett.* **2007**, *48*, 1049-1051.
- 133 Dallavalle, S.; Gattinoni, S.; Mazzini, S.; Scaglioni, L.; Merlini, L.; Tinelli, S.; Beretta, G. L.; Zunino, F. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 1484-1489.

- 
- 134 Squire, C. J.; Dickinson, J. M.; Ivanovic, I.; Baker, E. N. *Structure* **2005**, *13*, 541-550.
- 135 Martell, J. D.; Li, H.; Doukov, T.; Martásek, P.; Roman, L. J.; Soltis, M.; Poulos, T. L.; Silverman, R. B. *J. Am. Chem. Soc.* **2010**, *132*, 798-806.
- 136 Bell, I. M.; Gallicchio, S. N.; Abrams, M.; Beese, L. S.; Beshore, D. C.; Bhimnathwala, H.; Bogusky, M. J.; Buser, C. A.; Culberson, J. C.; Davide, J.; Ellis-Hutchings, M.; Fernandes, C.; Gibbs, J. B.; Graham, S. L.; Hamilton, K. A.; Hartman, G. D.; Heimbrook, D. C.; Homnick, C. F.; Huber, H. E.; Huff, J. R.; Kassahun, K.; Koblan, K. S.; Kohl, N. E.; Lobell, R. B.; Lynch, Jr., J. J.; Robinson, R.; Rodrigues, A. D.; Taylor, J. S.; Walsh, E. S.; Williams, T. M.; Zartman, C. B. *J. Med. Chem.* **2002**, *45*, 2388-2409.
- 137 Mulvihill, E. E.; Allister, E. M.; Sutherland, B. G.; Telford, D. E.; Sawyez, C. G.; Edwards, J. Y.; Markle, J. M.; Hegele, R. A.; Huff, M. W. *Diabetes* **2009**, *58*, 2198-2210.
- 138 Soobrattee, M. A.; Neergheen, V. S.; Luximon-Ramma, A.; Aruoma, O. I.; Bahorun, T. *Mutat. Res., Fundam. Mol. Mech. Mutagen.* **2005**, *579*, 200-213.
- 139 Fan, E.; Zhang, L.; Jiang, S.; Bai, Y. *J. Med. Food.* **2008**, *11*, 610-614.
- 140 Shankar, S.; Singh, G.; Srivastava, R. K. *Front Biosci.* **2007**, *12*, 4839-4854.
- 141 Gerhauser, C.; Alt, A. P.; Klimo, K.; Knauft, J.; Frank, N.; Becker, H. *Phytochem. Rev.* **2003**, *1*, 369-377.
- 142 Cuendet, M.; Guo, J.; Luo, Y.; Chen, S.; Oteham, C. P.; Moon, R. C.; van Breemen, R. B.; Marler, L. E.; Pezzuto, J. M. *Cancer. Prev. Res. Phila* **2010**, *3*, 221-232.

- 
- 143 Gerhauser, C.; Alt, A.; Heiss, E.; Gamal-Eldeen, A.; Klimo, K.; Knauff, J.; Neumann, I.; Scherf, H. R.; Frank, N.; Bartsch, H.; Becker, H. *Mol. Cancer Ther.* **2002**, *1*, 959-969.
- 144 Shon, Y. H.; Park, S. D.; Nam, K. S. *J. Biochem. Mol. Biol.* **2006**, *39*, 448-451.
- 145 Qian, Y. P.; Cai, Y. J.; Fan, G. J.; Wei, Q. Y.; Yang, J.; Zheng, L. F.; Li, X. Z.; Fang, J. G.; Zhou, B. *J. Med. Chem.* **2009**, *52*, 1963-1974.
- 146 Roy, P.; Kalra, N.; Prasad, S.; George, J.; Shukla, Y. *Pharm Res.* **2009**, *26*, 211-217.
- 147 Hsu, Y. L.; Chia, C. C.; Chen, P. J.; Huang, S. E.; Huang, S. C.; Kuo, P. L. *Mol. Nutr. Food Res.* **2009**, *53*, 826-835.
- 148 Deeb, D.; Gao, X.; Jiang, H.; Arbab, A. S.; Dulchavsky, S. A.; Gautam, S. *C. Anticancer Res.* **2010**, *30*, 3333-3339.
- 149 Sonmez, F.; Sevmezler, S.; Atahan, A.; Ceylan, M.; Demir, D.; Gencer, N.; Arslan, O.; Kucukislamoglu, M. *Bioorg. Med. Chem. Lett.* **2011**, *21*, 7479-7482.
- 150 Zanolì, P.; Zavatti, M. *J. Ethnopharmacol.* **2008**, *116*, 383-396.
- 151 Gerhauser, C. *Eur. J. Cancer* **2005**, *41*, 1941-1954.
- 152 Festa, M.; Capasso, A.; D'Acunto, C. W.; Masullo, M.; Rossi, A. G.; Pizza, C.; Piacente, S. *J. Nat. Prod.* **2011**, *74*, 2505-2513.
- 153 Sharma, P. *J. Chem. Pharm. Res.* **2011**, *3*, 403-423.
- 154 Petersen, M.; Simmonds, M. S. *J. Phytochemistry* **2003**, *62*, 121-125.
- 155 Akiyama, T.; Ishida, J.; Nakagawa, S.; Ogawara, H.; Watanabe, S.; Itoh, N.; Shibuya, M.; Fukami, Y. *J. Biol. Chem.* **1987**, *262*, 5592-5595.
- 156 Piacente, S.; Montoro, P.; Oleszek, W.; Pizza, C. *J. Nat. Prod.* **2004**, *67*, 882-885.

- 
- 157 Okura, A.; Arakawa, H.; Oka, H.; Yoshinari, T.; Monden, Y. *Biochem. Biophys. Res. Commun.* **1988**, *157*, 183-189.
- 158 Zhou, N.; Yan, Y.; Li, W.; Wang, Y.; Zheng, L.; Han, S.; Yan, Y.; Li, Y. *Int. J. Mol. Sci.* **2009**, *10*, 3255-3268.
- 159 Supporting Information of Ref.114
- 160 Wagner, J.; Matt, P. V.; Sedrani, R.; Albert, R.; Cooke, N.; Ehrhardt, C.; Geiser, M.; Rummel, G.; Stark, W.; Strauss, A.; Cowan-Jacob, S. W.; Beerli, C.; Weckbecker, G.; Evenou, J. P.; Zenke, G.; Cottens, S. *J. Med. Chem.* **2009**, *52*, 6193-6196.
- 161 Nagashima, K.; Shumway, S. D.; Sathyanarayanan, S.; Chen, A. H.; Dolinski, B.; Xu, Y.; Keilhack, H.; Nguyen, T.; Wiznerowicz, M.; Li, L.; Lutterbach, B. A.; Chi, A.; Paweletz, C.; Allison, T.; Yan, Y.; Munshi, S. K.; Klippel, A.; Kraus, M.; Bobkova, E. V.; Deshmukh, S.; Xu, Z.; Mueller, U.; Szewczak, A. A.; Pan, B.; Richon, V.; Pollock, R.; Blume-Jensen, P.; Northrup, A.; Andersen, J. N. *J. Biol. Chem.* **2011**, *286*, 6433-6448.
- 162 Stark, W.; Bitsch, F.; Berner, A.; Buelens, F.; Graff, P.; Depersin, H.; Fendrich, G.; Geiser, M.; Knecht, R.; Rahuel, J.; Rummel, G.; Schlaeppli, J. M.; Schmitz, R.; Strauss, A.; Wagner, J. [www.rcsb.org](http://www.rcsb.org); PDB code: 2JED.
- 163 Mellor, H.; Parker, P. J. *Biochem. J.* **1998**, *332*, 281-292.
- 164 Peifer, C.; Alessi, D. R. *Biochem. J.* **2009**, *417*, e5-e7.
- 165 Haughian, J. M.; Bradford, A. P. *J. Cell. Physiol.* **2009**, *220*, 112-118.
- 166 Ou, W.; Zhu, M.; Demetri, G. D.; Fletcher, C. D. M.; Fletcher, J. A. *Oncogene* **2008**, *27*, 5624-5634.
- 167 Blunt, J. W.; Copp, B. R.; Munro, M. H.; Northcote, P. T.; Prinsep, M. R. *Nat. Prod. Rep.* **2011**, *28*, 196-268.
- 168 Feling, R. H.; Buchanan, G. O.; Mincer, T. J.; Kauffman, C. A.; Jensen, P. R.; Fenical, W. *Angew. Chem., Int. Ed.* **2003**, *42*, 355-357.

- 
- 169 Manzanares, I.; Cuevas, C.; Garcia-Nieto, R.; Marco, E.; Gago, F. *Curr. Med. Chem.: Anti-Cancer Agents* **2001**, *1*, 257–276.
- 170 Mayer, A. M. S.; Glaser, K. B.; Cuevas, C.; Jacobs, R. S.; Kem, W.; Little, R. D.; McIntosh, J. M.; Newman, D. J.; Potts, B. C.; Shuster, D. E. *Trends Pharmacol. Sci.* **2010**, *31*, 255–265.
- 171 Taori, K.; Paul, V. J.; Luesch, H. *J. Am. Chem. Soc.* **2008**, *130*, 1806–1807.
- 172 Villa, F. A.; Gerwick, L. *Immunopharm. Immunotox.* **2010**, *32*, 228–237.
- 173 Plaza, A.; Keffer, J. L.; Bifulco, G.; Lloyd, J. R.; Bewley, C. A. *J. Am. Chem. Soc.* **2010**, *132*, 9069–9077.
- 174 Plaza, A.; Gustchina, E.; Baker, H. L.; Kelly, M.; Bewley, C. A. *J. Nat. Prod.* **2007**, *70*, 1753–1760.
- 175 Plaza, A.; Baker, H. L.; Bewley, C. A. *J. Nat. Prod.* **2008**, *71*, 473–477.
- 176 Matsunaga, S.; Fusetani, N.; Kato, Y.; Hirota, H. *J. Am. Chem. Soc.* **1991**, *113*, 9690–9692.
- 177 Harada, K.; Fujii, K.; Hayashi, K.; Suzuki, M.; Ikai, Y.; Oka, H. *Tetrahedron Lett.* **1996**, *37*, 3001–3004.
- 178 Rouhiainen, L.; Jokela, J.; Fewer, D. P.; Urmann, M.; Sivonen, K. *Chem. Biol.* **2010**, *17*, 265–273.
- 179 Walther, T.; Arndt, H. D.; Waldmann, H. *Org. Lett.* **2008**, *10*, 3199–3202.
- 180 Kobayashi, J.; Sato, M.; Ishibashi, M.; Shigemori, H.; Nakamura, T.; Ohizumi, Y. *J. Chem. Soc., Perkin Trans.* **1991**, *1*, 2609–2611.
- 181 Kobayashi, J.; Sato, M.; Murayama, T.; Ishibashi, M.; Walchi, M. R.; Kanai, M.; Shoji, J.; Ohizumi, Y. *J. Chem. Soc., Chem. Commun.* **1991**, *3*, 1050–1052.
- 182 Schmidt, E. W.; Harper, M. K.; Faulkner, D. J. *J. Nat. Prod.* **1997**, *60*, 779–782.

- 
- 183 Uemoto, H.; Yahiro, Y.; Shigemori, H.; Tsuda, M.; Takao, T.; Shimonishi, Y.; Kobayashi, J. *Tetrahedron* **1998**, *54*, 6719–6724.
- 184 Harada, K.; Fujii, K.; Shimada, T.; Suzuki, M.; Sano, H.; Adachi, K.; Carmichael, W. W. *Tetrahedron Lett.* **1995**, *36*, 1511–1514.
- 185 Bubik, A.; Sedmak, B.; Novinec, M.; Lenarcic, B.; Lah, T. T. *Biol. Chem.* **2008**, *389*, 1339–1346.
- 186 Gesner-Apter, S.; Carmeli, S. *J. Nat. Prod.* **2009**, *72*, 1429–1436.
- 187 Van Wagoner, R. M.; Drummond, A. K.; Wright, J. L. *Adv. Appl. Microbiol.* **2007**, *61*, 89–217.
- 188 Walther, T.; Renner, S.; Waldmann, H.; Arndt, H. D. *ChemBioChem* **2009**, *10*, 1153–1162.
- 189 Bunnage, M. E.; Blagg, J.; Steele, J.; Owen, D. R.; Allerton, C.; McElroy, A. B.; Miller, D.; Ringer, T.; Butcher, K.; Beaumont, K.; Evans, K.; Gray, A. J.; Holland, S. J.; Feeder, N.; Moore, R. S.; Brown, D. G. *J. Med. Chem.* **2007**, *50*, 6095–6103.
- 190 Rees, D. C.; Lipscomb, W. N. *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 5455–5459.
- 191 Rees, D. C.; Lipscomb, W. N. *J. Mol. Biol.* **1982**, *160*, 475–498.
- 192 Roy, J.; Laughton, C. *Biophys. J.* **2009**, *99*, 218–226.
- 193 Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theor. Comput.* **2008**, *4*, 435–447.
- 194 Harvey, M. J.; Giupponi, G.; De Fabritiis, G. *J. Chem. Theory Comput.* **2009**, *5*, 1632–1639.
- 195 Giupponi, G.; Harvey, M. J.; De Fabritiis, G. *Drug Discov. Today* **2008**, *13*, 1052–1058.
- 196 Buch, I.; Giorgino, T.; De Fabritiis, G. *Proc. Natl Acad. Sci. U. S. A.* **2011**, *108*, 10184–10189.

- 
- 197 Buch, I., Harvey, M. J.; Giorgino, T.; Anderson, D. P.; De Fabritiis, G. *J. Chem. Inf. Model.* **2010**, *50*, 397–403.
- 198 Voelz, V. A.; Bowman, G. R.; Beauchamp, K.; Pande, V. S. *J. Am. Chem. Soc.* **2001**, *132*, 1526–1528.
- 199 <http://www.acellera.com/products/acemd/>
- 200 Betzi, S.; Suhre, K.; Chetrit, B.; Guerlesquin, F.; Morelli, X. *J. Chem. Inf. Model.* **2006**, *46*, 1704–1712.
- 201 Charifson, P. S.; Corkery, J. J.; Murcko, M. A.; Walters, W. P. *J. Med. Chem.* **1999**, *42*, 5100–5109.
- 202 Oda, A.; Tsuchida, K.; Takakura, T.; Yamaotsu, N.; Hirono, S. *J. Chem. Inf. Model.* **2006**, *46*, 380–391.
- 203 Wall, I. D.; Leach, A. R.; Salt, D. W.; Ford, M. G.; Essex, J. W. *J. Med. Chem.* **1999**, *42*, 5142–5152.
- 204 Jones-Hertzog, D. K.; Jorgensen, W. L. *J. Med. Chem.* **1997**, *40*, 1539–1549.
- 205 Nicolotti, O.; Giangreco, I.; Miscioscia, T. F.; Convertino, M.; Leonetti, F.; Pisani, L.; Carotti, A. *J. Comput. Aided Mol. Des.* **2010**, *24*, 117–129.
- 206 Almlöf, M.; Carlsson, J.; Aqvist, J. *J. Chem. Theor. Comput.* **2007**, *3*, 2162–2175.
- 207 Rock, K. L.; Latz, E.; Ontiveros, F.; Kono, H. *Annu. Rev. Immunol.* **2010**, *28*, 321–342.
- 208 Ma, X.; Shah, Y. M.; Guo, G. L.; Wang, T.; Krausz, K. W.; Idle, J. R.; Gonzalez, F. J. *J. Pharmacol. Exp. Ther.* **2007**, *322*, 391–398.
- 209 Fiorucci, S.; Cipriani, S.; Baldelli, F.; Mencarelli, A. *Prog. Lipid Res.* **2010**, *49*, 171–185.

- 
- 210 Mencarelli, A.; Migliorati, M.; Barbanti, M.; Cipriani, S.; Palladino, G.; Distrutti, E.; Renga, B.; Fiorucci, S. *Biochem. Pharmacol.* **2010**, *80*, 1700-1707.
- 211 Xie, W.; Tian, Y. *Cell Metab.* **2006**, *4*, 177-178.
- 212 Zhou, C.; Tabb, M. M.; Nelson, E. L.; Grün, F.; Verma, S.; Sadatrafiei, A.; Lin, M.; Mallick, S.; Forman, B. M.; Thummel, K. E.; Blumberg, B. *J. Clin. Invest.* **2006**, *116*, 2280-2289.
- 213 Langmann, T.; Moehle, C.; Mauerer, R.; Scharl, M.; Liebisch, G.; Zahn, A.; Stremmel, W.; Schmitz, G. *Gastroenterology* **2004**, *127*, 26-40.
- 214 Day, A. S.; Gearry, R.B. *Dig. Dis. Sci.* **2010**, *55*, 877-879.
- 215 Cheng, J.; Shah, Y. M.; Ma, X.; Pang, X.; Tanaka, T.; Kodama, T.; Krausz, K. W.; Gonzalez, F. J. *J. Pharmacol. Exp. Ther.* **2010**, *335*, 32-41.
- 216 Jiang, Z. D.; Ke, S.; DuPont, H. L. *Int. J. Antimicrob. Agents* **2010**, *35*, 278-281.
- 217 Sepe, V.; Bifulco, G.; Renga, B.; D'Amore, C.; Fiorucci, S.; Zampella, A. *J. Med. Chem.* **2011**, *54*, 1314-1320.
- 218 De Marino, S.; Ummarino, R.; D'Auria, M. V.; Chini, M. G.; Bifulco, G.; Renga, B.; D'Amore, C.; Fiorucci, S.; Debitus, C.; Zampella, A. *J. Med. Chem.* **2011**, *54*, 3065-3075.
- 219 De Marino, S.; Sepe, V.; D'Auria, M. V.; Bifulco, G.; Renga, B.; Petek, S.; Fiorucci, S.; Zampella, A. *Org. Biomol. Chem.* **2011**, *9*, 4856-4862.
- 220 Sepe, V.; Ummarino, R.; D'Auria, M. V.; Chini, M. G.; Bifulco, G.; Renga, B.; D'Amore, C.; Debitus, C.; Fiorucci, S.; Zampella, A. *J. Med. Chem.* **2012**, *55*, 84-93.
- 221 Renga, B.; Mencarelli, A.; D'Amore, C.; Cipriani, S.; D'Auria, M. V.; Sepe, V.; Chini, M. G.; Monti, M. C.; Bifulco, G.; Zampella, A.; Fiorucci, S. *PLoS One* **2012**, *7*, e30443.

- 
- 222 Chini, M. G.; Jones, C. R.; Zampella, A.; D'Auria, M. V.; Renga, B.; Fiorucci, S.; Butts, C. P.; Bifulco, G. *J. Org. Chem.* **2012**, *77*, 1489-1496.
- 223 De Marino, S.; Ummarino, R.; D'Auria, M. V.; Chini, M. G.; Bifulco, G.; D'Amore, C.; Renga, B.; Mencarelli, A.; Petek, S.; Fiorucci, S.; Zampella, A. *Steroids* **2012**, *77*, 484-495.
- 224 Festa, C.; De Marino, S.; D'Auria, M. V.; Bifulco, G.; Renga, B.; Fiorucci, S.; Petek, S.; Zampella, A. *J. Med. Chem.* **2011**, *54*, 401-405.
- 225 Shah, Y. M.; Ma, X.; Morimura, K.; Kim, I.; Gonzalez, F. J. *Am. J. Physiol. Gastrointest. Liver Physiol.* **2007**, *292*, G1114-G1122.
- 226 Sepe, V.; Ummarino, R.; D'Auria, M. V.; Mencarelli, A.; D'Amore, C.; Renga, B.; Zampella, A.; Fiorucci, S. *J. Med. Chem.* **2011**, *54*, 4590-4599.
- 227 Rousseaux, C.; Lefebvre, B.; Dubuquoy, L.; Lefebvre, P.; Romano, O.; Auwerx, J.; Metzger, D.; Wahli, W.; Desvergne, B.; Naccari, G. C.; Chavatte, P.; Farce, A.; Bulois, P.; Cortot, A.; Colombel, J. F.; Desreumaux, P. *J. Exp. Med.* **2005**, *201*, 1205-1215.
- 228 Watkins, R. E.; Wisely, G. B.; Moore, L. B.; Collins, J. L.; Lambert, M. H.; Williams, S. P.; Willson, T. M.; Kliewer, S. A.; Redinbo, M. R. *Science* **2001**, *292*, 2329-2333.
- 229 Ekins, S.; Kortagere, S.; Iyer, M.; Reschly, E. J.; Lill, M. A.; Redinbo, M. R.; Krasowski, M. D. *PLoS Comput. Biol.* **2009**, *5*, e1000594, 1-12.
- 230 Xue, Y.; Moore, L. B.; Orans, J.; Peng, L.; Bencharit, S.; Kliewer, S. A.; Redinbo, M. R. *Mol. Endocrinol.* **2007**, *21*, 1028-1038.
- 231 Wang, K.; Damjanov, I.; Wan, Y. J. *Lab. Invest.* **2010**, *90*, 257-265.
- 232 Marek, C. J.; Tucker, S. J.; Konstantinou, D. K.; Elrick, L. J.; Haefner, D.; Sigalas, C.; Murray, G. I.; Goodwin, B.; Wright, M. C. *Biochem. J.* **2005**, *387*, 601-608.

- 
- 233 Renga, B.; Mencarelli, A.; Migliorati, M.; Cipriani, S.; D'Amore, C.; Distrutti, E.; Fiorucci, S. *Inflamm. Res.* **2011**, *60*, 577-587.
- 234 Fiorucci, S.; Antonelli, E.; Rizzo, G.; Renga, B.; Mencarelli, A.; Riccardi, L.; Orlandi, S.; Pellicciari, R.; Morelli, A. *Gastroenterology* **2004**, *127*, 1497-1512.
- 235 Fiorucci, S.; Antonelli, E.; Distrutti, E.; Severino, B.; Fiorentina, R.; Baldoni, M.; Caliendo, G.; Santagada, V.; Morelli A.; Cirino, G. *Hepatology* **2004**, *39*, 365-375.
- 236 Watkins, R. E.; Maglich, J. M.; Moore, L. B.; Wisely, G. B.; Noble, S. M.; Davis-Searles, P. R.; Lambert, M. H.; Kliewer, S. A.; Redinbo, M. R. *Biochemistry* **2003**, *42*, 1430-1438.
- 237 Xiao, L.; Nickbarg, E.; Wang, W.; Thomas, A.; Ziebell, M.; Prosser, W. W.; Lesburg, C. A.; Taremi, S. S.; Gerlach, V. L.; Le, H. V.; Cheng, K. C. *Biochem. Pharmacol.* **2011**, *81*, 669-679.
- 238 Mencarelli, A.; Renga, B.; Palladino, G.; Claudio, D.; Ricci, P.; Distrutti, E.; Barbanti, M.; Baldelli F.; Fiorucci, S. *Eur. J. Pharmacol.* **2011**, *668*, 317-324
- 239 Trauner, M.; Halilbasic, E. *Gastroenterology* **2011**, *140*, 1120-1125.
- 240 Kakizaki, S.; Takizawa, D.; Tojima, H.; Horiguchi, N.; Yamazaki, Y. Mori, M. *Front. Biosci.* **2011**, *17*, 2988-3005.
- 241 Haughton, E. L.; Tucker, S. J.; Marek, C. J.; Durward, E.; Leel, V.; Bascal, Z.; Onaghan, T.; Koruth, M.; Collie-Duguid, E.; Mann, D. A.; Trim, J. E.; Wright, M. C. *Gastroenterology* **2006**, *131*, 194-209.
- 242 Rizzo, G.; Disante, M.; Mencarelli, A.; Renga, B.; Gioiello, A.; Pellicciari, R.; Fiorucci, S. *Mol. Pharmacol.* **2006**, *70*, 1164-1173.
- 243 Heikkinen, S.; Auwerx, J.; Argmann, C. A. *Biochim. Biophys. Acta* **2007**, *1771*, 999-1013.

- 
- 244 Tontono, P.; Spiegelman, B. M. *Annu. Rev. Biochem.* **2008**, *77*, 289-312.
- 245 Pascual, G.; Fong, A. L.; Ogawa, S.; Gamliel, A.; Li, A. C.; Perissi, V.; Rose, D. W.; Willson, T. M.; Rosenfeld, M. G.; Glass, C. K. *Nature* **2005**, *437*, 759-763.
- 246 D'Auria, M. V.; Sepe, V.; Zampella, A. *Curr. Top. Med. Chem.* **2012**, *12*, 637-669.
- 247 Mora, F. D.; Jones, D. K.; Desai, P. V.; Patny, A.; Avery, M. A.; Feller, D. R.; Smillie, T.; Zhou, Y. D.; Nagle, D. G. *J. Nat. Prod.* **2006**, *69*, 547-552.
- 248 Kim, S. N.; Choi, H. Y.; Lee, W.; Park, G. M.; Shin, W. S.; Kim, Y. K. *FEBS lett.* **2008**, *582*, 3465-3472.
- 249 Rahm, F.; Hayes, P. Y.; Kitching, W. *Heterocycles* **2004**, *64*, 523-575.
- 250 Zampella, A.; Giannini, C.; Debitus, C.; D'Auria M. V. *Tetrahedron* **2001**, *57*, 257-263.
- 251 Campagnuolo, C.; Fattorusso, E.; Romano, A.; Tagliatela-Scafati, O.; Basilico, N.; Parapini, S.; Taramelli, D. *Eur. J. Org. Chem.* **2005**, 5077-5083.
- 252 Fattorusso, E.; Parapini, S.; Compagnuolo, C.; Basilico, N.; Tagliatela-Scafati, O.; Taramelli, D. *J. Antimicrob. Chemother.* **2002**, *50*, 883-888.
- 253 Kossuga, M. H.; Nascimento, A. M.; Reim~ao, J. Q.; Tempone, A. G.; Taniwaki, N. N.; Veloso, K.; Ferreira, A. G.; Cavalcanti, B. C.; Pessoa, C.; Moraes, M. O.; Mayer, A. M. S.; Hajdu, E.; Berlinck, R. G. S. *J. Nat. Prod.* **2008**, *71*, 334-339.
- 254 Fattorusso, E.; Tagliatela-Scafati, O. *Mar. Drugs* **2009**, *7*, 130-152.
- 255 Feng, Y.; Davis, R. A.; Sykes, M.; Avery, V. M.; Camp, D.; Quinn, R. J. *J. Nat. Prod.* **2010**, *73*, 716-719.

- 
- 256 Fattorusso, C.; Persico, M.; Calcinai, B.; Cerrano, C.; Parapini, S.; Taramelli, D.; Novellino, E.; Romano, A.; Scala, F.; Fattorusso, E.; Taglialatela-Scafati, O. *J. Nat. Prod.* **2010**, *73*, 1138-1145.
- 257 Mohammed, R.; Peng, J.; Kelly, M.; Yousaf, M.; Winn, E.; Odde, S.; Bie, Z.; Xie, A.; Doerksen, R. J.; Hamann, M. T. *Aust. J. Chem.* **2010**, *63*, 877-885.
- 258 Yong, K. W.; De Voss, J. J.; Hooper, J. N.; Garson, M. J. *J. Nat. Prod.* **2011**, *74*, 194-207.
- 259 Ueoka, R.; Nakao, Y.; Kawatsu, S.; Yaegashi, J.; Matsumoto, Y.; Matsunaga, S.; Furihata, K.; van Soest, R. W. M.; Fusetani, N. *J. Org. Chem.* **2009**, *74*, 4203-4207.
- 260 Kupchan, S. M.; Britton, R. W.; Ziegler, M. F.; Sigel, C. W. *J. Org. Chem.* **1973**, *38*, 178-179.
- 261 Stierle, D. B.; Faulkner, D. J. *J. Org. Chem.* **1980**, *45*, 3396-3401.
- 262 Compagnone, R. S.; Piña, I. C.; Rangel, H. R.; Dagger, F.; Suárez, A. I.; Reddy, M. V. R.; Faulkner, D. J. *Tetrahedron* **1998**, *54*, 3057-3068.
- 263 Shiraky, T.; Kamiya, N.; Shiki, S.; Kodama, T. S.; Kakizuka, A.; Jingami, H. *J. Biol. Chem.* **2005**, *280*, 14145-14153.
- 264 Itoh, T.; Fairall, L.; Amin, K.; Inaba, Y.; Szanto, A.; Balint, B. L.; Nagy, L.; Yamamoto, K.; Schwabe, J. W. R. *Nat. Struct. Mol. Biol.* **2008**, *15*, 924-931.
- 265 Waku, T.; Shiraki, T.; Oyama, T.; Fujimoto, Y.; Maebara, K.; Kamiya, N.; Jingami, H.; Morikawa, K. *J. Mol. Biol.* **2009**, *385*, 188-199.
- 266 Monti, M. C.; Margarucci, L.; Tosco, A.; Riccio, R.; Casapullo, A. *Food Funct.*, **2011**, *2*, 423-428.
- 267 Monti, M. C.; Chini, M. G.; Margarucci, L.; Riccio, R.; Bifulco, G.; Casapullo, A. *ChemBioChem*, **2011**, *12*, 2686-2691.

- 
- 268 Monti, M. C.; Casapullo, A.; Cavasotto, C. N.; Tosco, A.; Dal Piaz, F.; Ziemys, A.; Margarucci, L.; Riccio, R. *Chemistry, Eur. J.* **2009**, *15*, 1155-1163.
- 269 Davies, G. F.; McFie, P. J.; Khandelwal, R. L.; Roesler, W. J. *J. Pharmacol. Exp. Ther.* **2002**, *300*, 72-77.
- 270 Ciudin, A.; Hernandez, C.; Simo, R. *Curr. Top. Med. Chem.* **2012**, *12*, 585-604.
- 271 Hughes, T. S.; Chalmers, M. J.; Novick, S.; Kuruvilla, D. S.; Chang, M. R.; Kamenecka, T. M.; Rance, M.; Johnson, B. A.; Burris, T. P.; Griffin, P. R.; Kojetin, D. J. *Structure* **2012**, *20*, 139-150.
- 272 Raskin, P.; Rendell, M.; Riddle, M. C.; Dole, J. F.; Freed, M. I.; Rosenstock, J. *Diabetes Care* **2001**, *24*, 1226-1232.
- 273 Mattoo, V.; Eckland, D.; Widel, M.; Duran, S.; Fajardo, C.; Strand, J.; Knight, D.; Grossman, L.; Oakley, D.; Tan, M. *Clin. Ther.* **2005**, *27*, 554-567.
- 274 Nissen, S. E.; Nicholls, S. J.; Wolski, K.; Nesto, R.; Kupfer, S.; Perez, A.; Jure, H.; De Laroche, R.; Staniloae, C. S.; Mavromatis, K.; Saw, J.; Hu, B.; Lincoff, A. M.; Tuzcu, E. M. *JAMA* **2008**, *299*, 1561-1573.
- 275 Aronoff, S.; Rosenblatt, S.; Braithwaite, S.; Egan, J. W.; Mathisen, A. L.; Schneider, R. L. *Diabetes Care* **2000**, *23*, 1605-1611.
- 276 Diamond, G. A.; Bax, L.; Kaul, S. *Ann. Intern. Med.* **2007**, *147*, 578-581.
- 277 Singh, S.; Loke, Y. K.; Furberg, C. D. *JAMA* **2007**, *298*, 1189-1195.
- 278 Fiorucci S, Zampella A, Distrutti E. *Curr. Top. Med. Chem.* **2012**, *6*, 605-624.
- 279 Renga, B.; Mencarelli, A.; D'Amore, C.; Cipriani, S.; Baldelli, F.; Zampella, A.; Distrutti, E.; Fiorucci, S. *FASEB J.* **2012**, *26*, 3021-3031.
- 280 Rizzo, G.; Fiorucci, S. *Curr. Opin. Pharmacol.* **2006**, *6*, 421-427.

- 
- 281 Fiorucci, S.; Rizzo, G.; Antonelli, E.; Renga, B.; Mencarelli, A.; Riccardi, L.; Morelli, A.; Pruzanski, M.; Pellicciari, R. *J. Pharmacol. Exp. Ther.* **2005**, *315*, 58-68.
- 282 Elbrecht, A.; Chen, Y.; Adams, A.; Berger, J.; Griffin, P.; Klatt, T.; Zhang, B.; Menke, J.; Zhou, G.; Smith, R. G.; Moller, D. E. *J. Biol. Chem.* **1999**, *274*, 7913-7922.
- 283 Lee, G.; Elwood, F.; McNally, J.; Weiszmann J.; Lindstrom, M.; Amaral, K.; Nakamura, M.; Miao, S.; Cao, P.; Learned, R. M.; Chen, J. L.; Li, Y. *J. Biol. Chem.* **2002**, *277*, 19649-19657.
- 284 Pirat, C.; Farce, A.; Lebègue, N.; Renault, N.; Furman, C.; Millet, R.; Yous, S.; Speca, S.; Berthelot, P.; Desreumaux, P.; Chavatte, P. *J. Med. Chem.* **2012**, *55*, 4027-4061.
- 285 Wang, L. L. D.; Wang, C.; Zhang, Y.; Xu, J. *Mini. Rev. Med. Chem.* **2011**, *11*, 910-919.
- 286 Sun, H. D.; Huang, S. X.; Han, Q. B. *Nat. Prod. Rep.* **2006**, *23*, 673-698.
- 287 Tan, W.; Lu, J.; Huang, M.; Li, Y.; Chen, M.; Wu, G.; Gong, J.; Zhong, Z.; Xu, Z.; Dang, Y.; Guo, J.; Chen, X.; Wang, Y. *Chin. Med.* **2011**, *6*, 27-41.
- 288 Liu, J.; Huang, R.; Lin, D.; Wu, X.; Lu, H.; Pan, X.; Peng, J.; Li, M.; Lin, Q. *Comp. Clin. Path.* **2004**, *13*, 65-69.
- 289 Li, C. Y.; Wang, E. Q.; Cheng, Y.; Bao, J. K. *Int. J. Biochem. Cell. Biol.* **2011**, *43*, 701-704.
- 290 Ikezoe, T.; Yang, Y.; Bandobashi, K.; Saito, T.; Takemoto, S.; Machida, H.; Togitani, K.; Koeffler, H. P.; Taguchi, H. *Mol. Cancer. Ther.* **2005**, *4*, 578-586.
- 291 Li, X.; Li, X.; Wang, J.; Ye, Z.; Li, J. C. *Int. J. Biol. Sci.* **2012**, *8*, 901-912.
- 292 Rix, U.; Superti-Furga, G.; *Nat. Chem. Biol.* **2009**, *9*, 616-624.

- 
- 293 Lachance, H.; Wetzel, S.; Kumar, K.; Waldmann, H. *J. Med. Chem.* **2012**, *55*, 5989-6001.
- 294 Avonto, C.; Taglialatela-Scafati, O.; Pollastro, F.; Minassi, A.; Di Marzo, V.; De Petrocellis, L.; Appendino, G. *Angew. Chem. Int. Ed. Engl.* **2011**, *50*, 467-471.
- 295 Lee, I. S.; Shamon, L. A.; Chai, H. B.; Chagwedera, T. E.; Besterman, J. M.; Farnsworth, N. R.; Cordell, G. A.; Pezzuto, J. M.; Kinghorn, A. D. *Chem. Biol. Interact.* **1996**, *99*, 193-204.
- 296 McMorris, T. C.; Kelner, M. J.; Wang, W.; Moon, S.; Taetle, R. *Chem. Res. Toxicol.* **1990**, *3*, 574-579.
- 297 Weenen, H.; Nkunya, M. H.; Bray, D. H.; Mwasumbi, L. B.; Kinabo, L. S.; Kilimali, V. A.; Wijnberg, J. B. *Planta Med.* **1990**, *56*, 371-373.
- 298 Cetin, M.; Youn, Y. S.; Capan, Y.; Lee, K. C. *AAPS PharmSciTech* **2008**, *9*, 1191-1197.
- 299 Evans, C. G.; Chang, L.; Gestwicki, J. E. *J. Med. Chem.* **2010**, *53*, 4585-4602.
- 300 Dice, J. F. *Autophagy* **2007**, *3*, 295-299.
- 301 Wong, E.; Cuervo, A. M. *Cold. Spring. Harb. Perspect. Biol.* **2010**, *2*, a006734.
- 302 Powers, M. V.; Jones, K.; Barillari, C.; Westwood, I.; van Montfort, R. L.; Workman, P. *Cell. Cycle* **2010**, *9*, 1542-1550.
- 303 Leu, J. I.; Pimkina, J.; Frank, A.; Murphy, M. E.; George, D. L. *Mol. Cell.* **2009**, *36*, 15-27.
- 304 Fewell, S. W.; Smith, C. M.; Lyon, M. A.; Dumitrescu, T. P.; Wipf, P.; Day, B. W.; Brodsky, J. L. *J. Biol. Chem.* **2004**, *279*, 51131-51140.
- 305 Williams, D. R.; Ko, S. K.; Park, S.; Lee, M. R.; Shin, I. *Angew. Chem. Int. Ed. Engl.* **2008**, *47*, 7466-7469.

- 
- 306 Powers, M. V.; Clarke, P. A.; Workman, P. *Cancer Cell.* **2008**, *14*, 250–262.
- 307 Massey, A. J. *J. Med. Chem.* **2010**, *53*, 7280–7286.
- 308 Williamson, D. S.; Borgognoni, J.; Clay, A.; Daniels, Z.; Dokurno, P.; Drysdale, M. J.; Foloppe, N.; Francis, G. L.; Graham, C. J.; Howes, R.; Macias, A. T.; Murray, J. B.; Parsons, R.; Shaw, T.; Surgenor, A. E.; Terry, L.; Wang, Y.; Wood, M.; Massey, A. J. *J. Med. Chem.* **2009**, *52*, 1510–1513.
- 309 Chang, L.; Bertelsen, E. B.; Wisén, S.; Larsen, E. M.; Zuiderweg, E. R.; Gestwicki, J. E. *Anal. Biochem.* **2008**, *372*, 167–176.
- 310 Mayer, M. P.; Schröder, H.; Rüdiger, S.; Paal, K.; Laufen, T.; Bukau, B. *Nat. Struct. Biol.* **2000**, *7*, 586–593.
- 311 Mayer, M. P.; Bukau, B. *Cell. Mol. Life Sci.* **2005**, *62*, 670–684.
- 312 Carbone, D. L.; Doorn, J. A.; Kiebler, Z.; Sampey, B. P.; Petersen, D. R. *Chem. Res. Toxicol.* **2004**, *17*, 1459–1467.
- 313 Wisniewska, M.; Karlberg, T.; Lehtiö, L.; Johansson, I.; Kotenyova, T.; Moche, M.; Schüler, H. *Plos One* **2010**, *5*, e8625-e8632.
- 314 Zhuravleva, A.; Gierasch, L. M. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 6987–6992.
- 315 Hurley, J. H. *Annu. Rev. Biophys. Biomol. Struct.* **1996**, *25*, 137–162.
- 316 Wang, H.; Ye, Y.; Pan, S. Y.; Zhu, G. Y.; Li, Y.W.; Fong, D.W.; Yu, Z. L. *Phytomedicine* **2011**, *18*, 163–169.
- 317 Lu, X.; Xiao, L.; Wang, L.; Ruden, D. M. *Biochem. Pharmacol.* **2012**, *83*, 995–1004.
- 318 Powers, M. V.; Workman, P. *FEBS Lett.* **2007**, *581*, 3758–3769.
- 319 Mosser, D. D.; Morimoto, R. I. *Oncogene* **2004**, *23*, 2907–2918.
- 320 Brodsky, J. L.; Chiosis, G. *Curr. Top. Med. Chem.* **2006**, *6*, 1215–1225.

- 
- 321 Yerlikaya, A.; Okur, E.; Eker, S.; Erin, N. *Mol. Med. Report* **2010**, *3*, 333–339.
- 322 Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361–373.
- 323 Gaussian 09, Revision A.02, 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, Jr., 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, N. J.; 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, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian, Inc., Wallingford CT, 2009.
- 324 Number of cancer deaths: World Health Organization, mortality database <http://www.who.int/whosis/mort/download/en/index.html> accessed on 07/10/2012; Demographic data: United Nations, *World Population Prospects, the 2010 revision* <http://www.un.org/esa/population/unpop.html>.
- 325 Cohen, G. M. *Biochem J.* **1997**, *326*, 1-16.
- 326 Elmore, S. *Toxicol. Pathol.* **2007**, *35*, 495–516,
- 327 Haupt, S.; Berger, M.; Goldberg, Z.; Haupt Y. *J. Cell. Sci.* **2003**, *116*, 4077-4085.

- 
- 328 Uren, R. T.; Dewson, G.; Bonzon, C.; Lithgow, T.; Newmeyer, D. D.; Kluck, R.M. *J. Biol. Chem.* **2005**, 280, 2266-2274.
- 329 Grzasko, N. *Postepy. Hig. Med. Dosw.* **2004**, 58, 364-371.
- 330 Shen, Y.; White, E. *Adv. Cancer Res.* **2001**, 82, 55-84.
- 331 Levine, A. J. *Cell* **1997**, 88, 323-331.
- 332 Sherr, C. J. *Nat. Rev. Mol. Cell. Biol.* **2001**, 2, 731-737.
- 333 Shay, J. W.; Zou, Y.; Hiyama, E.; Wright, W. E. *Hum. Mol. Genet.* **2001**, 10, 677-685.
- 334 Hahn, W. C.; Meyerson, M. *Ann. Med.*, **2001**, 33, 123-129.
- 335 Shields, J. M.; Pruitt, K.; McFall, A.; Shaub, A.; Der, C. J. *Trends Cell. Biol.* **2000**, 10, 147-154.
- 336 Brazil, D. P.; Hemmings, B. A. *Trends Biochem. Sci.* **2001**, 26, 657-664.
- 337 Di Cristofano, A.; Pandolfi, P. P. *Cell.* **2000**, 18, 387-390.
- 338 Sellers, W. R.; Kaelin Jr, W. G. *J. Clin. Oncol.* **1997**, 15, 3301-3312.
- 339 Trimarchi, J. M.; Lees, J. A. *Nat. Rev. Mol. Cell. Biol.* **2002**, 3, 11-20.
- 340 Serrano, M.; Lin, A. W.; McCurrach, M. E.; Beach, D.; Lowe, S. W. *Cell.* **1997**, 88, 593-602.
- 341 Lloyd, A. C.; Obermüller, F.; Staddon, S.; Barth, C. F.; McMahon, M.; Land, H. *Genes. Dev.* **1997**, 11, 663-677.
- 342 Massagué, J. *Nat. Rev. Mol. Cell. Biol.* **2000**, 1, 169-178.
- 343 Bienz, M.; Clevers, H. *Cell* **2000**, 103, 311-320.
- 344 Verdu, J.; Buratovich, M. A.; Wilder, E.L.; Birnbaum, M. J. *Nat. Cell. Biol.* **1999**, 1, 500-506.