University of Florence

International Doctorate in Structural Biology

Cycle XXI (2006-2008)

"Drug Design Approaches For the Discovery of New Protein-Interacting Compounds"

Ph.D. thesis of Mattia Mori

Tutor ***Coordinator***

Prof. Ivano Bertini ***Prof. Claudio Luchinat***

S.S.D. CHIM/03

This thesis has been approved by the University of Florence, the University of Frankfurt and the Utrecht University

Table of Contents

Summary

This PhD project is focussed on the application of different strategies to develop new molecules able to interact with potential pharmacological targets. In particular, different approaches have been undertaken for the identification and optimization of hit and lead molecules. Computer-aided methodologies and computational tools, such as Molecular Mechanics and Quantum Mechanics, have been applied to better understand the chemical-physical features of the resulting protein-ligand adducts. Sample target proteins have been selected on the basis of their involvement in human diseases; more precisely, ligands have been designed to interact with these biomolecules, in such a way as to provide a therapeutic activity by interfering with the proteins function and related pathways. These ligands belong either to the broad class of enzyme inhibitors and to the less numerous class of the inhibitors of protein-protein interaction, according to the target protein's biological features. This computational investigation, flanked and supported by experimental data, highlights both the potentialities and the weakness of *in silico* approaches to drug design and discovery.

1. Introduction to drug design

The main therapeutic approach to human diseases is historically represented by the administration of synthetic and natural drugs. In the last decades, the need to increase the drugs' efficacy and to reduce their side effects resulted in the prevalent use of synthetic, *ad hoc* designed, molecules. [1] The fast-growing knowledge on biological systems has permitted a more efficient approach for designing ligands aimed at interacting with specific target macromolecules or metabolic and signalling pathways. Moreover, the complete sequencing of the human genome, [2,3] successfully carried out less than a decade ago, has provided great improvements in the field of life-sciences, and particularly in biotechnological research; the functional association between a gene and its related protein and pathways is a key point for approaching new strategies in drug discovery. [2,4] In the so called post-genomic era, efforts are now focused to understand the role of proteins in biological pathways and human diseases notably to derive new ligand compounds as therapeutics agents. A key component of these efforts is the use of protein structures in experimental design of drugs. [5] One of the major challenges is to translate all of the structural data into pharmaceutical knowledge, in particular to improve the pertinence and the specificity of the targets using genome-based drug discovery strategies. To achieve these objectives, many different methodologies have been developed, based on the use of protein structure data banks, combined with molecular modelling and exploitation of the energetic aspects of binding. [6,7]

Drug Design (DD) is the approach used to find drugs aimed at interacting with biological targets. It is currently one of the most reliable approaches for the identification of ligands with a certain biological activity; it can be considered as the logic evolution of the gene sequencing, technological progress, and the continuous need of, and search for innovative molecules to introduce in therapy. For these reasons, drug design approaches are nowadays widely undertaken in drug discovery by applying a large diversity of techniques and methodologies that span from chemistry to biology, and from *in vitro* experiments to *in silico* predictions of molecular properties. [5] Two main approaches common to all these strategies can be identified: Ligand-Based (LBDD) and Structure-Based (SBDD) Drug Design. The joint implementation of these two methods led to the Structure and Ligand-Based Drug Design (SLBDD). These names are rather self-explanatory. A fourth approach,

based on theoretical models for not yet structurally defined biological targets, has to be mentioned. [8] In the frame of each approach, a "rational" [9] or "irrational" strategy can be undertaken during the ligand selection process. Rational drug design is aimed at tailoring molecules beginning from the knowledge of some structural data and/or chemical response on the target organism, whereas with the term "irrational" are intended all that procedures that are generally free of structural foundations on ligand or target receptors. Indeed the irrational approach does not address any relationship between ligand's chemical structure and biological evidences, with respect to the rational drug design in which the so called Structure Activity Relationships (SAR) constitute a powerful guideline for the design of new molecules. These two philosophies can be also sequentially applied during the development of a large scale drug design project.

1.1. *Ligand-Based Drug Design.* Despite the great technological improvement in the field of protein spectroscopy and characterization, the 3D structure of about 75% of the validated target proteins has not yet been determined or made available. [8] However, many specific ligands are well known and described to bind these particular receptors, ranging from agonists to allosteric activators/inhibitors across antagonists and competitors. The ligand-based approach is aimed at extracting numerical properties for describing physical, chemical, and biological behaviours of the selected molecules towards that target proteins whose structure is not known, and for which is not possible to apply the structure-based approach. In particular, all those directly or indirectly measurable properties are defined as *molecular descriptors*. The pattern analysis between the variability of experimental observations and the correspondent molecular descriptor sets the bases to obtain Quantitative Structure Activity Relationships (QSAR) [10] that, by providing a 3D and physical description of the ligand's pharmacophores, can be exploited for designing new ligands. Due to the lack of the structure of target proteins, this approach is exclusively based on the chemical and biological properties of known and new ligands, and its reliability is proportional to the number of used molecules to asses the QSAR.

1.2. *Structure-Based Drug Design.* Technological progress in the fields of protein expression and spectroscopic characterization have made available a large and continuously growing number of 3D proteins structures. For target proteins whose structure

is available, the drug design approach is carried out by tailoring ligands on the basis of the spatial conformation of the target. [5,7] With respect to the ligand-based approach, the atomic detail of the intermolecular interaction constitute the driving force to optimize the binding properties of ligands. SBDD highlights the differences between rational and irrational approaches. Just as example, when the binding site of the target protein is defined, a rational approach can be undertaken in order to select and develop molecules that interact only in that specific protein region. When the binding site is not defined, an irrational approach can be planned to explore the protein region mainly involved in ligand binding. The logical joint implementation or these two approaches is represented by the time sequence of irrational and rational methods.

1.3. *Structure and Ligand-Based Drug Design.* Many target proteins are characterized in their ligand and/or substrate binding conformation. Thus, the knowledge of the chemical structure and binding pose for some specific ligands can be exploited for the design of new, "further-generation" of molecules. [11] Moreover, the knowledge of the binding conformation of the target protein generally provides higher reliability in ligand design. This is mainly a rational approach, and the availability of a starting compound constitutes a great advantage in drug design.

1.4. *Model-Based approach.* An undefined number of potential target proteins are still uncharacterized. In case no ligands have been tested or described towards these proteins, and *in vitro* assays are unavailable for testing their biological response to ligand binding, the above described approaches cannot be undertaken. For this completely irrational approach, the most common procedure requires the computer-aided building of a homology model of the biological target on which to test, by virtual simulations, a set of ligands. The reliability of these models is proportional to the sequence homology between the target protein and the reference structure. It has been estimated that the structure of about 20% of validated target proteins, still not characterized, could be modelled on the basis of the structure of highly homologous proteins available on protein databases. [8] This model-based approach is more frequently carried out for tailoring ligands towards membrane proteins, ion channels and new, unstructured, target proteins. [12,13] However, it is not always feasible because many potential drug targets, inside the cell, are constituted by oligomers of protein

and receptors, whose complex structures are quite difficult to be modelled with these bioinformatics approaches right now available.

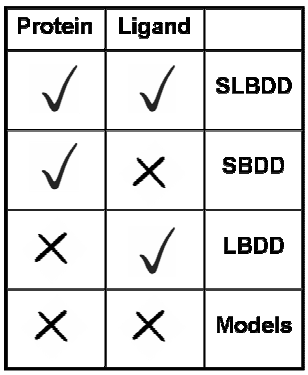


Figure 1. Schematic representation of the dependence between the described approach (right column) and the availability of the starting molecules.

2. Computer Aided Drug Design

The above described approaches are schematized in Figure 1. Despite the different methodologies that can be undertaken during each approach, the aim of drug design is to discover lead molecules to be developed and optimized as candidate drugs. It has to be considered that the probability of success for a general lead discovery project is proportional to the number of starting molecules. The PubChem Database, as example, is one of the most complete database of existing organic molecules and comprises about 19 million of compounds and 41 million of substances, [14] whereas the ZINC database [15] contains over 8 million purchasable molecular hits. As they appear, these databases represent an excessive amount of molecules to be managed during a drug discovery project. Drug design is thus aimed at drastically reducing the number of putative ligands for each target, down to a reasonable and handy number of leads, by filtering the initial large set of molecules by specific filters chosen on the basis of the selected methodology. A filter can be represented in a first instance by the experimentally measured, or *in silico* predicted, biological activity; many other filters can be added, such as toxicity evaluation or prediction, analysis of the synthetic perspectives, molecular weight or other chemical features related to the Lipinski's rules (see chapter 5.2).

The huge amount of starting molecules that can be putatively considered in drug design, however, represents a limiting factor for the time duration of the entire process, especially when approached by sophisticated techniques as X-Ray crystallography and NMR spectroscopy. To get around the problem, in the last decades computer aided procedures have been developed. In this way, the initial phases of each approach are supplied by *in silico* simulations of the ligands and/or of protein-ligand adducts behaviour, focussing the experimental efforts only to the most promising ligands.

The development of focussed software has greatly contributed to the widespread usage of computer aided drug design. [16-18] This method makes use of computational chemistry to simulate the molecular behaviour of both ligands and macromolecular target, in order to discover and optimize new candidate drugs aimed at interacting with biologically active macromolecule such as proteins or DNA polymers. This *in silico* methodology includes simple molecular modelling, using molecular mechanics, molecular dynamics, semi-empirical methods and *ab initio* quantum chemistry to select the most promising hits and

leads on which to concentrate the time-consuming experimental procedures. By the way, *hits* are defined as promising ligands, generally characterized by a molecular weight lower than 300 daltons, that provide a certain activity in preliminary tests, whereas with the term *leads* are considered all the molecules with a molecular weight generally within the range 300-500 daltons, optimized towards the selected target according to the pharmacological criteria of efficacy, safety and ADME properties (Absorption, Distribution, Metabolism, Elimination). Moreover, hits and leads follow the Lipinski's "rule of 3" and "rule of 5" respectively, as further discussed.

Computer aided drug design needs to be complemented by experimental data, especially in developing new methods and to provide confirmations about the reliability of the selected approach.

This *in silico* methodology has been selected to carry out the project described in this thesis. Molecular modelling procedures were investigated, as further described, with the aim to identify hit or lead compounds towards the selected target proteins. More precisely, docking calculations were performed during the preliminary ligands screening steps, whereas molecular dynamics and quantum mechanics were applied for the refinement and optimization of the docking results. During this project both synthetic molecules and peptides were accounted as candidate drugs.

3. Computational methods

This PhD project is mainly focussed on *in silico* drug design. In particular, due to the availability of the 3D structure of the selected target proteins, the structure based approach has been carried out in both rational and irrational ways to design ligands. Computational techniques and theoretical methods were applied to model the molecular behaviour of ligands, targets and their resulting complexes and adducts.[18] Here is reported a brief and basic description of the computational methods used in this work. The techniques listed below were exploited in a case-specific manner, on the basis of the level of refinement requested for the results.

3.1. *Molecular Mechanics (MM) and Docking.* Under the name "molecular mechanics" are collected all the simulations that refer to the use of Newton's mechanics equations to approximate molecular properties and behaviour. Accordingly, atoms are considered as solid spheres with mass proportional to their atomic weight and a radius corresponding, in general, to the Van-der-Waals atomic radius. Electrons are not accounted for; hence, atom charges and polarization properties cannot be computed during MM calculations and have to be retrieved from quantum chemistry or experimental data, where available. Bonded interactions are treated as springs, with strengths and distances assigned from available experimental or quantum chemistry data. The functional application of molecular mechanics is the *force field* equation that describes all the terms involved in the energy calculation. [19,20] In a generic force field, bonds, bond angles and dihedral angles are considered as harmonic potentials centred around the equilibrium values, that are experimentally calculated or theorized by *ab initio* computations. The implementation of force field equations by the addition of non-bonding terms (VdW, electrostatic and solvation/desolvation) completes the chemical and physical description of the molecular environment. These equations are generally applied in the energy minimization process of a single molecule. However, force fields can even be used for the energy minimization of a complex of molecules such as ligand-protein adducts; one of the methods for exploring this aspect of binding is docking. Docking is the easiest process by which two molecules fit together in 3D space. It is a MM-based approach to molecular modelling commonly applied for the prediction of binding conformations, and relative energies, of molecules

towards a target receptor. The docking methodology can be approached for the prediction of protein-protein conformation and energy of interaction (protein docking), or to simulate the interaction between a protein and a ligand (molecular docking). Molecular docking has been largely used in the present work; it represents the first and most intuitive tool for the *in silico* structure-based drug design, allowing the prediction of preferred orientations, and their relative binding energies, of small molecules towards target proteins.

$$\Delta G = \Delta G_{vdw} \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right) + \Delta G_{hbond} \sum_{i,j} E(t) \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} + E_{hbond} \right) + \\ + \Delta G_{elec} \sum_{i,j} \frac{q_i q_j}{\varepsilon(r_{ij})^2} + \Delta G_{tor} N_{tor} + \Delta G_{solv} \sum_{i,j} \left(S_i V_j + S_j V_i \right) e^{\frac{-r_{ij}^2}{2\sigma^2}} \quad (1)$$

Equation 1 – AutoDock force field equation for the calculation of the Free Energy of Binding

In equation 1 is described the general formula for the calculation of the free energy of binding, in the way it is used by AutoDock program. [21-23] This energy is considered as the sum of different contributions that are determined by linear regression analysis of complexes with known 3D-structures and known binding free energies. More precisely, these free energy weights are: Van-der-Waals energy (calculated by the 12-6 Lennard-Jones equation, solved for each pair of i and j atoms), hydrogen bond energy (considered as a 12-10 Lennard-Jones potential plus an energetic adjustment related to the angle between acceptor and donor), electrostatic energy (calculated by the Coulomb's equation implemented with the dielectric constant), ligand's torsional energy and solvation energy (parameterized as the contribution of the full complexes minus the only protein solvation energy). In Lennard-Jones equations for non-bonding contribution to the total energy of the system, coefficients A, B, C and D, reported in equation 1, correspond to $4\varepsilon\sigma^n$ where ε is the depth of the potential well and σ is the (finite) distance at which the interparticle potential is zero. The *n* exponent refers to the order of the specific term. Just as example, the Van-der-Walas coefficients A and B for carbon-carbon interaction are 373712.5 and 182.5, whereas the same coefficients for oxygen-carbon interaction are 121872.1 and 112.0, respectively.

According to the standard operating mode of AutoDock 3.0.5, the structure of the target protein is kept unchanged during the simulation of the ligands' pose. In this way, rigid docking corresponds to the early pharmaceutical concept of "lock-and-key" for ligand-protein interactions, where the protein can be thought of as the "lock" and the ligand can be thought of as a "key". The docking procedure considers the target receptor as a rigid body on which to bind the flexible ligand. This mechanism, however, introduces an approximation relative to the protein-ligand reciprocal adaptation that occurs in a physiological environment. The lock-and-key concept has been progressively dismissed in favor of the "hand-and-glove" or "induced-fit" mechanism for which both protein and ligand modify their conformation due to a reciprocal adaptation. With the aim to increase the reliability of Docking simulations, this problem can be circumvented, for example, by taking a particular care in the choice of the target protein's conformation. Interestingly, this approximation on the protein behaviour determines a faster solution of equation 1 in which the atoms *i* and *j* involved in intra-protein interactions are not accounted for.

Programs for flexible docking have been also developed to permit the movements of small residues' side chains, in order to moderately adapt the protein conformation to the ligand binding. Backbone atoms and side chains of bulky residues such as phenylalanine, tyrosine and arginine, are kept fixed as in the rigid docking protocol; these residues, however, greatly participate to the ligand interaction and stabilization introducing a limit to this methodology. Moreover, the partial flexibility of the protein determines a higher number of atoms to be computed in the force field equation, resulting in increased time of calculation, and attributing to flexible docking a limited applicability. Nevertheless, rigid docking represents a powerful tool for the hit selection and lead optimization processes, especially thanks to some of its advantages, such as the speed of calculation, the graphical rendering of the results and the reliability. For these reason docking procedures are largely applied for virtual screening.

In the present PhD project docking procedures were applied for the irrational virtual screening of libraries and for rational drug design towards the selected target proteins, with the aim to provide preliminary information about the molecular scaffolds suitable for optimization and their relative energy of binding.

3.2. *Molecular Dynamics (MD).* One of the possible approaches to refine and optimize docking results is represented by Molecular Dynamics. MD is again a molecular mechanics technique but, with respect to docking, can take into account the reciprocal movements and the time evolution of protein-ligand or protein-protein complexes. Indeed, with this method the movements of the target receptor can be accounted during the reciprocal adaptation and arrangements of its complex with a ligand; for this reason Molecular Dynamics represents the most reliable tool to switch from the "lock-and-key" like approximation observed with rigid docking simulations, to a "induced-fit" like simulation of the interaction. In a general view, the MD process consists in supplying a stressing amount of energy to the system with the purpose to analyze the conformational variation during the recovering at the energetic state of minimum. So, even MD is a procedure for energy minimization and, such as docking, energy is calculated by an equation that describe the relative force field (equation 2, Basic Amber Force Field).

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

Equation 2 – Amber force field for the potential energy calculation. The energy of a system is calculated as the sum of bond, angles, dihedrals, Van-der-Waals and electrostatic contributions. Weights and coefficients are calculated in the same way as above described for AutoDock.

During the present project the Amber package [24-26] (versions 8, 9 and 10) was used in order to refine the conformation of the most interesting ligand-protein complexes, as obtained from preliminary docking simulations. The reciprocal conformation of target protein and ligand at the energy convergence, together with other energetic and geometric variables, was taken into account in the results analysis. MD can offer more reliability than docking in predicting the conformational behaviour with time of a molecular system. However, it requires a greater amount of informatics resources, determining a longer computational time with respect to the simplest docking procedure. For this reason MD can be approached for lead optimization or, at maximum, at the hit selection step of the structure-based drug design workflow further described.

3.3. *Quantum Mechanics (QM).* Quantum chemistry is that branch of physical chemistry that applies quantum mechanics and field theory to understand the electronic behaviour and the reactivity of atoms and molecules. With respect to MM, electrons are explicitly considered. Thus, atomic orbitals are described as electron clouds in which the relative probability to find electrons is about 90%. Molecular orbital calculations carried out accordingly, are even defined as electronic structure calculations and are based on solving the Schrödinger's equation with the electronic molecular hamiltonian.

For a molecule based approach, given a set of nuclear coordinates, all the other properties such as atom connectivity, point charges, polarizability and bond strength are directly calculated by the programs and do not need to be pre-assigned as in molecular mechanics. This level of theory implementation provides a higher reliability in predicting molecular properties and behaviour, even if the time for calculation remarkably rise up with respect to MM approaches. For this reason, by a quantum approach molecules or molecular systems constituted by less than 200 atoms can be studied, whereas, by MM approaches, the properties of systems constituted by thousands of atoms can be predicted. However, QM approach allows the study of not standard molecular properties and interactions related to the molecular electronic behaviour. For example, stacking interactions, chemical reactivity and electronic induction and transfer are not suitable for MM studies, but require a quantum methodology. [27,28] Finally, these properties have to be generally accounted for in the lead optimization phase, whereas they can be generally neglected in the initial screening phase of the SBDD workflow described below. The Gaussian03 program [29] was selected for molecular orbital calculations in our work. In the frame of the interaction between a target protein and a candidate drug, the QM approach has been applied for the fine prediction of electronic-related properties. In particular, desolvation contributions to the binding and π-stacking interactions were accounted for by quantum chemistry approaches.

3.4*.Integrated Approach*. According to the molecular modelling overview presented above, the structure-based drug design approach carried out in this PhD work can be schematized as reported in Figure 2. The approach mainly used to simulate the interaction of small molecules towards target proteins is the docking calculation (dot line in Figure 2). This methodology, in fact, permits the analysis of thousands of molecules in a time much lower

than MD and QM based calculations; for this reason, docking simulation was used for preliminary virtual screening and rational drug design. In case of very large ligands, docking analysis was replaced by molecular dynamics calculations (see cap. 5.4). By the way, more sophisticated calculations were performed to optimize the most interesting hit and lead compounds. In particular, the necessity to explore the reciprocal protein-ligand adaptation and to analyze some fine molecular properties (especially those related to the electron transfer and polarization), has led to the introduction of two refinement cycles into the workflow. MD and QM were the preferred methodologies for these refinements, finalized to increase the success of the selected drug design approach. This general strategy has been adapted to the project carried out here and the predictions have been confirmed by the experimental results.

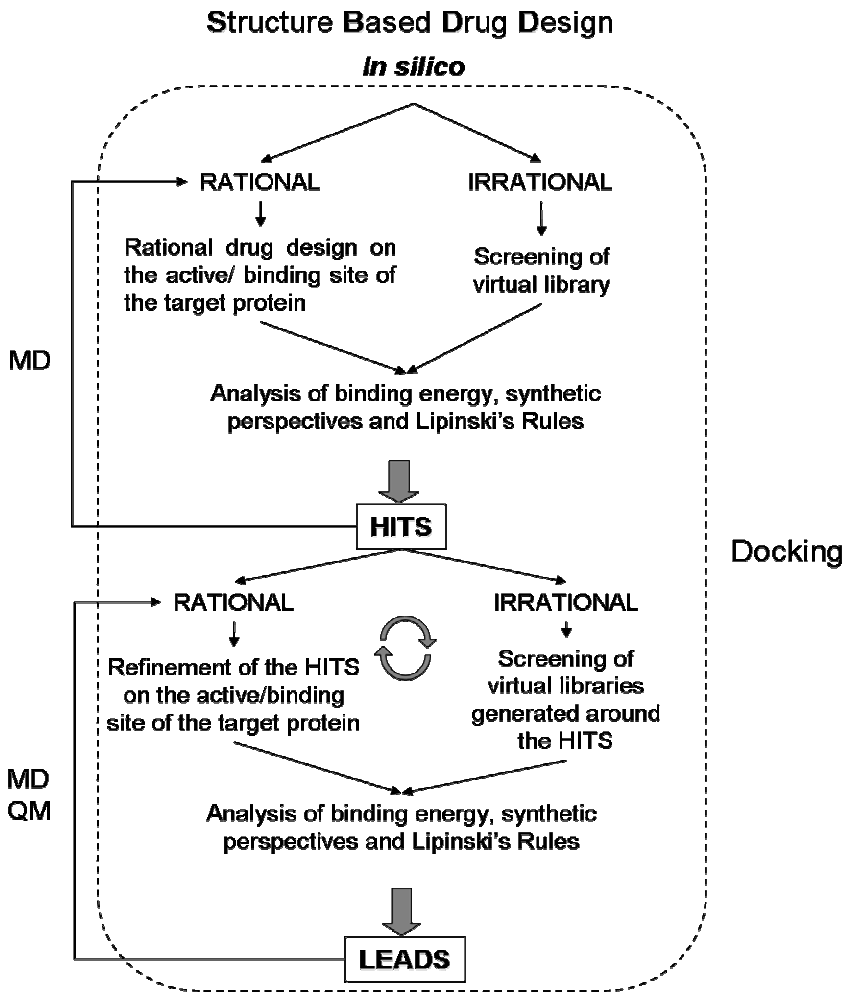


Figure 2. SBDD workflow used in the presented work

4. Selected Target Proteins

In this project, proteins belonging to three main families have been taken into account as drug targets. Here a description of the main features for each protein family and targeted ligands is provided.

4.1. *S100s*

S100 proteins are a highly conserved family of calcium-binding proteins found in vertebrates, and increasingly considered potential drug targets. [30-33] The name is hystorically derived from the fact that these proteins are 100% soluble in ammonium sulphate at neutral pH. More than 20 different members of the family are present in humans. S100 proteins are low-molecular weight protein dimers, Figure 3, where each monomer is constituted by an EF-hand domain, containing two helix-loop-helix motifs interconnected by a linker (hinge loop) and capable of binding two calcium ions in the loops, in a characteristic way.

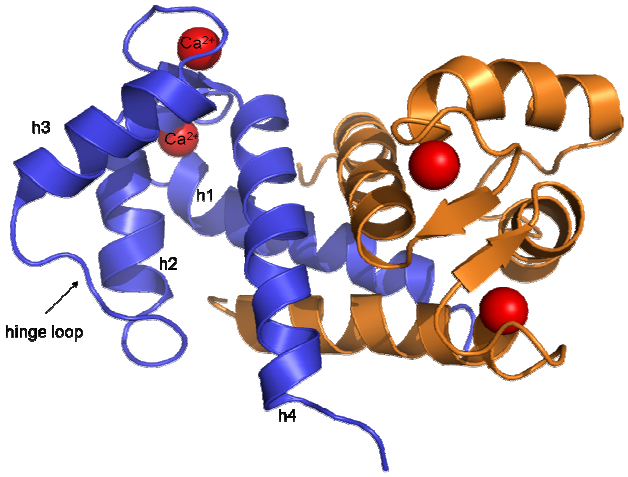


Figure 3. cartoon representation of a S100 protein (S100B). In blue and orange are evidenced the two monomer, in red spheres the calcium ions. The description of the helices is reported for only one monomer.

The EF-hand domain in these proteins is characterized by the first loop being noncanonical (14 aa long) and by the first and last helices preceded and followed by residues with no

secondary structure. [34-36] Different S100 proteins are expressed in a tissue-specific and cell type-specific pattern. [37] In particular, their expression occurs in many different tissues including those of the nervous system, musculature, skin, adipose tissues, reproductive system, gastrointestinal system, respiratory system, and urinary system. [38-41] Importantly, these proteins regulate intracellular processes such as cell growth and mobility, cell cycle regulation, transcription, and differentiation. This suggests that different S100 proteins have different functions, and as they do not have any catalytic activity of their own it is likely that they regulate the activity of other proteins. Many of these regulatory pathways involve a direct interaction of a specific S100 protein with a particular target protein, so it is reasonable to expect that different members of the S100 family have quite different physiological roles. Misregulation of any of these interactions can thus cause pathologies, which makes S100 proteins potential drug targets. It is widely appreciated that inhibiting protein-protein interactions is a more ambitious goal than inhibiting, for examples, enzymes. [42] However, at variance with general enzyme inhibitors, S100 protein family members have a variety of partners and, therefore, ligand selectivity towards these proteins can be expected to be more easily achieved. [43] This family of proteins has been studied during the present project by following two different strategies. A first approach, based on molecular modelling procedures, was carried out in order to identify the S100B's residues involved in substrate and ligand binding. In fact, only few ligands are reported to weakly interact with S100B, [44,45] and for these molecules structural detail of the interaction are not available. A second step was aimed at the *in silico* and NMR based identification of fragments able to interact with S100B and, separately, with S100A13.

4.2*.MMPs*

MMPs are a family of extracellular zinc and calcium dependent endopeptidases. [46] 23 members of the family plus a variable number of isoenzymes have been identified in the human genome and classified in 4 groups on the basis of their substrate specificity: i) Gelatinases (MMP-2 and MMP-9) are capable to digest type IV collagen and gelatin; they expose the gelatin binding region as a folding element close to the catalytic zinc binding motifs. ii) Collagenases (MMP-1, MMP-8, MMP-13) cleave collagen fibers at about ¼ of their length and are considered the only proteins able to degrade collagen fibers in

mammalian bones and cartilage. iii) Stromelysins (MMP-3, MMP-10, MMP-11) are not able to cleave collagen fibers but display a broad capability to cleave other extracellular matrix proteins. iv) mixed groups of MMPs that include elastase (MMP-12), matrilysin (MMP-7), enamelysin (MMP-20), membrane-type MMPs (MMP-14, MMP-15, MMP-16, MMP-17, MMP-24, MMP-25) and all the other members still not assigned to one of previously described groups. [47] From a structural point of view, all MMPs, with the only exception of MMP-7, are constituted by a catalytic domain connected to a haemopexin-like domain by a flexible linker (also called hinge region) that constitutes the recognition site for antibodies towards these proteins. MMP-2 and MMP-9 present an additional fibronectin-like domain.

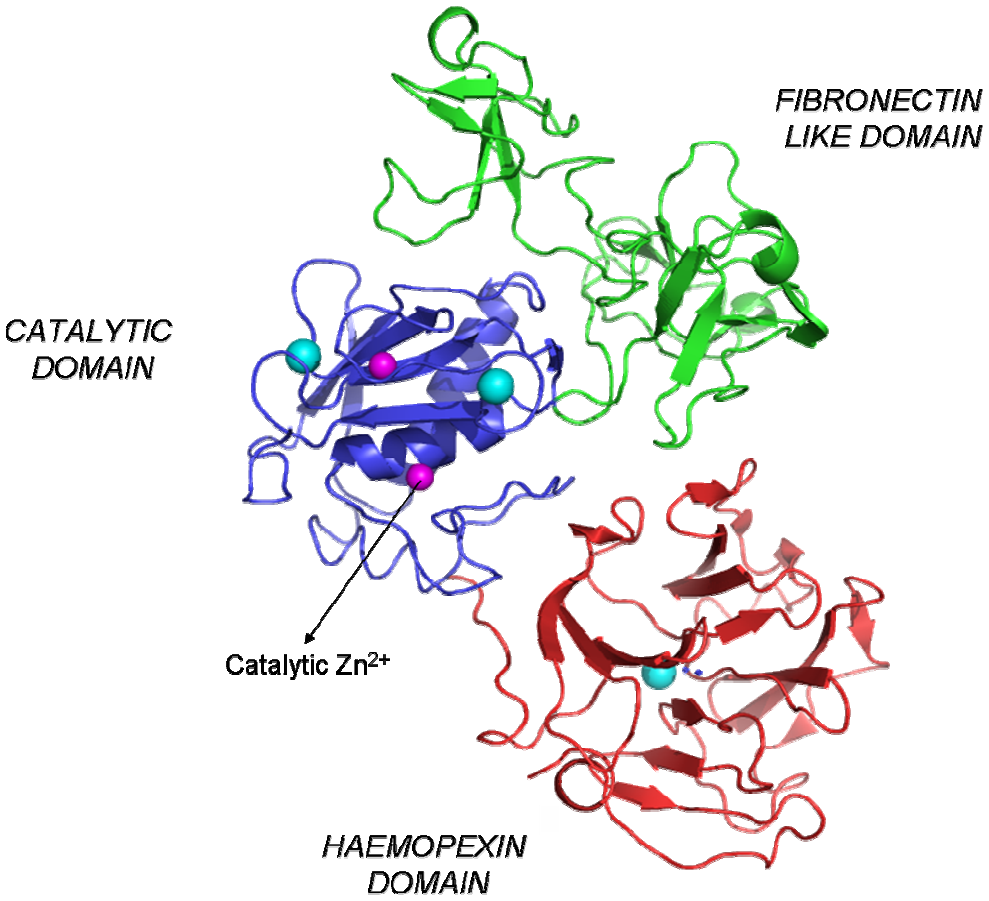


Figure 4. Cartoon representation of full length MMP-2, constituted by the catalytic domain (blue) that contains two zinc ions (magenta) and a variable number of calcium ions (cyan); the haemopexin domain (red) and the fibronectin like domain (green). The tertiary structure of gelatinase MMP-2 could be considered as representative for the entire family of MMPs, according to the specificity described in the text.

The haemopexin domain is reported to be fundamental for substrate recognition and, particularly, for collagen processing by presenting a single helix of collagen to the catalytic domain. Fibronectin-like domain presented by MMP-2 and MMP-9 participates in the interaction with gelatin substrate. [48] Moreover MMPs incorporate two zinc ions (one constitutive and one catalytic) in the catalytic domain, plus a numbers of calcium ions variable from one to three both in the catalytic and the haemopexin domain, as reported in Figure 4.

These proteins are expressed in an inactive form on which a prodomain peptide is included in the structure to inhibit catalytic activity; the activation process is due to many different factors, including the mutual activation one, and is aimed at freeing the active site to the catalysis, by proteolytically cleaving the prodomain peptide. [49,50] Thus, MMPs are physiologically involved in degradation and remodeling of the ExtraCellular Matrix (ECM) components, cell movement, proliferation, and tissue expansion. In physiological conditions MMPs present a low expression profile and their activity is well regulated by TIMPs (Tissue Inhibitors of Metallo Proteinases), so that the ECM presents always the physiological organization and architecture. Due to an imbalance in the MMPs/TIMPs homeostatic system, often referred to an MMPs over-expression, the architecture of the ECM is altered, leading the tissue to a pathological condition that needs to be repaired. [51-54] In some cases, the tissues themselves are able to repair the occurred damage, but when MMPs levels are too high or the damage is extended in time, tissues cannot recover. Accordingly, the application of synthetic MMPs Inhibitors (MMPIs) could represents a therapeutic opportunity for approaching all those disorders in which the restoration of the ECM organization is needed. In the last decades, many MMP inhibitors have been deeply investigated both in pre-clinical and clinical trials for their anticancer activity; however, none of them reached the market. In the last few years the interest for MMPs has resurged, and these proteins still represent a noticeable class of therapeutic targets.

MMPs were extensively investigated during the present PhD project and constituted the core of the research activity carried out. Most part of the research in this field has been carried out in collaboration with ProtEra Srl, an Italian biotech company involved in drug discovery and focused on the development of MMPs inhibitors as therapeutic agents. In detail, two different approaches were carried out: a first and deep computational analysis for ligand design purposes, and a *in vitro* and *in vivo* evaluation of a MMP-9 inhibitor

therapeutic activity. In detail, a new scaffold for MMPs inhibition has been developed by the fine computational-based optimization of an already validated scaffold. The second approach was mainly focussed on the efficacy evaluation of MMPs inhibitors, previously studied by docking analysis, in the therapy of Dry Eye Syndrome. [55] In the field of MMP proteins a complete discovery activity was performed, starting from the early phases of the ligand selection (see workflow of Figure 2), until the preliminary analysis of the therapeutic activity and toxicity for the selected lead molecules.

4.3*. Tubulin and Microtubules***.**

Microtubules are major components of the cytoskeleton of all eukaryotic cells. These polymers participate to the intracellular organization and maintenance of cellular integrity. They are essential for cellular life owing to their implication in numerous cellular processes, such as determination of cellular shape, position of organelles, cellular motility, movement of intracellular organelles and formation of the mitotic spindle during cell division. [56,57] Tubulin is a dimer of α/β subunits that consist of a core formed by ten β-sheets surrounded by 12 α-helices. The most significant differences between subunits are situated in loop M (microtubule loop), which participates to lateral interaction between protofilaments. The most variable domains are located in the C-terminal of subunits that undergo post-translational modifications and are involved in protein-protein interactions. Due to structural and functional differences the α-subunit is always bound to GTP, whereas the β-subunit can also be loaded with GDP. Microtubules are cylindrical structures composed of α and β tubulin subunits, non-covalently bound together, that can assemble and disassemble in response to different cellular stimuli. The inherent microtubule dynamics is provided by hydrolysis of GTP and closely related with microtubule functions. Cellular microtubules appear as hollow cylinders of about 25 nm in diameter. The cylinders within cells are formed by the lateral association of 13 protofilaments on average. Each protofilament is a string-like structure, Figure 5, which consists of tubulin subunits that are arranged in a head-to-tail orientation, a property that confers a polarity to microtubules. [58] The “head-to-tail” orientation of αβ-tubulin in microtubules results in a different subunit exposition at their ends: α-subunits are exposed to solvent at the minus end (-) and β- at the

plus end (+). Microtubule ends were named in this manner because “plus” end is more dynamic than “minus” end

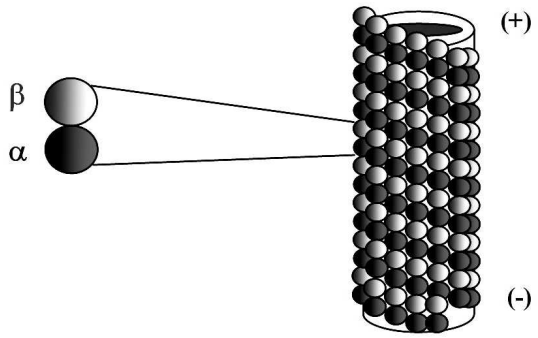


Figure 5. Schematic representation of microtubule organization. Microtubule polarity indicates that β-subunits are exposed at «plus» end, whereas α-subunits on «minus» end of the microtubule.

These macromolecules have been targeted for cancer therapies aimed at stopping the mitotic division process of cancer cells. Numerous natural microtubule-targeting drugs were found in plants and animals that use these toxic molecules as passive defence. Nowadays, some of these drugs serve in medicine as effective anti-cancer agents that arrest the cell cycle and inhibit proliferation of cells. Microtubules become more dynamic when cells enter mitosis. The fundamental reasons of this phenomenon are still obscure; however, this higher dynamics is necessary for the formation of the mitotic spindle and proper chromosomal segregation. [59] The failure of mitotic spindle formation or its stabilization at the stage of chromosome segregation results in mitotic arrest. The most important drugs that act on spindle formation are colchicine, vinca alkaloids and paclitaxel (taxol). [60,61] It has been demonstrated that colchicine and vinca alkaloids bind alpha tubulin monomer in a pocket that is close to the interface of binding with beta tubulin. [61] The binding of these natural substances stabilizes a curved conformation of the tubulin heterodimer, Figure 6, which impedes longitudinal contacts in protofilament and leads to microtubule depolymerization. On the other hand it is accepted that taxol binds to β-tubulin on the inner surface of microtubule [62] and probably stabilize lateral contacts between protofilaments blocking mitosis even at concentrations several hundred times less than that of cellular tubulin.

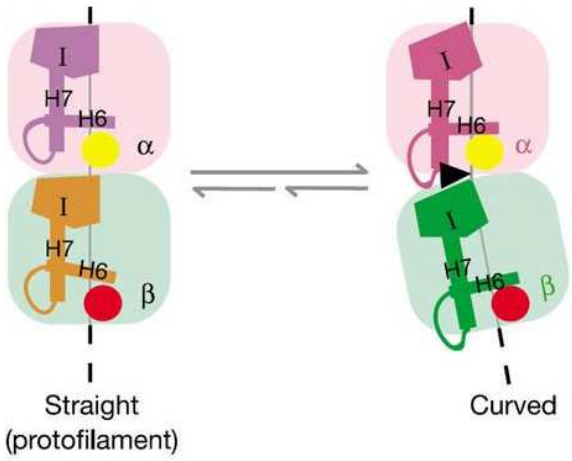


Figure 6. Schematic representation for stabilization of curved heterodimer conformation by colchicines (shown as a black triangle). Modified from Ravelli et al., 2004. [61]

In addiction, many small proteins and peptides have been found to interfere with the tubulin polymerization process. Several of them belong to the stathmin family. It was discovered that the presence of stathmin destabilize microtubules, an effect first attributed to an increase of the transitions between growth and depolymerization phases (catastrophes). It was soon demonstrated that this effect also results from the sequestration of tubulin in a tertiary complex made of two tubulin heterodimers per stathmin molecule, reported in Figure 7. The stathmin molecule stabilizes a curved conformation of the pair of tubulin heterodimer in a similar way as observed for vinca alkaloids.

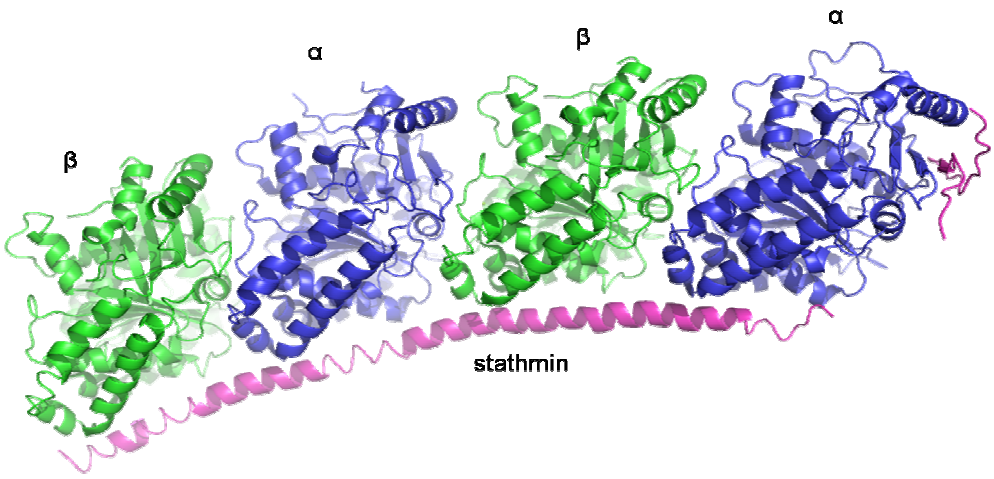


Figure 7. X-Ray structure of the stathmin-tubulin complex (PDB entry 1SA0) at 3.58 Å resolution. In magenta stathmin, in green β-tubulin, in blue α-tubulin. Stathmin protein stabilizes the two heterodimer in a curved conformation. Moreover, a β-hairpine peptide of stathmin protein binds on α-tubulin, impeding the polimerization.

Moreover, it was observed that the amino-terminal end of stathmin binds to the α-tubulin subunit at the longitudinal interdimer interface. Both dimer curvature and steric hindrance of α-tubulin subunit by the N-terminal part of stathmin prevent incorporation of tubulin dimer inside the microtubule lattice. [63-66] Recent data show that stathmin strongly increases the catastrophe frequency at the minus ends due to direct interaction with pre-polymerized microtubules. [67] These data propose co-existence of two different mechanisms for regulation of microtubule dynamics by stathmin.

These proteins were investigated during the development of an European Project (FP6, acronym: Nano4Drugs) aimed at the discovery and optimization of a drug delivery system for carrying protein-based anticancer drugs targeting tubulins and microtubules. Starting from an already known binding peptide derived from the amino terminal end of stathmin, computational tools were applied to identify the most promising mutations in terms of interaction energy towards alpha-tubulin and ligand stabilization. By the sequential application of homology modelling, molecular dynamics and free energy calculation by the solution of the Poisson-Boltzmann equation, we have selected new peptides theoretically able to bind alpha-tubulin with higher affinity with respect to the reference peptide. According to the aim of the project these peptide should constitute, in a first attempt, the protein-based drugs that have to be linked to a fluorescent nanodiamond in order to realize a traceable drug-delivery system.

5. Projects

Here are described the projects carried out during this PhD training. For each project, a small introduction with rational for the development of ligands, a brief experimental session, the discussion of the results obtained and the description of the personal contribution are provided.

5.1. Theoretical Study on S100B.

This theoretical work has been aimed at the identification of S100B residues involved in ligand and substrate binding. This protein was selected within the S100 family due its clear involvement in human pathologies such as melanoma and neurodegenerative disorders. [68-73] S100B is one of the most promising S100 protein for drug designing purposes because it is reported to interact with the C-terminal negative regulatory domain of p53, a physiologic tumor suppressor protein that signals for cycle arrest and apoptosis. S100B is involved in the down-regulation of p53 and, as the inactivation of normal p53 functions is found in over half of human cancers, restoration of normal p53 functions through the destruction or prevention of S100B-p53 complex represents a possible approach for the development of new drugs. [74-76] Despite the potential application of S100B ligands, the limited availability of experimental information relative to S100B's ligands binding conformation represents the main difficulty for drug design purposes. Only few ligands have been actually demonstrated to interact with S100B with micromolar affinity. [45] For these weakly binding molecules, experimentally determined structural information about the conformation of binding are not available. In this work, we have applied computational methodologies to available S100B-substrate complexes to provide information about the protein residues involved in ligand and substrate binding. Moreover, we have provided a docking based binding model for a organic compound, pentamidine, that is currently one of the few molecules reported to interfere with S100B-p53 adduct. [44] Molecular dynamics simulations [77] were carried out for the refinement of the three available structures of S100B bound to TRTK-12 and p53 peptides. [78-81] Solution structures of S100B complexes were retrieved from the Brookhaven Protein Database. Calcium loaded structures have been selected due the evidence that S100 proteins are activate by two calcium ions; the metal binding process provides conformational modifications that expose the hydrophobic binding core of the

protein towards the physiological environment, by increasing the angle between helices 3 and 4 until about 85° (Figure 8). This mechanism of activation determines an increased affinity of S100 proteins towards their substrates.

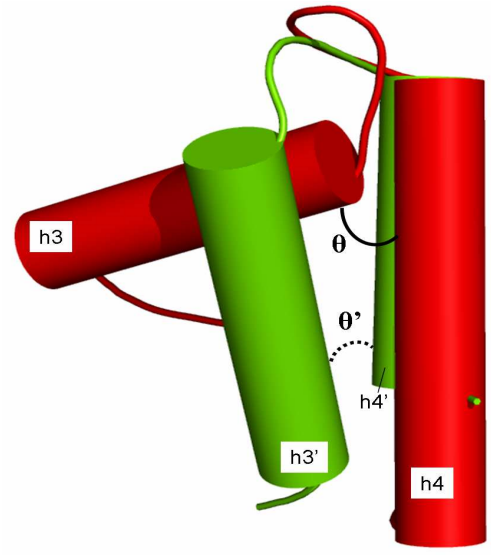


Figure 8. cylindrical representation of helices 3 and 4 of calcium loaded S100B (red) and helices 3' and 4' of apo S100B (green). Helices 4 and 4' of both structures have been manually over imposed, and vertically aligned in the figure in order to highlight the different orientation of the helix 3 and 3'. It is evident how the binding of calcium determines an angle θ between h3 and h4 of about 85 °, whereas h3' and h4' in the apo form are quite parallel (θ')

The computational analysis of side chain reciprocal movements in the selected complexes has revealed that only few residues play an important role in substrate stabilization and binding. In particular, F76 and V56 constitute a hydrophobic entrance to the substrate; moreover, F76 is involved in the stacking-based stabilization of aromatic side chains of these peptides. A further lipophilic interaction is mediated by a cluster of leucine and isoleucine residues distributed on helices 2 and 3 (L32, I36, L60). These residues' side chains belong to the hydrophobic binding core of the S100B and correspond to the part of proteins that is exposed to the external environment during the calcium dependent activation. As a confirmation of the relationship between the 3D conformation of the protein and its binding affinity towards substrate peptides, the distance between c-alpha of F76 (h4) and V56 (h3) was measured. A distance comprised between 8.2 and 9.7 Å was found to be crucial for the interaction to occurs. All these information about the binding

conformation of substrate peptides on S100B have been used to describe the putative binding pocket on which to target ligands.

To complete the characterization of this binding pocket, the interaction between pentamidine and S100B has been studied *in silico*. Pentamidine is a member of aromatic diamidines, currently in widespread clinical use for the treatment of infections determined by *Pneumocisti carinii Pneumonia*. Isethionate salt, as approved by FDA, was experimentally observed to bind S100B at low micromolar concentrations. [45] With the aim to propose a computational-based binding model for this ligand on S100B, a molecular mechanics approach has been undertaken. The protein coordinates for the building of the potential grid map have been retrieved from the Protein Data Bank. The solution structure of calcium loaded human S100B in complex with TRTK-12 high affinity peptide (PDB 1MQ1) [79,80] was selected because it describes the only available complex between human S100B protein and a substrate. After removing the coordinates of the binding peptide from the adduct, a short MD step has been applied to the backbone-restrained S100B protein in order to stabilize the new side chains conformation to their corresponding local minimum level. To describe the conformational accessibility of pentamidine on the binding core of S100B, two different simulations were performed. A first molecular dynamics approach has been carried out for the energy minimization (50K, 3ps) of the complex between S100B reference structure and pentamidine, manually placed in the above described binding pocket. In a second step, AutoDock program was applied for the automatic prediction of pentamidine binding conformation, and relative interaction energy. Due to the structural symmetry of this protein, both the simulations were performed in a single monomer. With those approaches, two main conformations of binding were identified. The interaction in the binding pocket has been predicted in the same way by AMBER and AutoDock, whereas some differences have been observed in binding the protein surface. However, the protein residues involved in pentamidine binding are the same observed to be involved in the interaction with substrates. According to these results, in the first surface conformation, predicted by AMBER package, a benzamidine group of pentamidine interacts with the N-terminal region of helix 3 close to the hinge loop. The main residues involved in this interaction are E51 (salt bridge) and V52 (hydrophobic interaction). In the conformation predicted by AutoDock the benzamidine group binds in the central region of helix 4. In that case V80 and A83 (hydrophobic interactions) were the mainly involved

residues, Figure 9. In conclusion, other than the above described residues involved in substrates binding, E51, V52, V56 and A83 seem to be crucial for ligands interaction on the surface of S100B, whereas F76 play a key role in the stacking stabilization of aromatic rings of both ligands and substrates. All these results have been used to describe a possible S100B binding pocket, Figure 10, providing the bases for a further drug design approach.

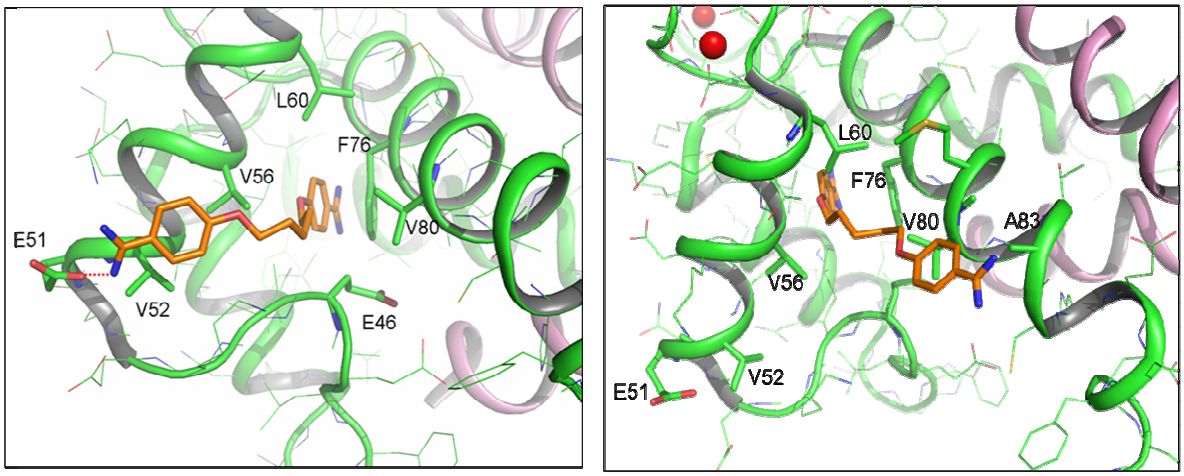


Figure 9 – Binding conformation of pentamidine predicted by AMBER (left side) and AutoDock (right side). In green cartoon, sticks and lines is reported one monomer of S100B. In red spheres are highlighted the calcium ions. In violet is reported the second monomer of the S100B. The binding conformation of pentamidine was studied in only one monomer. It is observable that one of the two benzamidine group binds preferentially in a pocket described by F76, V56 and L60, whereas the other benzamidine group binds in the end of helix3 (amber) and in the end of helix 4 (AutoDock).

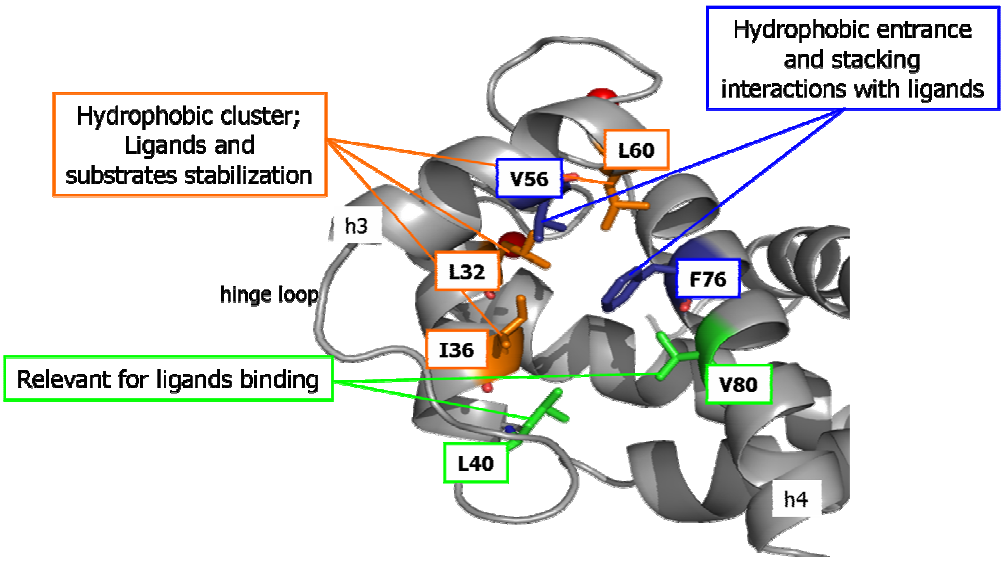


Figure 10 – In stick: S100B residues involved in ligands and substrates binding

In this work I have contributed with the docking calculations of Pentamidine and with the molecular dynamics based analysis of the interaction between S100B and the high affinity peptides, available in the literature. Artur Gieldon has contributed with the Molecular Dynamics studies of the complexes between S100B and the peptides and between S100B and pentamidine.

These results have been published in the Journal of Molecular Modelling 2007, 13, 1123-1131 (Gieldon A., Mori M., Delconte R.) and the original publication is reported in the attachment (appendix) A.

5.1.1. *NMR Study on Pentamidine Binding*.

Several attempts aimed to obtain the X-ray structure of the pentamidin in complex with S100B failed. Therefore, to obtain information on the binding mode of this micromolar ligand on S100B, the interaction with pentamidine was investigated following the chemical shift perturbations on ^{15}N-^{1}H HSQC spectra using ^{15}N labeled protein samples. The chemical shift is a sensitive indicator of protein-ligand interactions and can be use to identify the binding site when the resonance assignment is available. [82,83] The complete assignment of the unbound form of S100B is available in our laboratory (results unpublished). Increasing amounts of pentamidine dissolved in DMSO (stock solution of 25 mM) were added stepwise to 200 µM solutions of the protein up to a concentration of 1 mM. Although several NH signals were shifted by adding the ligand, most of cross-peaks belonging to protein adducts with pentamidine, where easily re-assigned.

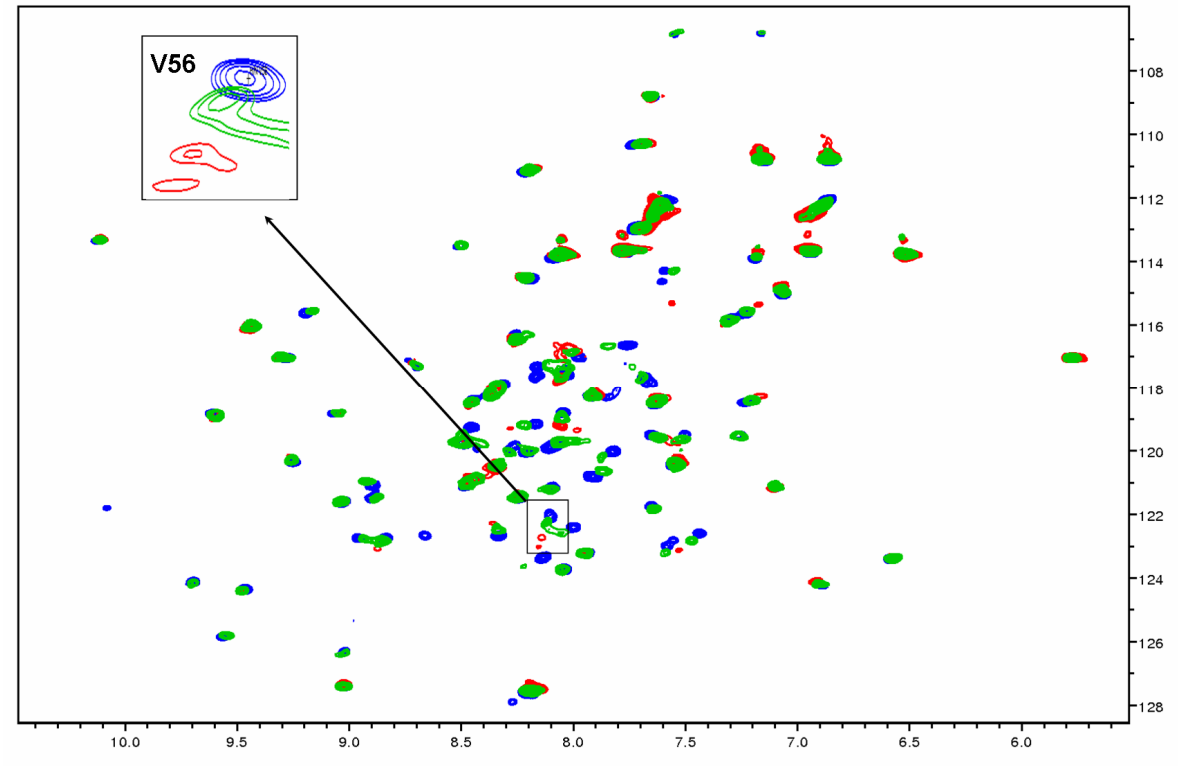


Figure 11. ^{15}N-HSQC spectrum of calcium loaded S100B (blue), calcium loaded S100B + pentamidine 0.2mM (green) and 1mM (red). In the box at the top-left is magnified the peak of V56, as example of a residue exhibiting strong chemical shift perturbations.

The binding site of pentamidine, was analyzed and compared by reporting the Garrett values [Δδ(NH)] on a cartoon and stick representation of the S100B structure (Figure 12).

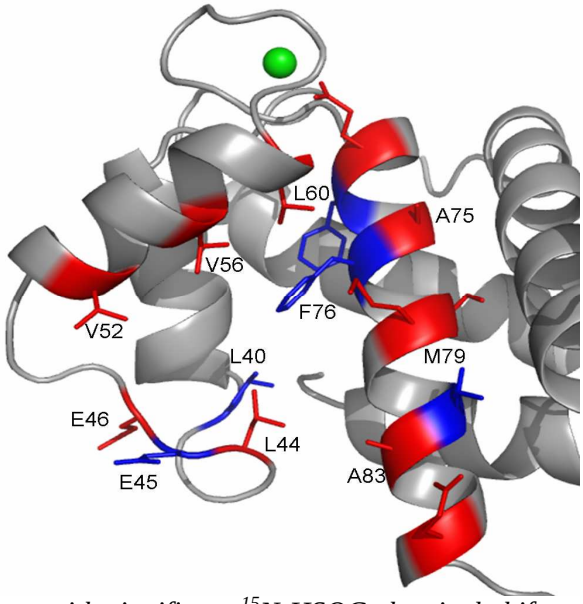


Figure 12 – S100B residues with significant ^{15}N-HSQC chemical shift perturbations (CSPs). Red: residues exhibiting Δδ>0.1 ppm; Blue: residues exhibiting 0.1>Δδ>0.04 ppm; green: calcium ions.

As it is evident from Figure 12, the region with significant ^{15}N-^{1}H HSQC CPSs involves the protein region described by helices 3 and 4, with a particular involvement of the helices' residues that constitute the above described hydrophobic binding core of S100B. The analysis of the Garrett values rules out relevant additive interactions of the pentamidine with different protein regions, suggesting two different binding mode as obtained by the MM calculations. The NMR-based data are in agreement with the theoretically predicted binding mode where the benzamidine group of pentamidine interacts with F76 (Δδ=0.04 ppm). At the same time the distribution of CSPs supports the MM calculations where the second benzamidine group switch between two different binding sites, one on helix 3 and the other on helix 4. However, the NMR data suggest a slightly different interaction between the ligand and the amino acids on helix 3, with respect to what predicted by the MM calculation, since only E46 instead of E51 experiences a chemical shift alteration (Δδ=0.13 ppm).

Interestingly, the predicted binding mode for pentamidine has been supported by some experimental data, confirming the possibility for this ligand to exploit the two predicted interaction with S100B (Figure 13). In conclusion, these results set the bases for the further drug design approach.

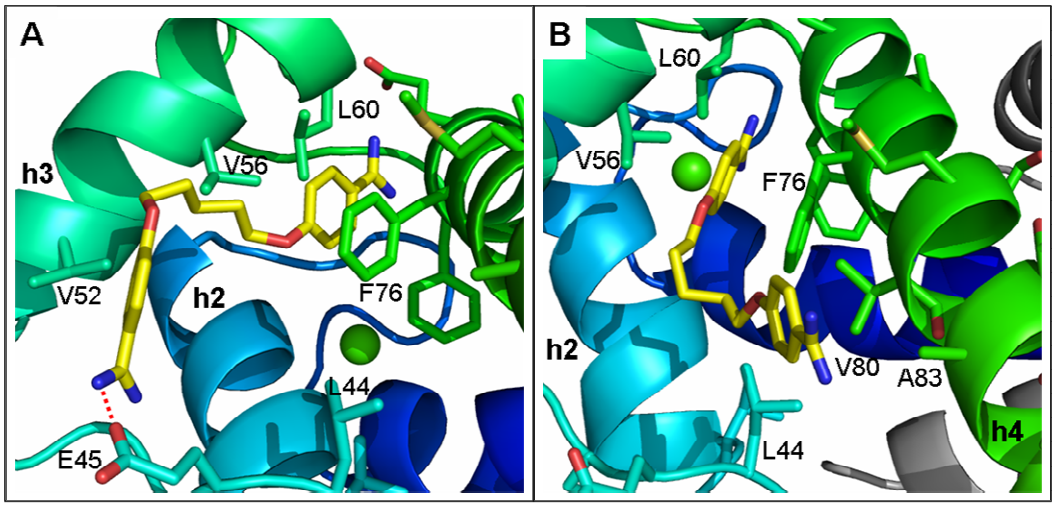


Figure 13 – Binding mode of pentamidine modelled on the basis of the ^{15}N-HSQC chemical shift perturbations. A: pentamidine close to h3. B: pentamidine close to h4. In yellow stick pentamidine; in stick are highlighted the residues identified by ^{15}N-HSQC experiments.

The results of this NMR-based experiments are unpublished and were used to provide experimental confirmations to the previously described theoretical work on S100B. I contributed to this experimental investigation with the analysis of the HSQC spectra and the mean shift difference calculation.

***5.2**. Fragment Docking to S100 Proteins Reveals a Wide Diversity of Weak Interaction Sites.*

The results obtained by the theoretical approach to S100B and the experimental NMR-based binding analysis for pentamidine provided useful information about the substrates and ligand binding mode of this protein. These results were exploited for a structure-based drug design approach aimed at the identification of new fragments and/or ligands able to bind S100B and S100A13. S100B has been taken into account as drug target for the reasons above described. On the other hand, S100A13 has been recently designated as a new marker for angiogenesis in human astrocytic glioma and as a regulator of the FGF-1 release. S100A13 is widely expressed in various types of tissues, and shows a high expression level in thyroid gland. In smooth muscle cells, this protein is co-expressed with other family members in the nucleus and in stress fibers, suggesting diverse functions in signal transduction. [84]

In this work a combined NMR and docking-based approach was undertaken for the identification of the fragments binding sites on the surface of these proteins. This approach gave us the possibility to address some differences in ligand binding affinity and specificity between these two member of the S100 proteins family. Structure-based drug design and NMR-based screening were combined to identify the fragments binding sites on the surface of S100B and S100A13. To identify the binding sites we carried out an irrational approach based on the *in vitro* screening of a fragment library, whereas *in silico* tools were used for to predict the conformation of binding of most interesting hits. A fragment library of 430 commercially available compounds was initially screened by NMR spectroscopy on each calcium loaded protein. This library does not completely agree with the Lipinski's rules, [85,86] according to which a candidate drug molecule should adopt the following criteria:

-Not more than 5 hydrogen bond donors
-Not more than 10 hydrogen bond acceptors
-A molecular weight under 500 daltons (from 160 to 480)
-A partition coefficient log*P* less than 5 (in -0.4 to 5.6 range)
-(number of heavy atoms from 20 to 70)

In bracket are reported some recently validated improvements to these rules. [87] However, this library of compounds with an intermediate molecular weight between fragments and ligands was selected to obtain preliminary binding information, because it was representative for the most chemical diversity as possible.

A two step NMR approach was performed in the screening process. The ligand based WaterLOGSY [88] (Water Ligand Observed Gradient Spectroscopy) technique was applied for the initial screening in order to identify, for each protein, the set of promising hits showing a certain interaction with the protein. Water-LOGSY is a NMR-based experiment for the preliminary qualitative evaluation of the ligand binding capability towards a target protein. This technique is based on the magnetization transfer between irradiated water molecules and ligands, that are in the equilibrium state between the solvent and the protein bound form. By this experiment 56 and 47 fragments has been observed to interact with S100A13 and S100B respectively, and the most interesting evidence obtained by this preliminary approach is that the two proteins show a particular affinity towards the fragments. Only 13 fragments, in facts, bind both proteins. These two sets of fragments were then tested by ^{15}N-HSQC experiments, with the aim to identify the relative protein residues involved in the binding of each fragment. The chemical shift perturbation (Garrett values) of uniformly ^{15}N-labeled protein samples upon complex formation with ligands, provides a sensitive tool for the identification of binding sites. [82] The results analysis revealed that there are two main fragments binding regions on S100B, showed as cubes in Figure 14; the first (A) is located between the hinge loop and the C-terminal part of helix 4, and the second one (B) involves the hinge loop and helix 3. The hinge loop represents a bridge between these two areas, playing an important role in the interaction of ligands. Interestingly, these results are in good agreement with the *in silico* prediction of the S100B ligand binding area provided by the theoretical work described above, and by the NMR-based study of pentamidine.

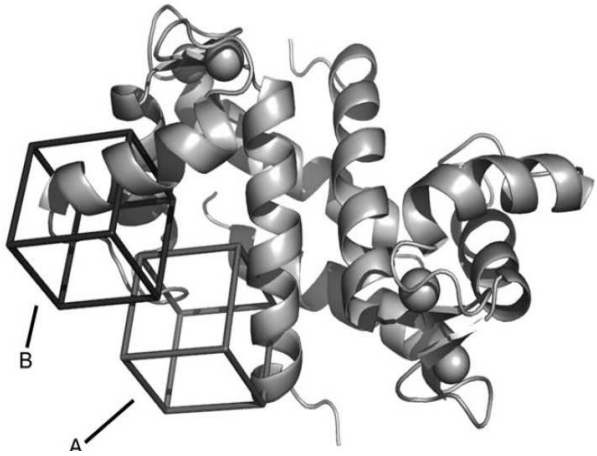


Figure 14 – The two binding areas identified on the surface of S100B (black and grey cubes). Area A is located between the C-terminal of helx 4 and the hinge loop, whereas area B involves mainly the C–terminal of helix 3.

On the other hand, three interaction areas have been found on S100A13, as reported in Figure 15; the first area (A) involves the second calcium binding site at the C-terminal part of helix 1, the second one (B) is located in the hinge loop region between helix 2 and helix 3, including the C-terminal part of helix 4 and N-terminal part of helix 1, and the third area (C) is placed in the region of interaction of the monomers, mainly involving helix 4 and helix 1.

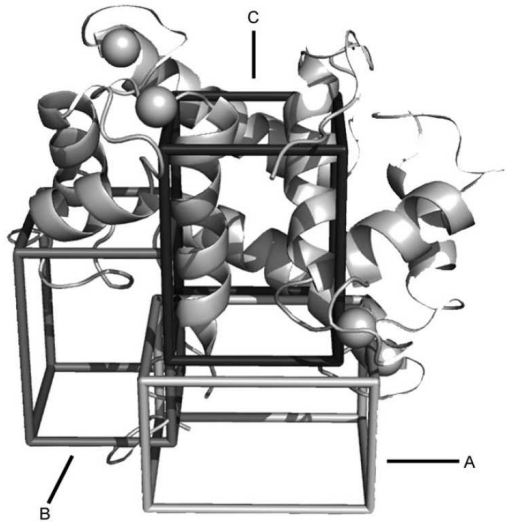


Figure 15 – The tree interaction areas identified on the surface of S100A13 (in black, grey and light grey cubes). Area A is located in proximity to the second calcium binding site; area B includes the hinge loop between helix 2 and helix 3, whereas area C is placed in the region of interaction of the monomers.

By this first analysis these two protein displayed a different interaction path towards the same set of ligands, These results support the hypothesis of a selective ligands interaction between the family of S100 proteins.

Despite the great contribution given by ^{15}N-HSQC experiments in determining the protein residues involved in the fragments binding, and its fundamental application in mapping the above described binding areas, this method does not provides information about the conformation of binding for the tested ligands. Docking simulations have been performed for all those molecules that showed interaction in Water-LOGSY experiments, with the aim to understand the fragments conformation of binding in the areas identified by the NMR-based screening. Protein structures identified by PDB entries 1YUU [89] and 1DT7 [81] were selected for grid parameterization on S100A13 and S100B respectively. The grid maps calculations was performed on each NMR determined calcium loaded protein, selecting the structure with the lowest energy and the least restraint violations, and removing the eventually present binding peptides such as the C-terminal peptide of p53 protein. The entire surface of both the proteins has been considered during the docking calculations even if the grid maps were centered around a single monomer. The Lamarckian Genetic Algorithm (LGA) has been used for the automatic docking procedure and the global search method has been applied on each compound. Results clearly show that the fragment binding regions identified by the docking approach are consistent with the binding areas identified by ^{15}N-HSQC experiments in both proteins. Moreover, each fragment binds preferentially in only one of the three areas identified for S100A13 and the two identified on S100B. The good quality of these results is even supported by the fine correlation of the predicted conformation of binding with the experimental HSQC data.

The lead fragments providing the lower NMR-determined Kd values for the interaction to each protein were deeper analyzed. In particular, cromolyn (cromoglicic acid) [90,91] for S100A13 and alpha naphtol for S100B provided the better results in terms of binding energy and agreement of the docking prediction with experimental data. One of the possible parameters to evaluate the quality of a ligand is the ligand efficiency (LE), [92-95] that is calculated as the ratio between Delta Energy of Binding towards the target protein and the ligand's number of heavy atoms. Cromolyn is an antiallergic drug reported to interact with S100A13. To investigate the molecular bases of this interaction, a ^{15}N-HSQC titration on S100A13 was performed and a Kd value between 100 μM and 500 μM was

obtained, providing a LE≈ 0.14. The strongest chemical shift perturbation due to the cromolyn binding was observed for residues within the binding site described by the interaction of the two monomer in correspondence of helix 4 and helix 1 (Area C in Figure 8). These predicted data provide the structural bases for the interaction between cromolyn and S100A13 (Figure 16A).

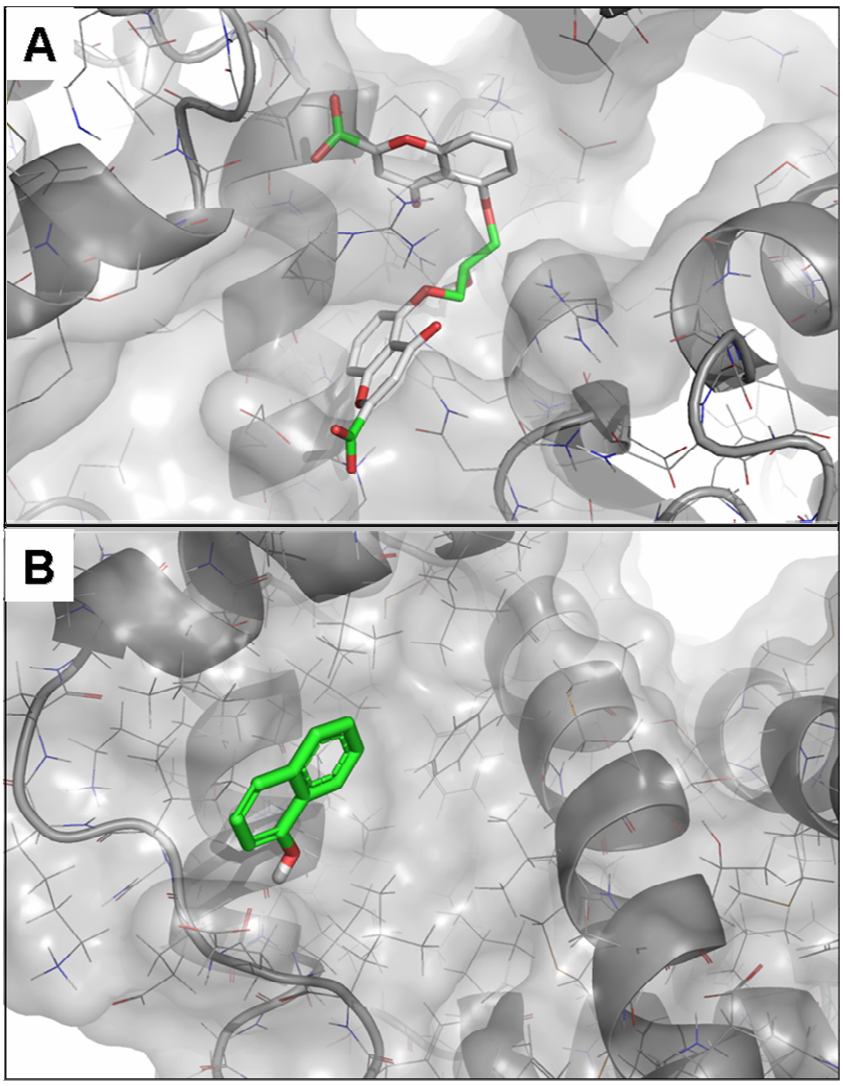


Figure 16 – A: binding conformation of cromolyn on the binding area located at the S100A13 monomer interaction surface, between the two helices 4. B: binding conformation of alpha naphtol on the binding area close to the S100B's hinge loop.

Alpha naphtol (Figure 16B) is the strongest interacting fragment on S100B, providing a NMR determined Kd value between 50 μM and 200 μM and a LE≈ 0.5. To check the ability of this fragments in preventing the pathologically relevant interaction between

S100B and p53, the ^{15}N labeled protein was titrated with this ligand in presence of a p53 mimic peptide derived from the C-terminal regulatory domain of p53 protein (sequence SHLKSKKGQSTSRHKKLMFKTE). The titration was monitored by HSQC spectra, observing the effect on the S100B resonances upon the addition of the ligands. ^{15}N labeled S100B was first titrated with p53 mimic peptide up to a 1:2 molar ratio with respect to the S100B monomer; then the alpha naphtol was added stepwise up to a molar ratio of 1:2 :4, respectively. By comparing the spectra it was observed that the addiction of alpha naphtol affects several peaks. Some of them are restored to the original position corresponding to the S100B-alpha naphtol complex; some other peaks were shifted to a new position that is different from that observed in S100B-p53 or S100B-alpha naphtol adducts (Figure 17).

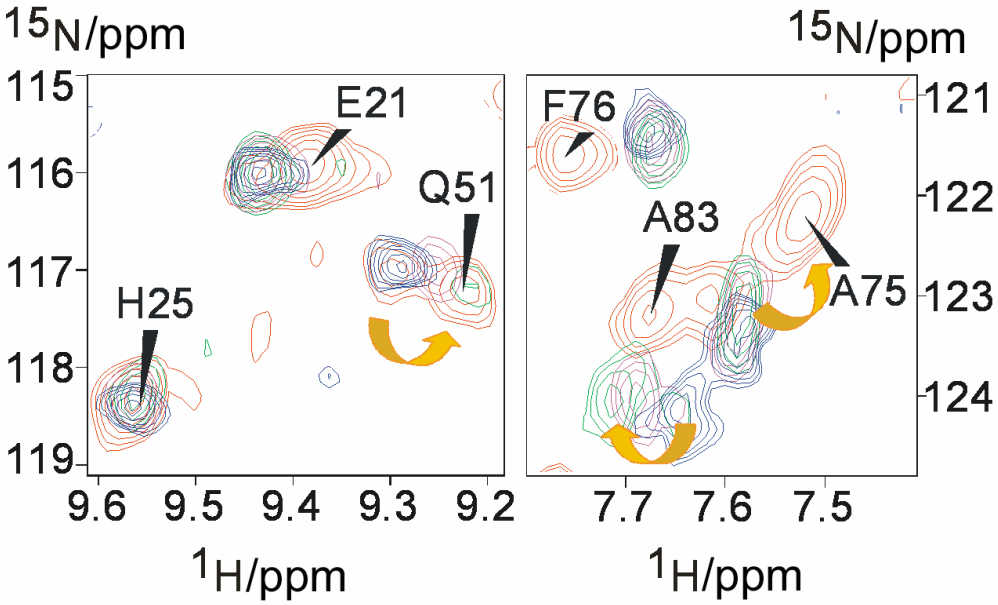


Figure 17 - Superimposition of two areas of ^{15}N-HSQC spectra of 1) S100B + 5B adduct at 1:2 ratio (red); 2) S100B+ p53 at 1:2 (blue); 3) S100B + p53 + 5B at 1:2 :2 (magenta); 4) S100B + p53 + 5B at 1:2 :4, (green).

This points to alpha naphtol replacing a portion of the bound peptide molecule with a mixed competitive/noncompetitive behaviour. The other portion of the p53 mimic peptide presumably rearranges in a new conformation to avoid conflict with the surface portion occupied by the fragment. The peaks that are restored clearly identify the region of the protein where alpha naphtol binds, in good agreement with the previously described binding areas.

In conclusion, the combination of *in silico* prediction and NMR experiments provided the structural bases for the identification of two binding areas on the surface of S100B, and three binding areas on S100A13. Although the two proteins have quite similar quaternary structure and a common binding area (around the hinge loop), our results showed that they have only a few ligands in common, suggesting that selective ligand could be developed starting from the different lead fragments with high LE values, identified herein. Coupling of the experimental data with docking results indicated the most probable docked conformations of the ligands on the proteins surface. The study on the antiallergic drug cromolyn provided for the first time structural information for the interaction, and demonstrated that cromolyn has a unique binding site on the protein surface. Finally, competition experiments conducted with a-naphtol on the S100B–p53 peptide complex showed that even if the ligands identified by the NMR screening bind weakly on the protein surface, they are able to significantly interfere with the interaction with other proteins or peptides.

These results constitute a good starting point for the *in silico* and experimental development of new ligands that could play an important role in the protection of p53 from S100B-dependent down regulation. S100 proteins in general are becoming targets of pharmaceutical interest because of their involvement in cellular functions. The present data suggest that selective and high efficacy modulators of their activity are within reach of current lead development strategies.

In this work I have done all docking calculations on both S100 proteins and all necessary parameterizations related to the docking procedure.

These results have been published on ChemMedChem journal, 2007, 2, 1648-1654 (Arendt Y., Bhaumik A., Delconte R., Luchinat C., Mori M., Porcu M.). The original publication is reported in Appendix B

***5.3**. Computational Approach to the Optimization of The Classic Aryl-Sulphonamide Scaffold for MMPs Inhibition.*

Many proteins such as MMPs have been structurally characterized in complex with synthetic molecules and substrates. In these cases the ligand conformation of binding can be exploited for the design of new focussed molecules approaching a structure and ligand-based drug design strategy. As above described, the availability of the 3D structures of target proteins in their complex conformation represents a great advantage for drug design. However, not always these ligands perform the most of the favourable interactions with the protein, and a lack of ligand optimization can be observed. In these cases, a refined *in silico* approach can be undertaken with the aim to improve the ligands efficacy. In this work we have applied computational tools to refine and optimize a validated scaffold for MMP inhibitors.

MMPs are a family of proteins considered as pharmaceutical targets due to their involvement in various human diseases. Other than cancer, MMPs are involved in many different pathologies such as periodontitis, dry eye syndrome, skin aging and inflammation that could be managed in a topical way, especially in the attempt to avoid all the side effects observed during the systemic administration of MMPs inhibitors and related to their lack of selectivity. [96] In this field, the development of strong MMPs inhibitors is still representing a possible therapeutic opportunity. Many experimental structures of MMPs are available and most of them describe their complex with synthetic inhibitors. A fine analysis of these 3D structures, however, reveals that most of these molecules lacks of optimization towards their target protein. In particular, for the molecules belonging to the most speculated hydroxamic aryl sulphonamide scaffold, small structural changes could be hypothesized in order to increase their binding affinity.

During this phase of the PhD project, computational tools have been applied for the optimization and refinement of hits. According to the workflow described in Figure 2, this strategy properly corresponds to a structure and ligand-based drug design approach finalized to the lead discovery and optimization. The work started with the analysis of the available PDB structures that describe the complexes between MMPs and hydroxamic aryl sulphonamide inhibitors. [97-100] The 3D conformational analysis at atomic level of detail revealed two main portion of the reference scaffold on which to focus some chemical modifications. All these molecules, in fact, did not perform a geometrically correct stacking

interaction with a highly conserved aromatic residues of MMPs (zinc-binding histidine); hence, a linker between the sulphonamide moiety and the so called P1' aromatic portion can be introduced in order to better perform this interaction. Moreover, the amino acidic block of the reference scaffold can be modified by introducing an additional carbon chain, with the aim to decrease the solvation energy of free inhibitors. From a thermodynamic point of view, during the ligand protein interaction process many variable can play a key role in determining the ligand's affinity. One of this is exactly the solvation/desolvation effect by which a lower ligand solvation energy provides lower free energy of binding, with respect to more water solvated ligands. [101-103] Therefore an improving in ligands efficiency can be achieved by playing with water solvation/desolvation effects. Thus, the amino acid residue of the reference scaffold has been modified by substituting the classic glycine with a D-proline; in addiction, a carbon linker was inserted between the sulphonamide group and the so called P1' aromatic portion of inhibitors. D-proline was selected for its cyclic restrained features in order to avoid any possible energy indetermination related to the side chain's fluctuation of non cyclic residues. On the other hand, the dimension of the linker has been chosen to maintain the binding conformation as much similar as possible to the reference scaffold, keeping the P1' aromatic portion inside the so called S1' lipophilic pocket characteristic of MMP proteins. By a comparative docking analysis of molecules equipped with one, two and three carbon linker, the two carbon based was selected for further investigations. By the combination of these modifications, an initial small library of 12 compounds was generated. Docking procedures were applied to this library in order to estimate the effective advantage in MMPs binding with respect to the compounds belonging to the classic scaffold. These preliminary docking calculations were performed towards MMP-12 in order to have a direct comparison with the further presented *in vitro* and crystallographic data, easily available for this protein. Moreover, the modifications to the scaffold affect the protein binding in some particular region on which MMPs display a high sequence and structure homology; [104] for this reason MMP-12 can be considered as representative for the entire family of proteins. The docking results showed that there is an advantage in free energy of binding due to the D-proline based introduction, whereas the linker effect seems to be not equally performing. Crystallographic data and fluorimetric assays provided experimental confirmations of the predicted binding pose and relative affinity trend towards MMP-12. In particular, these results showed that the linker is able to

push the fist aromatic ring more deep inside the S1' pocket, positioning the ring itself in a parallel displaced conformation for stacking interaction towards the aromatic imidazole ring of the highly conserved zinc binding histidine of MMPs (Figure 18).

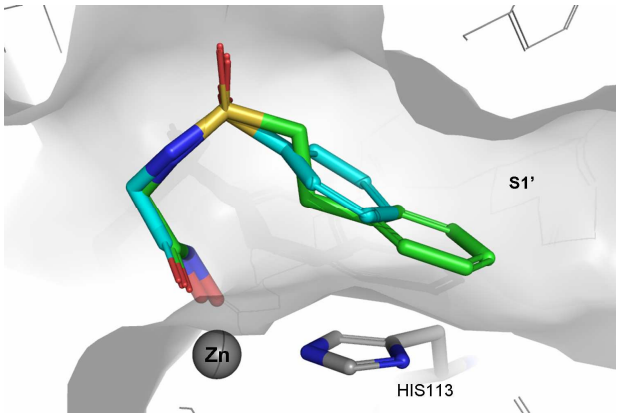


Figure 18 – Overlap of the binding conformation of a classic (cyan) and a optimized (green) inhibitor. In grey stick is highlighted the aromatic side chain of Zinc binding histidine (His113).

It is known that this conformation provides lower interaction energy than the sandwich like conformation displayed by the classic scaffold. [105,106] Moreover, fluorimetric assays have pointed out the higher inhibitory activity of molecules bearing the linker with respect to the classic ligands. The fine analysis of the single atom energy contribution to the enthalpy of interaction, carried out by docking refinement of the X-Ray structures, was not able, however, to pick out the molecular determinants responsible for the observed binding improvement. The same limit has been observed during the automated docking-based energy calculation. Therefore a more sophisticated analysis was required due to the limits of molecular mechanics in correctly parameterise these systems. The most reliable approach to predict both solvation effect and, in particular, stacking interaction energy is represented by the quantum mechanics. A*b initio* second order Møller Plesset perturbation theory (MP2) and SCRF-PCM methods were applied to calculate the π-π interaction energy and the solvation energy of each isolated compounds respectively. [106-108] Self-Consistent Reaction Field methods were exploited to calculate the energy of a system in solution (in this case a water solution), applying the Polarizable Continuum Model in which the cavity where the solute is placed is created via a series of overlapping spheres of solvent within the reaction field. MP2 method was selected for a more accuracy calculation of the

interaction energy than simple Hartree-Fock theory, due to the introduction of the electronic perturbation of molecular orbitlas. The results of solvation energy calculations carried out with SCRF(PCM) methods at the Hartree-Fock level of theory with a moderately large basis set (6-311++G(d,p)), clearly show that the D-proline residue determines a massive decrease of the ligand solvation energy with respect to glycine. On the other hand, the linker introduction does not determines the same effect, providing higher values for solvation energy that are included within the standard error for the selected method (see Table 1 and 2). [107]

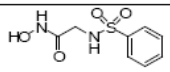
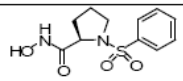
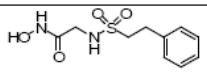
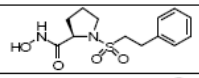
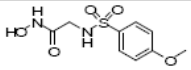
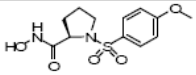
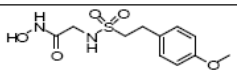
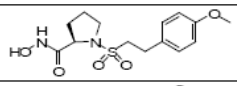
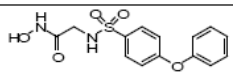
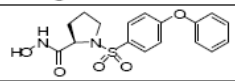
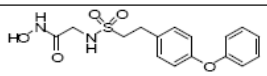
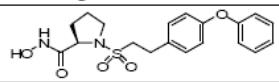
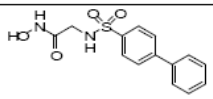
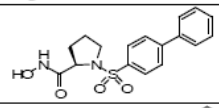
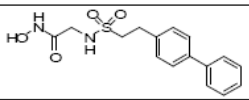
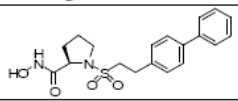
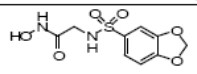
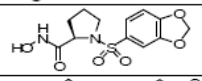
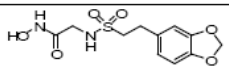
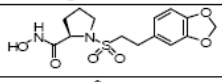
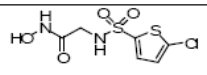
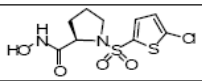
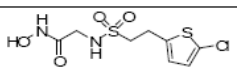
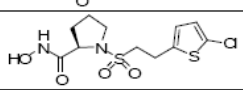
Glycine	ΔG_{solv} Kcal/mol	$\Delta\Delta G_{solv}$	ΔG_{solv} Kcal/mol	**D-Proline**
	-29.01	**-10.09**	-18.92	
	-27.52	**-10.31**	-17.21	
	-30.34	**-10.64**	-19.70	
	-29.45	**-10.59**	-18.86	
	-28.43	**-11.03**	-17.40	
	-27.50	**-10.52**	-16.88	
	-25.41	**-9.73**	-15.68	
	-24.57	**-9.36**	-15.21	
	-33.80	**-9.26**	-24.54	
	-33.70	**-10.01**	-23.69	
	-28.94	**-8.67**	-20.27	
	-28.64	**-8.49**	-20.15	

Table 1 – Solvation Energy differences between classic and new molecules, calculated with Gaussian (PCM method). $\Delta\Delta G_{solv}$ values are significant for the different energy contribution provided by glycine (left column) and D-proline (right column).

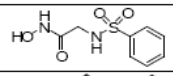
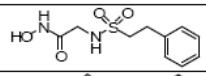
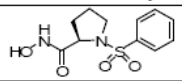
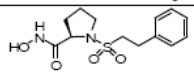
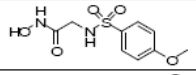
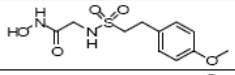
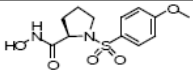
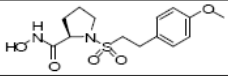
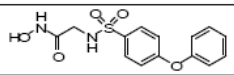
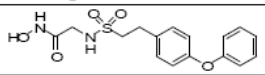
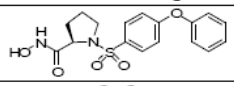
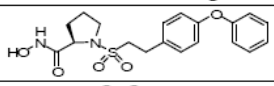
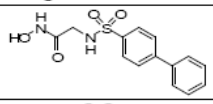
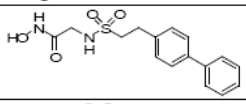
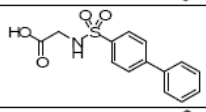
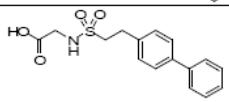
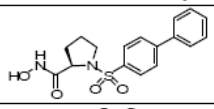
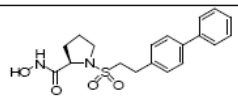
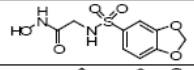
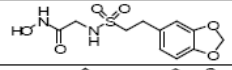
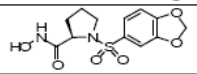
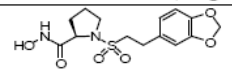
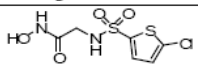
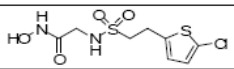
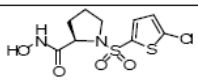
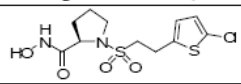
Classic	ΔG_{solv} Kcal/mol	$\Delta\Delta G_{solv}$	ΔG_{solv} Kcal/mol	Second Generation
	-29.01	**-1.49**	-27.52	
	-18.92	**-1.71**	-17.21	
	-30.34	**-0.89**	-29.45	
	-19.70	**-0.84**	-18.86	
	-28.43	**-0.93**	-27.50	
	-17.40	**-0.52**	-16.88	
	-25.41	**-0.84**	-24.57	
	-26.00	**-1.18**	-24.82	
	-15.68	**-0.47**	-15.21	
	-33.80	**-0.10**	-33.70	
	-24.54	**-0.85**	-23.69	
	-28.94	**-0.30**	-28.64	
	-20.27	**-0.12**	-20.15	

Table 2 – Solvation Energy differences between glycine-based and proline-based molecules, calculated with Gaussian (PCM method). $\Delta\Delta G_{solv}$ *values are not significant and included within the standard error of the method.*

This evidence, however, set the bases for the further QM approach aimed at the calculation of stacking interaction energies for the aromatic systems in protein/inhibitors adducts. In fact, the PCM calculations evidenced that the experimentally observed positive effect of the linker to the binding can not be attribute to the solvation/desolvation effect. The main conformational difference between classic and the new scaffold can be found in the S1' pocket. In particular, the X-Ray determined different position of the inhibitor's P1' aromatic ring towards the imidazole side chain of the zinc binding histidine reflects, as already assumed, a different conformation for π-stacking interaction. To evaluate the energetic contribution of these two stacking conformations, MP2 interaction energy

calculations were performed applying a moderately large basis set (6-311++G(d,p)) on coordinate's systems isolated from the correspondent X-Ray structures. These calculations were performed on the lowest molecular weight compound of each scaffold. The studied systems were constituted by the catalytic Zn(II) ion, the three imidazole rings of zinc binding histidines (on which the Cβ have been transformed into methyl groups), and the inhibitor, as reported in Figure 19.

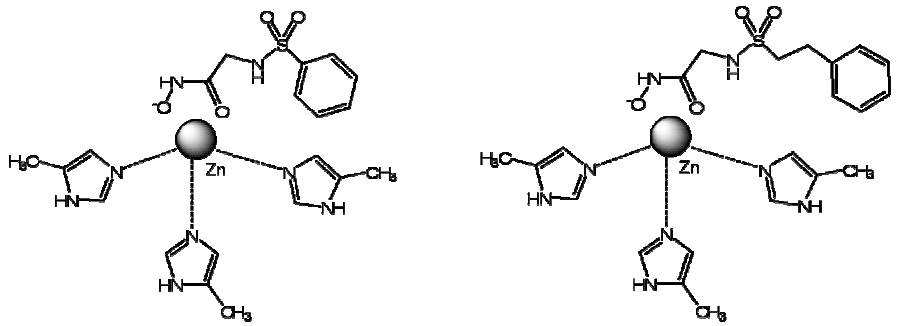


Figure 19 – The two systems extrapolated from X-Ray determined adducts of classic (left) and optimized inhibitors (right).

Interaction energy between Zn-his(3) and inhibitor was calculated as the difference between the total energy of the system and the sum of the energy of each component. BSSE was corrected during the calculations at the counterpoise level. [109] The results clearly show that the linker introduction determines an effective decrease of the interaction energy, with respect to the classic scaffold. From the evidence that the only difference in those systems is represented by the reciprocal position of the two aromatic rings in performing π-π interaction, this energetic difference can be attributed to the different stacking conformation. Thus, as above assumed, different conformations for π-π interaction correspond to different interaction energy. The energy values calculated with this approach reflect the higher stability of the parallel displaced reciprocal position of the aromatic rings, with respect to the sandwich like conformation observed for classic scaffold. This approach has been applied for the optimization of the interaction between ligands and MMP-12, but can easily be extended to other MMPs due the high sequence and structural conservation of the protein residues involved in the binding of these ligands. [104] In conclusion, both linker introduction and D-proline substitution at the same time in the hydroxamic aryl sulphonamide scaffold for MMPs inhibitors provide higher affinity towards these target

proteins. The computational approach here investigated was able to identify the ligand's molecular determinants responsible for the observed binding improvement.

This computational work on MMP proteins was mainly carried out by me.

These results are going to be submitted for publication, and the draft is reported in appendix C

***5.4**. Design and Virtual Screening of a Focussed Library of Peptides Towards alpha Tubulin.*

In the frame of an European project (FP6, acronym: Nano4Drugs), *in silico* methodologies have been applied for the discovery of new tubulin binding peptides as candidate protein-based drugs. According to the overview described above (see chapter 4.3), the dynamics of tubulins and microtubules is affected by molecules, such as natural alkaloids or stathmin protein, that interact with alpha tubulin stabilizing a curved conformation of the heterodimer constituted by α and β tubulin. This particular conformation impedes the polymerization of tubulins into microtubules, and determines cellular apoptosis. Thus, the development of molecules able to interfere with the dynamics of tubulins and microtubules could represent a new therapeutic approach towards cancer.

Using available structures of tubulin, microtubules and microtubule associated proteins, we have applied a preliminary docking based virtual screening of the "NCI-DIVERSITY Set" (about 2000 compounds representing the broader chemical space of the 140,000 or so compounds in its holdings) to define and characterise the individual molecular targets present in these proteins. [110] The resulting data allowed us to optimize protein-based drug candidates in terms of selectivity and potency through the analysis of structure and energetic aspects of binding. One of the putative binding areas correspond to the region of interaction of a small peptide (19aa) derived from stathmin. This peptide, also called I19L, has been demonstrated to interfere with the tubulins polymerization and dynamics; for that reason it constituted the reference scaffold for the generation of the initial focused virtual library of peptides by applying single point mutations and double consecutive-point mutations.

The initial library was screened by *in silico* procedures. In particular, a first homology modelling step on the reference 3D structure of alpha tubulin bound to the I19L peptide (unpublished data) was carried out. The 2D sequences of single point mutants and double consecutive point mutants were modelled on the 3D structure of I19L, in its complex conformation, until their state of energy minimum; the corresponding 3D adducts with alpha tubulin were then refined by a short molecular dynamics step. At the end of the MD calculations the peptide energy and the delta energy of binding between each peptide and tubulin were calculated by solving the Poisson Boltzmann equation. [111] *Internal Energy* and Δ *Energy* values were taken into account for the scoring analysis on the mutants, as

described in Table 3. *Internal Energy* describes the peptide energy in the binding conformation, whereas Δ *Energy* correspond to the interaction energy between peptides and tubulin, and is calculated by solving the Poisson-Boltzmann equation towards these systems.

Single point mutations to I19L	Internal energy	Δ energy	**Double points mutations to I19L**	Internal energy	Δ energy
I1M	-804.89	-96.09	I1E_Q2R	-956.95	-103.28
V3Y	-771.27	-91.27	Q2S_V3H	-697.27	-103.10
K4R	-952.24	-94.58	K4S_E5K	-702.23	-105.92
E5M	-705.55	-100.53	E5R_L6R	-1054.86	-106.33
K8Q	-829.63	-104.65	E5T_L6T	-727.30	-109.81
A10Q	-849.28	-92.75	E7I_K8H	-691.46	-105.80
Q13L	-723.09	-90.64	Q13L_A14Y	-749.77	-109.13
A14I	-759.52	-90.83	**Q13R_A14R**	-1102.20	-102.42
A14R	-986-59	-92.75	A14E_F15W	-821.88	-102.96
E16K	-742.48	-95.14	E16S_L17M	-708.43	-100.02
L19K	-756.21	-97.77	L17M_I18K	-781.34	-103.89
L19T	-808.33	-93.45	I18C_L19R	-950.20	-104.44
			I18M_L19Q	-805.82	-103.97

Table 3 – Best results obtained by the single and double point mutation of I19L peptide. In red are highlighted the peptides tested in vitro. Internal Energy is the total energy of each peptide in the complex conformation and Δ energy is the free energy of interaction between each peptide and alpha tubulin. All values are expressed in Kcal mol-1 and have been extrapolated from the molecular dynamics output file by solving the Poisson-Boltzmann equation.

The most promising hits come out from this preliminary analysis were proposed for *in vitro* testing. Efficacy of these new hit peptides was quantified by their effects on microtubule mass and dynamics *in vitro*. The most efficient mutant was the peptide called K4R that was selected for a further *in silico* refinement cycle on which it was subjected to single point sequence mutations, homology modelling and MD energy minimization, as previously described. The final Energy calculation step pointed out the large interaction energy improvement due to substitutions in positions 10, 14 and 19 of the beta-hairpine peptide scaffold. Moreover, the ligand energy was surprisingly decreased in the cases of arginine substitution applied in the previously mentioned positions.

Single Point Mutations to K4R	Internal energy	Δ energy
A10R	-1090.12	-102.63
A14H	-982.11	-100.02
L19R	-1127.50	-98.12
A14I	-946.73	-98.05
L19H	-925.16	-97.10
A14E	-1005.68	-96.49
E7R	-1041.93	-96.30
S11Y	-966.08	-95.46
V3Q	-1003.22	-95.18
L17E	-1000.21	-94.40

Single Point Mutations to K4R	Internal energy	Δ energy
A14R	-1143.68	-86.75
L19R	-1127.50	-98.12
G12R	-1120.05	-74.79
K8R	1116.49	-82.24
F15R	-1115.21	-81.51
I18R	1106.96	-76.79
L17R	-1106.47	-94.39
L6R	-1106.01	-90.48
V3R	-1094.42	-75.20
S11R	-1090.62	-90.32

Tables 4 and 5 – Best results sorted by Delta Energy of binding (upper table) and Ligand Energy (down table), as obtained from K4R refinement cycle. These energetic values have been calculated in the way as above described.

The conformational analysis of the *in silico* predicted complexes between alpha tubulin and peptides of tables 4 and 5 revealed that arginine enriched peptides are able to interact with a cluster of -1 charged residues that are present in the binding pocket of tubulin. In particular, the position 10, 14 and 19 are the most suitable for arginine substitution due to their close proximity and correct directionality with respect to glutamate and aspartate residues of tubulin.

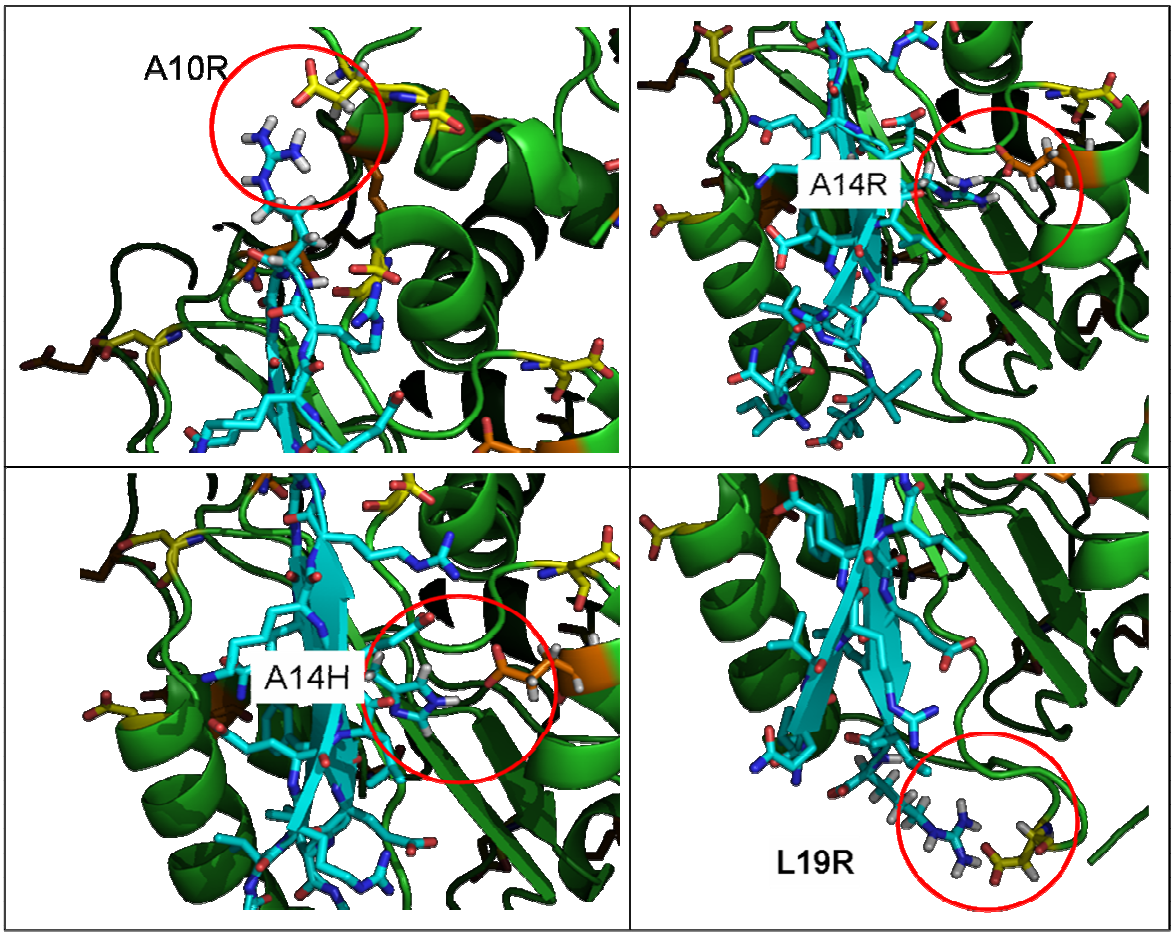


Figure 20 – Conformational details of arginine enriched peptides in complex with α tubulin.

The linear combination of these mutation represented the final step of the peptides optimization process. The above described protocol was applied to the mutants reported in the table below. As observable, both the energetic parameters remarkably decrease.

Multiple Point Mutations to K4R	Internal energy	Δ energy
A10R_A14R	-1331.45	-98.8
A10R_A14H	-1111.79	-85.27
A10R_A14R_L19R	-1489.56	-121.41
A10R_A14H_L19R	-1314.75	-99.99

Table 6 – Energy results obtained from the combination of the favourable single point substitution to K4R peptide. Results are expressed in Kcal mol^{-1}, and have been calculated as above described.

The conformational analysis also revealed that all arginine substitutions are functional for targeting aspartate and glutamate residues placed in the selected binding region of alpha tubulin. The histidine substitution performed in position 14 is not able to provide low energy values as observed in case of arginine substitution; thus, arginine-based peptides have been proposed for a *in vitro* test that is currently under development from one member of the Nano4Drugs consortium. Despite the increased binding affinity and ligand stabilization, these peptides display a higher values for theoretical Isoelectric Point due to the increased total charge. This effect could negatively affects the cell membranes penetration of the drug peptide. Nevertheless, it has been reported that arginine enriched peptides are generally more active than others to cross the biological membranes. [112-114] A10R_A14R and A10R_AA14R_L19R mutants have been proposed, together with the set of tables 4 and 5, as the lead peptides for the final *in vitro* and *in vivo* efficacy evaluation. In case that *in silico* predictions will be experimentally confirmed, these peptides would probably represent the protein-based candidate drugs aimed at stabilize/destabilize microtubules, according to the consortium achievements.

I have contributed to this work with most of calculations on the initial library, focussed on I19L peptide, and with all calculations on the second library focussed on K4R peptide, until the selection on the lead peptides.

***5.5**. In Vivo Efficacy Evaluation of a MMP-9 inhibitor in Dry Eye Syndrome.*

Computational methods are generally employed in drug discovery pipelines. Despite their precision and accuracy in predicting molecular behaviour, the results obtained by these approaches need to be experimentally confirmed. In this way i*n vitro* and *in vivo* experiments are aimed at providing more precise information about the efficacy, toxicity and bioavailability of the selected lead molecules.

In this part of the project we have preliminarily *in vitro* and *in vivo* evaluated the preclinical efficacy and safety of a MMP-9 inhibitor, called PES_103, in the therapy of Dry Eye Syndrome (DES). [115] MMP-9 is a member of gelatinases group and is involved in many different human diseases such as cancer and systemic syndromes. [116] To evaluate the efficacy of a synthetic inhibitor selected by computational and *in vitro* approaches, experimental studies have been performed. In particular, the inhibitory activity of this molecule in the therapy of Dry Eye Syndrome (DES) was evaluated.

Dry Eye Syndrome or Keratoconjunctivitis sicca (KCS) is an eye disease caused by the chronic lack of lubrication and moisture in the eye. In humans it can be caused by a decrease in tears production or an increase in tear film evaporation. Both eyes are in general affected and the physical damages occurs mainly at the corneal level. [117] Causes of this syndrome may be attributed to many different factors and, according to a recent statistic, more than 20 million Americans may be suffering from Dry Eye. [118] The 75% of the population over the age of sixty-five years is affected by this syndrome that is currently managed by the topical administration of artificial tears (AT). This symptomatic approach is right now the most efficient commercially available management for dry eye but it requires frequent application, without providing long term efficacy.

One of the pathological consequences of eye dryness is reported to be the over-expression of MMP-9 (Collagenases B) at corneal level. This protein is a marker of eye dryness; nevertheless it has been demonstrated that knock-out mice for the gene coding for MMP-9 protein are not able to develop experimental dry eye, leading to the evidence that this protein is somehow involved in the dryness process itself. [115] Accordingly, a possible asymptomatic therapeutic approach to DES could be represented by the eye topical administration of a MMPs inhibitor with a particular affinity towards MMP-9. The aim of the presented investigation was the preclinical evaluation of an *in house* developed MMPs

inhibitor, coded as PES_103, in the treatment of dry eye. This molecules has a high affinity towards the target protein (Ki=14 nM) and a high water solubility (more than 40 mg/ml).
This experimental work started with a *in vitro* analysis of the inhibitory activity of the selected inhibitor. HT1080 are sarcoma cells with a high secretion of MMP-2 and MMP-9. [119] HT1080 conditioned medium was then analyzed by gelatine zymography in presence of increasing concentrations of PES_103. The gelatin zymogram of HT1080 culture supernatant is presented in Figure 21. Untreated supernatant (left hand column) produces tree bands of molecular weights 92 kDa, 82 kDa and 72 kDa representing pro-MMP-9, active MMP-9 and fused band for MMP-2 that include both the pro-domain and active forms, respectively. The incubation of 150 µl medium for 24h together with increasing concentrations of PES_103 (1, 3, 25, 50 and 75 mM) provided a decreased intensity of the band associated to active MMP-9, as a probable consequence of PES_103 inhibitory activity.

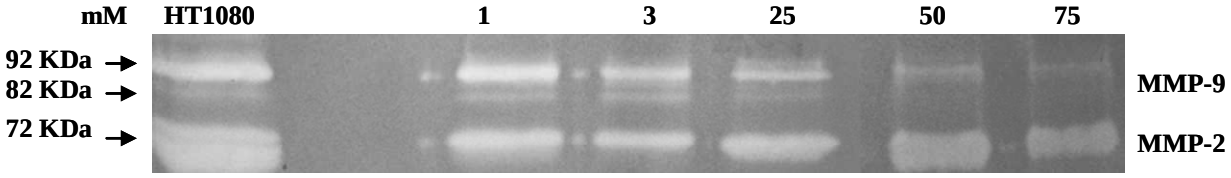


Figure 21. Gelatin zymogram of equal volumes of medium conditioned by HT1080 cells; The numbers give millimolar (mM) concentration of PES_103 added in the culture.

Once achieved the *in vitro* confirmation of the inhibitory activity of PES_103 towards MMP-9, the biggest challenge was the demonstration of the MMP-9 involvement in the dryness process, and then the evaluation of the efficacy of PES_103 on the adopted animal model for DES. To evaluate the over-expression of MMP-9 in association with eye dryness, Experimental Dry Eye (EDE) was induced in C57BL/6 mice by transdermal application of scopolamine patch. As described by Dursun et al. [120] scopolamine resulted in a significant decrease in tears production due to the cholinergic-based blockade of the tear gland's muscles. The effective dryness was verified by measuring the tears production with phenol red–impregnated cotton threads, while the active MMP-9 over-production was checked by gelatine zymography after tears collection. As reported in Figure 22, the ocular dryness condition occurs after about 24 hours from the transdermal scopolamine

administration. Tears volumes remain at low levels (dry eye) until about 60 hours; after this time they quickly recover until the physiological level (black histograms; control mice). On the other hand, zymographyc plots clearly show that high levels of active MMP-9 in tears fluid are observable starting from 24h until 48h from the SCP administration.

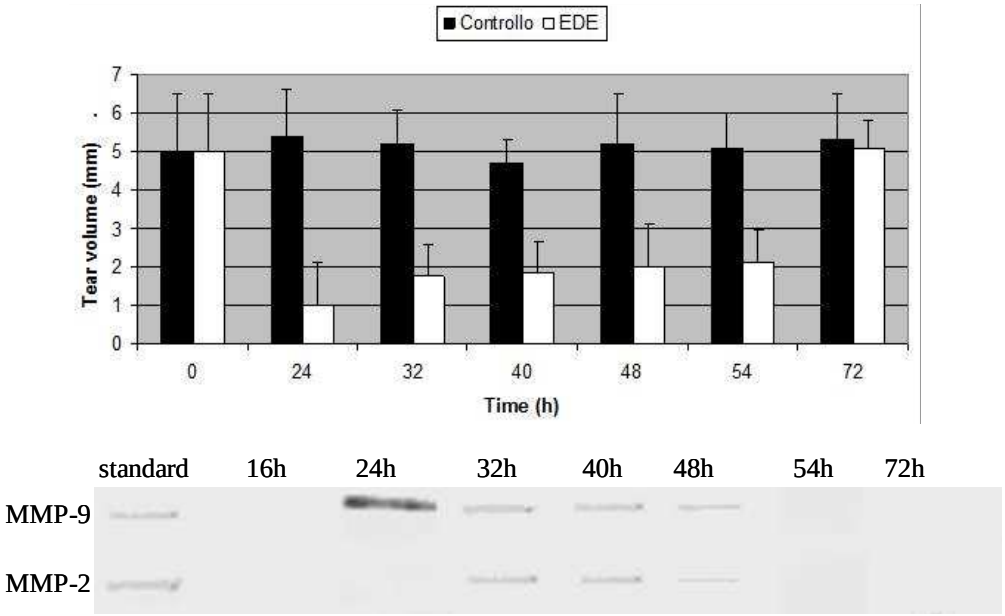


Figure 22 – Plot of the tear production (top) and active MMP-9 expression levels (down)

In conclusion, the most important feature merged from this experiment is the relationship between MMP-9 over-expression and dry eye induction that allows the possibility to evaluate the efficacy of the selected inhibitor in the asymptomatic treatment of dry eye.

Before to proceed with the efficacy test, the safety of PES_103 was tested. Toxicity was preliminarily checked by *in vivo* and histological assays on treated and control mice. PES_103 was applied in sterile physiological buffer solution at 0.01%, 0.1% and 1% concentrations. Fluoresceine plus UV light and lissamine green were applied on the mice's eyes to verify the complete absence of corneal ulceration and keratinisation processes, respectively. [121,122] Histological analysis of eye sections after H&E and Tunel colouring treatments revealed the absence of tissue alteration or increased number of apoptotic cells with respect to the control, untreated, mice. The lack of toxicity on mice was considered as a milestone step before to proceed with the efficacy evaluation.

The final part of this work was aimed at the evaluation of the efficacy of the selected molecule on the validated animal model. Scopolamine patches were applied on the mice's midtail in order to produce eye dryness. Then, EDE mice were treated for 5 days, twice a day, with PES_103 0.1% (3 mM). This specific concentration of MMPs inhibitor was selected by a dose finding study on which the above described concentrations were preliminarily tested on the selected animal model (data not presented). The efficacy of the selected MMP-9 inhibitor was finally evaluated by the comparison of the tears volume of PES_103 treated EDE mice with respect to the same variable measured in control, only EDE, and AT treated EDE mice. Tear volumes were measured immediately before and after 30' and 60' of each administration. The plot of the tears volume for each group, Figure 23, clearly evidences the higher efficacy of PES_103 in restoring the tears production (60%), with respect to EDE mice (25%) and AT treated EDE mice (40%).

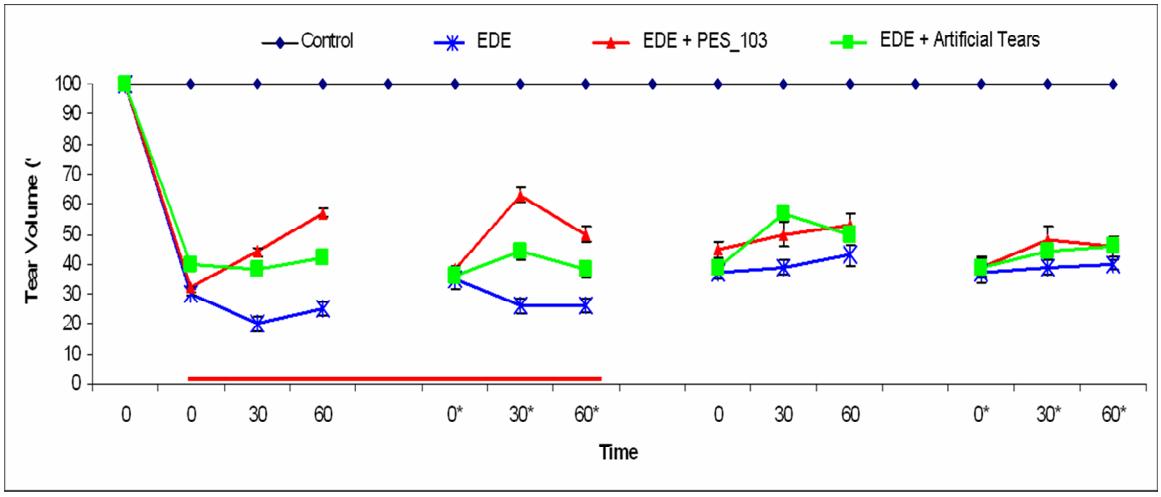


Figure 23 – plotting of the tear volume (percentage with respect to control mice) over the time of the four measurements session. MMP.9 levels are detectable from the beginning of the first session until the end of the second, as reported by the red line.

Interestingly, the effects of PES_103 seems to be related to the corneal level of MMP-9 whereas AT, for example, display always a constant profile. When the MMP-9 level are no more detectable by zymography (beginning of the third session in Figure 23), the efficacy of PES_103 is quite similar to those observed for artificial tears. This PES_103's behaviour could account for an asymptomatic mechanism of action.

In conclusion, in this work we have verified that inhibiting MMP-9 at corneal level could represent a therapeutic opportunity to manage Dry Eye Syndrome. Moreover, we provided

support to the asymptomatic and therapeutic mechanism of action of PES_103, evidencing that its potency in restoring the tears production is proportional to the corneal MMP-9 levels. Finally, we have been able to verify the good quality of the *in silico* prediction for the tested molecules towards MMP proteins.

I have physically participated to the *in vivo* tests of the MMP-9 inhibitor, and I have essentially planned and contributed to execute the necessary experiments for toxicity and efficacy evaluation of the selected molecule.

These results are going to be submitted for publication, and a early draft is included in appendix D.

6. Conclusions

In this work are presented the results obtained in carrying out projects in the frame of structure-based drug design. Computer aided methodologies were applied to simulate the biological behaviour and physical/chemical properties of ligands during the protein binding process. In detail, ligands belonging to the classes of enzyme inhibitors, protein signalling pathway modulators, and protein-protein interaction inhibitors were analysed. Computational tools were applied to simulate the interaction between ligands and the potential pharmacological target proteins, chosen on the basis of their involvements in human diseases. Molecular mechanics methodologies were applied for ligand virtual screening and preliminary docking simulations of the interaction between ligands and the selected target protein; molecular dynamics was also applied in the hits refinement step, with the aim to simulate the reciprocal adaptation of protein and ligands during the interaction. Quantum mechanics methods were finally used for the fine analysis of electron correlated ligands properties, such as solvation/desolvation effects and π-stacking energy calculations. All those procedures were applied in a drug discovery pipeline with the aim to identify the set of promising hit or lead compounds on which to concentrate the *in vitro* and *in vivo* analysis.

Finally, an *in vivo* experimental session was carried out to evaluate the quality of the *in silico* prediction. More precisely, the preliminary therapeutic activity of a MMP-9 inhibitor was succesfully verified in an animal model of Dry Eye Syndrome, confirming the capability of the applied computational protocol in predicting the biological behaviour of this class of molecules. The results herein presented highlight the advantages and the worsening of applying *in silico* structure-based drug design approach in a drug discovery pipeline.

Reference List

1. Yeh, B. J.; Lim, W. A. Synthetic biology: lessons from the history of synthetic organic chemistry. *Nat. Chem Biol.* **2007,** *3* (9), 521-525.

2. Bentley, D. R. Decoding the human genome sequence. *Hum. Mol. Genet.* **2000,** *9* (16), 2353-2358.

3. Venter, J. C.; Adams, M. D.; Myers, E. W.; Li, P. W.; Mural, R. J.; Sutton, G. G.; Smith, H. O.; Yandell, M.; Evans, C. A.; Holt, R. A.; Gocayne, J. D.; Amanatides, P.; Ballew, R. M.; Huson, D. H.; Wortman, J. R.; Zhang, Q.; Kodira, C. D.; Zheng, X. H.; Chen, L.; Skupski, M.; Subramanian, G.; Thomas, P. D.; Zhang, J.; Gabor Miklos, G. L.; Nelson, C.; Broder, S.; Clark, A. G.; Nadeau, J.; McKusick, V. A.; Zinder, N.; Levine, A. J.; Roberts, R. J.; Simon, M.; Slayman, C.; Hunkapiller, M.; Bolanos, R.; Delcher, A.; Dew, I.; Fasulo, D.; Flanigan, M.; Florea, L.; Halpern, A.; Hannenhalli, S.; Kravitz, S.; Levy, S.; Mobarry, C.; Reinert, K.; Remington, K.; bu-Threideh, J.; Beasley, E.; Biddick, K.; Bonazzi, V.; Brandon, R.; Cargill, M.; Chandramouliswaran, I.; Charlab, R.; Chaturvedi, K.; Deng, Z.; Di, F., V; Dunn, P.; Eilbeck, K.; Evangelista, C.; Gabrielian, A. E.; Gan, W.; Ge, W.; Gong, F.; Gu, Z.; Guan, P.; Heiman, T. J.; Higgins, M. E.; Ji, R. R.; Ke, Z.; Ketchum, K. A.; Lai, Z.; Lei, Y.; Li, Z.; Li, J.; Liang, Y.; Lin, X.; Lu, F.; Merkulov, G. V.; Milshina, N.; Moore, H. M.; Naik, A. K.; Narayan, V. A.; Neelam, B.; Nusskern, D.; Rusch, D. B.; Salzberg, S.; Shao, W.; Shue, B.; Sun, J.; Wang, Z.; Wang, A.; Wang, X.; Wang, J.; Wei, M.; Wides, R.; Xiao, C.; Yan, C.; Yao, A.; Ye, J.; Zhan, M.; Zhang, W.; Zhang, H.; Zhao, Q.; Zheng, L.; Zhong, F.; Zhong, W.; Zhu, S.; Zhao, S.; Gilbert, D.; Baumhueter, S.; Spier, G.; Carter, C.; Cravchik, A.; Woodage, T.; Ali, F.; An, H.; Awe, A.; Baldwin, D.; Baden, H.; Barnstead, M.; Barrow, I.; Beeson, K.; Busam, D.; Carver, A.; Center, A.; Cheng, M. L.; Curry, L.; Danaher, S.; Davenport, L.; Desilets, R.; Dietz, S.; Dodson, K.; Doup, L.; Ferriera, S.; Garg, N.; Gluecksmann, A.; Hart, B.; Haynes, J.; Haynes, C.; Heiner, C.; Hladun, S.; Hostin, D.; Houck, J.; Howland, T.; Ibegwam, C.; Johnson, J.; Kalush, F.; Kline, L.; Koduru, S.; Love, A.; Mann, F.; May, D.; McCawley, S.; McIntosh, T.; McMullen, I.; Moy, M.; Moy, L.; Murphy, B.; Nelson, K.; Pfannkoch, C.; Pratts, E.; Puri, V.; Qureshi, H.; Reardon, M.; Rodriguez, R.; Rogers, Y. H.; Romblad, D.; Ruhfel, B.; Scott, R.; Sitter, C.; Smallwood, M.; Stewart, E.; Strong, R.; Suh, E.; Thomas, R.; Tint, N. N.; Tse, S.; Vech, C.; Wang, G.; Wetter, J.; Williams, S.; Williams, M.; Windsor, S.; Winn-Deen, E.; Wolfe, K.; Zaveri, J.; Zaveri, K.; Abril, J. F.; Guigo, R.; Campbell, M. J.; Sjolander, K. V.; Karlak, B.; Kejariwal, A.; Mi, H.; Lazareva, B.; Hatton, T.; Narechania, A.; Diemer, K.; Muruganujan, A.; Guo, N.; Sato, S.; Bafna, V.; Istrail, S.; Lippert, R.; Schwartz, R.; Walenz, B.; Yooseph, S.; Allen, D.; Basu, A.; Baxendale, J.; Blick, L.; Caminha, M.; Carnes-Stine, J.; Caulk, P.; Chiang, Y. H.; Coyne, M.; Dahlke, C.; Mays, A.; Dombroski, M.; Donnelly, M.; Ely, D.; Esparham, S.; Fosler, C.; Gire, H.; Glanowski, S.; Glasser, K.; Glodek, A.; Gorokhov, M.; Graham, K.; Gropman, B.; Harris, M.; Heil, J.; Henderson, S.; Hoover, J.; Jennings, D.; Jordan, C.; Jordan, J.; Kasha, J.; Kagan, L.; Kraft, C.; Levitsky, A.; Lewis, M.; Liu, X.; Lopez, J.; Ma, D.; Majoros, W.; McDaniel, J.; Murphy, S.; Newman, M.; Nguyen, T.; Nguyen, N.; Nodell, M. The sequence of the human genome. *Science* **2001,** *291* (5507), 1304-1351.

4. Grenet, O. Significance of the human genome sequence to drug discovery. *Pharmacogenomics. J.* **2001,** *1* (1), 11-12.

5. Whittle, P. J.; Blundell, T. L. Protein structure--based drug design. *Annu. Rev. Biophys. Biomol. Struct* **1994,** *23*, 349-375.

6. Walkinshaw, M. D. Protein targets for structure-based drug design. *Med. Res. Rev.* **1992,** *12* (4), 317-372.

7. Colman, P. M. Structure-based drug design. *Curr. Opin. Struct Biol.* **1994,** *4* (6), 868-874.

8. Overington, J. P.; Al-Lazikani, B.; Hopkins, A. L. How many drug targets are there? *Nat. Rev. Drug Discov.* **2006,** *5* (12), 993-996.

9. Hopkins, A. L.; Mason, J. S.; Overington, J. P. Can we rationally design promiscuous drugs? *Curr. Opin. Struct Biol.* **2006,** *16* (1), 127-136.

10. Bacilieri, M.; Moro, S. Ligand-based drug design methodologies in drug discovery process: an overview. *Curr. Drug Discov. Technol.* **2006,** *3* (3), 155-165.

11. Ferenczy, G. [Structure-based drug design]. *Acta Pharm. Hung.* **1998,** *68* (1), 21-31.

12. Gago, F. Modelling and simulation: a computational perspective in anticancer drug discovery. *Curr. Med. Chem. Anticancer Agents* **2004,** *4* (5), 401-403.

13. Capener, C. E.; Kim, H. J.; Arinaminpathy, Y.; Sansom, M. S. Ion channels: structural bioinformatics and modelling. *Hum. Mol. Genet.* **2002,** *11* (20), 2425-2433.

14. Kaiser, J. Science resources. Chemists want NIH to curtail database. *Science* **2005,** *308* (5723), 774.

15. Irwin, J. J.; Shoichet, B. K. ZINC--a free database of commercially available compounds for virtual screening. *J. Chem Inf. Model.* **2005,** *45* (1), 177-182.

16. Joseph-McCarthy, D. Computational approaches to structure-based ligand design. *Pharmacol. Ther.* **1999,** *84* (2), 179-191.

17. Apostolakis, J.; Caflisch, A. Computational ligand design. *Comb. Chem. High Throughput. Screen.* **1999,** *2* (2), 91-104.

18. Bohm, H. J. Computational tools for structure-based ligand design. *Prog. Biophys. Mol. Biol.* **1996,** *66* (3), 197-210.

19. Huang, N.; Jacobson, M. P. Physics-based methods for studying protein-ligand interactions. *Curr. Opin. Drug Discov. Devel.* **2007,** *10* (3), 325-331.

20. Lybrand, T. P. Molecular simulation and drug design. *J. Pharm. Belg.* **1991,** *46* (1), 49-54.

21. Goodsell, D. S.; Morris, G. M.; Olson, A. J. Automated docking of flexible ligands: applications of AutoDock. *J. Mol. Recognit.* **1996,** *9* (1), 1-5.

22. Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated Docking Using a Lamarckian Genetic Algorithm and Empirical Binding Free Energy Function. *Journal Computational Chemistry* **1998,** *19*, 1639-1662.

23. Olson, A. J.; Goodsell, D. S. Automated docking and the search for HIV protease inhibitors. *SAR QSAR. Environ. Res.* **1998,** *8* (3-4), 273-285.

24. DePristo, M. A.; de Bakker, P. I.; Lovell, S. C.; Blundell, T. L. Ab initio construction of polypeptide fragments: efficient generation of accurate, representative ensembles. *Proteins* **2003,** *51* (1), 41-55.

25. Guvench, O.; MacKerell, A. D., Jr. Comparison of protein force fields for molecular dynamics simulations. *Methods Mol. Biol.* **2008,** *443*, 63-88.

26. Ponder J.W.; Case D.A. Force Field for Protein Simulations. *Advances in Protein Chemistry* **2003,** *66*, 27-85.

27. Raha, K.; Peters, M. B.; Wang, B.; Yu, N.; Wollacott, A. M.; Westerhoff, L. M.; Merz, K. M., Jr. The role of quantum mechanics in structure-based drug design. *Drug Discov. Today* **2007,** *12* (17-18), 725-731.

28. Peters, M. B.; Raha, K.; Merz, K. M., Jr. Quantum mechanics in structure-based drug design. *Curr. Opin. Drug Discov. Devel.* **2006,** *9* (3), 370-379.

29. M.J.Frisch; .W.Trucks; H.B.Schlegel; G.E.Scuseria; M.A.Robb; J.R.Cheeseman; J.A.Montgomery; Jr., T. V.; K.N.Kudin; .C.Burant; J.M.Millam; S.S.Iyengar; J.Tomasi; V.Barone; B.Mennucci; M.Cossi; G.Scalmani; N.Rega; G.A.Petersson; H.Nakatsuji; M.Hada; M.Ehara; K.Toyota; R.Fukuda; J.Hasegawa; M.Ishida; T.Nakajima; Y.Honda; O.Kitao; H.Nakai; M.Klene; X.Li; J.E.Knox; H.P.Hratchian; J.B.Cross; V.Bakken; C.Adamo; J.Jaramillo; R.Gomperts; R.E.Stratmann; O.Yazyev; A.J.Austin; R.Cammi; C.Pomelli; J.W.Ochterski; P.Y.Ayala; K.Morokuma; G.A.Voth; P.Salvador; J.J.Dannenberg; .G.Zakrzewski; S.Dapprich; A.D.Daniels; M.C.Strain; O.Farkas; D.K.Malick; A.D.Rabuck; K.Raghavachari; J.B.Foresman; J.V.Ortiz; Q.Cui; A.G.Baboul; S.Clifford; J.Cioslowski; B.B.Stefanov; G.Liu; A.Liashenko; P.Piskorz; I.Komaromi; R.L.Martin; D.J.Fox; T.Keith; M.A.Al-Laham; C.Y.Peng; A.Nanayakkara; M.Challacombe; P.M.W.Gill; B.Johnson; W.Chen; M.W.Wong; C.Gonzalez; J.A.Pople . Gaussian 03. Gaussian, Inc . 2004.

30. Arendt, Y.; Bhaumik, A.; Del, C. R.; Luchinat, C.; Mori, M.; Porcu, M. Fragment docking to S100 proteins reveals a wide diversity of weak interaction sites. *ChemMedChem.* **2007,** *2* (11), 1648-1654.

31. Asano, T.; Mori, T.; Shimoda, T.; Shinagawa, R.; Satoh, S.; Yada, N.; Katsumata, S.; Matsuda, S.; Kagamiishi, Y.; Tateishi, N. Arundic acid (ONO-2506) ameliorates delayed ischemic brain damage by preventing astrocytic overproduction of S100B. *Curr. Drug Targets. CNS. Neurol. Disord.* **2005,** *4* (2), 127-142.

32. Hsieh, H. L.; Schafer, B. W.; Sasaki, N.; Heizmann, C. W. Expression analysis of S100 proteins and RAGE in human tumors using tissue microarrays. *Biochem. Biophys. Res. Commun.* **2003,** *307* (2), 375-381.

33. Heizmann, C. W.; Ackermann, G. E.; Galichet, A. Pathologies involving the S100 proteins and RAGE. *Subcell. Biochem.* **2007,** *45*, 93-138.

34. Donato, R. Intracellular and extracellular roles of S100 proteins. *Microsc. Res. Tech.* **2003,** *60* (6), 540-551.

35. Donato, R. S100: a multigenic family of calcium-modulated proteins of the EF-hand type with intracellular and extracellular functional roles. *Int. J. Biochem. Cell Biol.* **2001,** *33* (7), 637-668.

36. Donato, R. Functional roles of S100 proteins, calcium-binding proteins of the EF-hand type. *Biochim. Biophys. Acta* **1999,** *1450* (3), 191-231.

37. Donato, R. S-100 proteins. *Cell Calcium* **1986,** *7* (3), 123-145.

38. Kubista, H.; Donato, R.; Hermann, A. S100 calcium binding protein affects neuronal electrical discharge activity by modulation of potassium currents. *Neuroscience* **1999,** *90* (2), 493-508.

39. Ridinger, K.; Ilg, E. C.; Niggli, F. K.; Heizmann, C. W.; Schafer, B. W. Clustered organization of S100 genes in human and mouse. *Biochim. Biophys. Acta* **1998,** *1448* (2), 254-263.

40. Allore, R. J.; Friend, W. C.; O'Hanlon, D.; Neilson, K. M.; Baumal, R.; Dunn, R. J.; Marks, A. Cloning and expression of the human S100 beta gene. *J. Biol. Chem.* **1990,** *265* (26), 15537-15543.

41. Zimmer, D. B.; Cornwall, E. H.; Landar, A.; Song, W. The S100 protein family: history, function, and expression. *Brain Res. Bull.* **1995,** *37* (4), 417-429.

42. Johnson, S. L.; Pellecchia, M. Structure- and fragment-based approaches to protease inhibition. *Curr. Top. Med. Chem.* **2006,** *6* (4), 317-329.

43. Cuniasse, P.; Devel, L.; Makaritis, A.; Beau, F.; Georgiadis, D.; Matziari, M.; Yiotakis, A.; Dive, V. Future challenges facing the development of specific active-site-directed synthetic inhibitors of MMPs. *Biochimie* **2005,** *87* (3-4), 393-402.

44. Markowitz, J.; MacKerell, A. D., Jr.; Weber, D. J. A search for inhibitors of S100B, a member of the S100 family of calcium-binding proteins. *Mini. Rev. Med. Chem.* **2007,** *7* (6), 609-616.

45. Markowitz, J.; MacKerell, A. D., Jr.; Carrier, F.; Charpentier, T. H.; Weber, D. J. Design of Inhibitors for S100B. *Curr. Top. Med. Chem.* **2005,** *5* (12), 1093-1108.

46. Ennis, B. W.; Matrisian, L. M. Matrix degrading metalloproteinases. *J. Neurooncol.* **1994,** *18* (2), 105-109.

47. Baramova, E.; Foidart, J. M. Matrix metalloproteinase family. *Cell Biol. Int.* **1995,** *19* (3), 239-242.

48. Cockett, M. I.; Birch, M. L.; Murphy, G.; Hart, I. R.; Docherty, A. J. Metalloproteinase domain structure, cellular invasion and metastasis. *Biochem. Soc. Trans.* **1994,** *22* (1), 55-57.

49. Fu, X.; Kassim, S. Y.; Parks, W. C.; Heinecke, J. W. Hypochlorous acid oxygenates the cysteine switch domain of pro-matrilysin (MMP-7). A mechanism for matrix metalloproteinase activation and atherosclerotic plaque rupture by myeloperoxidase. *J. Biol. Chem.* **2001,** *276* (44), 41279-41287.

50. Van Wart, H. E.; Birkedal-Hansen, H. The cysteine switch: a principle of regulation of metalloproteinase activity with potential applicability to the entire matrix metalloproteinase gene family. *Proc. Natl. Acad. Sci. U. S. A* **1990,** *87* (14), 5578-5582.

51. Bode, W.; Fernandez-Catalan, C.; Grams, F.; Gomis-Ruth, F. X.; Nagase, H.; Tschesche, H.; Maskos, K. Insights into MMP-TIMP interactions. *Ann. N. Y. Acad. Sci.* **1999,** *878*, 73-91.

52. Denhardt, D. T.; Feng, B.; Edwards, D. R.; Cocuzzi, E. T.; Malyankar, U. M. Tissue inhibitor of metalloproteinases (TIMP, aka EPA): structure, control of expression and biological functions. *Pharmacol. Ther.* **1993,** *59* (3), 329-341.

53. Ray, J. M.; Stetler-Stevenson, W. G. The role of matrix metalloproteases and their inhibitors in tumour invasion, metastasis and angiogenesis. *Eur. Respir. J.* **1994,** *7* (11), 2062-2072.

54. Vincenti, M. P. The matrix metalloproteinase (MMP) and tissue inhibitor of metalloproteinase (TIMP) genes. Transcriptional and posttranscriptional regulation, signal transduction and cell-type-specific expression. *Methods Mol. Biol.* **2001,** *151*, 121-148.

55. Pflugfelder, S. C.; Solomon, A.; Dursun, D.; Li, D. Q. Dry eye and delayed tear clearance: "a call to arms.". *Adv. Exp. Med. Biol.* **2002,** *506* (Pt B), 739-743.

56. Amos, L. A.; Schlieper, D. Microtubules and maps. *Adv. Protein Chem.* **2005,** *71*, 257-298.

57. Chernov, K. G.; Mechulam, A.; Popova, N. V.; Pastre, D.; Nadezhdina, E. S.; Skabkina, O. V.; Shanina, N. A.; Vasiliev, V. D.; Tarrade, A.; Melki, J.; Joshi, V.; Baconnais, S.; Toma, F.; Ovchinnikov, L. P.; Curmi, P. A. YB-1 promotes microtubule assembly in vitro through interaction with tubulin and microtubules. *BMC. Biochem.* **2008,** *9*, 23.

58. Chernov, K. G.; Curmi, P. A.; Hamon, L.; Mechulam, A.; Ovchinnikov, L. P.; Pastre, D. Atomic force microscopy reveals binding of mRNA to microtubules mediated by two major mRNP proteins YB-1 and PABP. *FEBS Lett.* **2008,** *582* (19), 2875-2881.

59. Kline-Smith, S. L.; Walczak, C. E. Mitotic spindle assembly and chromosome segregation: refocusing on microtubule dynamics. *Mol. Cell* **2004,** *15* (3), 317-327.

60. Gigant, B.; Wang, C.; Ravelli, R. B.; Roussi, F.; Steinmetz, M. O.; Curmi, P. A.; Sobel, A.; Knossow, M. Structural basis for the regulation of tubulin by vinblastine. *Nature* **2005,** *435* (7041), 519-522.

61. Ravelli, R. B.; Gigant, B.; Curmi, P. A.; Jourdain, I.; Lachkar, S.; Sobel, A.; Knossow, M. Insight into tubulin regulation from a complex with colchicine and a stathmin-like domain. *Nature* **2004,** *428* (6979), 198-202.

62. Nogales, E.; Wolf, S. G.; Khan, I. A.; Luduena, R. F.; Downing, K. H. Structure of tubulin at 6.5 A and location of the taxol-binding site. *Nature* **1995,** *375* (6530), 424-427.

63. Curmi, P. A.; Andersen, S. S.; Lachkar, S.; Gavet, O.; Karsenti, E.; Knossow, M.; Sobel, A. The stathmin/tubulin interaction in vitro. *J. Biol. Chem.* **1997,** *272* (40), 25029-25036.

64. Gigant, B.; Curmi, P. A.; Martin-Barbey, C.; Charbaut, E.; Lachkar, S.; Lebeau, L.; Siavoshian, S.; Sobel, A.; Knossow, M. The 4 A X-ray structure of a tubulin:stathmin-like domain complex. *Cell* **2000,** *102* (6), 809-816.

65. Jourdain, L.; Curmi, P.; Sobel, A.; Pantaloni, D.; Carlier, M. F. Stathmin: a tubulin-sequestering protein which forms a ternary T2S complex with two tubulin molecules. *Biochemistry* **1997,** *36* (36), 10817-10821.

66. Manna, T.; Thrower, D.; Miller, H. P.; Curmi, P.; Wilson, L. Stathmin strongly increases the minus end catastrophe frequency and induces rapid treadmilling of bovine brain microtubules at steady state in vitro. *J. Biol. Chem.* **2006,** *281* (4), 2071-2078.

67. Manna, T.; Thrower, D.; Miller, H. P.; Curmi, P.; Wilson, L. Stathmin strongly increases the minus end catastrophe frequency and induces rapid treadmilling of bovine brain microtubules at steady state in vitro. *J. Biol. Chem* **2006,** *281* (4), 2071-2078.

68. Rothermundt, M.; Peters, M.; Prehn, J. H.; Arolt, V. S100B in brain damage and neurodegeneration. *Microsc. Res. Tech.* **2003,** *60* (6), 614-632.

69. Mocellin, S.; Zavagno, G.; Nitti, D. The prognostic value of serum S100B in patients with cutaneous melanoma: a meta-analysis. *Int. J. Cancer* **2008,** *123* (10), 2370-2376.

70. Harpio, R.; Einarsson, R. S100 proteins as cancer biomarkers with focus on S100B in malignant melanoma. *Clin. Biochem.* **2004,** *37* (7), 512-518.

71. Hauschild, A.; Engel, G.; Brenner, W.; Glaser, R.; Monig, H.; Henze, E.; Christophers, E. S100B protein detection in serum is a significant prognostic factor in metastatic melanoma. *Oncology* **1999,** *56* (4), 338-344.

72. Schulte-Herbruggen, O.; Hortnagl, H.; Ponath, G.; Rothermundt, M.; Hellweg, R. Distinct regulation of brain-derived neurotrophic factor and noradrenaline in S100B knockout mice. *Neurosci. Lett.* **2008,** *442* (2), 100-103.

73. Michelakakis, H.; Kariyannis, C.; Moraitou, M.; Dimitriou, E.; Sarafidou, J.; Papassotiriou, I. Serum S100B levels in X-linked adrenoleukodystrophy and Gaucher disease. *J. Inherit. Metab Dis.* **2007,** *30* (5), 822.

74. Lin, J.; Yang, Q.; Yan, Z.; Markowitz, J.; Wilder, P. T.; Carrier, F.; Weber, D. J. Inhibiting S100B restores p53 levels in primary malignant melanoma cancer cells. *J. Biol. Chem.* **2004,** *279* (32), 34071-34077.

75. Lin, J.; Blake, M.; Tang, C.; Zimmer, D.; Rustandi, R. R.; Weber, D. J.; Carrier, F. Inhibition of p53 transcriptional activity by the S100B calcium-binding protein. *J. Biol. Chem.* **2001,** *276* (37), 35037-35041.

76. Wilder, P. T.; Rustandi, R. R.; Drohat, A. C.; Weber, D. J. S100B(betabeta) inhibits the protein kinase C-dependent phosphorylation of a peptide derived from p53 in a Ca2+-dependent manner. *Protein Sci.* **1998,** *7* (3), 794-798.

77. Pearlman, D. A.; Connelly, P. R. Determination of the differential effects of hydrogen bonding and water release on the binding of FK506 to native and Tyr82-->Phe82 FKBP-12 proteins using free energy simulations. *J. Mol. Biol.* **1995,** *248* (3), 696-717.

78. Inman, K. G.; Yang, R.; Rustandi, R. R.; Miller, K. E.; Baldisseri, D. M.; Weber, D. J. Solution NMR structure of S100B bound to the high-affinity target peptide TRTK-12. *J. Mol. Biol.* **2002,** *324* (5), 1003-1014.

79. McClintock, K. A.; Shaw, G. S. A novel S100 target conformation is revealed by the solution structure of the Ca2+-S100B-TRTK-12 complex. *J. Biol. Chem.* **2003,** *278* (8), 6251-6257.

80. McClintock, K. A.; Shaw, G. S. Assignment of 1H, 13C and 15N resonances of human Ca2+-S100B in complex with the TRTK-12 peptide. *J. Biomol. NMR* **2002,** *23* (3), 255-256.

81. Rustandi, R. R.; Baldisseri, D. M.; Weber, D. J. Structure of the negative regulatory domain of p53 bound to S100B(betabeta). *Nat. Struct. Biol.* **2000,** *7* (7), 570-574.

82. Grzesiek, S.; Bax, A.; Clore, G. M.; Gronenborn, A. M.; Hu, J. S.; Kaufman, J.; Palmer, I.; Stahl, S. J.; Wingfield, P. T. The solution structure of HIV-1 Nef reveals an unexpected fold and permits delineation of the binding surface for the SH3 domain of Hck tyrosine protein kinase. *Nat. Struct. Biol.* **1996,** *3* (4), 340-345.

83. Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. Discovering high-affinity ligands for proteins: SAR by NMR. *Science* **1996,** *274* (5292), 1531-1534.

84. Hayrabedyan, S.; Kyurkchiev, S.; Kehayov, I. FGF-1 and S100A13 possibly contribute to angiogenesis in endometriosis. *J. Reprod. Immunol.* **2005,** *67* (1-2), 87-101.

85. Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews* **2001,** *46* (1-3), 3-26.

86. Oprea, T. I.; Davis, A. M.; Teague, S. J.; Leeson, P. D. Is There a Difference between Leads and Drugs? A Historical Perspective. *J. Chem. Inf. Comput. Sci.* **2001,** *41* (5), 1308-1315.

87. Ghose, A. K.; Viswanadhan, V. N.; Wendoloski, J. J. A Knowledge-Based Approach in Designing Combinatorial or Medicinal Chemistry Libraries for Drug Discovery. 1. A Qualitative and Quantitative Characterization of Known Drug Databases. *J. Comb. Chem* **1999,** *1* (1), 55-68.

88. Dalvit, C.; Fogliatto, G.; Stewart, A.; Veronesi, M.; Stockman, B. WaterLOGSY as a method for primary NMR screening: practical aspects and range of applicability. *J. Biomol. NMR* **2001,** *21* (4), 349-359.

89. Arnesano, F.; Banci, L.; Bertini, I.; Fantoni, A.; Tenori, L.; Viezzoli, M. S. Structural interplay between calcium(II) and copper(II) binding to S100A13 protein. *Angew. Chem Int. Ed Engl.* **2005,** *44* (39), 6341-6344.

90. Oyama, Y.; Shishibori, T.; Yamashita, K.; Naya, T.; Nakagiri, S.; Maeta, H.; Kobayashi, R. Two distinct anti-allergic drugs, amlexanox and cromolyn, bind to the same kinds of calcium binding proteins, except calmodulin, in bovine lung extract. *Biochem. Biophys. Res. Commun.* **1997,** *240* (2), 341-347.

91. Shishibori, T.; Oyama, Y.; Matsushita, O.; Yamashita, K.; Furuichi, H.; Okabe, A.; Maeta, H.; Hata, Y.; Kobayashi, R. Three distinct anti-allergic drugs, amlexanox, cromolyn and tranilast, bind to S100A12 and S100A13 of the S100 protein family. *Biochem. J.* **1999,** *338 (Pt 3)*, 583-589.

92. Hopkins, A. L.; Groom, C. R.; Alex, A. Ligand efficiency: a useful metric for lead selection. *Drug Discov. Today* **2004,** *9* (10), 430-431.

93. Kuntz, I. D.; Chen, K.; Sharp, K. A.; Kollman, P. A. The maximal affinity of ligands. *Proc. Natl. Acad. Sci. U. S. A* **1999,** *96* (18), 9997-10002.

94. Carr, R. A.; Congreve, M.; Murray, C. W.; Rees, D. C. Fragment-based lead discovery: leads by design. *Drug Discov. Today* **2005,** *10* (14), 987-992.

95. bad-Zapatero, C.; Metz, J. T. Ligand efficiency indices as guideposts for drug discovery. *Drug Discov. Today* **2005,** *10* (7), 464-469.

96. Corbitt, C. A.; Lin, J.; Lindsey, M. L. Mechanisms to Inhibit Matrix Metalloproteinase Activity: Where are we in the Development of Clinically Relevant Inhibitors? *Recent Patents. Anticancer Drug Discov.* **2007,** *2* (2), 135-145.

97. Lovejoy, B.; Welch, A. R.; Carr, S.; Luong, C.; Broka, C.; Hendricks, R. T.; Campbell, J. A.; Walker, K. A.; Martin, R.; Van, W. H.; Browner, M. F. Crystal structures of MMP-1 and -13 reveal the structural basis for selectivity of collagenase inhibitors. *Nat. Struct. Biol.* **1999,** *6* (3), 217-221.

98. Bertini, I.; Calderone, V.; Cosenza, M.; Fragai, M.; Lee, Y. M.; Luchinat, C.; Mangani, S.; Terni, B.; Turano, P. Conformational variability of matrix metalloproteinases: beyond a single 3D structure. *Proc. Natl. Acad. Sci. U. S. A* **2005,** *102* (15), 5334-5339.

99. Alcaraz, L. A.; Banci, L.; Bertini, I.; Cantini, F.; Donaire, A.; Gonnelli, L. Matrix metalloproteinase-inhibitor interaction: the solution structure of the catalytic domain of human matrix metalloproteinase-3 with different inhibitors. *J. Biol. Inorg. Chem* **2007,** *12* (8), 1197-1206.

100. Pikul, S.; Dunham, K. M.; Almstead, N. G.; De, B.; Natchus, M. G.; Taiwo, Y. O.; Williams, L. E.; Hynd, B. A.; Hsieh, L. C.; Janusz, M. J.; Gu, F.; Mieling, G. E. Heterocycle-based MMP inhibitors with P2' substituents. *Bioorg. Med. Chem Lett.* **2001,** *11* (8), 1009-1013.

101. Olsson, T. S.; Williams, M. A.; Pitt, W. R.; Ladbury, J. E. The Thermodynamics of Protein-Ligand Interaction and Solvation: Insights for Ligand Design. *J. Mol. Biol.* **2008**.

102. Riley, K. E.; Merz, K. M., Jr. Role of solvation in the energy stabilization inside the hydrophobic core of the protein rubredoxin. *J. Phys. Chem B* **2006,** *110* (32), 15650-15653.

103. Vajda, S.; Weng, Z.; Rosenfeld, R.; DeLisi, C. Effect of conformational flexibility and solvation on receptor-ligand binding free energies. *Biochemistry* **1994,** *33* (47), 13977-13988.

104. Aureli, L.; Gioia, M.; Cerbara, I.; Monaco, S.; Fasciglione, G. F.; Marini, S.; Ascenzi, P.; Topai, A.; Coletta, M. Structural bases for substrate and inhibitor recognition by matrix metalloproteinases. *Curr. Med. Chem* **2008,** *15* (22), 2192-2222.

105. McGaughey, G. B.; Gagne, M.; Rappe, A. K. pi-Stacking interactions. Alive and well in proteins. *J. Biol. Chem* **1998,** *273* (25), 15458-15463.

106. Wang, W.; Hobza, P. Theoretical study on the complexes of benzene with isoelectronic nitrogen-containing heterocycles. *Chemphyschem.* **2008,** *9* (7), 1003-1009.

107. Carlos Alemán; Sergio E.Galembeck Solvation of chromone using combined Discrete/SCRF models. *Chemical Physics* **1998,** *232* (1-2), 151-159.

108. Reddy, M. R.; Singh, U. C.; Erion, M. D. Ab initio quantum mechanics-based free energy perturbation method for calculating relative solvation free energies. *J. Comput. Chem.* **2007,** *28* (2), 491-494.

109. Galano, A.; varez-Idaboy, J. R. A new approach to counterpoise correction to BSSE. *J. Comput. Chem* **2006,** *27* (11), 1203-1210.

110. Seifert, M. H.; Kraus, J.; Kramer, B. Virtual high-throughput screening of molecular databases. *Curr. Opin. Drug Discov. Devel.* **2007,** *10* (3), 298-307.

111. Fogolari, F.; Brigo, A.; Molinari, H. The Poisson-Boltzmann equation for biomolecular electrostatics: a tool for structural biology. *J. Mol. Recognit.* **2002,** *15* (6), 377-392.

112. Hou, Y. W.; Chan, M. H.; Hsu, H. R.; Liu, B. R.; Chen, C. P.; Chen, H. H.; Lee, H. J. Transdermal delivery of proteins mediated by non-covalently associated arginine-rich intracellular delivery peptides. *Exp. Dermatol.* **2007,** *16* (12), 999-1006.

113. Abes, R.; Arzumanov, A.; Moulton, H.; Abes, S.; Ivanova, G.; Gait, M. J.; Iversen, P.; Lebleu, B. Arginine-rich cell penetrating peptides: design, structure-activity, and applications to alter pre-mRNA splicing by steric-block oligonucleotides. *J. Pept. Sci.* **2008,** *14* (4), 455-460.

114. Wang, Y. H.; Hou, Y. W.; Lee, H. J. An intracellular delivery method for siRNA by an arginine-rich peptide. *J. Biochem. Biophys. Methods* **2007,** *70* (4), 579-586.

115. Pflugfelder, S. C.; Farley, W.; Luo, L.; Chen, L. Z.; de Paiva, C. S.; Olmos, L. C.; Li, D. Q.; Fini, M. E. Matrix metalloproteinase-9 knockout confers resistance to corneal epithelial barrier disruption in experimental dry eye. *Am. J. Pathol.* **2005,** *166* (1), 61-71.

116. Hulkkonen, J.; Pertovaara, M.; Antonen, J.; Pasternack, A.; Hurme, M.; Pollanen, P.; Lehtimaki, T. Matrix metalloproteinase 9 (MMP-9) gene polymorphism and MMP-9 plasma levels in primary Sjogren's syndrome. *Rheumatology. (Oxford)* **2004,** *43* (12), 1476-1479.

117. Niederkorn, J. Y.; Stern, M. E.; Pflugfelder, S. C.; de Paiva, C. S.; Corrales, R. M.; Gao, J.; Siemasko, K. Desiccating stress induces T cell-mediated Sjogren's Syndrome-like lacrimal keratoconjunctivitis. *J. Immunol.* **2006,** *176* (7), 3950-3957.

118. Eagle Vision-Yankelovich Partner Survey. 1997.

119. Park, M. T.; Cha, H. J.; Jeong, J. W.; Kim, S. I.; Chung, H. Y.; Kim, N. D.; Kim, O. H.; Kim, K. W. Glucocorticoid receptor-induced down-regulation of MMP-9 by ginseng components, PD and PT contributes to inhibition of the invasive capacity of HT1080 human fibrosarcoma cells. *Mol. Cells* **1999,** *9* (5), 476-483.

120. Dursun, D.; Wang, M.; Monroy, D.; Li, D. Q.; Lokeshwar, B. L.; Stern, M. E.; Pflugfelder, S. C. A mouse model of keratoconjunctivitis sicca. *Invest Ophthalmol. Vis. Sci.* **2002,** *43* (3), 632-638.

121. CALMETTES; DEODATTI; AMALRIC [The fluoresceine test in eye diseases.]. *Bull. Soc. Ophtalmol. Fr.* **1954,** *6*, 597-601.

122. Franck, C.; Palmvang, I. B. Break-up time and lissamine green epithelial damage in 'office eye syndrome'. Six-month and one-year follow-up investigations. *Acta Ophthalmol. (Copenh)* **1993,** *71* (1), 62-64.

Appendices

Appendix A

Original Paper

Theoretical study on binding of S100B protein

Artur Gieldon[1, 2], Mattia Mori[1] and Rebecca Del Conte[1]

(1) Protera S. r. l., Viale delle Idee, 22, 50019 Sesto Fiorentino (Fi), Italy

(2) Institute of Biochemistry, Johan Wolfgang Goethe University, Max-von-Laue-Str. 7, 60438 Frankfurt a.M., Germany

Received: 27 March 2007 **Accepted:** 29 June 2007 **Published online:** 23 August 2007

Abstract S100B protein is one of the factors involved in the down-regulation of tumor suppressor protein p53, a transcription activator that signals for cycle arrest and apoptosis. As the inactivation of normal p53 functions is found in over half of human cancers, restoration of normal p53 functions through the destruction or prevention of S100B - p53 complexes represents a possible approach for the development of anti-cancer drugs. The aim of this work was to propose the S100B binding interface through an examination of the literature and use of molecular modeling (MM) techniques with AutoDock program and the AMBER force field. We propose two residues in the S100B binding pocket (Val56, Phe76) and two residues on the protein surface (Val52, Ala83) are essential for ligand binding. The data presented here indicate that interactions with these four residues are necessary for a reduction in the incidence of the S100B - p53 complex. Additionally, we have tried to explain a mechanism for the action of pentamidine, the best-known S100B ligand, and have proposed two S100B - pentamidine structures. The results presented here may be useful for the efficient design of new S100B ligands.

Keywords AMBER - AutoDock - Pentamidine - p53 - S100B - TRTK-12

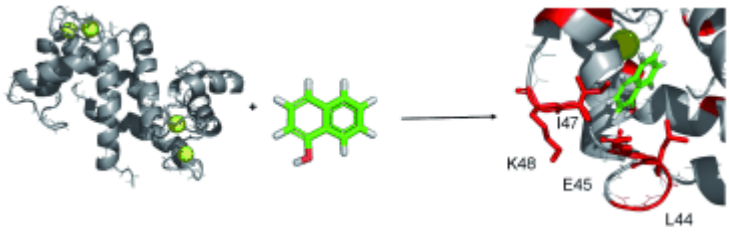
Appendix B

Fragment Docking to S100 Proteins Reveals a Wide Diversity of Weak Interaction Sites (p 1648-1654)

Yvonne Arendt, Anusarka Bhaumik, Rebecca Del Conte, Claudio Luchinat, Mattia Mori, Marco Porcu

Published Online: Aug 20 2007 1:52AM

DOI: 10.1002/cmdc.200700096



NMR screening of two S100 proteins which are potential drug targets has been performed with a fragment library. A relatively large variety of interaction regions for various ligands for the two S100 proteins were identified even applying computer-aided drug design. Our results show that they have only few ligands in common, suggesting that selective leads could be developed starting from the different hits identified in the present work.

Appendix C

In preparation for submission

"Computational approach to the optimization of the classic aryl-sulphonamide scaffold for MMPs inhibition"

(Tentative Title)

Abstract

In this work we present a new computational approach to the design of new potent inhibitors of matrixmetalloproteinases (MMPs) by optimizing validated scaffolds. Starting from the analysis of X-Ray adducts of MMP-12 with classic aryl-sulphonamide inhibitors, computational tools have been applied in order to estimate the advantage in binding affinity, determined by two selected modifications: i) linkage of the sulphonamide group and P1' aromatic ring with a carbon linker and ii) D-proline substitution instead of glycine residue in the amino acid portion. Theoretical results have been experimentally confirmed by fluorimetric assays and X-Ray crystallography. Since the intermolecular interaction energy calculated by a standard docking program didn't provide accurate explanations for the observed binding improvements, *ab initio* PCM and MP2 calculations have been carried out to evaluate: i) the proline desolvation contribution, and ii) the contribution of π-stacking interactions between the inhibitors and aromatic residues in the MMPs active site. Here we demonstrate that the introduction of the linker and D-proline residues at the same time in the reference scaffold, generates ligands with improved potency.

Appendix D

Preliminary draft.

In preparation for submission

In Vivo Study on the Efficacy of an MMP-9 Inhibitors in the Therapy of Dry Eye Syndrome

(Tentative Title)