Scuola Dottorale in Biologia
Sezione “Biologia applicata alla salute dell’uomo”
Ciclo di Dottorato XXVIII

Virtual screening for protein function annotation and drug discovery

Candidato:
Dott.ssa Elena Di Muzio

Docente guida:
Prof. Fabio Polticelli

Coordinatore:
Prof. Paolo Visca

ANNO ACCADEMICO 2014/2015

SUMMARY

SUPPLEMENTARY MATERIALS

Reprints of papers published

TITOLO DELLA TESI:

Screening virtuale finalizzato all'annotazione funzionale di proteine e alla scoperta di farmaci

RIASSUNTO

Il docking molecolare è una procedura computazionale che tenta di predire in maniera efficiente la modalità di legame non covalente tra due macromolecole o, più frequentemente, tra una macromolecola (recettore) e una piccola molecola (ligando), partendo dalle rispettive strutture non complessate. Questo approccio è ampiamente usato nello *screening* virtuale, anche detto *screening in silico*, divenuto sempre più popolare nella ricerca farmaceutica per l'identificazione di possibili candidati, i cosiddetti *lead compounds*, o composti guida. L'obiettivo di base dello *screening* virtuale è la riduzione del vasto spazio chimico di piccole molecole organiche a un numero gestibile di composti che hanno la più alta probabilità di condurre all'identificazione di un candidato farmaco. Lo *screening* virtuale applicato alla scoperta di nuovi farmaci prevede l'uso del docking molecolare, quindi il posizionamento computazionale dei composti nel sito di interesse di una proteina, l'assegnazione di un punteggio e la classificazione dei composti in base al punteggio assegnato. In questo contesto, le banche dati di composti chimici e *drug-like* sono molto utili perché contengono, nella maggior parte dei casi, i composti in formati pronti per essere utilizzati nella procedura di docking (struttura tridimensionale).

Il presente lavoro è stato finalizzato allo studio del protocollo di docking, ed in particolare del docking proteina-ligando, e alla sua applicazione a sistemi biologici rilevanti da un punto di vista biomedico, come ad esempio l'albumina sierica umana (HSA), con lo scopo di fornire evidenze a supporto degli studi sperimentali. A tal fine, è stato utilizzato AutoDock Vina, un programma *open-source* per l'identificazione di potenziali farmaci, per il docking molecolare e per lo *screening* virtuale, che offre capacità *multi-core* (ossia la possibilità di utilizzare contemporaneamente tutti i processori presenti nella macchina utilizzata), prestazioni elevate e una relativa facilità d'uso. Sostanzialmente sono state eseguite simulazioni di docking sull'albumina per investigare l'interazione tra questa proteina e alcuni ligandi, nello specifico il farmaco antileucemico imatinib, l'acido retinoico e il retinolo. I risultati ottenuti con queste simulazioni sono in accordo con i dati sperimentali e inoltre hanno permesso di identificare il sito preferenziale di legame dei composti oggetto di studio e di analizzare l'interazione a livello atomico.

Simultaneamente alle analisi di docking, è stata sviluppata una piattaforma di docking e *screening* virtuale automatizzato, successivamente testata su strutture sperimentali di potenziali

bersagli proteici di *Pseudomonas aeruginosa* per l'identificazione di nuovi farmaci con potenziale attività antibatterica. E' noto infatti che la resistenza batterica è una minaccia crescente e ancora pochi nuovi antibiotici attivi contro i batteri multiresistenti sono stati identificati. La combinazione di meccanismi di resistenza batterici e l'uso irrazionale e sconsiderato di antibiotici ha portato ad una situazione allarmante in cui alcune infezioni non hanno alcuna cura. In questo scenario, identificare un nuovo bersaglio batterico, oppure identificare nuovi composti in grado di ridurre o inibire la virulenza batterica, può tradursi in risultati significativi.

Dato l'elevato numero di composti da esaminare (1466 composti approvati dalla *Food and Drug Administration* (FDA) per uso farmacologico nell'uomo, disponibili nella banca dati DrugBank), utilizzando il linguaggio di programmazione Python, è stata sviluppata una procedura automatizzata che permette di calcolare automaticamente i parametri di input per la successiva fase di docking, di classificare i composti secondo la rispettiva energia di legame predetta e di visualizzare il relativo sito di legame sulla proteina di interesse. Tale procedura è stata incorporata in un'applicazione di facile utilizzo denominata "DockingApp". DockingApp permette di effettuare le suddette analisi anche ad utenti privi di alcuna esperienza in bioinformatica. Infatti l'applicazione è dotata di un'interfaccia grafica (*graphic user interface* o GUI) per guidare l'utente attraverso tutti i passaggi necessari per eseguire le simulazioni. In particolare, l'input dell'applicazione è una proteina con funzione non nota oppure un obiettivo (*target*) di interesse farmacologico. L'utente deve solamente indicare il file relativo alla struttura tridimensionale del *target* (in formato PDB); l'input viene poi elaborato per ottenere una struttura con tutte le informazioni necessarie per l'algoritmo di docking (atomi di idrogeno, cariche atomiche). Il nucleo dell'algoritmo è una simulazione automatica di docking proteina-ligando o di *virtual screening* contro una libreria di composti, nello specifico contro il sottoinsieme di composti presenti in DrugBank approvati dalla FDA, che viene fornito all'utente nella cartella di download e che viene impostato direttamente come database di input nell'interfaccia grafica. Tuttavia, l'utente può personalizzare la libreria, o utilizzarne una propria, indicando semplicemente la cartella relativa nella GUI. L'output è un elenco (memorizzato anche in un file di testo) contenente i migliori risultati derivanti dallo *screening*, sulla base delle affinità di legame calcolate. Questo file contiene anche alcuni dettagli, come i residui del recettore in contatto con il ligando a soglie di distanza diverse e informazioni circa i composti analizzati. Ovviamente, in questo file vengono memorizzati anche i complessi tra la proteina input e i migliori composti con le rispettive coordinate.

Per eseguire l'applicazione l'utente deve installare MGLtools (ADT) e Autodock Vina. In particolare, MGLtools è utile nella preparazione di recettore e ligando. Infatti i relativi file PDB devono essere convertiti in formato PDBQT, necessario per eseguire una simulazione di docking con Vina. Attraverso l'interfaccia grafica utente ADT è possibile realizzare questa conversione, nonché impostare lo spazio di ricerca desiderata. Tuttavia, per un utente senza (o con limitate) competenze bioinformatiche questi passaggi potrebbero risultare impegnativi. Per superare questa limitazione in questo lavoro sono stati utilizzati alcuni scripts Python forniti dagli sviluppatori di Vina, che consentono la preparazione automatica dei file PDBQT. Per quanto riguarda la definizione dello spazio di ricerca (griglia), cioè della porzione del recettore in cui l'algoritmo di docking cerca le migliori conformazioni del ligando, è stata sviluppata una funzione specifica che consente la regolazione automatica di tale spazio, sia locale, cioè relativo ad un numero limitato di residui, che esteso, per coprire l'intera struttura del recettore.

Sebbene alcuni test siano attualmente in corso, la versione di *screening* virtuale del programma è stata già testata su 5 bersagli proteici di *Pseudomonas aeruginosa* che sono coinvolti nella virulenza batterica o sono essenziali per la crescita batterica (PqsE, PqsR, LasR, WspR, MurG), con l'intento di individuare composti approvati dalla FDA che potrebbero avere attività antibatterica. I risultati ottenuti sono ancora in fase di analisi ma sono promettenti perché nella maggior parte dei casi le migliori soluzioni, in termini di energia di legame, relative a ciascun composto analizzato, sono posizionate nel sito di legame del bersaglio con un'affinità di legame migliore se confrontata con il ligando co-cristallizzato.

Riassumendo, con questa applicazione si possono raggiungere due obiettivi: in primo luogo, a partire dalle informazioni derivate dai risultati di docking si può cercare di predire la funzione delle proteine. Ad esempio, per enzimi a funzione non nota, la predizione di probabili substrati basata sulla complementarità strutturale diviene interessante quando il target non ha correlazione strutturale con ortologhi a funzione nota, rendendo tale predizione non realizzabile altrimenti. Da questo punto di vista, l'identificazione di alcuni composti che si legano al bersaglio con la più alta probabilità, cioè l'identificazione di potenziali candidati substrati/inibitori, potrebbe portare all'annotazione della funzione proteica.

In secondo luogo, e sicuramente scopo centrale del presente lavoro, l'applicazione dello *screening* virtuale implementato in questa procedura può condurre alla scoperta e/o al riposizionamento (repurposing) di farmaci, attività che si propone di individuare nuove indicazioni terapeutiche per farmaci già esistenti. Infatti, lo *screening* automatizzato contro una libreria di composti approvati dalla FDA riduce il costo dello *screening* sperimentale iniziale e accelera la scoperta di composti guida.

ABSTRACT

Molecular docking is a computational procedure that attempts to efficiently predict non-covalent binding of macromolecules or, more frequently, of a macromolecule (receptor) and a small molecule (ligand), starting with their unbound structures. This approach is widely used in virtual screening, an *in silico* procedure that has become increasingly popular in the pharmaceutical research for lead identification. The basic goal of the virtual screening is the reduction of the massive virtual chemical space of small organic molecules, to be screened against a specific target protein, to a manageable number of compounds that have the highest chance to lead to a drug candidate. Virtual screening applied to the discovery of new drugs involves docking, computational fitting of structures of compounds to the active site of a protein and scoring and ranking of each compound. In this context, databases of chemical and drug-like compounds are very helpful because, in most cases, contain many compounds in ready-to-dock formats (three-dimensional structures).

The present work was aimed at studying molecular docking protocols, and specifically protein-ligand docking, and its application to biological systems of biomedical relevance, such as human serum albumin (HSA), to provide complementary and supporting evidence to experimental studies. For this purpose was used the program AutoDock Vina, an open-source program for drug discovery, molecular docking and virtual screening, offering multi-core capability (or, in other words, the possibility to use at the same time all the processors of the computer used), high performance, enhanced accuracy, and ease of use. Docking simulations on HSA were carried out to investigate the interaction of this protein with the drug imatinib and with retinoids. Results obtained in these simulations are in agreement with available experimental data and allowed both to identify the preferential binding sites of the ligand/drug, and to investigate the interaction at atomic level.

In addition, a user-friendly procedure was developed, which allows to perform automatic molecular docking simulations and virtual screening analyses. This method was tested on experimental structures of potential *Pseudomonas aeruginosa* protein targets for the identification of novel lead compounds with potential antibacterial activity. In fact, bacterial resistance is a growing threat and, at present, very few new antibiotics active against multi-resistant bacteria are available. A combination of falling profits, regulatory mechanisms and irrational and injudicious use of antibiotics has led to an alarming situation where some infections have no cure. Finding novel targets, possibly with no cross-resistance, and/or

identification of novel compounds that could reduce or inhibit bacterial virulence, has elements of uncertainty but successful outcomes may translate into significant results.

Given the huge number of compounds to be screened (1466 FDA-approved compounds from DrugBank), an automated procedure for docking studies was developed, using the Python programming language, which allows to calculate the input parameters to perform the docking phase, to rank the compounds according to their predicted binding energy and to visualize the predicted drug binding site on the protein of interest. This procedure has been incorporated in an application named "DockingApp". DockingApp allows to perform this kind of analyses also to users without any expertise in bioinformatics. In fact, a Java graphical user interface (GUI) was developed to guide the user through all the steps required to perform the docking simulations and analyze the results. In particular, the input of the virtual screening/docking application is a protein with still unknown function or a target of interest from a pharmaceutical viewpoint. The user needs only the three-dimensional structure of the target (*i.e*. the PDB file); the input is then processed to obtain a structure with all the information needed by the docking algorithm (hydrogen atoms, atomic charges). The core of the algorithm is an automatic protein-ligand docking procedure or a virtual screening analysis against a library of compounds, in particular the subset of FDA-approved compounds retrieved from DrugBank, provided in the download folder. In addition, the user can customize the library, or use his own, simply indicating the relative folder in the GUI. The output is a list (stored in a text file) containing the best hits resulting from the screening, based on the binding affinity calculated. This file also contains some details, such as residues in contact with the ligand at different distance thresholds and details about the compounds. Obviously, in the output are stored also the complexes between the input protein and the best hit compounds with their respective atomic coordinates. To run the application the user must install MGLtools (ADT) and download Autodock Vina. In particular, MGLtools is quite useful in receptor and ligand preparation. In fact, PDB files of the receptor and the ligand must be converted in PDBQT file format, required by Vina to run docking simulations. Through the ADT graphical user interface, it is possible to obtain this conversion as well as to set the desired search space; however, for a user without any (or low) expertise in bioinformatics these steps could be difficult. To overcome this limitation, some Python scripts provided by Vina developers were used in this work, allowing the automatic preparation of PDBQT files. Regarding the search space (docking grid), that is the portion of the receptor that Vina will explore to place the ligand, a specific function was developed to allow to automatically set the search space, either local, relative to a limited number of residues, or extended to cover the entire structure of the receptor.

Although tests are still in progress, the virtual screening version of the application was already tested on five protein targets from *Pseudomonas aeruginosa* that are involved in bacterial virulence or are essential for bacterial growth (PqsE, PqsR, LasR, WspR, MurG), with the aim to identify FDA-approved compounds with potential antibacterial activity.

The results obtained are still under investigation but they are promising because in most cases the best poses of each compound are placed in the target's binding site with a better affinity if compared with the co-crystallized ligand.

Summarizing, with this application it should be possible to achieve two objectives.

First, starting from the information derived from docking results it may be possible to predict protein function. For enzymes of unknown function, substrate prediction based on structural complementarity becomes attractive when the target enzyme has little relationship to orthologues of known activity, making inference unreliable. From this viewpoint, the identification of few compounds that bind the target with the highest probability, that is the identification of potential substrate/inhibitor candidates, could lead to protein function annotation.

Then, drug discovery and/or repositioning, that is the main goal of the present project, could be achieved. In fact, the automated virtual screening against a library of FDA-approved compounds reduces the cost of the initial experimental screening and accelerates lead compounds discovery. Moreover, this analysis can lead to drug repositioning, which aims to identify new therapeutic indications for existing drugs.

1. INTRODUCTION

1.1 Protein-ligand interactions and drug discovery

Molecular recognition and binding performed by proteins are the background of all biochemical processes in a living cell. In particular, the identification of interactions between drugs and target proteins is a key area in drug discovery. Such interactions with ligands can modulate the function of many classes of pharmaceutically useful protein targets (Yamanishi *et al*., 2008).

Enzyme-substrate, drug-protein, drug-nucleic acid, protein-nucleic acid, and protein-protein interactions play important roles in many biological processes such as signal transduction, cell regulation, and macromolecular complexes assembly. Therefore, determination of the binding mode and affinity between the interacting molecules is crucial to understand the interaction mechanisms and to design therapeutic interventions (Huang and Zou, 2010).

In recent years, the genome-wide detection of compound-protein interactions has become a key issue in chemical genomics research, and gained great importance in drug design. Indeed, it could lead to the successful identifications of new drug leads and therapeutic targets for known diseases. Although biological assays are developing rapidly and becoming more accessible, experimental determination of compound–protein interactions or potential drug–target interactions remains very challenging. For these reasons, effective *in silico* prediction methods need to be developed, with the aim to reduce the time of the initial steps of the drug design process (Figure 1.1).

In silico pharmacology (also known as computational therapeutics or computational pharmacology) includes the development of techniques for using software to retrieve, analyze and integrate biological and medical data from many diverse sources. More specifically, it defines the use of this information in the creation of computational models or simulations that can be used to make predictions, suggest hypotheses, and ultimately provide discoveries or advances in medicine and therapeutics (Ekins *et al*., 2007).

In a drug discovery process, computer-aided drug discovery/design (CADD) is usually used for the following three major purposes. 1) Filter large compound libraries into smaller sets of predicted active compounds that can be tested experimentally. 2) Guide the optimization of lead compounds, whether to increase their affinity or optimize drug metabolism and pharmacokinetics (DMPK). 3) Design novel compounds, either by "growing" starting

molecules one functional group at a time or by piecing together fragments into novel chemotypes. (Figure 1.2) (Aparoy *et al*., 2012, Sliwoski *et al*., 2014)

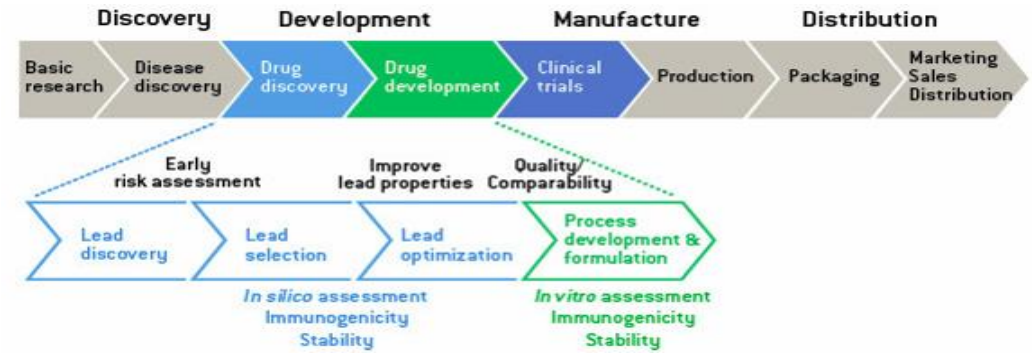


Figure 1.1: **Drug design process**. Bioinformatics tools are used to reduce the time of the initial steps of the drug design process. The results obtained can provide complementary and supporting evidence to experimental studies.

.

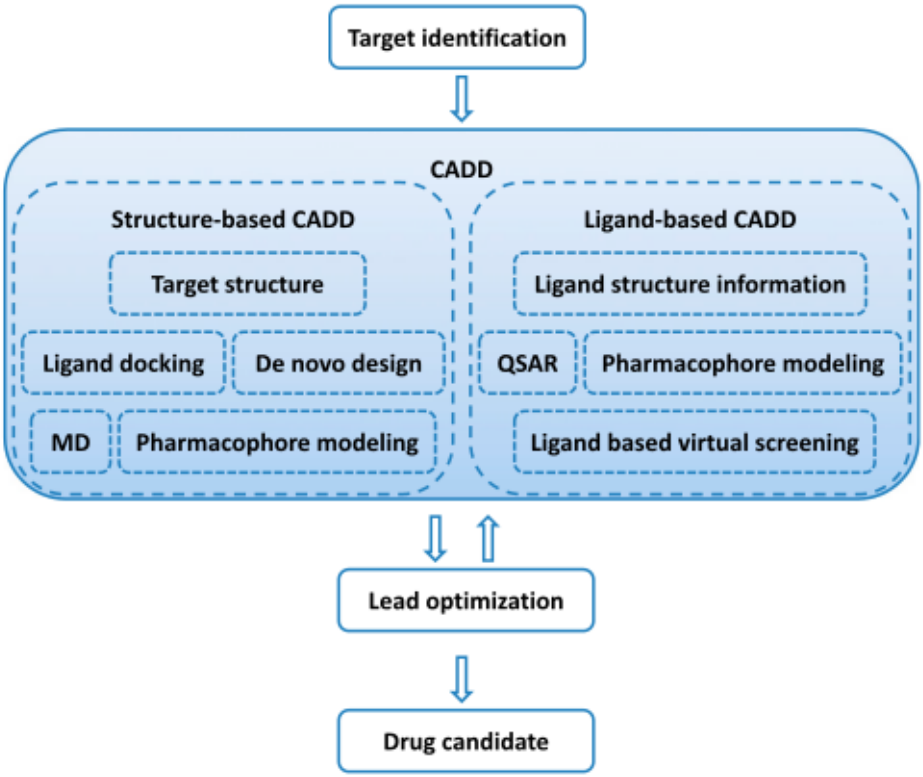


Figure 1.2: **CADD in drug discovery/design pipeline.** A therapeutic target is identified against which a drug has to be developed. Depending on the availability of structure information, a structure-based approach or a ligand-based approach is used (Sliwoski *et al*., 2014).

Computer-aided drug discovery has already been used in the discovery of compounds that passed clinical trials and became novel therapeutics in the treatment of a variety of diseases.

Some examples of approved drugs, whose discovery is in large part due to computer-aided drug discovery methods, include: carbonic anhydrase inhibitor dorzolamide, approved in 1995 (Vijayakrishnan, 2009); the angiotensin-converting enzyme (ACE) inhibitor captopril, approved in 1981 as an antihypertensive drug (Talele *et al*., 2010); the three therapeutics for the treatment of human immunodeficiency virus (HIV) saquinavir (approved in 1995), ritonavir, and indinavir (both approved in 1996) (Van Drie, 2007).

Computer-aided drug discovery methods are broadly classified as either structure-based or ligand-based methods. Structure-based methods are in principle analogous to high-throughput screening (HTS) in that both the target and the ligand structure information is mandatory. Structure-based approaches include ligand docking, pharmacophore, and ligand design methods. Ligand-based methods use only ligand information for predicting activity depending on its similarity/dissimilarity to previously known active ligands (Figure 1.3) (Aparoy *et al*., 2012, Sliwoski *et al*., 2014).

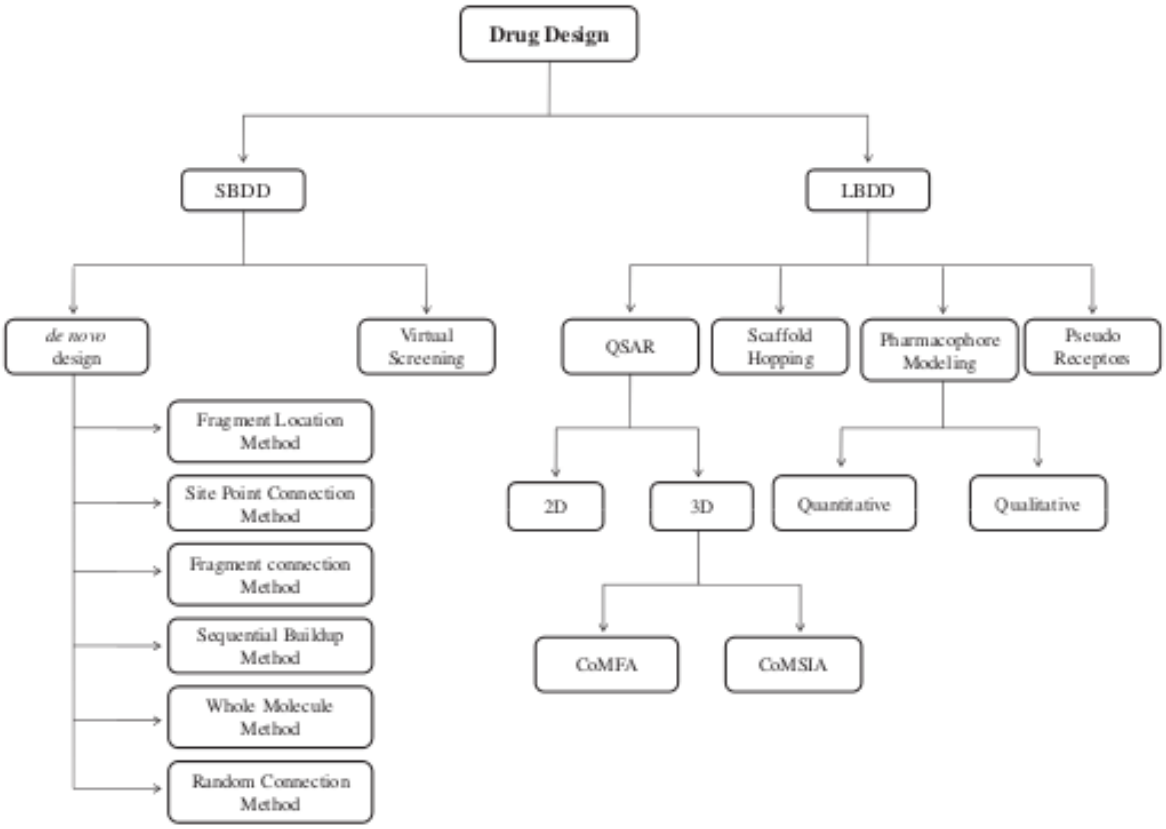


Figure 1.3: **Basic principles and types of drug design.** The term SBDD indicates various approaches wherein the structural knowledge of the drug target is exploited and LBDD indicates the strategies wherein the information from existing ligands of a drug target is utilized. QSAR, Quantitative Structure Activity Relationship; CoMFA, Comparative Molecular Field Analysis; CoMSIA, Comparative Molecular Similarity Index Analysis (Aparoy *et al*., 2012).

1.2 Virtual screening

Virtual screening is one of the commonly used approaches in the lead identification step and is seen as a complementary approach to experimental high throughput screening to improve the speed and efficiency of the drug discovery and development process. Virtual screening is generally regarded as the top computer-aided drug discovery tool to screen large libraries of chemical structures and reduce them to a key set of likely drug candidates for a specific protein target. It falls into the category of structure-based methods and involves molecular docking (a process to predict the binding mode) of each ligand to the binding site of the target, and scoring (a process to predict the binding affinity). The compounds in the databases screened are ranked to select and experimentally test a small subset of them for biological activity on a given receptor (Aparoy *et al*., 2012). A typical workflow of a receptor-based virtual screening is illustrated in Figure 1.4.

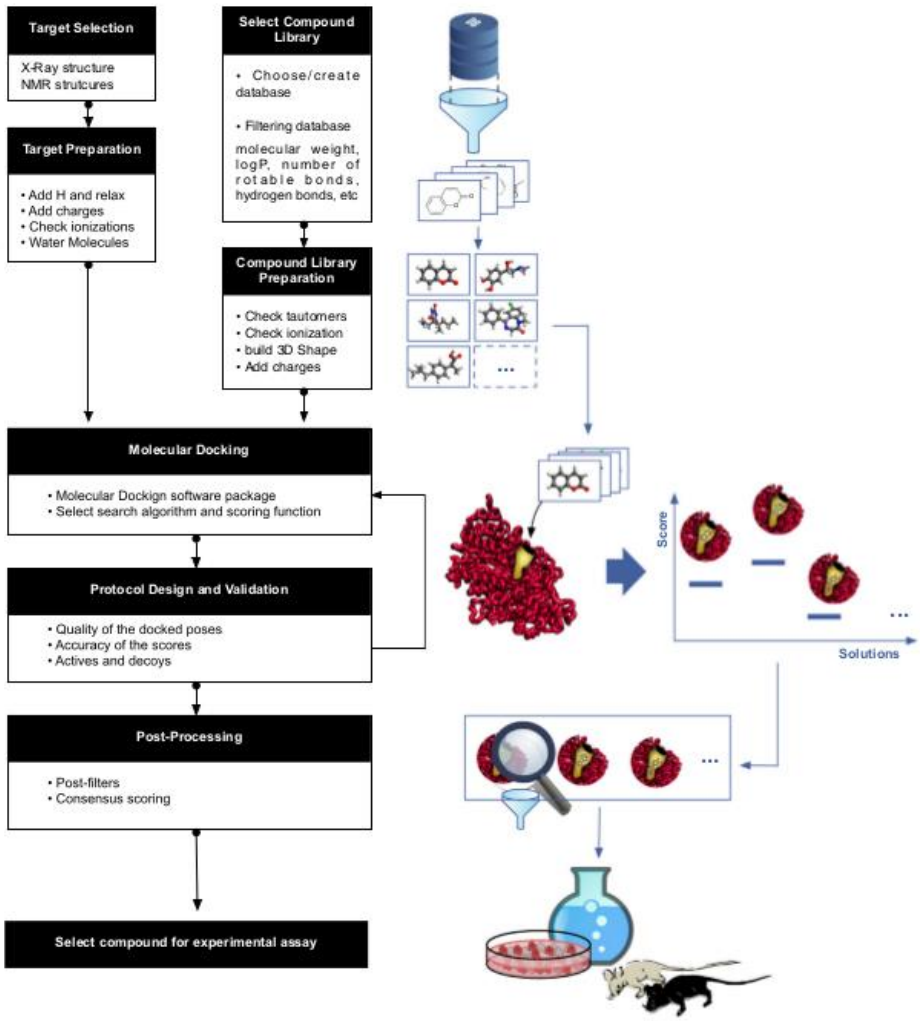


Figure 1.4: **General workflow of a receptor-based virtual screening.** The typical workflow consists of a preparation phase for the database and the target, followed by a molecular docking phase, and concluded with the post-processing and compounds selection phases (Cerqueira *et al*., 2015).

1.2.1 Molecular docking

Molecular docking is a computational procedure that tries to predict non-covalent binding of macromolecules or, more frequently, of a macromolecule (receptor) and a small molecule (ligand) efficiently, starting with their unbound structures. The goal is to predict the bound conformations and the binding affinity (Trott and Olson, 2009). Pioneered during the early 1980s by Kuntz and colleagues (Kuntz *et al*., 1982), it remains a field of vigorous research, having become a useful tool in drug discovery efforts, and a primary component in many drug discovery programs (Sousa *et al*., 2006). Some of the most widely used docking programs in high-throughput screening, listed in Figure 1.5, are DOCK (Ewing *et al*., 2001), FlexX (Rarey *et al*, 1996), GOLD (Jones *et al*., 1997), Glide (Halgren *et al*., 2004), and AutoDock (Morris *et al*., 2009) (Table 1.1). These are among the most cited protein-ligand docking programs, as shown in Figure 1.6 (Sousa *et al*., 2015).

Brief overview of some of the widely used docking software

Software	Platform	Fees	Applications
DOCK	Linux, Unix, Windows, Mac OS X	No (academics), yes (for-profits)	Protein–ligand docking, virtual screening, protein–protein docking, protein–nucleic acid docking, structure-based design
AutoDock	Linux, Unix, Mac OS X, Windows	No (academics), yes (for-profits)	Protein–ligand docking, virtual screening, combinatorial library design, protein–protein docking, structure-based design
FlexX	Linux, Unix, Sun Solaris, HP-UX, Windows	Yes	Protein–ligand docking, binding mode prediction, virtual screening, structure-based design
GOLD	Linux, Unix, Windows, Sun Solaris IBM, AIX	Yes	Dock ligands into protein, DNA and RNA structures, virtual screening, structure-based design
Glide	Linux, Unix, IBM, AIX	Yes	Protein–ligand docking, binding-mode prediction, virtual screening, structure-based design
Fred	Linux, Unix, Windows, OS X, AIX, HP-UX, Solaris, Tru64UNIX	Yes	Protein–ligand docking, binding-mode prediction, virtual screening
ICM	Linux, Unix, Windows, Mac OS X	Yes	Protein–ligand docking, protein–protein docking, virtual screening, structure-based design

Figure 1.5: **Overview of some of the most widely used docking software** (Ghosh *et al*., 2006).

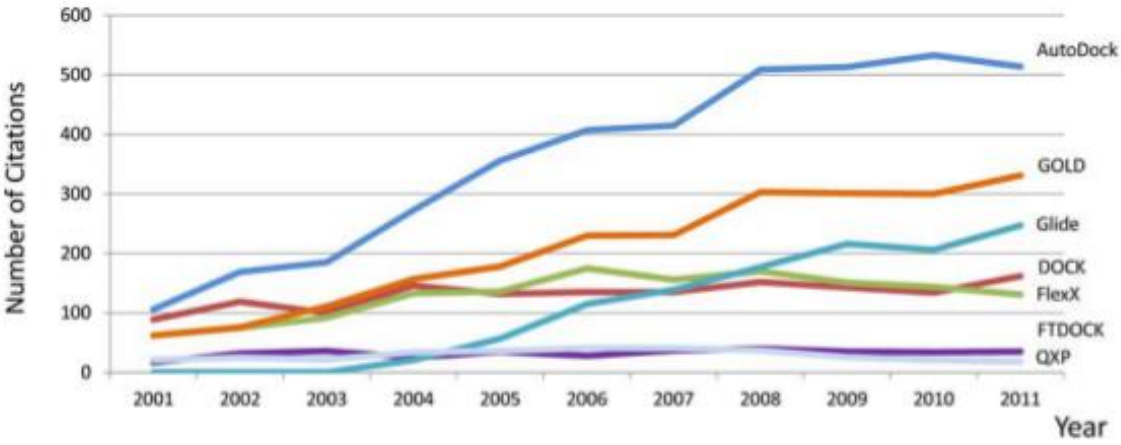


Figure 1.6: **Evolution of the number of citations per year for the seven most cited protein-ligand docking programs over the period 2001-2011** (Sousa *et al*., 2015).

An important type of molecular docking is protein-ligand docking because of its biomedical applications in modern structure-based drug design. In fact, the prediction of binding of small molecules to proteins is of particular practical importance because it is used to screen virtual libraries of drug-like molecules to obtain leads for further drug development.

Generally speaking, molecular docking comprises the process of generating a model of a complex based on the known three-dimensional structures of its components, free or complexed with other species. In terms of protein–ligand docking methods, the docking problem can be rationalized as the search for the precise ligand conformations and orientations (commonly referred as posing) within a given targeted protein when the structure of the protein is known or can be predicted. The binding affinity prediction problem addresses the question of how well the ligands bind to the protein (scoring) (Sousa *et al*., 2006).

The success of a docking algorithm in predicting a ligand binding pose is normally measured in terms of the root-mean-square deviation (RMSD) between the experimentally observed heavy-atom positions of the ligands and the one(s) predicted by the algorithm.

Given two structures, *a* and *b*, of an identical molecule, RMSD is defined as follows:

$$\mathrm{RMSD}_{ab} = \max(\mathrm{RMSD}'_{ab}, \mathrm{RMSD}'_{ba}),$$

where

$$\mathrm{RMSD}'_{ab} = \sqrt{\frac{1}{N}\sum_i \min_j r_{ij}^2},$$

and the summation is over all N heavy atoms in structure a, the mininum is over all atoms in structure b with the same element type as atom i in structure a.

Docking methods can be classified as rigid-body docking and flexible docking applications depending on the degree to which they consider ligand and protein flexibility during the docking process (Halperin *et al*., 2002). Rigid body docking methods consider only static geometric/physicochemical complementarities between the ligand and the target and ignore flexibility and induced-fit binding models. More advanced algorithms consider several possible conformations of the ligand or the receptor, or both at the same time according to the conformational selection paradigm (Changeux and Edelstein, 2011). Rigid docking simulations are generally preferred when time is critical, *i.e.* when a large number of compounds are to be docked during an initial high-throughput screening. However, flexible docking methods are still needed for refinement and optimization of the poses obtained from an initial rigid docking procedure. With the evolution of computational resources and efficiency, flexible docking methods are becoming increasingly used (Sliwoski *et al.*, 2014).

The flexibility of the system in a docking protocol remains a major challenge in the search for the pose that best matches the receptor configuration: indeed the number of degrees of freedom included in the conformational search is a central aspect that determines the searching efficiency (Muegge and Rarey, 2001). In a real biological system, the system would include at least the ligand, the macromolecular receptor, and the solvent molecules. Because of the huge number of degrees of freedom associated with the solvent molecules, they are normally excluded from the simulation, or in special cases implicitly considered in the scoring functions as a way to address the solvent effect (Sousa *et al*., 2006).

The other critical aspect of every docking protocol is the evaluation and ranking of the ligand conformations predicted based on the search algorithm. Being capable of generating the right conformation is not enough; it is also necessary to be able to recognize it. An effective scoring function, which can be seen as an attempt to approximate the standard chemical potentials of the system, should enable the distinction between the true binding modes and all the other alternative modes explored, or between active and random compounds. However, a very rigorous scoring function would be computationally too expensive, rendering the analysis of several binding modes unfeasible. Hence, a number of assumptions and simplifications have to

be used to reduce the complexity of the scoring functions, with a natural cost in terms of accuracy (Sousa *et al*., 2006). The scoring functions normally employed in protein–ligand docking can be divided into three major classes. Force field-based scoring functions, which take into account the interaction energy between the receptor and the ligand through a combination of Van der Waals and electrostatic energy terms (Kramer *et al*., 1999). Empirical scoring functions, which take advantage of experimentally determined binding energies (Wang *et al*., 2002). Knowledge-based scoring functions, which focus on information statistically derived from large datasets, and thus rely on existing well-determined structures (Gohlke *et al.*, 2000). The imperfections in the scoring function continue to be the major limiting factor of the docking procedure, in terms of speed and accuracy. In fact, a number of physical phenomena known to be determinant in molecular recognition are completely neglected or at least not fully accounted for; for example entropic effects, solvation, and long-range non-bonded interactions (Kitchen *et al*., 2004).

1.2.2 Virtual screening libraries

Before starting a virtual screening analysis, it is necessary to collect all the structures to be tested. If the search space is very limited, and it is known which kind of molecules is likely to bind the target, it is possible to draw our own structures and readily start the docking process. However, for most campaigns this is not the case. Usually, before docking, it is necessary to build a library of potential ligands which can contain thousands of structures that have to be tested.

In recent years, several databases of chemical structures have been developed, and are easily accessible, which not only store the structure of these molecules, but also many chemical and biologically relevant information (Cerqueira *et al*., 2015, Sliwoski *et al*., 2014). Some widely used databases are listed in Figure 1.7.

Database	Type	Size
PubChem (Wheeler et al., 2006)	Biologic activities of small molecules	~40,000,000
Accelrys Available Chemicals Directory (ACD) (Accelrys, 2012)	Consolidated catalog from major chemical suppliers	~7,000,000
PDBeChem (Dimitropoulos, 2006)	Ligands and small molecules referred in PDB	14,572
Zinc (Irwin and Shoichet, 2005)	Annotated commercially available compounds	~21,000,000
LIGAND (Goto et al., 2002)	Chemical compounds with target and reactions data	16,838
DrugBank (Wishart et al., 2006)	Detailed drug data with comprehensive drug target information	6711
ChemDB (Chen et al., 2005, 2007)	Annotated commercially available molecules	~5,000,000
WOMBAT Data base (World of Molecular BioAcTivity) (Ekins et al., 2007; Hristozov et al., 2007)	Bioactivity data for compounds reported in medicinal chemistry journals	331,872
MDDR (MDL Drug Data Report) (Hristozov et al., 2007)	Drugs under development or released; descriptions of therapeutic	180,000
3D MIND (Mandal et al., 2009).	Molecules with target interaction and tumor cell line screen data	100,000

Figure 1.7: **Widely used chemical compounds libraries.** Information about the classes of compounds they host and the size of the repositories are reported (Sliwoski *et al*., 2014).

Over the past decade, in order to accelerate the drug development process with reduced risk of failure and relatively lower costs, pharmaceutical companies have adopted drug repositioning as an alternative. This strategy involves exploration of drugs that have already been approved for treatment of other diseases and/or whose targets have already been discovered. Existing collections, such as DrugBank (see Figure 1.7), provide access to most FDA-approved and marketed drugs. This approach has the potential to lower the risk of drug development and offers shorter drug development times and lower investment costs than traditional drug discovery methods (Baek *et al*., 2015, Padhy and Gupta, 2011).

2. AIM OF THE WORK

The present work was aimed at studying and applying protein-ligand docking protocol to biological systems highly relevant from a biomedical point of view, to provide complementary and supporting evidence to experimental studies. Using experimentally characterized macromolecular systems, a tuning of the algorithm parameters and a validation of the scoring system were carried out, by changing the default running parameters, the size of the search space, and the conformational flexibility of the receptor side-chains.

Another objective of this work was that of developing a user-friendly procedure which would allow users without any expertise in bioinformatics to perform molecular docking simulations and virtual screening analyses.

This automated virtual screening/docking platform, which falls in the category of structure-based methods, was tested on experimental structures of *Pseudomonas aeruginosa* protein targets for the identification of novel lead compounds with potential interest from a pharmaceutical viewpoint. More in general, this procedure was aimed at achieving two primary objectives. First, starting from the information derived from docking results, make it possible to predict the function of an uncharacterized protein. In fact, for enzymes of unknown function, substrate prediction based on structural complementarity becomes attractive when the target enzyme has little relationship to orthologues of known activity, making inferences unreliable. From this viewpoint, the identification of few compounds that bind the target with the highest probability, that is the identification of potential substrate/inhibitor candidates, could lead to protein function annotation.

Second, facilitate drug discovery and/or repositioning activities, that is the main goal of the present project. In fact, the automated screening against a library of FDA-approved compounds would reduce the cost of the initial experimental screening and accelerate lead compounds discovery.

3. METHODS

3.1 The Protein Data Bank

The Protein Data Bank (PDB) (Berman *et al.*, 2000) is an archive for biological macromolecular structures such as proteins and nucleic acids. The data contained in the archive include atomic coordinates, crystallographic structure factors and NMR experimental data. In addition to coordinates, each entry also includes the names of the molecules, primary and secondary structure information, sequence database references, ligand and biological assembly information, details about data collection and structure solution, and bibliographic citations. The primary information stored in the PDB archive consists of coordinate files for biological molecules. These files list the atoms in each protein, and their spatial coordinates. A typical PDB formatted file includes a large "header" section of text that summarizes the information on the protein, followed by the sequence and a list of the atoms and their coordinates. ATOMs and HETATMs are the entries in which each atom is stored. The ATOM keyword is used to identify proteins or nucleic acid atoms, and the keyword HETATM is used to identify atoms of small molecules, metal ions or water molecules. Following this keyword, there is a list of the information about the atom, including its name, its number in the file, the name and number of the residue it belongs to, one letter to specify the chain (in oligomeric proteins), its x, y, and z coordinates, and occupancy and temperature factor values. In addition, another information provided in a PDB file that allows to understand the quality of a protein structure, is the resolution. The resolution is a measure of the level of detail present in the diffraction pattern and the level of detail that will be seen when the electron density map is calculated. High-resolution structures, with resolution values around 1.0 Å, are of high quality and the position of every atom in the electron density map is determined precisely.

It should be noticed that X-ray crystallography cannot resolve hydrogen atoms in most protein crystals, thus in PDB files, hydrogen atoms are generally absent. If required (*e.g.* in docking simulations) hydrogens are added by modeling. On the other hand, hydrogen atoms are always present in PDB files resulting from NMR analysis, and usually present in theoretical models.

3.2. AutoDock Vina

AutoDock Vina (also referred to as Vina; Trott and Olson, 2009) is the program for protein-ligand docking and virtual screening used in this work. Its multi-core capability, high performance and enhanced accuracy, ease of use and free-availability have contributed to its extremely fast dissemination through the docking community (Sousa *et al*., 2013).

Vina's scoring function combines some advantages of knowledge-based potentials and empirical scoring functions: it extracts empirical information from both the conformational preferences of the receptor-ligand complexes and the experimental affinity measurements from the PDBbind data set (Wang *et al.*, 2004). The optimization algorithm, used to find the local minimum of the scoring function, is the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method, which is an efficient quasi-Newtonian method (Nocedal and Wright, 1999).

The number of steps in a run is determined adaptively, according to the apparent complexity of the problem, and several runs starting from random conformations are performed. The number of runs is determined by the "exhaustiveness" parameter, that is the accuracy of the global search, roughly proportional to the execution time. In Vina, these runs can be performed concurrently, using multithreading. This allows to take advantage of the shared-memory hardware parallelism, such as the now ubiquitous multicore CPUs.

The optimization algorithm maintains a set of diverse significant minima found that are then combined from the separate runs and used during the structure refinement and clustering stage.

The performance of Vina has been compared to that of AutoDock 4.0.1 (Morris *et al*., 2009) on a set of 190 protein-ligand complexes that had been used as a training set for the AutoDock scoring function (Morris *et al*., 1998). In this set, the receptors are treated as rigid, and the ligands are treated as flexible molecules with the number of active rotatable bonds ranging from 0 to 32.

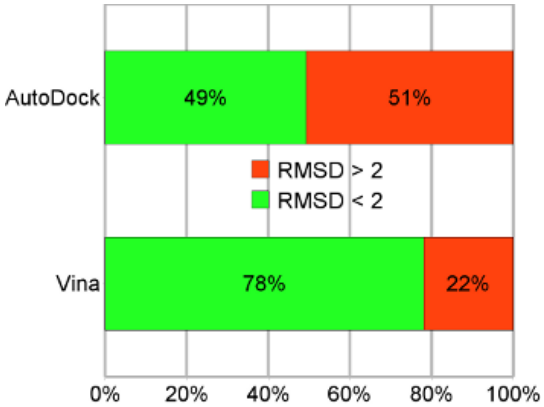


Figure 3.1: **The fraction of the 190 test complexes for which RMSD< 2Å was achieved by AutoDock and Vina.** In about 80% of cases, Vina was able to place the ligand in a similar position with respect to experimentally determined one.

The RMSD cutoff of 2Å between the experimentally observed ligand position on the receptor and conformation and the one(s) predicted by the algorithm, is often used as a criterion of the correct bound structure prediction (Bursulaya *et al*., 2003). Using the same cutoff value, the fractions of accurate predictions for AutoDock and Vina are reported in Figure 3.1. From the figure it is easy to see the increase in precision of Vina with respect to AutoDock.

Additionally and independently, AutoDock Vina has been tested against a virtual screening benchmark called the Directory of Useful Decoys (Huang *et al*., 2006) by the Watowich group and was found to be "a strong competitor against the other programs (*i.e.* GLIDE HTVS, ICM, DOCK, FlexX), and at the top of the pack in many cases".

See https://drugdiscovery.tacc.utexas.edu/#/faq for details.

Vina's design philosophy is not to require the user to understand its implementation details, tweak obscure search parameters, cluster results or know advanced algebra. All that is required is the structures of the molecules to be docked (*i.e.* receptor and ligand) and the specification of the search space including the putative binding site. Vina requires the ligand file to be written in PDBQT format. PDBQT format is very similar to PDB format but it includes partial charges ('Q') and AutoDock 4 (AD4) atom types ('T'). There is one line for each atom in the ligand, plus special keywords indicating which bonds, if any, are required to be rotatable during the docking experiment.

Preparing the receptor and ligand involves ensuring that their atoms are assigned the correct AutoDock atom types, adding Gasteiger charges, merging non-polar hydrogens, detecting aromatic carbons if any, and setting up the 'torsion tree' (only for the ligand). For most atoms, the AD4 atom type is the same as its element; the exceptions are "OA", "NA", "SA" for hydrogen-bond acceptor O, N and S atoms; "HD" for hydrogen-bond donor H atoms; "N" for non-hydrogen bonding nitrogens, and "A" for carbon atoms in aromatic rings.

In particular, starting from the PDB file of the receptor, the conversion to PDBQT file involves:

- the deletion of the co-crystallized ligand, if any (HETATM field in PDB file);
- the deletion of the water molecules;
- the addition of polar hydrogen atoms;
- the addition of partial charges.

If the user wants to allow the flexibility of some residues in the receptor structure, a third file has to be generated, with the information about rotatable bonds in the side chain of each residue.

In addition, the steps required to prepare the ligand PDBQT file are:

- the addition of polar hydrogen atoms;
- the addition of partial charges;

- the detection of all active torsions within the ligand structure.

In general, the more rotatable bonds in the ligand, the more difficult it will be to find good binding modes in repeated docking experiments. Once the above mentioned steps are concluded, it is necessary to set the search space (*i.e*. the docking grid), that is the portion of the receptor in which Vina will search for the best orientation and conformation of the ligand. The smaller the search space, the easier it is for the docking algorithm to explore it. On the other hand, it will not explore ligand and flexible side chain atom positions outside the search space. Search spaces bigger than 30 x 30 x 30 (27,000) $Å^3$ should be avoided, unless the "exhaustiveness" parameter is also increased.

These results can be achieved by using AutoDock Tools (ADT, also referred as MGLTools), an auxiliary software developed at the Molecular Graphics Laboratory (MGL) of The Scripps Research Institute for visualization and analysis of molecular structures, for preparing the files, choosing the search space, and viewing the results (Morris *et al*., 2009). In Figure 3.2, is reported a snapshot of the ADT interface, by which the user can set the search space extension through the specification of its length along the X, Y and Z-axes and of its center (the “grid box” center). All these parameters are then reported in a configuration file (Figure 3.3) which is needed to run the simulation.

The main output of the program is the binding energy value (ΔG): based on this calculated value, the poses obtained for the ligand are ranked and reported in the output file of the program. This file can then be analyzed with many molecular graphics programs, to visualize the preferential binding sites of the ligand and to analyse the interactions at the atomic level.

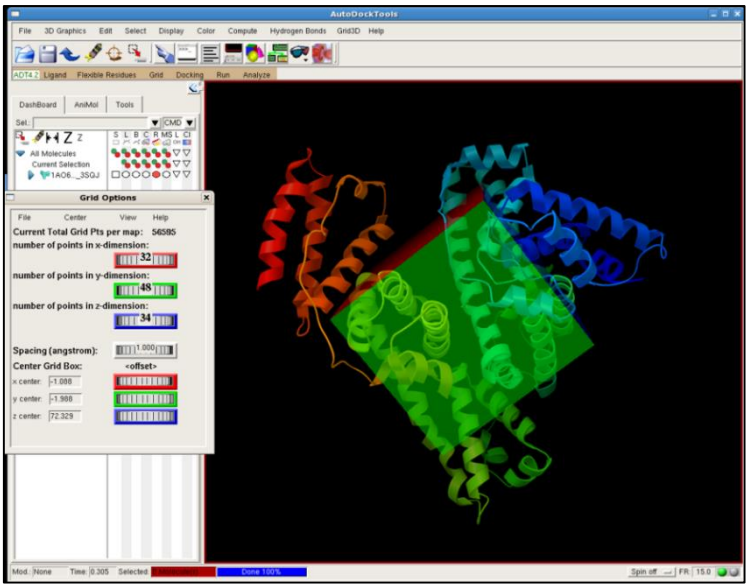


Figure 3.2: **ADT interface with the window to setup the search space.** Through this interface it is possible to prepare PDBQT files, choose the search space and visualize the results.

```
receptor = 1AO6.pdbqt
ligand = imatinib.pdbqt

center_x = -1.088
center_y = -1.988
center_z = 72.329

size_x = 32
size_y = 48
size_z = 34

exhaustiveness = 24
log = log.txt
```

Figure 3.3: **Example of Vina's configuration file.** Here are contained all the information needed by Vina to run the docking simulation.

3.3 DrugBank

The DrugBank database (Wishart *et al.*, 2006) is a unique bioinformatics and cheminformatics resource that combines detailed drug (*i.e.* chemical, pharmacological and pharmaceutical) data with comprehensive drug target (*i.e.* sequence, structure, and pathway) information. The database contains 8312 drug entries including 2036 FDA-approved small molecule drugs, 233 FDA-approved biotech (protein/peptide) drugs, 93 nutraceuticals and over 6000 experimental drugs. Additionally, 4317 non-redundant protein (*i.e.* drug target/enzyme/transporter/carrier) sequences are linked to these drug entries. Each DrugCard entry contains more than 200 data fields with half of the information concerning drug/chemical data and the other half related to drug target or protein data (Table 3.1).

Each set of compounds (approved, experimental, nutraceutical, illicit, withdrawn and investigational) is freely reachable at the website (http://www.drugbank.ca/) as SDF (Structure Data Format) file, that is a chemical file format to represent multiple chemical structure records, delimited by lines consisting of four dollar signs ($$$$), and associated data fields. Obviously, SDF files contain also the coordinates of each compound. In this work, the subset of FDA-approved compounds was used. In fact, a drug that is approved is said to be safe and effective when used as directed. Therefore, the screening against this subset should reduce the risk of low quality hits and, more important, if in the initial screening a drug displays a sufficient affinity for the target, it could be immediately tested in patients.

Drug or compound information	Drug target or receptor information
Generic name	Target name
Brand name(s)/synonyms	Target synonyms
IUPAC name	Target protein sequence
Chemical structure/sequence	Target no. of residues
Chemical formula	Target molecular weight
PubChem/KEGG/ChEBI Links	Target pI
Swiss-Prot/GenBank Links	Target gene ontology
FDA/MSDS/RxList Links	Target general function
Molecular weight	Target specific function
Melting point	Target pathways
Water solubility	Target reactions
pKa or pI	Target Pfam domains
LogP or hydrophobicity	Target signal sequences
NMR/MS spectra	Target transmembrane regions
MOL/SDF/PDF text files	Target essentiality
MOL/PDB image files	Target GenBank protein ID
SMILES string	Target Swiss-Prot ID
Indication	Target PDB ID
Pharmacology	Target cellular location
Mechanism of action	Target DNA sequence
Biotransformation/absorption	Target chromosome location
Patient/physician information	Target locus
Metabolizing enzymes	Target SNPs/mutations

Table 3.1: **Summary of the data fields or data types found in each DrugCard**. A more complete listing is provided on the DrugBank home page (Wishart *et al.*, 2006).

3.4 Molconverter

Molconverter is a command line program in Marvin Beans and JChem Suites (http://www.chemaxon.com/products/marvin/molconverter/) that converts between various coordinates file types. Operating in batch mode users can specify various documents, molecules files, graphic and compression/encoding formats. When handling chemical structures users can specify how to handle chemical features relevant to molecule file formats.

In this work, this program was used to convert a two-dimensional SDF file from DrugBank to a three-dimensional PDB file (low energy conformer).

3.5 UCFS Chimera

UCSF Chimera (also referred simply to as Chimera) is a molecular graphics program developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (Pettersen *et al*., 2004). In this work, molecular graphics and analyses of docking solutions were performed with this program, including receptor-ligand complex visualization and detailed atomic study of the relative interactions.

4. RESULTS

4.1 Overview

Results obtained in this work are divided into two sections: first, docking analyses on human serum albumin (HSA) are reported, carried out to investigate the interaction of this protein with the drug imatinib and with retinoids. Experimental data relative to these macromolecular systems were provided by Professor Paolo Ascenzi from the Biochemistry Lab of the Department of Science, University of Roma Tre. Data relative to these analyses were published in two papers (Di Muzio *et al*., 2014, Di Muzio *et al*., 2015) whose reprints can be found in Supplementary Material. Subsequently, a user-friendly procedure implemented within the present Ph.D. project is described, which allows to perform molecular docking simulations and virtual screening analyses also to users without any expertise in bioinformatics. Although results obtained are still under analysis, the virtual screening version of the program was tested on five targets from *Pseudomonas aeruginosa* that are involved in bacterial virulence or are essential for bacterial growth with the aim to identify those FDA-approved compounds that could display antibacterial activity.

4.2 Docking of imatinib to HSA and HSA-heme-Fe(III)

Human serum albumin (HSA), the most abundant protein in plasma, is a monomeric multidomain macromolecule, representing the main determinant of plasma oncotic pressure and the main modulator of fluid distribution between body compartments. HSA displays an extraordinary ligand binding capacity, representing a depot and carrier for many endogenous and exogenous compounds. Indeed, HSA represents the main carrier for fatty acids (FA), being able to bind up to nine equivalents of long chain FAs, which represent the primary physiological ligands at multiple binding sites (*i.e*., the FA1–FA9 sites). Moreover, HSA affects pharmacokinetics of many drugs, provides the metabolic modification of some ligands, renders potential toxins harmless, accounts for most of the anti-oxidant capacity of human plasma, and displays (pseudo-)enzymatic properties (Fanali *et al*., 2012).

Imatinib is a potent and selective kinase inhibitor approved in the treatment of chronic myelogenous leukemia (CML) and metastatic gastrointestinal stromal tumor (GIST). Imatinib acts by blocking the tyrosine kinase activity of key proteins involved in the pathogenesis of

CML and GIST (Hartmann *et al*., 2009). Imatinib is approximately 95% bound to plasma proteins, a1-acid glycoprotein (AGP) being the primary carrier. Although the plasma level of AGP is markedly elevated in malignant diseases, in contrast to the decline of HSA level (Gupta and Lis, 2010), half of the normal plasma levels of AGP have been found in patients with hepatic cirrhosis and hepatitis (Kremer *et al*., 1988); therefore, in these pathological states, HSA may serve as a secondary carrier for imatinib.

Docking simulations of imatinib binding to HSA were performed using the crystal structures of the ligand-free HSA (PDB ID: 1AO6) (Sugio *et al*., 1999) and of HSA-heme-Fe(III) (PDB ID: 1N5U) (Wardell *et al*., 2002).

Simulations were carried out using the docking program Autodock Vina with a search space (docking grid) that included the whole protein, in order to carry out "blind" predictions of the imatinib binding site(s). Additional simulations were also carried out restricting the search space only to the FA2, FA6 and FA7 sites. The grid spacing was set to 1 Å per grid unit and the exhaustiveness parameter was increased from the default value of 8 to 24 as suggested by Vina developers for grid sizes larger than 27,000 $Å^3$, which is the case for HSA simulations.

Initially the number of solutions (docking poses) generated by Vina was set to the default value of 9, but simulations were also carried out with a number of solutions of 20. The simulations were carried out both by keeping all protein residues rigid and by allowing flexibility of the residues building up the walls of the sites from FA1 to FA7 (Fanali *et al*., 2012, Fanali *et al*., 2012a). Dihedral angles involving single bonds of the flexible residues are, by default, varied by a 50 degrees increment during Vina execution. Residues for which flexibility was allowed are reported in Table 4.1. Rotatable bonds of the imatinib lowest energy resonance structure were kept flexible in all the simulations, as suggested by Vina developers (see http://vina.scripps.edu/tutorial.html).

FA binding site	**Residues**
FA1	Arg117, Leu182, Tyr161
FA2	Tyr150, Arg257, Ser287, Ala254
FA3	Ser342, Arg485, Arg348, Leu453
FA4	Tyr411, Ser489, Arg410, Leu457, Phe488, Val415, Leu460
FA5	Tyr 401, Lys525, Met548, Phe551
FA6	Arg209, Lys351, Ser480, Asp324, Glu354, Ala213
FA7	Lys199, Arg218, Arg222, His242, Val241

Table 4.1: Residues for which flexibility was allowed in docking simulations of imatinib binding to HSA and HSA-heme-Fe(III).

4.2.1 Docking analyses of imatinib binding to ligand-free HSA

Docking simulations of imatinib binding to HSA, with the search space extended to the whole protein (Figure 4.1), indicated the preferential binding of the drug to the FA1 and FA7 sites with similar binding energy. The number of complexes observed in FA1 and FA7 in docking simulations with a maximum of 9 poses was 4 and 2, respectively. By increasing to 20 the number of possible poses, 9 and 3 complexes were observed in FA1 and FA7, respectively. The highest ranking complexes of imatinib in FA1 and FA7 are shown in Figure 4.2. In simulations carried out restricting the docking search space only to FA2, FA6 and FA7 sites, imatinib was found to bind with the best binding energies and in the majority of the poses (7 out of 9) in FA7 rather than in FA6 site (2 out of 9), while binding in FA2 was not observed. These results suggest that in ligand free-HSA imatinib binds preferentially to the FA1 and FA7 sites.

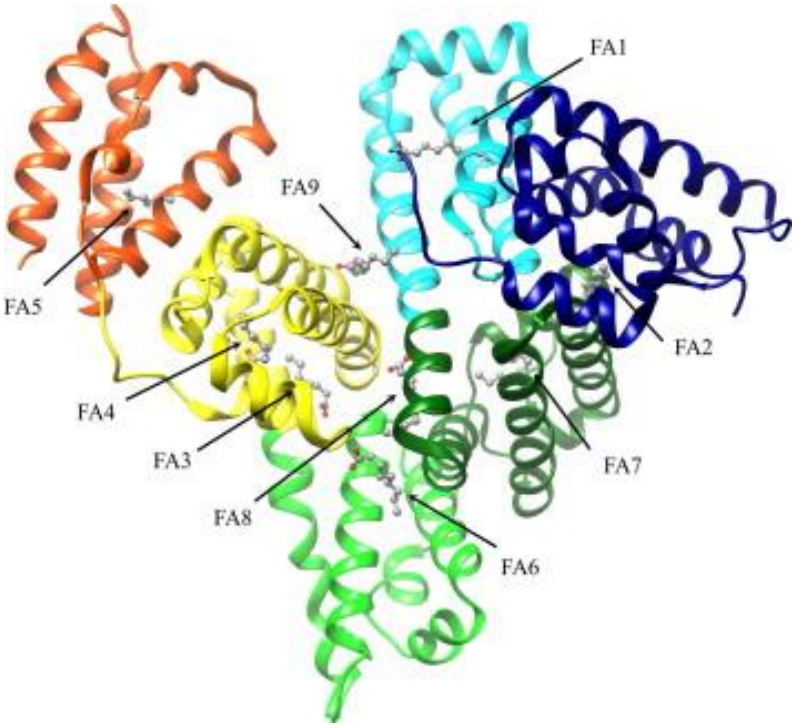


Figure 4.1: **Ribbon representation of the three-dimensional structure of HSA.** The subdomains of HSA are rendered with different colors (domain IA, in blue; domain IB, in cyan; domain IIA, in dark green; domain IIB, in light green; domain IIIA, in yellow; and domain IIIB, in red). The FA1–FA9 sites are occupied by capric acid (in ball-and-stick representation). Atomic coordinates were taken from the PDB entry 1E7E (Bhattacharya *et al.*, 2000). The FA binding sites are numbered according to literature (Bhattacharya *et al.*, 2000).

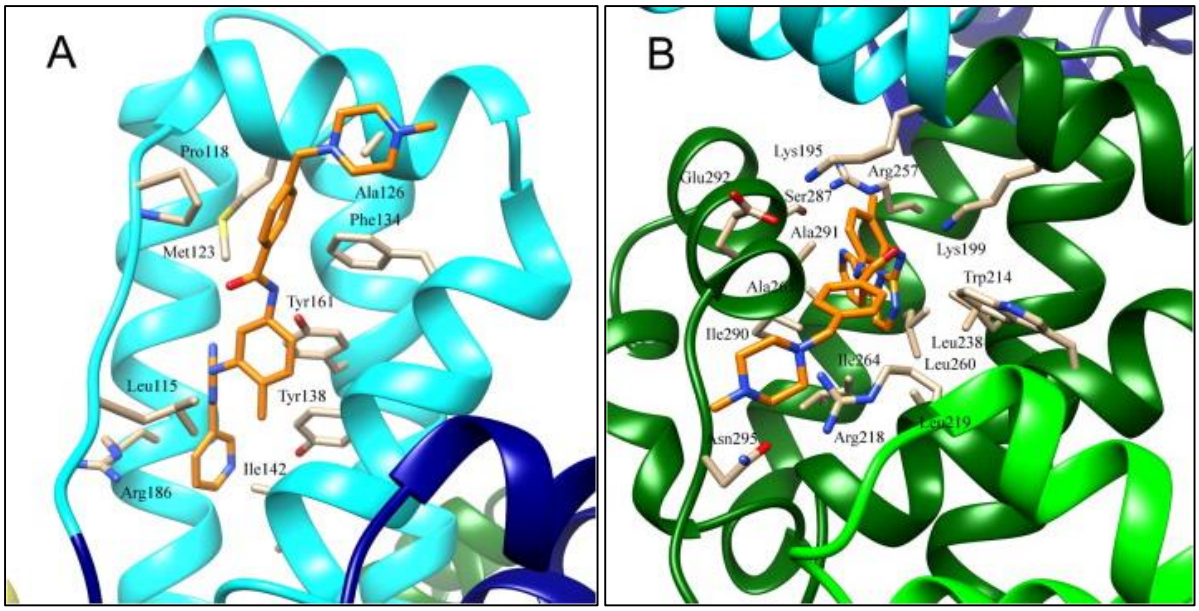


Figure 4.2: **Atomic details of the highest ranking complexes obtained by docking simulations between imatinib and ligand-free HSA.** (A) Imatinib bound to FA1. (B) Imatinib bound to FA7. Atomic coordinates were taken from the PDB entry 1AO6 (Sugio *et al*., 1999).

4.2.2 Docking analyses of imatinib binding to HSA-heme-Fe(III)

Docking simulations of imatinib binding to HSA-heme-Fe(III), with the search space extended to the whole protein, indicated that the most likely binding sites are FA2 and FA7, binding energy values being of the same order of magnitude. When the maximum number of docking poses was set to 9, the number of complexes observed in FA2 and FA7 was 2 and 6, respectively. Setting the maximum number of docking poses to 20 resulted in 1 and 9 complexes in FA6 and FA7, respectively. When the search space was restricted only to FA2 and FA7 sites, 1 complex was observed in FA2 and 8 in FA7. These results suggest that in HSA-heme-Fe(III), imatinib binds preferentially to the FA2 and FA7 sites as illustrated in Figure 4.3.

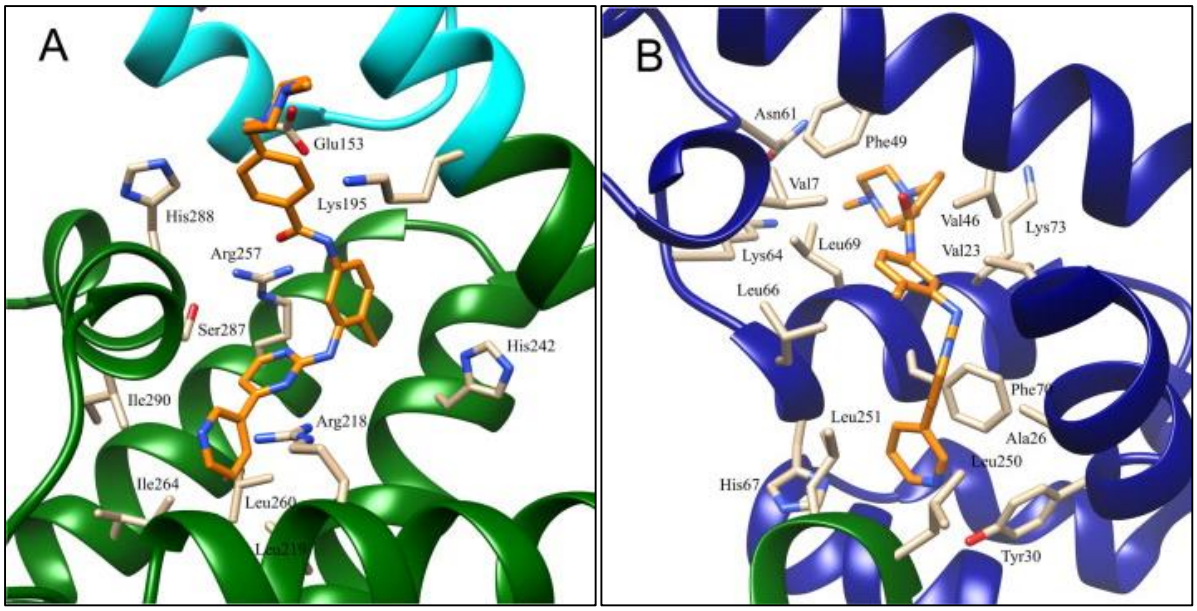


Figure 4.3: **Atomic details of the highest ranking complexes obtained by docking simulations between imatinib and HSA-heme-Fe(III)**. (A) Imatinib bound to FA7. (B) Imatinib bound to FA2. Atomic coordinates were taken from the PDB entry 1N5U (Wardell *et al*., 2002).

4.2.3 Experimental data of imatinib binding to HSA and HSA-heme-Fe(III)

Experimental analyses were carried out to investigate thermodynamics of imatinib binding to full-length HSA and its recombinant Asp1-Glu382 truncated form (containing only the FA1, FA2, FA6, and FA7 binding sites; trHSA), in the absence and presence of ferric heme (heme-Fe(III)). Imatinib binding to HSA was followed by competitive inhibition of dansyl-arginine and dansyl-sarcosine association, whereas imatinib binding to trHSA was followed by competitive inhibition of dansyl-arginine association. Competitive inhibition of dansyl-arginine and dansyl-sarcosine association reflects ligand binding to the FA7 site and the FA3–FA4 cleft, respectively. As resulted from these analyses, imatinib affects only dansyl-arginine binding to HSA and trHSA, thus suggesting that imatinib binds selectively to the FA7 site of HSA, which accommodates dansyl-arginine, but not to the FA3–FA4 cleft that binds dansyl-sarcosine. Moreover, from the experimental studies resulted that imatinib affects HSA-heme-Fe(III) and trHSA-heme-Fe(III) reactivity, *i.e*. peroxynitrite isomerization. Even if, at present, the available data do not allow to identify unambiguously the low affinity imatinib binding site(s) modulating the HSA-heme-Fe(III) reactivity, possible candidates are the FA2 and/or FA6 sites. This is because no single chromophore is available to probe both sites at the same time. However, docking simulations (Figure 4.3, panel B) suggest that FA2 is the most probable imatinib low affinity binding site. For further details on the experimental characterization of imatinib binding to HSA, see the reprint of the paper Di Muzio *et al.* (2014) in the Supplementary Materials.

4.3 Docking of all-*trans*-RA and all-*trans*-ROL to HSA

Dietary vitamin A and its natural and synthetic analogs (named retinoids) are naturally-occurring, fat-soluble, unsaturated isoprenoids present in all living organisms. Retinoids mainly influence cell growth, differentiation and death, the deregulation of retinoid signaling pathways being linked to tumorigenesis. For these reasons, they soon emerged as potential therapeutic agents for several diseases, including cancer (di Masi *et al*., 2015). Vitamin A, also named all-*trans*-retinol (all-*trans*-ROL), is converted into two classes of biologically active retinoids, *i.e.* 11-*cis*-retinoids and acidic retinoids. Among acidic retinoids, all-*trans*-retinoic acid (all-*trans*-RA) and 9-*cis*-retinoic acid (9-*cis*-RA) represent the main metabolic products.

RA, ROL, and all-trans-retinal bind to specific and aspecific proteins, which solubilize, protect, and detoxify retinoids in the extracellular environment (di Masi *et al*., 2015, Folli *et al*., 2010, O'Byrne and Blaner, 2013). Among the extracellular retinoid-binding proteins (RBP), the epididymal retinoid-binding protein (ERBP), the interphotoreceptor matrix retinoid-binding protein (IRBP), and the retinoid binding protein 4 (RBP4) play a central role in ROL transport, whereas lipocalin-type prostaglandin D synthase (also known as β-trace) and HSA transport preferentially all-trans-RA (Kuruvilla *et al*., 1991, Zahn *et al*., 1993, O'Byrne and Blaner, 2013). This suggests that RBPs may function as alterative carriers of ROL and RA in health and disease. Indeed, HSA may act as a secondary carrier in human diseases associated with reduced levels of IRBP (*e.g.*, in the early stages of diabetic retinopathy) (Garcia-Ramírez *et al*., 2009) or RBP4 (*e.g.*, in prolonged dietary vitamin A deficiency, in migraine, in HIV, in diabetic patients treated with anti-TNFa-therapy, and in end-stage renal disease after kidney transplantation) (Zhang *et al*., 2014, Kotzé *et al.*, 2015).

Docking simulations of all-*trans*-RA and all-*trans*-ROL binding to HSA were performed using the crystal structure of ligand-free HSA (PDB ID: 1AO6) (Sugio *et al*., 1999), and of FA-bound HSA (PDB ID: 3SQJ) (He *et al*., 2011). All-*trans*-RA and all-*trans*-ROL three-dimensional structures were obtained respectively from the crystal structures of the all-*trans*-RA-bound ligand binding domain of RAR-γ (PDB ID: 2LBD) (Renaud *et al*., 1995) and of the all-*trans*-ROL-bound retinol-binding protein (PDB ID: 1KT6) (Calderone *et al*., 2003).

Simulations were carried out using Vina with a search space (docking grid) that included the whole protein, in order to carry out “blind” predictions of the all-*trans*-RA and all-*trans*-ROL binding sites. Additional simulations were also carried out restricting the search space only to the FA1 site, in order to obtain more accurate details on all-*trans*-RA and all-*trans*-ROL interaction with HSA.

The simulations were carried out both by keeping all protein residues rigid and by allowing flexibility only of the residues building up the walls of the FA sites (FA1 to FA9). Residues for which flexibility was allowed are those reported in Table 4.1. In addition, flexibility was allowed for residues Lys195, Asp451, Ser454 belonging to the FA8 site, and for residues Asp187, Lys432 belonging to the FA9 site. Rotatable bonds of the all-*trans*-RA ad ROL lowest energy resonance structures were kept flexible in all the simulations.

4.3.1 Docking analysis of retinoids binding to FA-free HSA

Docking simulations of all-*trans*-RA and all-*trans*-ROL binding to HSA, with the search space extended to the whole protein, indicated the preferential binding of both compounds to the FA1 site, in which retinoids bind in the vicinity of the FA1 site residue Tyr161. Interestingly, this residue coordinates the iron ion in the HSA-heme-Fe(III) derivative (Wardell *et al*., 2002). An overall view of the nine lowest energy retinoids-HSA complexes and atomic details of the lowest energy complex obtained for both compounds are shown in Figure 4.4. The number of complexes observed in FA1 in docking simulations with a maximum of 9 poses was 5 for all-*trans*-RA and 6 for all-*trans*-ROL. The binding affinity of the best all-*trans*-RA and all-*trans*-ROL poses was −8.1 kcal/mol in both cases. In other poses of this simulation, all-*trans*-RA was located in the FA6 site (3 poses) and in the FA9 site (1 pose), while all-*trans*-ROL in the FA9 site (3 poses). The simulation with the search space reduced to the FA1 site provided more reliable details of the interaction: in this case, the binding affinity of all-*trans*-RA and all-*trans*-ROL for the FA1 site of HSA was −8.8 kcal/mol and −8.0 kcal/mol, respectively. The highest ranking complexes of all-*trans*-RA and all-*trans*-ROL bound to the FA1 site of HSA are shown in Figure 4.4 (panels C and D).

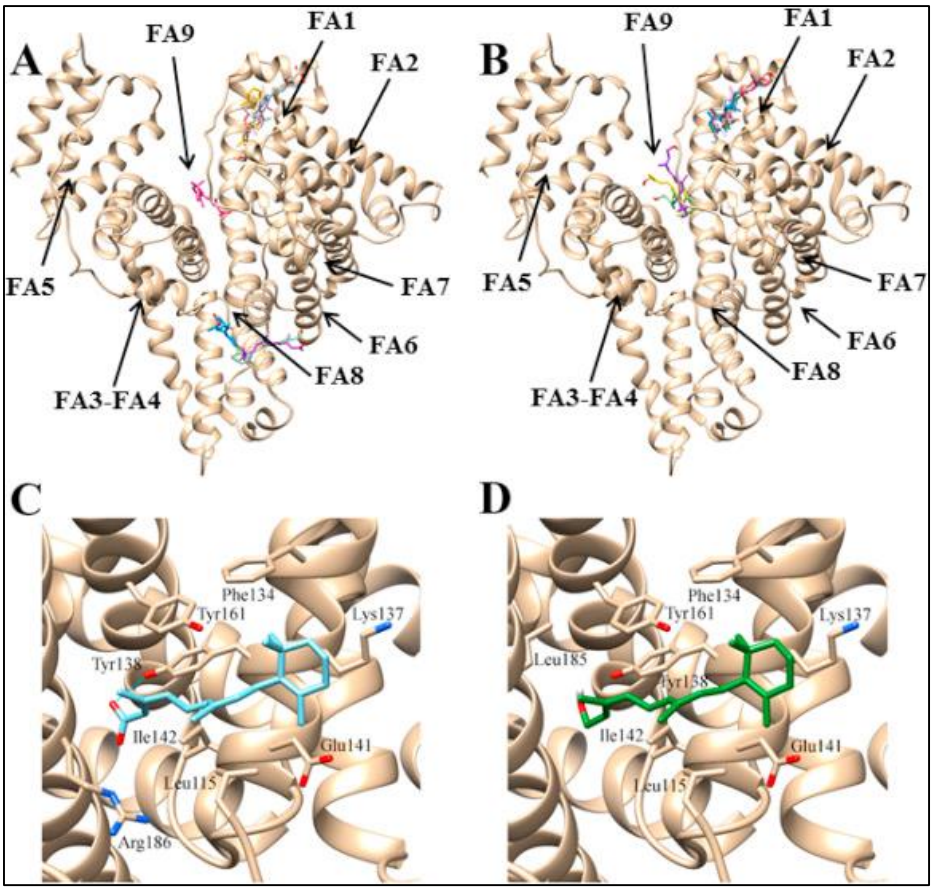


Figure 4.4: **Schematic representation of all-*trans*-RA and all-*trans*-ROL complexes with FA-free HSA as obtained by docking simulations**. Overall view of the nine lowest energy all-*trans*-RA-HSA complexes (panel A). Overall view of the nine lowest energy all-*trans*-ROL-HSA complexes (panel B). Atomic details of the lowest energy all-*trans*-RA-HSA complex (panel C). Atomic details of the lowest energy all-*trans*-ROL-HSA complex (panel D). Both all-*trans*-RA and all-*trans*-ROL bind in the vicinity of the FA1 site residue Tyr161, which coordinates the iron in HSA-heme-Fe(III) derivative.

4.3.2 Docking analysis of retinoids binding to FA-bound HSA

Docking simulations of all-*trans*-RA and all-*trans*-ROL binding to FA-bound HSA, with the search space extended to the whole protein, indicated the binding of all-*trans*-RA and all-*trans*-ROL in three different sites. In Figure 4.5 is shown the overall view of the complexes obtained. It should be noted that in the FA-HSA structure used in this simulations, the FA1 to FA7 sites are occupied by myristic acid (He *et al*., 2011).

In the first-ranking solution both ligands were placed in the FA8 site, located at the base of the crevice between subdomains IA-IB-IIA on one side and subdomains IIB-IIIA-IIIB on the other side. The FA8 site is relevant for ligand recognition only when the FA1 to FA7 sites of HSA

are occupied by long-chain FAs such as myristic acid (Fanali *et al*., 2012, Bhattacharya *et al*., 2000).

The binding affinity of the best all-*trans*-RA and all-*trans*-ROL poses in the FA8 site was −8.5 kcal/mol and −8.6 kcal/mol, respectively. In other poses, all-*trans*-RA and all-*trans*-ROL bind near the FA1 site occupied by myristate with an apparent free energy of −8.4 kcal/mol and −8.3 kcal/mol, respectively.

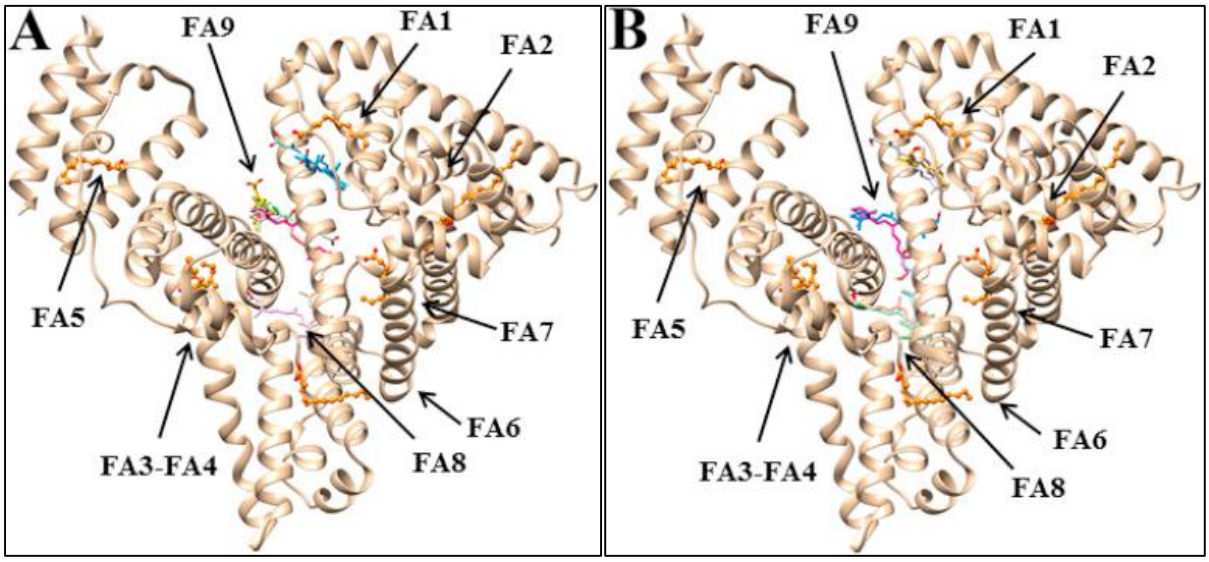


Figure 4.5: **Schematic view of the all-*trans*-RA and all-*trans*-ROL interaction with FA-bound HSA as obtained by docking simulations**. Overall view of the nine lowest energy all-*trans*-RA-HSA complexes (panel A). Overall view of the nine lowest energy all-*trans*-ROL-HSA complexes (panel B). Myristate molecules occupying the FA1 to FA7 sites (PDB ID: 3SQJ, He *et al*., 2011) are shown in ball-and-stick representation and colored in orange.

The binding affinity of the best all-*trans*-RA and all-*trans*-ROL poses in the FA8 site was −8.5 kcal/mol and −8.6 kcal/mol, respectively. In other poses, all-*trans*-RA and all-*trans*-ROL bind near the FA1 site occupied by myristate with an apparent free energy of −8.4 kcal/mol and −8.3 kcal/mol, respectively. Similar co-binding of aromatic drugs or drug-like molecules such as indomethacin and triiodobenzoic acid in the FA1 site was already reported (Curry *et al*., 1998, Ghuman *et al*., 2005) . Lastly, all-*trans*-RA and all-*trans*-ROL bind to the FA9 site of HSA, located in the upper region of the cleft built up by subdomains IA-IB-IIA on one side and subdomains IIB-IIIA-IIIB on the other side. Of note, the FA9 site is an additional binding pocket of HSA relevant for ligand recognition only when long-chain FAs, such as myristic acid, occupy the FA1 to FA7 sites of HSA (Bhattacharya *et al*., 2000). All-*trans*-RA and all-*trans*-ROL bind to the FA9 site of HSA in a fashion similar to that observed for thyroxine recognition when long-chain FAs, such as myristic acid, occupy the FA1 to FA7 sites (Petitpas *et al*., 2003).

The binding affinity of the best RA and ROL poses in this cleft was −8.0 kcal/mol and −7.6 kcal/mol.

Summarizing, docking analyses here reported allowed both to identify the preferential binding sites of the ligand(s)/drug, and to investigate the interaction at atomic level. Moreover, results obtained are in agreement with experimental data simultaneously obtained (see below; Di Muzio *et al.*, 2014, Di Muzio *et al*., 2015).

4.3.3 Experimental data of all-*trans*-RA and all-*trans*-ROL binding to HSA

Thermodynamics of all-trans-RA and all-trans-ROL (*i.e.*, retinoid) binding to the FA1 site, the FA3-FA4 cleft, and the FA7 site of HSA was followed by competitive inhibition of heme-Fe(III), dansylsarcosine, and dansyl-arginine association, respectively. Functional studies demonstrated that, in the absence of FAs, all-trans-RA and all-trans-ROL bind to the FA1 site of HSA, impairing competitively heme-Fe(III) association. On the other hand, all-trans-RA and all-trans-ROL do not bind to the FA3-FA4 cleft and the FA7 site, respectively. In fact, retinoids do not affect the association of dansyl-sarcosine and dansyl-arginine with HSA. Noteworthy, the FA7 site, the FA3-FA4 cleft, and the FA1 pocket (located in subdomains IIA, IIIA, and IB, respectively) represent the first, the second, and the third major ligand binding region of HSA, respectively. These data are in full agreement with the docking studies and thus the complexes obtained by docking simulations results can be used to infer the nature of the atomic interactions leading to retinoids binding to HSA. For further details on the experimental characterization of retinoids binding to HSA, see the reprint of the paper Di Muzio *et al.* (2016) in the Supplementary Materials.

4.4 DockingApp

"DockingApp" is the name given to the procedure implemented in Python programming language, which allows to perform molecular docking simulations and virtual screening analyses also to users without any expertise in bioinformatics. In Figure 4.6 is reported the "splash screen" of the application GUI.

Figure 4.6: **Splash screen of the DockingApp application**. It will be soon available at our lab website whose URL is reported in the figure.

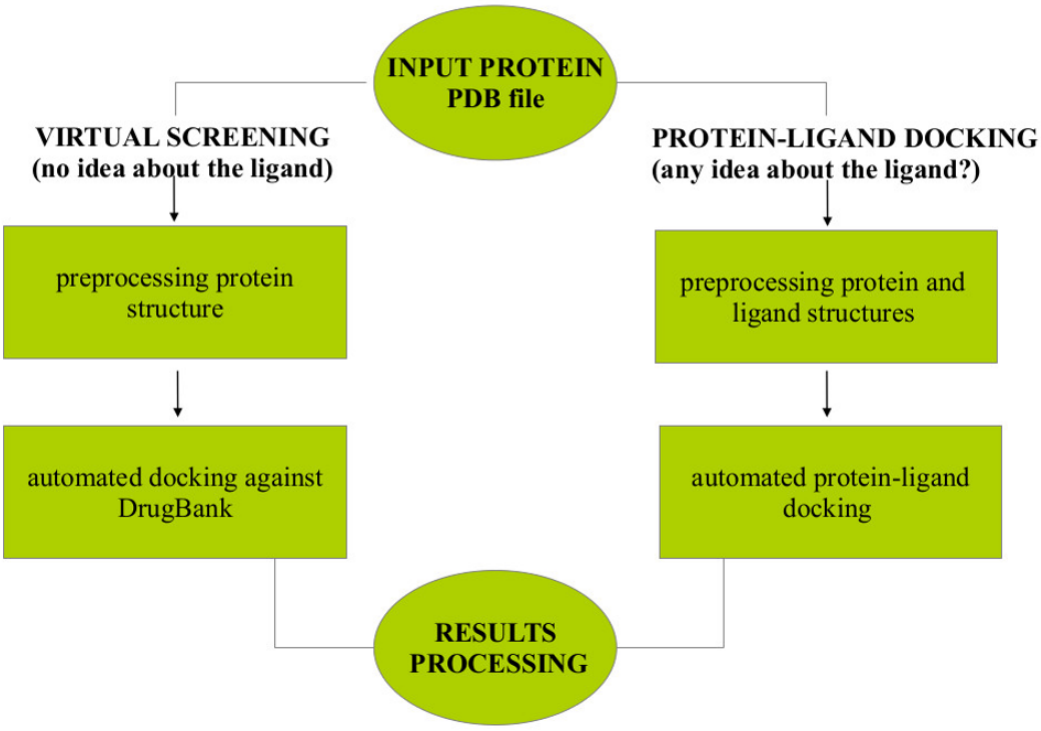


Figure 4.7: **Algorithm overview.** With DockingApp it is possible to carry out both docking and virtual screening analyses.

In Figure 4.7 is reported the algorithm flow. Briefly, the input of the virtual screening/docking application is a protein with still unknown function or a target of interest from a pharmaceutical viewpoint. The user needs only the three-dimensional structure of the target (*i.e.* the PDB file); the input is then automatically processed to obtain a structure with all information needed by the docking algorithm (hydrogen atoms, atomic charges, etc., see Methods for details). The core of the algorithm is an automatic docking run against a library of compounds, in particular in the case of this work, the subset of FDA-approved compounds retrieved from DrugBank. The output is a list (stored in a text file) containing the best hits resulting from the screening, on the basis of the binding affinity calculated. This file also contains some details, such as residues in contact with the ligand at different distance thresholds and details about the docked compounds. Obviously, in the output are stored also the coordinates of the complexes between the input protein and the best hit compounds.

In addition, a Java graphical user interface was developed, to guide the user through all the steps required to launch the application. To run the application the user must install MGLtools (ADT) and Autodock Vina. As explained in the Methods section (see Par. 3.2), the PDB files of the receptor and the ligand must be converted in PDBQT file format, required by Vina to run docking simulation. Through the ADT graphical user interface, it is possible to obtain this conversion as well as to set the desired search space, but for a user without any (or low) expertise in bioinformatics these steps can be difficult. To overcome this limitation, some Python scripts provided by Vina developers were used in this work, allowing the automatic preparation of PDBQT files. Regarding the grid box setting, a specific function was developed to permit the automatic setting of the search space, both local and extended to cover the entire structure of the receptor. These procedures are explained in the following paragraphs.

4.4.1 Docking files preparation

The automatic conversion of receptor file from PDB to PDBQT was achieved with the pyhon script **"prepare_receptor4.py"**, which can be found in AutoDockTools/Utilities24 folder. Typing the name of the script in that directory at a shell prompt results in printing Usage, which is an overview of how to use the script, illustrated in Figure 4.8. To use this file it is necessary to download and install MGLTools at http://mgltools.scripps.edu/downloads. This will install the graphical user interface and the Python shell scripts for ADT. The input is the PDB receptor file, passed to the script, which should contain only one molecule, the one to be formatted as the receptor. All hydrogens must be added ('hydrogens' repair) to the receptor, if they have not

already been added. The output default filename is the input file stem plus .pdbqt. For example, '1ao6.pdb' by default is written to '1ao6.pdbqt'.

If flexibility in the receptor is going to be included in the Vina simulation, the residues which have to be considered flexible must be written in a separate file from the one containing the residues to be considered rigid. Thus, a third file has to be generated which includes special keywords defining the torsional flexibility. This file is generated using the script **"prepare_flexreceptor4.py"**. The input file for this script should be in PDBQT format and output files from the script are in PDBQT format. Optional arguments can be used to forbid rotation around specified types of bonds or between specified atoms (see Figure 4.9).

Finally, the preparation of the ligand file is performed using the **"prepare_ligand4.py"** script. The input file should contain only one molecule, which will be formatted as the ligand. The output default filename is the input file stem plus .pdbqt. For example, 'imatinib.pdb' by default is written to 'imatinib.pdbqt'. In Figure 4.10 the description of the script options is reported.

```
prepare_receptor4.py -r filename
    Description of command...
         -r   receptor_filename
    Optional parameters:
        [-v]  verbose output (default is minimal output)
        [-o pdbqt_filename]  (default is 'molecule_name.pdbqt')
        [-A]  type(s) of repairs to make:
             'bonds_hydrogens': build bonds and add hydrogens
             'bonds': build a single bond from each atom with no bonds to its closest neighbor
             'hydrogens': add hydrogens
             'checkhydrogens': add hydrogens only if there are none already
             'None': do not make any repairs
             (default is 'None': do not to make any repairs)
        [-C]  preserve all input charges ie do not add new charges
             (default is addition of gasteiger charges)
        [-p]  preserve input charges on specific atom types, eg -p Zn -p Fe
        [-U]  cleanup type:
             'nphs': merge charges and remove non-polar hydrogens
             'lps': merge charges and remove lone pairs
             'waters': remove water residues
             'nonstdres': remove chains composed entirely of residues of
                     types other than the standard 20 amino acids
             'deleteAltB': remove XX@B atoms and rename XX@A atoms->XX
             (default is 'nphs_lps_waters_nonstdres')
        [-e]  delete every nonstd residue from any chain
              'True': any residue whose name is not in this list:
                   ['CYS', 'ILE', 'SER', 'VAL', 'GLN', 'LYS', 'ASN',
                    'PRO', 'THR', 'PHE', 'ALA', 'HIS', 'GLY', 'ASP',
                    'LEU', 'ARG', 'TRP', 'GLU', 'TYR','MET']
              will be deleted from any chain. NB: there are no
              nucleic acid residue names at all in the list.
             (default is False which means not to do this)
        [-M]  interactive
             (default is 'automatic': outputfile is written with no further user input)
```

Figure 4.8: **Options of the Python script "prepare_receptor4.py".**

```
prepare_flexreceptor4.py -r receptor_filename -s list_of_names_of_residues_to_move
    Description of command...
         -r   receptor_filename
         -s    list of names of residues to move
    Optional parameters:
        [-v]    verbose output
        [-N]    type(s) of bonds to disallow:
        [-P]    pairs of atom names bonds between which to disallow:
        [-g pdbqt_filename] (rigid output filename)
        [-x pdbqt_filename] (flexible output filename)
```

Figure 4.9: **Options of the Python script "prepare_flexreceptor4.py".**

```
prepare_ligand4.py -l filename

    Description of command...
         -l     ligand_filename

    Optional parameters:
        [-v]    verbose output
        [-o pdbqt_filename] (output filename)
        [-d]    dictionary to write types list and number of active torsions
        [-A]    type(s) of repairs to make:
                 bonds_hydrogens, bonds, hydrogens
        [-C]    do not add charges
        [-p]    preserve input charges on atom type, eg -p Zn
        [-U]    cleanup type:
                 nphs_lps, nphs, lps, ''
        [-B]    type(s) of bonds to allow to rotate
        [-R]    index for root
        [-F]    check for and use largest non-bonded fragment (False)
        [-M]    interactive (default is automatic)
        [-I]    string of bonds to inactivate composed of
                   of zero-based atom indices eg 5_13_2_10
                   will inactivate atoms[5]-atoms[13] bond
                               and atoms[2]-atoms[10] bond
                       (default is '')
        [-Z]    inactivate all active torsions
                       (default is leave active)
```

Figure 4.10: **Options of the Python script "prepare_ligand4.py".**

Essentially, these three scripts are invoked within the Python script that launches the application, thus allowing the user to automatically prepare the input PDBQT files, starting from the PDB files of the receptor and the ligand that the user indicates through the *ad hoc* GUI that as been developed (see below section 4.4.4).

4.4.2 Search space setting

The other mandatory step is the definition of the search space, that is the portion of the receptor in which Vina will search the ligand binding site. An in-house function to automatically set this space was developed and allows the user to avoid the use of the ADT graphical user interface.
The starting point is the three-dimensional structure of the receptor (*i.e.*, the PDB file already used to prepare the PDBQT file), on which two type of grid boxes can be defined: either an extended or a local one. In both cases, the function automatically detects the centroid (or the geometric center) starting from all the atoms of the structure (in the case of the extended grid), or from a set of atoms selected using the receptor residues indicated by the user though the GUI (in the case of a local grid).
Once detected the centroid, the function calculates the mono-dimensional distances between the centroid and each atom along the three axes x, y and z). The size of the grid box along the x, y and z axes is then calculated by doubling the maximum value of the relative mono-dimensional distances along each dimension, adding to these values 5 Å to ensure that the grid box encloses all the atoms of the receptor structure (or of the subset of residues chosen by the user in the case of a local grid box).

4.4.3 FDA-approved compounds preparation

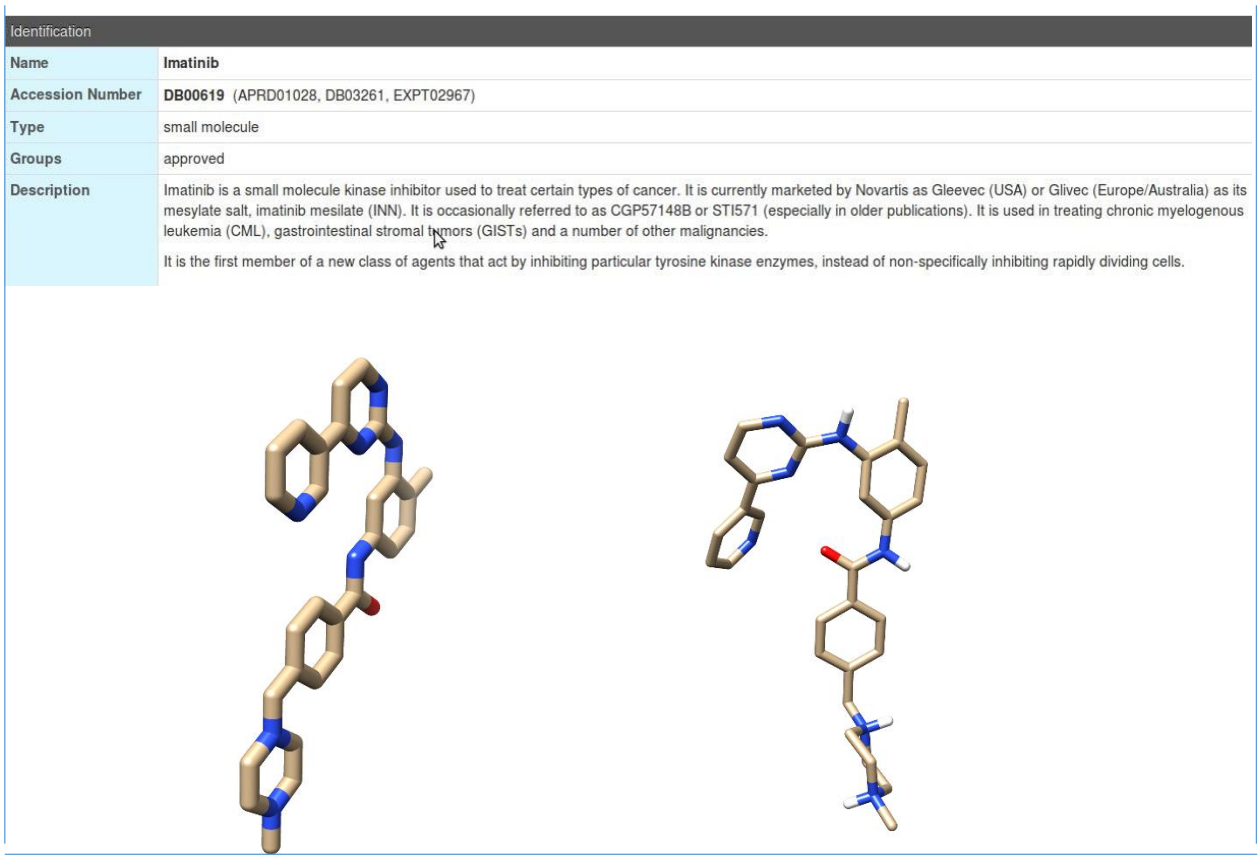
The subset of FDA-approved compounds used in this work has been downloaded from DrugBank (as of July 2014) and is composed by 1584 small molecule drugs. The subset of FDA-approved compounds is stored in a single SDF file. One problem encountered was that each compound is represented in two dimensions. In other words, atomic coordinates of the compounds are given in a two-dimensional reference system instead of a three-dimensional one. Thus, a procedure to convert the two-dimensional coordinates to three-dimensional ones, an essential step to perform docking simulations, was performed. For this purpose was used the MolConverter utility.
Here follows an example of the command from the Linux shell which converts a two-dimensional SDF file of DrugBank compound DB00091 to a three-dimensional PDB coordinates file (low energy conformer):

```
[Elena@biochimica7 ~]$./molconvert pdb -3:S{fine} DB00619.sdf -
o DB00619.pdb
```

This command was executed in an iterative and automatic way, including it in a script, and after this procedure each compound is in the correct structural format and conformation needed by Vina to process it (Figure 4.11).

The set used in this work is composed of 1466 molecules (of the initial 1584), because during the processing with the ADT Python script some errors were encountered (mainly an "unknown atom type" error, *e.g.* for platinum and arsenic atoms), that did not allow the generation of a correct PDBQT file for 118 compounds.

The FDA-approved set thus generated is ready to dock and is provided within the input folder of DockingApp to be used for virtual screening analyses.



Identification	
Name	Imatinib
Accession Number	DB00619 (APRD01028, DB03261, EXPT02967)
Type	small molecule
Groups	approved
Description	Imatinib is a small molecule kinase inhibitor used to treat certain types of cancer. It is currently marketed by Novartis as Gleevec (USA) or Glivec (Europe/Australia) as its mesylate salt, imatinib mesilate (INN). It is occasionally referred to as CGP57148B or STI571 (especially in older publications). It is used in treating chronic myelogenous leukemia (CML), gastrointestinal stromal tumors (GISTs) and a number of other malignancies. It is the first member of a new class of agents that act by inhibiting particular tyrosine kinase enzymes, instead of non-specifically inhibiting rapidly dividing cells.

Figure 4.11: **Example of conversion of a DrugBank entry from two-dimensional to three-dimensional coordinates.** Top: overview of the drug imatinib details (DrugBank ID: DB00619) reported in DrugBank website. Bottom: on the left, view of the molecule as downloaded from the website; on the right, view of the three-dimensional structure of the same molecule after conversion using Molconverter.

4.4.4 Graphical User Interface

The Graphical User Interface (GUI) of DockingApp has been developed in collaboration with Dr. Daniele Toti from the Engineering Department the University of Roma Tre. The GUI was designed to be as simple as possible, to guide the user through all steps required to launch the application.

At the first execution of the application, within the initial settings panel (Figure 4.12), the number of CPU cores to be used is set by default as the half of the total CPU cores detected. This is done because Vina automatically detects the number of CPU cores of the computer and uses all of them to run the simulation. However, if this was the case, since Vina is very heavy from a computational viewpoint, other computer's functionalities could became very slow.

At the same time, the installation directory of MGL Tools is detected and set, thus allowing the use of the Python scripts described in Section 4.4.1. The user can also modify these parameters to indicate different initial settings.

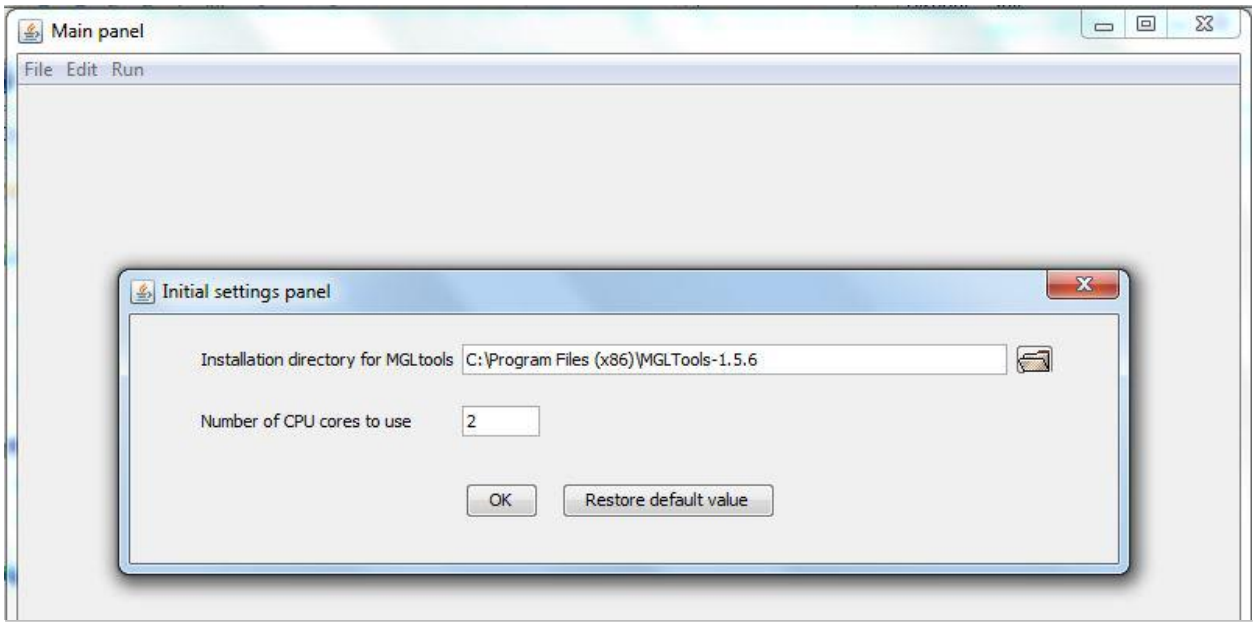


Figure 4.12: **Initial settings panel of the DockingApp GUI.** In this panel the number of CPU cores to use and the installation directory of MGL Tools are set.

The following step is the choice of the application mode, *i.e.* virtual screening or protein-ligand docking. In the first case, the user only needs to indicate the PDB filename of the receptor (target), the way to compute the docking grid, if automatic (extended) or local, and the virtual screening type, if rigid or flexible docking. In fact, the set of ligands to be screened has been

already prepared (see Section 4.4.3) and is provided within the input folder. Thus, it is automatically set as input database in the GUI (Figure 4.13).

The docking panel (Figure 4.14) differs from the latter only for the specification of the ligand to be docked on the target structure. If the user wants to set a local grid, he has to indicate the residues to be used to compute the grid box dimensions and position relative to the target. This is possible using the drop-down menu illustrated in Figure 4.15.

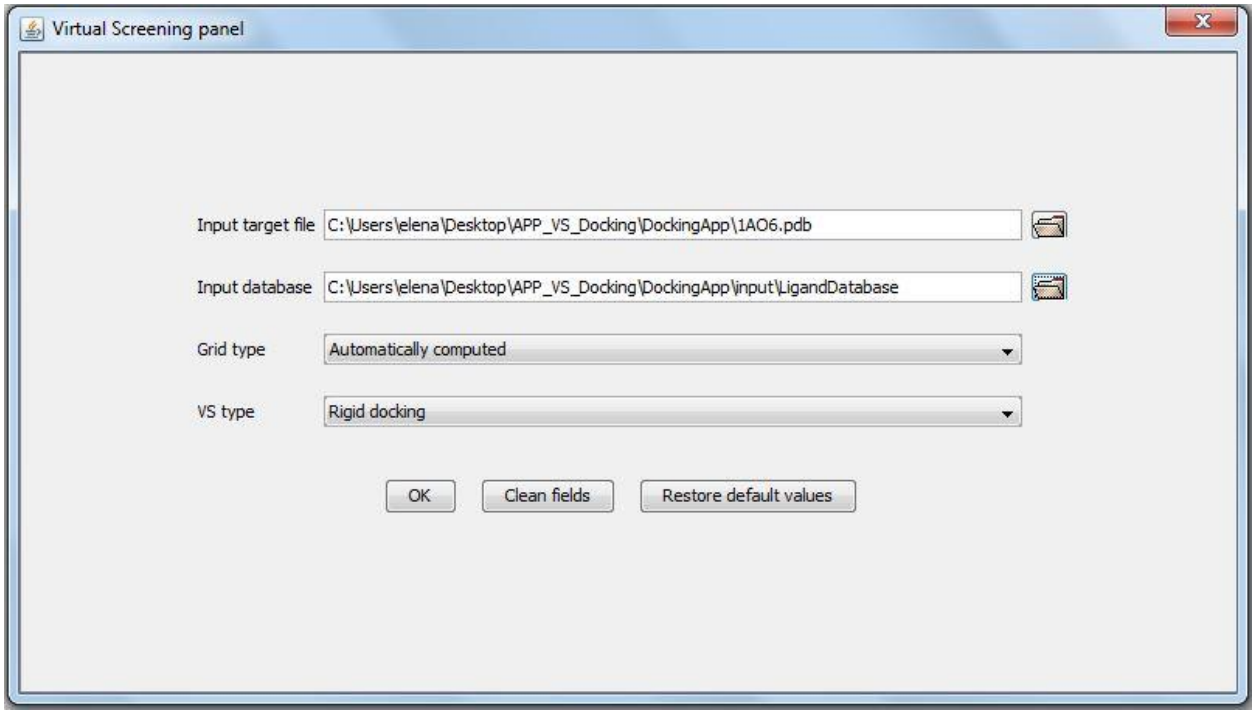


Figure 4.13: **Virtual screening panel of DockingApp.** The input database is already provided and set by default in this panel. The user has only to indicate the input PDB file of the target, the way to compute the docking grid, if automatic (extended) or local, and the virtual screening type, rigid or flexible docking.

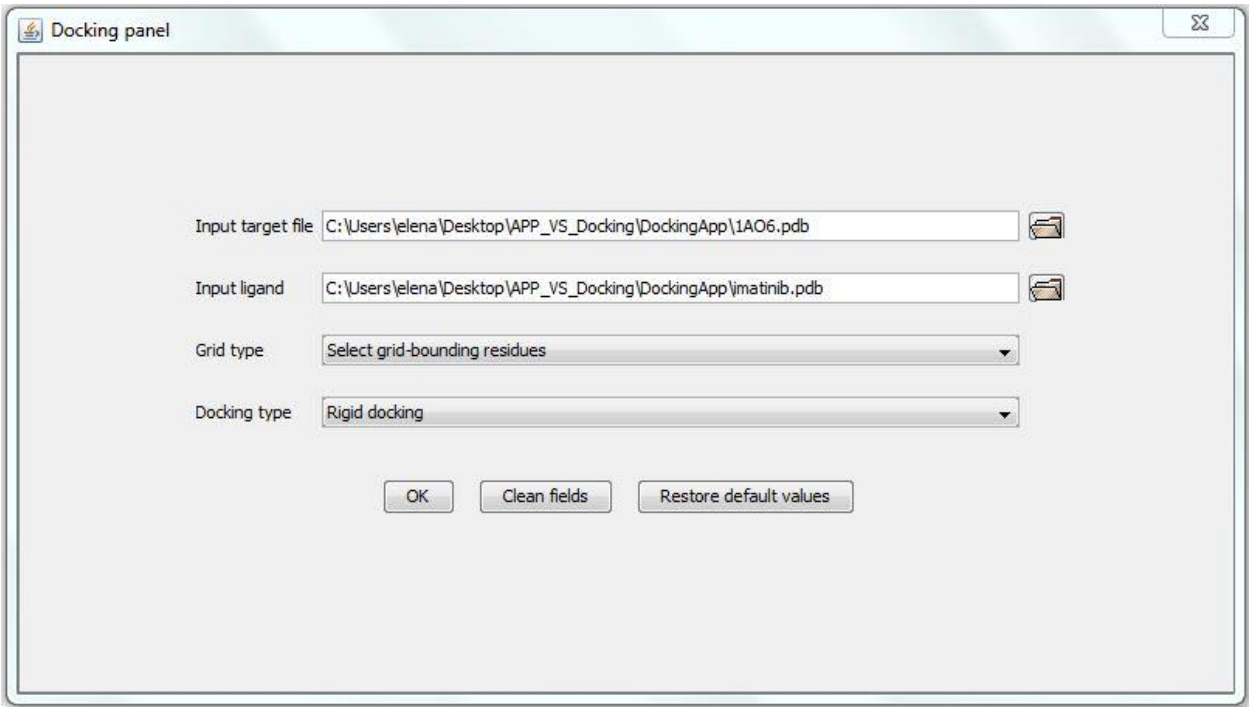


Figure 4.14: **Docking panel of DockingApp.**

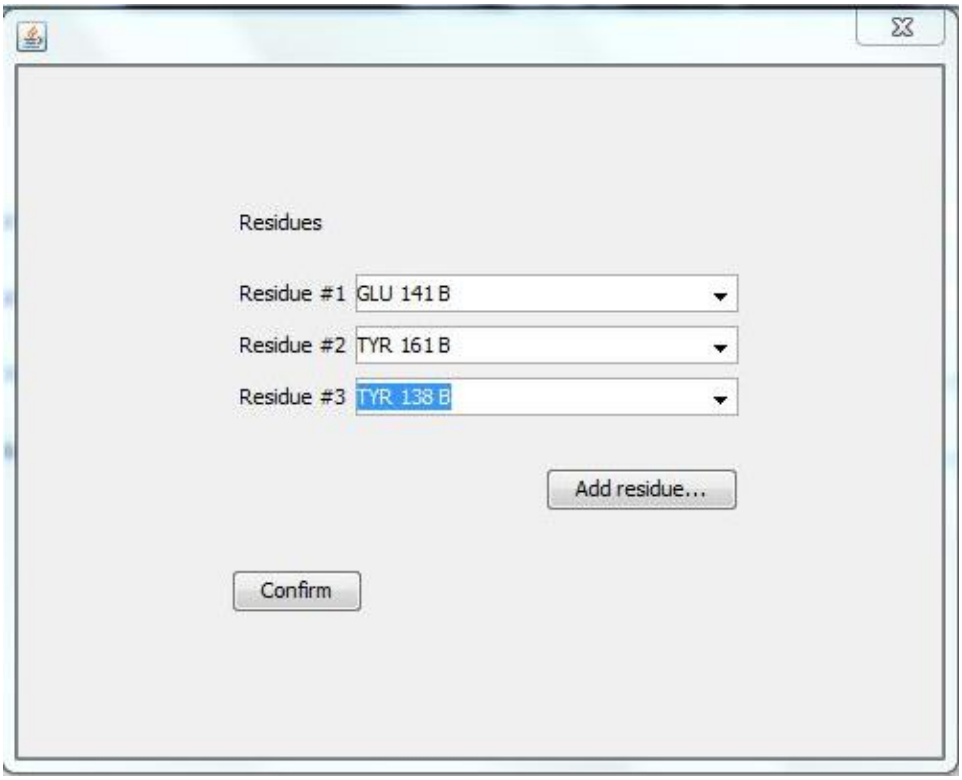


Figure 4.15: **Drop-down menu to specify the residues to be used to calculate the grid box dimensions and position relative to the target**. With a similar menu, the user can also set the target residues for which he wants to allow flexibility.

After confirming all parameters, the docking or virtual screening simulation is launched.

The output of the program is stored in a file that the user can find in the execution directory. The final output panel of the GUI retrieves the information from this file and visualizes the ligand poses obtained for the relative target, which are ranked based on the binding affinity (ΔG). An example of docking output on HSA is reported in Figure 4.16, where nine poses of the ligand REA are shown. In virtual screening analysis, instead, because of the huge number of ligands to be analyzed, only the first pose for each compound (*i.e.* the best in term of binding energy) is reported and ranked with respect to the other compounds.

Within other panels the user can manage target and ligand structures or visualize/hide ligand poses based on some filters indicated in the panel on the right: the user can choose to hide ligands whose predicted binding energy is above a user-defined threshold or the maximum number of ligands to visualize, options that are quite useful in analyzing virtual screening output.

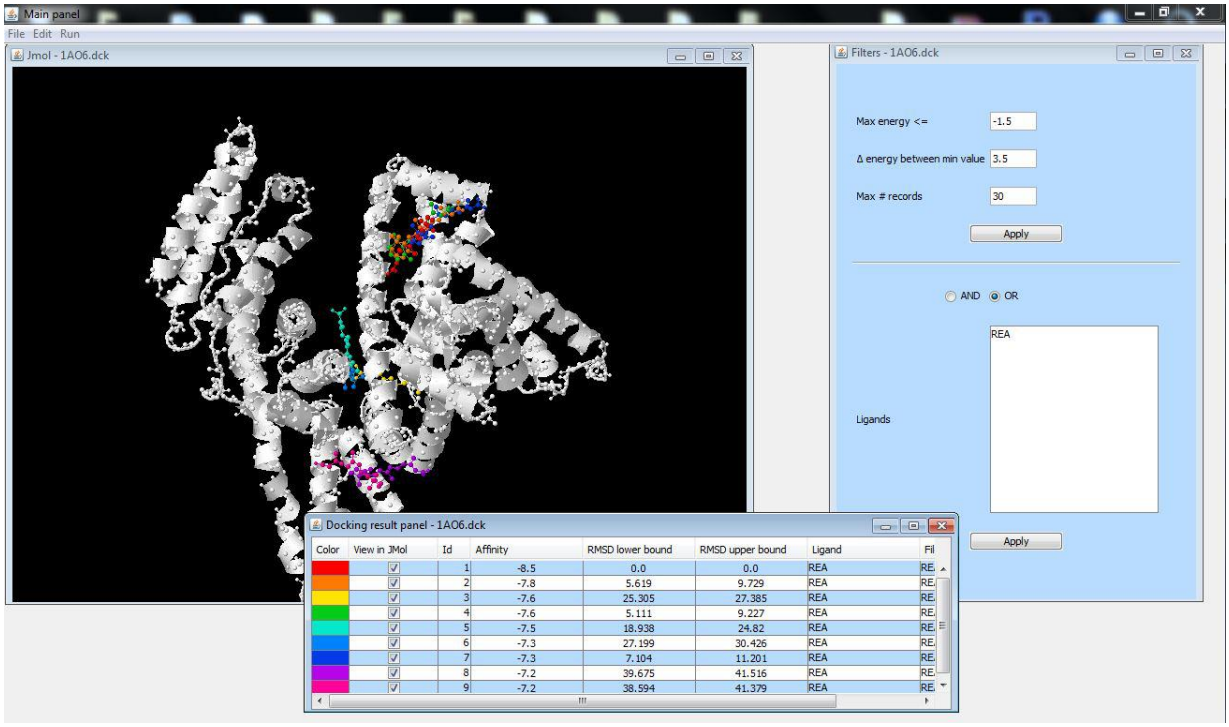


Figure 4.16: **Output panel of DockingApp.** This output was obtained after a simple docking run to analyze the interaction between human serum albumin and the ligand REA.

4.4.5 Communication between Python scripts and GUI

The Python scripts used to prepare the PDBQT files for docking simulation and the function developed to calculate the search space, described in paragraphs 4.4.1 and 4.4.2, were incorporated in the main program that launches the simulation. Two software levels were implemented, that are the control on input parameters, and the output processing, in which there is a communication between the main program and the GUI. When the user sets the initial parameters through the GUI, there is a control on the paths relative to the input target file and to the input ligand (docking version of DockingApp, see Figure 4.14), to verify if these files exist or not in the computer (in the virtual screening version of DockingApp (see Figure 4.13), the control is on the input database's path). If the control ends with a positive result, hence the files exists and the paths are correct, these parameters are written in the 'input_parameters' text file (Figure 4.17), by which the GUI communicates with the main program. Then, the program reads the informations stored in this file and automatically performs all the steps required before Vina docking run (PDBQT files preparation, search space setting, etc.), and automatically launches the simulation.

```
receptor_file = C:\Users\elena\Desktop\APP_VS_Docking\DockingApp
\1AO6.pdb

ligand_folder = C:\Users\elena\Desktop\APP_VS_Docking\DockingApp
\imatinib.pdb

docking = true

automaticGrid = false

gridResidues = 1AO6:B:GLU141_TYR161_TYR138

flexibility = false
```

Figure 4.17: **Example of 'input_parameters' file**. In this example, the user is going to perform a rigid ("flexibility = false") docking simulation ("docking = true") between HSA (see "receptor_file") and imatinib (see "ligand folder"). The search space is set as local (automaticGrid = false), hence residues on which build the grid are set by the user (see gridResidues).

Once the simulation is concluded, two output files are generated, one relative to the ligand conformations (*e.g.*, 'imatinib_out.pdbqt'), the other relative to the binding affinity values calculated by Vina (*e.g.*, 'imatinib_log.txt'). Finally, these files are used to visualize the output through the GUI (see Figure 4.17).

4.5 Application of the procedure to *Pseudomonas aeruginosa* targets

The Gram-negative bacterium *Pseudomonas aeruginosa* is a main cause of infection in hospitalized, burned, immunocompromised, and cystic fibrosis patients. The high intrinsic antibiotic resistance, mostly due to low membrane permeability and secondary resistance mechanisms, such as active drug efflux and antibiotic modification, renders *P. aeruginosa* infections difficult to treat (Breidenstein *et al*., 2011, Lister *et al*., 2009). In this scenario, finding novel compounds that could reduce or inhibit bacterial virulence has elements of uncertainty but successful outcomes may translate into significant results. Five proteins, involved in bacterial virulence or essential for bacterial growth of *P. aeruginosa*, were chosen to perform virtual screening analyses: PqsR, PqsE, LasR, WspR, MurG, all being potential targets for novel antibacterial agents. For these proteins, experimental data were provided by Professor Giordano Rampioni, from the Microbiology Lab of the Department of Science, University of Roma Tre. Moreover, information about their co-crystallized ligand/inhibitor is available, thus allowing the evaluation of the method efficiency in finding the correct binding site of the proteins.

For each target, only the ten highest ranking compounds were analyzed and, to have additional information about each ligand (for example in the case in which two compounds display the same binding affinity value), the SILE parameter was used. SILE stands for Size-Independent Ligand Efficiency (Nissink, 2009) and is derived from the Ligand Efficiency (LE) parameter; the latter is used for optimizing compounds from hit-finding to lead-optimization stages, and is defined as:

$$\mathrm{LE} = \frac{\mathrm{affinity}}{N}$$

where N denotes the number of heavy atoms in the molecule. LE depends strongly on the molecule size, so that the measure tends to decrease with increasing number of atoms, something that can introduce a strong bias particularly for very small molecules. On the other hand, this has implications for applications in drug design and fragment-based drug design in particular, as drug size tends to increase during hit-to-lead optimization. The size-dependency effect equally well applies to *in silico* binding energy estimates. The SILE is the size-independent correction of LE, and is defined as:

$$\mathrm{SILE} = \frac{\mathrm{affinity}}{N^{0.3}}$$

where the exponent 0.3 was estimated to produce a size-independent LE value through the analysis of large set of K_i values collected from the BindingDB database (Chen *et al*., 2002, Liu *et al*., 2007). The correction is expected to be of use in applications like protein-ligand docking, where calculated affinities and scores tend to show a strong dependency on compound size.

4.5.1 PqsR

The LysR-type transcriptional regulator (LTTR) protein PqsR (MvfR) is a key component of alkyl-quinolone (AQ)-dependent *quorum sensing* (QS) in *P. aeruginosa*, a cell-to-cell communication mechanism employing diffusible signal molecules (Ilangovan *et al*., 2013). PqsR is activated by 2-alkyl-4-quinolones including the *Pseudomonas* quinolone signal (PQS; 2-heptyl-3-hydroxy-4(1H)-quinolone, produced from genes encoded in the *pqs* operon, *PqsABCD*), its precursor 2-heptyl-4-hydroxyquinoline (HHQ) and their C9 congeners, 2-nonyl-3-hydroxy-4(1H)-quinolone (C9-PQS) and 2-nonyl-4-hydroxyquinoline (NHQ). These drive the autoinduction of AQ biosynthesis and the up-regulation of key virulence determinants as a function of bacterial population density. Consequently, PqsR constitutes a potential target for novel antibacterial agents, able to attenuate the infection through the reduction of the virulence (Ilangovan *et al*., 2013, Jimenez *et al.*, 2012).

Virtual screening analysis was performed using the crystal structure of PqsR co-inducer binding domain (PDB ID: 4JVC) (Ilangovan *et al*., 2013). In the PDB are available also the crystal structure of PqsR with the ligand NHQ (PDB ID: 4JVD) and with the inhibitor 3-amino-7-chloro-2-nonylquinazolin-4(3H)-one (3NH2-7Cl-C9, ligand ID: QZN) (PDB ID: 4JVI), used in this work to compare experimental and docking results.

Prof. Rampioni provided experimental data obtained from *in vivo* screening of 1600 FDA-approved compound of the Pharmakon library against the QS system *pqs*. In this screening system if a compound inhibits PqsR the activity of a specific biosensor system will be lower.

In Table 4.2 are reported the top ten compounds identified through virtual screening with the FDA-approved set of compounds from Drugbank, *i.e*. those that are predicted to bind PqsR with a higher affinity value (lower ΔG). It resulted that among the top ten probable inhibitors of PqsR, only five are also present in the Pharmakon library, used for the *in vivo* screening. In

particular, Suramin has no effect on the biosensor, while Ergotamin, Pimozide, Dihydroergotamine and Risperidone cause a reduction of the activity of the specific reporter system, even though only at high concentration (200 µM). However, at the same concentration, Dihydroergotamine and Risperidone cause a reduction of the activity of a control reporter system, too. This system in not dependent on PqsR, thus indicating that the effect of Dihydroergotamine and Risperidone is probably not specific.

PqsR				**Residual activity of specific reporter system (%)**		**Residual activity of control reporter system (%)**	
	Binding affinity (kcal/mol)	**SILE**	**Pharmakon**	20 µM	200 µM	20 µM	200 µM
Voacamine	-12,0	-3.668	ABSENT	-	-	-	-
Dutasteride	-11,2	-3.791	ABSENT	-	-	-	-
Ergotamine	-10,9	-3.527	13 H10	93,08	**66,80**	101,35	**102,48**
Eltrombopag	-10,8	-3.783	ABSENT	-	-	-	-
Pimozide	-10,7	-3.715	7 A4	100,05	**66,01**	123,47	**114,99**
Dihydroergotamine	-10,5	-3.397	2 E10	97,57	74,43	75,91	62,33
Suramin	-10,4	-2.733	7 F3	121,18	116,14	88,41	78,78
Canaglifozin	-10,2	-3.641	ABSENT	-	-	-	-
Conivaptan	-10,1	-3.391	ABSENT	-	-	-	-
Risperidone	-10,0	-3.605	15 D3	104,71	78,96	69,94	79,18

Table 4.2: **Top ten compounds from Drugbank identified through virtual screening on PqsR.** Only five compounds from DrugBank are contained in the Pharmakon library used for *in vivo* screening.

Since the effect of Ergotamine and Pimozide was not so marked and being Pimozide an antipsychotic agent, additional studies on these compounds were not carried out. Moreover, no experimental data were available supporting the efficacy of other compounds (highlighted in green in Table 4.2) identified through *in silico* analysis.

In virtual screening analysis, PqsR inhibitor QZN and PqsR ligand NHQ were included in the simulation with the other 1466 FDA-approved compounds from Drugbank, resulting 480th and 524th in the final ranking based on binding energy, with ΔG values of -7.4 e -7.3 kcal/mol, respectively.

In Figure 4.18 is reported the PqsR structure in complex with its ligand NHQ (PDB ID: 4JVD, chain A) and with Voacamine, the compound with the best affinity value. As shown in figure, Voacamine structure was placed in the binding site of PqsR with a better binding affinity value

(-12,0 kcal/mol, see table 4.2). Similarly, the other nine compounds, except Suramin, were placed in the active site of the target.

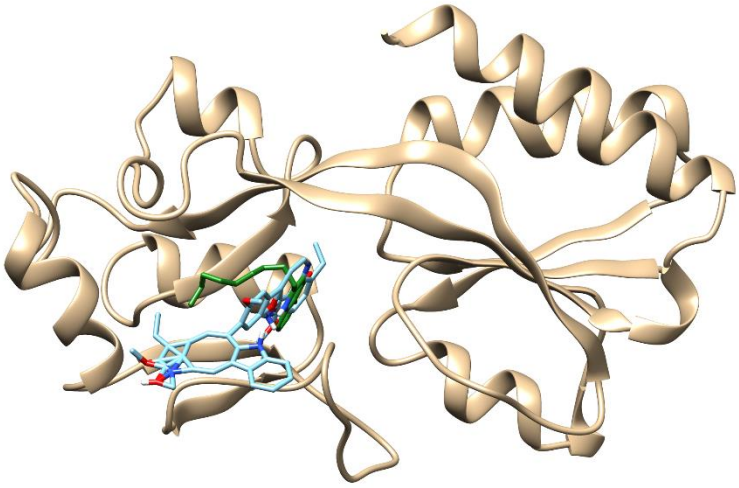


Figure 4.18: **Schematic representation of the PqsR-NHQ-Voacamine complex**. Voacamine (cyan), the compound with the best affinity value, is placed in the binding site of PqsR. The natural ligand NHQ is also shown and coloured in green.

4.5.2 PqsE

In addition to the biosynthetic enzymes PqsABCD, *pqs* operon contains a fifth gene, *pqsE*, that is not required for production of PQS but whose mutation leads to loss of signal transduction in several but not all *pqs* operon dependent processes. PqsE was hence termed "PQS response protein", but its exact mechanism of action is unknown (Yu *et al*., 2009).

Virtual screening analysis was performed using the crystal structure of PqsE with a copurified ligand assigned as benzoate (PDB ID: 2Q0J) (Yu *et al*., 2009).

Pharmakon FDA-approved compound library was used for *in vivo* screening, in which if a compound inhibits PqsE, its effect will result in increasing the activity of a specific reporter system. In table 4.3 are reported the top ten compounds identified through virtual screening with the FDA-approved set from Drugbank. Among the top ten probable inhibitors of PqsE, only four are also present in the Pharmakon library. These compounds do not display any effect on the specific reporter system activity, thus indicating that they are not able to inhibit PqsE. Regarding other compounds (highlighted in green in Table 4.3), no experimental data are available supporting their efficacy.

PqsE				**Residual activity of specific reporter system (%)**		**Residual activity of control reporter system (%)**	
	Binding affinity (kcal/mol)	**SILE**	**Pharmakon**	**20 µM**	**200 µM**	**20 µM**	**200 µM**
Risperidone	-11,2	-4.037	15 D3	114,09	111,17	69,94	79,18
Nandrolone phenpropionate	-11,1	-4.001	ABSENT	-	-	-	-
Voacamine	-11,1	-3.392	ABSENT	-	-	-	-
Eltrombopag	-10,9	-3.818	ABSENT	-	-	-	-
Atovaquone	-10,8	-4.064	11 C11	102,36	106,53	89,43	52,36
Dutasteride	-10,8	-3.656	ABSENT	-	-	-	-
Paliperidone	-10,7	-3.819	8 B9	105,02	81,79	80,30	91,15
Flumethasone Pivalate	-10,6	-3.648	3 A9	96,66	79,51	104,7	89,07
Dabrafenib	-10,6	-3.648	ABSENT	-	-	-	-
Lurasidone	-10,6	-3.648	ABSENT	-	-	-	-

Table 4.3: **Top ten compounds from Drugbank identified through virtual screening on PqsE.**

In Figure 4.19 is reported the complex between PqsE and both its copurified ligand and Risperidone, the compound with the best predicted binding affinity value (-11,2 kcal/mol, see table 4.3) that was placed in the active site of the target.

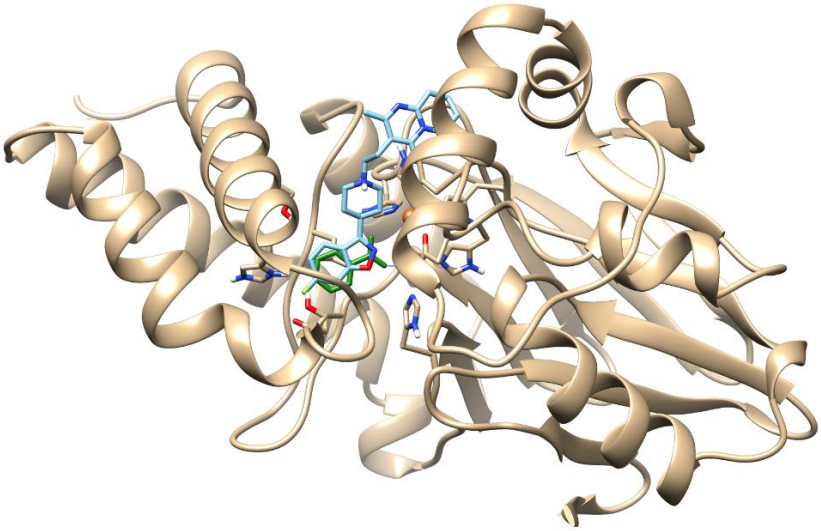


Figure 4.19: **Schematic representation of the crystal structure of PqsE with a cocrystallized ligand assigned as benzoate (PDB ID: 2Q0J, chain A) and Risperidone.** Risperidone (cyan), the compound with the best predicted binding affinity value, is placed in the binding site of PqsE. The cocrystallized ligand is coloured in green.

4.5.3 LasR

The LasI synthase of *P. aeruginosa* constitutively produces the signal 3-oxo-C12-HSL (*N*-3-oxododecanoyl-L-homoserine lactone), which accumulates with population growth and activates LasR, a transcriptional regulator. When activated by 3-oxo-C12-HSL, LasR dimers bind target genes promoters and activate the transcription of many toxic virulence factors, including exoproteases, exotoxins, and secondary metabolites. Activated LasR also participates in the maturation of biofilms, which commonly result in persistent pathogenic infections (Bottomley *et al*., 2007).

Virtual screening analysis was performed using the structure of LasR ligand-binding domain bound to its autoinducer 3-oxo-C12-HSL (ligand ID: OHN) (PDB ID: 2UV0) (Bottomley *et al*., 2007).

The results obtained from the *in silico* screening were compared to those obtained in 2013 by Dr. Francesco Imperi (personal communication), which performed a screening of the Prestwick library (1120 FDA-approved molecules), to identify inhibitors against LasR.

LasR				**Residual activity of specific reporter system (%)**		**Residual activity of control reporter system (%)**	
	Binding affinity (kcal/mol)	**SILE**	**Prestwick**	**10 µg/mL**	**100 µg/mL**	**20 µM**	**200 µM**
Droperidol	-12,4	-4.563	5D 11	127,86	326,94	75,88	54,91
Amlexanox	-11,4	-4.510	ABSENT	-	-	-	-
Dapiprazole	-11,3	-4.355	ABSENT	-	-	-	-
Yohimbine	-11,3	-4.252	8 C5	95,52	251,18	86,92	87,14
Nepafenac	-11	-4.547	ABSENT	-	-	-	-
Minaprine	-11	-4.352	1 G7	153,36	285,64	88,58	96,59
Mitiglinide	-10,9	-4.255	ABSENT	-	-	-	-
Voacamine	-10,8	-3.291	ABSENT	-	-	-	-
Thalidomide	-10,7	-4.423	ABSENT	-	-	-	-
Tretinoin	-10,6	-4.193	ABSENT	-	-	-	-

Table 4.4: **Top ten compounds from Drugbank identified through virtual screening on LasR.**

As shown in table 4.4, among the top ten putative inhibitors of LasR, only three are present in the Prestwick library. These three compounds had not effect on the activity of the specific

reporter system and as regards the other compounds (highlighted in green in the table), no experimental data are available supporting their efficacy.

The best pose obtained from the virtual screening analysis is that of Droperidol, which is predicted to bind to LasR in its active site, in a similar way with respect to the autoinducer compound OHN, as can be seen in Figure 4.20. The other nine compounds, except Voacamine, were placed in the active site of the target.

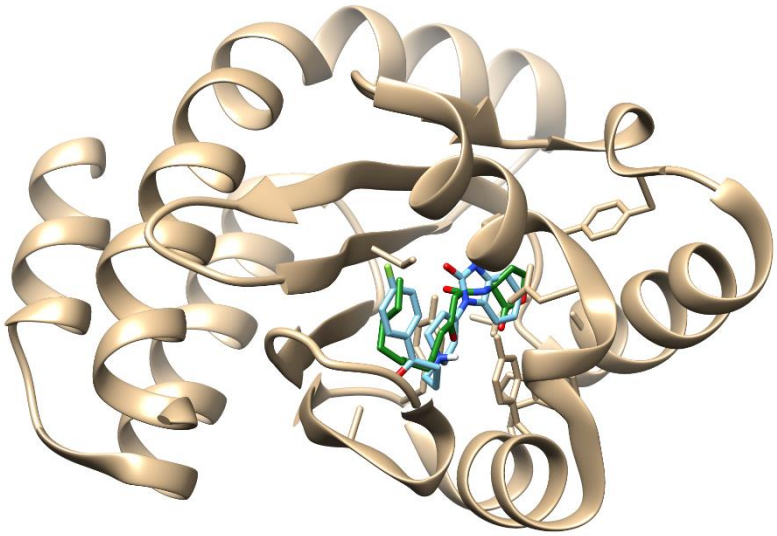


Figure 4.20: **Schematic representation of the crystal structure of LasR ligand-binding domain bound to its autoinducer OHN (PDB ID: 2UV0, chain G) and Droperidol (cyan) .** Droperidol (cyan), the compound with the best affinity value is placed in the binding site of LasR. The copurified ligand is coloured in green.

4.5.4 WspR

WspR is a diguanylate cyclase that contains a GGDEF domain known to catalyze the formation of the cytoplasmic signaling molecule cyclic diguanylate (c-di-GMP) (Hickman *et al*., 2005), a second-messenger signal that plays a key role in bacterial biofilm formation. Overexpression of WspR causes hyperbiofilm formation, whereas loss-of-function mutants show reduced biofilm formation and cytotoxicity, suggesting that WspR is a potent switch controlling virulence mechanisms (De *et al.*, 2008). Moreover, diguanylate cyclases are absent in mammals and hence are considered attractive molecular targets for the development of antibiofilm agents (Fernicola *et al*., 2015).

Virtual screening analysis was performed using the structure of WspR bound to cyclic diguanosine monophosphate (ligand ID: C2E) (PDB ID: 3BRE) (De N. *et al*., 2008).

As shown in table 4.5, among the top ten putative inhibitors of WspR, only six are present in the Prestwick library. These compounds had not effect on the activity of the specific reporter system and regarding other compounds (highlighted in green), no experimental data are available supporting their efficacy.

WspR				**Residual activity of specific reporter system (%)**		**Residual activity of control reporter system (%)**	
	Binding affinity (kcal/mol)	**SILE**	**Pharmakon**	**20 µM**	**200 µM**	**20 µM**	**200 µM**
Voacamine	-9,6	-2.934	ASSENTE	-	-	-	-
Tubocurarine	-9,5	-3.032	5 H7	100,75	103,52	104,14	111,57
c-di-GMP	-8,7	-2.759	ASSENTE	-	-	-	-
Dihydroergotamine	-8,6	-2.230	2 E10	75,91	62,33	97,57	74,34
Dactinomycin	-8,6	-2.783	ASSENTE	-	-	-	-
Nystatin	-8,5	-2.430	4 D6	82,2	87,86	100,19	77,34
Teniposide	-8,4	-2.498	11 A10	111,82	108,87	82,31	126,51
Acetyldigitoxin	-8,4	-2.401	ASSENTE	-	-	-	-
Sirolimus	-8,4	-2.663	11 A7	117,81	110,77	114,48	100,48
Itraconazole	-8,3	-2.362	16 A7	124,38	76,58	98,98	103,45

Table 4.5: **Top ten compounds from Drugbank identified through virtual screening on WspR.**

With respect to other target analyzed, at variance with the previous cases, the WspR substrate (c-di-GMP) resulted 3rd in the final ranking based on binding energy, with a ΔG value of -8.7 kcal/mol. However, the analysis of the complex obtained with docking simulation (Figure 4.21), revealed that this molecule was placed in another site if compared with the crystallographic structure. Similarly, also Voacamine is predicted to bind to WspR in a site different from the protein's active site.

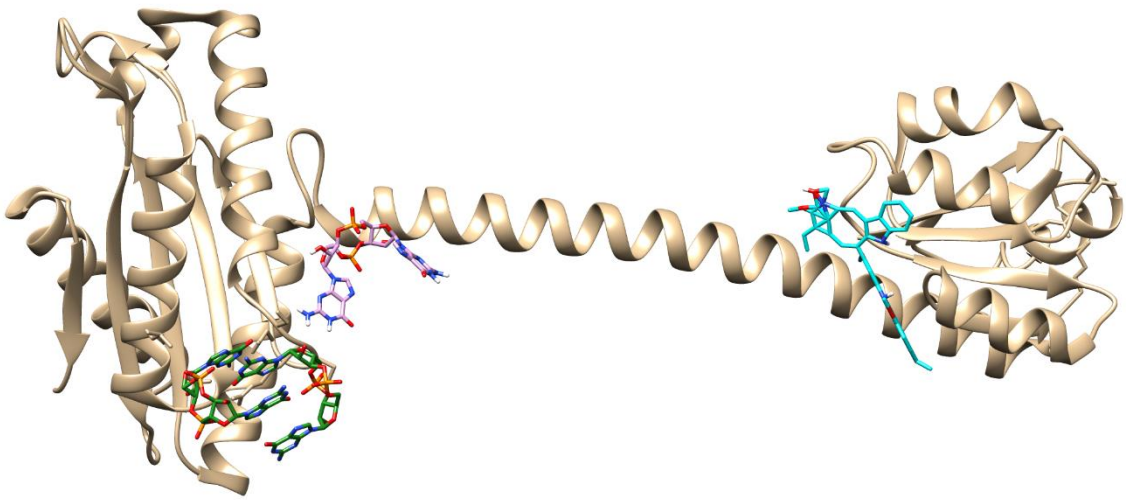


Figure 4.21: **Schematic representation of the crystal structure of WspR bound to cyclic diguanosine monophosphate (c-di-GMP) (PDB ID: 3BRE, chain A) and Voacamine (cyan).** The co-crystallized ligand is shown in green, while the c-di-GMP pose obtained through docking simulations is coloured in magenta.

4.5.5 MurG

MurG is an essential bacterial glycosyltransferase enzyme in *P. aeruginosa* performing one of the key steps of peptidoglycan synthesis, catalyzing the transfer of N-acetyl glucosamine (GlcNAc) from its donor substrate, UDP-GlcNAc, to the acceptor substrate Lipid I (Brown *et al*., 2013). Thus, MurG is essential for bacterial growth and, for this reason, could be a potential target for novel antibacterial agents.

Virtual screening analysis was performed using the structure of MurG in complex with its substrate UDP-GlcNAc (ligand ID: UD1) (PDB ID: 3S2U) (Brown *et al*., 2013).

As reported in Table 4.6, among top ten putative inhibitors identified through *in silico* screening, only four are present in the Pharmakon library. These three compounds had not effect on MurG activity, as resulted from three different screenings carried out by Prof. Rampioni.

MurG				Residual Growth (%)	
	Binding affinity (kcal/mol)	**SILE**	**Pharmakon**	**20 µM**	**200 µM**
Digoxin	-11,1	-3.336	2 E9	94,27	100,58
Voacamine	-10,7	-3.270	ABSENT	-	-
Suramin	-10,5	-2.760	7 F3	93,29	99,11
Dihydroergotamine	-10,4	-3.365	2 E10	95,04	92,33
Diosmin	-10,4	-3.365	18 F9	97,52	116,24
Conivaptan	-10,2	-3.425	ABSENT	-	-
Dolutegravir	-10	-3.503	ABSENT	-	-
Eltrombopag	-10	-3.605	ABSENT	-	-
Ticagrelor	-10	-3.171	ABSENT	-	-
Zafirlukast	-10	-3.282	ABSENT	-	-

Table 4.6: **Top ten compounds from Drugbank identified through virtual screening on MurG.**

Moreover, no data are available supporting the efficacy of the other compounds highlighted in green in table 4.6.

The analysis of the complex obtained by docking simulation (Figure 4.22), revealed that both Digoxin and the re-docked substrate UDP-GlcNAc, were placed in the active site of MurG. The other nine compounds, except Voacamine and Dihydroergotamine, were placed in the active site of the target as well.

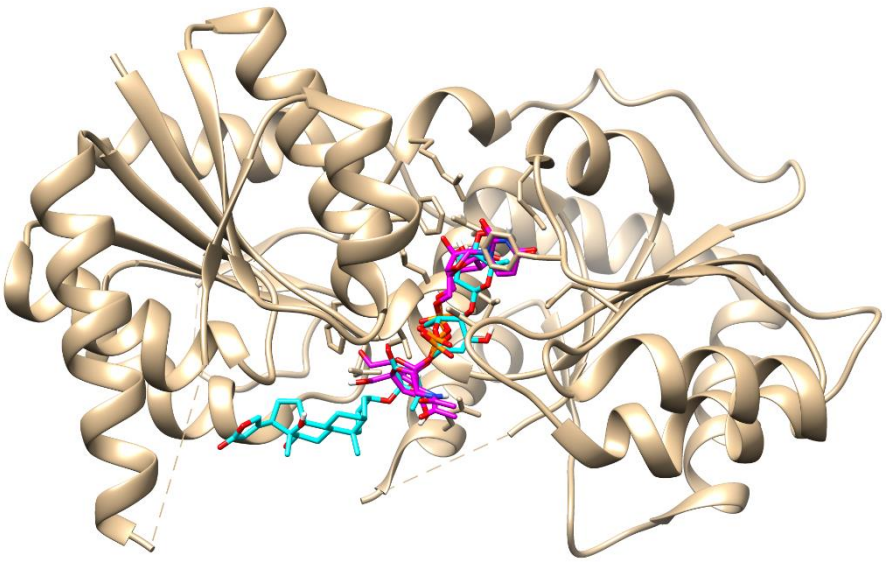


Figure 4.22: **Schematic representation of the crystal structure of MurG in complex with its substrate UDP-GlcNAc (ligand ID: UD1) (PDB ID: 3S2U).** Both Digoxin (cyan) and the re-docked substrate (magenta), were placed in the active site.

5. DISCUSSION

In this work a protein-ligand docking protocol application to biological systems, such as human serum albumin (HSA) and its complex with drug imatinib and with retinoids, is reported. Moreover, a user-friendly procedure which allows to perform molecular docking simulations and virtual screening analyses also to users without any expertise in bioinformatics, is described.

Regarding the analyses on HSA, docking simulations were carried out using AutoDock Vina to investigate the interaction of this protein with the drug imatinib and with retinoids. These simulations were performed both to test the dependency of the results on the specific target structure (ligand-free HSA, fatty acids bound HSA, heme-HSA) and on the simulation parameters (exhaustiveness, maximum number of poses, etc.), and to compare the results obtained with experimental data obtained in the framework of a collaboration with Prof. Paolo Ascenzi of the Department of Sciences, University of Roma Tre.

As resulted from these analyses, imatinib preferentially binds, with similar binding energy, to the FA1 and FA7 sites of ligand-free-HSA (Paragraph 4.2.1 and Figure 4.2), and to the FA2 and FA7 sites of HSA-heme-Fe(III) (Paragraph 4.2.2 and Figure 4.3). On the other hand, docking studies undertaken to complement the experimental characterization of retinoic acid and retinol binding to HSA indicate a preferential binding of these compounds in the FA1 site in FA-free HSA(Paragraph 4.3.1 and Figure 4.4), and in FA8 and FA9 sites in the myristate-HSA complex (Paragraph 4.3.2 and Figure 4.5).

Docking analyses reported in this work provided complementary and supporting evidence to experimental studies, and allowed both to confirm the preferential binding sites of the ligands analysed, and to investigate their interaction with the target protein at an atomic level.

In the second part of this work, the development of a user-friendly application, named “DockingApp”, is described, which allows to perform molecular docking simulations and virtual screening analyses also to users without previous knowledge of docking simulations programs/protocols. A Java graphical user interface for this application was developed (Paragraph 4.4.4), to guide the user through all the steps required to launch the application, *i.e.* docking files preparation (conversion from PDB to PDBQT), both for the receptor and the ligand, search space setting and, finally, docking\virtual screening simulation launching.

Within the virtual screening version of the application, the FDA-approved set of compounds from DrugBank was prepared (Paragraph 4.4.3) and provided within the input folder of DockingApp. The availability of a user-friendly interface for docking simulations, together with a library of ready-to-dock known drugs is seen as particularly useful for the biomedical community. In fact, it is well known that the automated screening against a library of FDA-approved compounds reduces the cost of the initial experimental screening and accelerates lead compounds discovery (Padhy *et al*., 2011).

The virtual screening version of the program was tested with five targets from *Pseudomonas aeruginosa* that are involved in bacterial virulence or are essential for bacterial growth, with the aim to identify those FDA-approved compounds that could have antibacterial activity.

In particular, in this work are reported results obtained for PqsR (Paragraph 4.5.1), PqsE (Paragraph 4.5.2), LasR (Paragraph 4.5.3), WspR (Paragraph 4.5.4) and MurG (Paragraph 4.5.5). In general, the top ten compounds identified for the five targets have no effect on the activity of *Pseudomonas aeruginosa* targets, except in the case of PqsR, for which Ergotamine and Pimozide were identified as possible inhibitors both from the *in vivo* and the *in silico* screening (Paragraph 4.5.1., Table 4.2). Since the effect of these compounds was not so marked and being Pimozide an antipsychotic agent, additional experimental studies on these compounds were not carried out. It should be remarked that the high intrinsic antibiotic resistance, mostly due to low membrane permeability and secondary resistance mechanisms, such as active drug efflux and antibiotic modification, cannot be taken into account during docking simulations. Thus, the *in vivo* inefficacy of many of the compounds analyzed does not imply *per se* an inefficacy of the *in silico* analyses.

Moreover, most of the ligands identified through the *in silico* screening are not present in the library used for the *in vivo* screening, hence no data are available to support or disprove their efficacy. Therefore, it would be interesting to perform additional analyses on these compounds to test their potential antibacterial activity. In fact, from virtual screening analyses resulted that in most cases the best poses (in terms of binding energy) of FDA-approved set of compounds screened are placed in the target's binding site with a better affinity if compared with the co-crystallized ligand.

From this viewpoint, it must be remarked that the virtual screening presented in this work has been carried out keeping all the residues of the targets fixed. The results should thus be refined using a limited number of the top scoring compounds (*e.g*. ten compounds) and allowing the flexibility of the residues building up the walls of the targets' binding sites.

Interestingly, the virtual screening analyses evidenced that seven molecules are putative ligands for different targets (see Table 5.1).

	PqsR	PqsE	LasR	WspR	MurG
Voacamine	-12,0	-11,1	-10,8	-9,6	-10,7
Dihydroergotamine	-10,5	-	-	-8,6	-10,4
Eltrombopag	-10,8	-10,9	-	-	-10,0
Dutasteride	-11,2	-10,8	-	-	-
Risperidone	-10,0	-11,2	-	-	-
Conivaptan	-10,1	-	-	-	-10,2
Suramin	-10,4	-	-	-	-10,5

Table 5.1: **Drugs found by virtual screening as probable ligands for different target**. Reported values indicate the binding affinity calculated (kcal/mol).

Thus, it should be evaluated, through additional studies, if these molecules are interesting from a pharmaceutical point of view, or if they could be involved in non-specific interactions, because of their binding with different proteins. An alternative factor that must be taken into account is the compounds size. In fact, many of the top ranking compounds are relatively big. This could be the result of a bias towards high molecular weight compounds that, for their dimensions, can establish a higher number of intermolecular interactions. In this regard, the compounds could be re-ranked using the SILE parameter, which was used in this work only to re-rank the compounds that displayed an identical value of binding energy.

As a final remark, Voacamine was always among the top ten scoring ligands for all the targets. This compound is an alkaloid isolated from the bark of the *Pescheria fuchsiae folia* tree. It is an antimalarial drug approved for use in several African countries and is under investigation for use in modulating multidrug-resistance in tumor cells. Therefore, it would be interesting to test its activity *in vivo* against *Pseudomonas aeruginosa*.

6. CONCLUSIONS

In the present work, a docking protocol was exploited to study macromolecular systems, such as human serum albumin in complex with the drug imatinib and with retinoids, and to apply virtual screening approach for the identification of compounds that could display activity against *Pseudomonas aeruginosa*. Through the development of an automatic procedure, named DockingApp, the FDA-approved set of compounds from DrugBank was screened, and this allowed to identify some compounds to be tested in further experimental studies. Therefore, this work could represent the starting point for the discovery of new antibacterial molecules which could be exploited to develop novel therapeutics and drugs anti *Pseudomonas aeruginosa*.

Regarding future perspectives, the performance and efficiency of this structure-based method can be improved in various ways. For example, the simulations could be refined used a limited set of high-scoring compounds and repeating the docking procedure allowing the flexibility of the residues building up the identified binding sites. Further, in cases in which it is known which class of molecules is likely to bind the target of interest, the number of the initial compounds could be reduced to those with similar size and readily start the docking process, with advantage in terms of execution time.

Finally, the developed procedure could also be used in protein function prediction when a protein whose function is unknown has no significant similarity to any annotated protein. This could be achieved using databases of biologically relevant small molecules. In this regard it is worthwhile to mention that such a database has already been setup in the lab in which this work has been carried out, extracting more than 20000 small molecule ligands from the experimentally determined complexes available in the PDB (Viet Hung *et al.*, 2015).

7. REFERENCES

Aparoy P., Reddy K.K., Reddanna P., 2012. Structure and ligand based drug design strategies in the development of novel 5- LOX inhibitors. *Curr Med Chem*., 19, 3763-78.

Baek M.C., Jung B., Kang H., Lee H.S., Bae J.S., 2015. Novel insight into drug repositioning: Methylthiouracil as a case in point. *Pharmacol Res*., 99, 185-93.

Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N., Bourne P.E., 2000. The Protein Data Bank. *Nucleic Acids Research*, 28, 235-242.

Bhattacharya A.A., Grüne T., Curry S., 2000. Crystallographic analysis reveals common modes of binding of medium and long-chain fatty acids to human serum albumin. *J Mol Biol.*, 303, 721-32.

Bottomley M.J., Muraglia E., Bazzo R., Carfì A., 2007. Molecular insights into quorum sensing in the human pathogen *Pseudomonas aeruginosa* from the structure of the virulence regulator LasR bound to its autoinducer. *J Biol Chem*., 4, 282, 13592-600.

Breidenstein E.B., de la Fuente-Núñez C., Hancock R.E., 2011. *Pseudomonas aeruginosa*: all roads lead to resistance. *Trends Microbiol*. 19, 419-26.

Brown K., Vial S.C., Dedi N., Westcott J., Scally S., Bugg T.D., Charlton P.A., Cheetham G.M., 2013. Crystal structure of the *Pseudomonas aeruginosa* MurG: UDP-GlcNAc substrate complex. *Protein Pept Lett.*, 20, 1002-8.

Bursulaya B.D., Totrov M., Abagyan R., Brooks C.L. 3rd. 2003. Comparative study of several algorithms for flexible ligand docking. *J Comput Aided Mol Des*.17,755-63.

Calderone V., Berni R., Zanotti G., 2003. High-resolution structures of retinol-binding protein in complex with retinol: pH-induced protein structural changes in the crystal state. J Mol Biol., 329, 841-50.

Cerqueira N.M., Gesto D., Oliveira E.F., Santos-Martins D., Brás N.F., Sousa S.F., Fernandes P.A., Ramos M.J., 2015. Receptor-based virtual screening protocol for drug discovery. *Arch Biochem Biophys*., 582, 56-67.

Changeux J.P. and Edelstein S., 2011. Conformational selection or induced fit? 50 years of debate resolved. *F1000 Biol Rep*, 3, 19.

Chen X., Lin Y., Liu M., Gilson M.K., 2002. The Binding Database: Data Management and Interface Design. *Bioinformatics*, 18, 130-139.

Curry S., Mandelkow H., Brick P., Franks N., 1998. Crystal structure of human serum albumin complexed with fatty acid reveals an asymmetric distribution of binding sites. *Nat Struct Biol*., 5, 827-35.

De N., Pirruccello M., Krasteva P.V., Bae N., Raghavan R.V., Sondermann H., 2008. Phosphorylation-independent regulation of the diguanylate cyclase WspR. PLoS Biol., 6, e67.

di Masi A., Leboffe L., De Marinis E., Pagano F., Cicconi L., Rochette-Egly C., Lo-Coco F., Ascenzi P., Nervi C., 2015. Retinoic acid receptors: from molecular mechanisms to cancer therapy. *Mol Aspects Med*., 41, 1-115.

Di Muzio E., Polticelli F, di Masi A, Fanali G., Fasano M., 2015. Retinoic acid and retinol binding to the FA1 site of human serum albumin competitively inhibits heme-Fe(III) association. *Arch Biochem Biophys*., 590, 56-63.

Di Muzio E., Polticelli F., Trezza V., Fanali G., Fasano M., Ascenzi P., 2014. Imatinib binding to human serum albumin modulates heme association and reactivity. *Arch Biochem Biophys*., 560,100-12.

Ekins S., Mestres J., Testa B., 2007. In silico pharmacology for drug discovery: methods for virtual ligand screening and profiling. *Br J Pharmacol*., 152, 9-20.

Ewing T.J., Makino S., Skillman A.G., Kuntz I.D., 2001. DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *J Comput Aided Mol Des*., 15, 411-28.

Fanali G., Ascenzi P., Bernardi G., Fasano M., 2012a. Sequence analysis of serum albumins reveals the molecular evolution of ligand recognition properties. *J Biomol Struct Dyn*., 29, 91-701.

Fanali G., di Masi A, Trezza V., Marino M., Fasano M., Ascenzi P., 2012. Human serum albumin: from bench to bedside. *Mol Aspects Med.*, 33, 209-90.

Fernicola S., Paiardini A., Giardina G., Rampioni G., Leoni L., Cutruzzolà F., Rinaldo S.,2015. In Silico Discovery and In Vitro Validation of Catechol-Containing Sulfonohydrazide Compounds as Potent Inhibitors of the Diguanylate Cyclase PleD. *J Bacteriol*., 198, 147-56.

Folli C., Calderone V., Ramazzina I., Zanotti G., Berni R., 2002. Ligand binding and structural analysis of a human putative cellular retinol-binding protein. *J Biol Chem*., 277, 41970-7.

Garcia-Ramírez M., Hernández C., Villarroel M., Canals F., Alonso M.A., Fortuny R., Masmiquel L., Navarro A., García-Arumí J., Simó R., 2009. Interphotoreceptor retinoid-binding protein (IRBP) is downregulated at early stages of diabetic retinopathy. *Diabetologia*. 52, 2633-41.

Ghosh S., Nie A., An J., Huang Z., 2006. Structure-based virtual screening of chemical libraries for drug discovery. *Curr Opin Chem Biol*., 10, 194-202.

Ghuman J., Zunszain P.A., Petitpas I., Bhattacharya A.A., Otagiri M., Curry S., 2005. Structural basis of the drug-binding specificity of human serum albumin. *J Mol Biol*., 353, 38-52.

Gohlke H., Hendlich M., Klebe G., 2000. Knowledge-based scoring function to predict protein-ligand interactions. *J. Mol. Biol.*, 295:337-356.

Gupta D. and Lis C.G., 2010. Pretreatment serum albumin as a predictor of cancer survival: a systematic review of the epidemiological literature. *Nutr J*., 22, 9, 69.

Halgren T.A., Murphy R.B, Friesner R.A., Beard H.S., Frye L.L., Pollard W.T., Banks J.L., 2004. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J Med Chem*., 47, 1750-9.

Halperin I., Ma B., Wolfson H., and Nussinov R., 2002. Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins*, 47, 409–443.

Hartmann J.T., Haap M., Kopp H.G., Lipp H.P., 2009. Tyrosine kinase inhibitors - a review on pharmacology, metabolism and side effects. *Curr Drug Metab*., 10, 470-81.

He Y., Ning T., Xie T., Qiu Q, Zhang L., Sun Y., Jiang D., Fu K., Yin F., Zhang W., Shen L., Wang H., Li J., Lin Q., Sun Y., Li H., Zhu Y., Yang D., 2001. Large-scale production of functional human serum albumin from transgenic rice seeds. *Proc Natl Acad Sci U S A*., 108, 19078-83.

Hickman J., Tifrea D., Harwood C., 2005. A chemosensory system that regulates biofilm formation through modulation of cyclic diguanylate levels. *Proc Natl Acad Sci U S A*, 102, 14422–14427.

Huang N., Shoichet B.K., Irwin J.J., 2006. Benchmarking Sets for Molecular Docking. *J. Med. Chem.*, 49 , 6789–6801.

Huang S.Y. and Zou X, 2010. Advances and challenges in protein-ligand docking. *Int J Mol Sci*., 11, 3016-34.

Ilangovan A., Fletcher M., Rampioni G., Pustelny C., Rumbaugh K., Heeb S., Cámara M., Truman A., Chhabra S.R., Emsley J., Williams P., 2013. Structural basis for native agonist and synthetic inhibitor recognition by the *Pseudomonas aeruginosa* quorum sensing regulator PqsR (MvfR). *PLoS Pathog*., 9,e1003508.

Imperi F., Massai F., Ramachandran Pillai C., Longo F., Zennaro E., Rampioni G., Visca P., Leoni L., 2013. New life for an old drug: the anthelmintic drug niclosamide inhibits *Pseudomonas aeruginosa* quorum sensing. *Antimicrob Agents Chemother* 57, 996-1005.

Jimenez P.N., Koch G., Thompson J.A., Xavier K.B., Cool R.H., Quax W.J., 2012. The multiple signaling systems regulating virulence in *Pseudomonas aeruginosa*. *Microbiol. Mol. Biol Rev.* 76, 46–65.

Jones G., Willett P., Glen R.C., Leach A.R., Taylor R.D., 1997. Development and validation of a genetic algorithm for flexible docking. *J Mol Biol*., 267, 727-48.

Kitchen D.B., Decornez H., Furr J.R., Bajorath J., 2004. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug. Discov.*, 3, 935-948.

Kotzé S.R., Zinyama-Gutsire R., Kallestrup P., Benn C.S., Gomo E., Gerstoft J., van Dam G., Mortensen O.H., Ullum H., Erikstrup C., 2015. HIV and schistosomiasis in rural Zimbabwe: the association of retinol-binding protein with disease progression, inflammation and mortality. *Int J Infect Dis*., 33, 159-64.

Kramer B., Rarey M., Lenaguer T., 1999. Evaluation of the FlexX incremental construction algorithm for protein-ligand docking. *Proteins*, 37, 228-241.

Kremer J.M., Wilting J., Janssen L.H., 1988. Drug binding to human alpha-1-acid glycoprotein in health and disease. *Pharmacol Rev*., 40, 1-47.

Kuntz I.D., Blaney J.M., Oatley S.J., Langridge R., Ferrin T.E., 1982. A geometric approach to macromolecule–ligand interactions. *J. Mol. Biol.*, 161, 269-288.

Kuruvilla A.P., Hochwald G.M., Ghiso J., Castaño E.M., Pizzolato M., Frangione B., 1991. Isolation and amino terminal sequence of beta-trace, a novel protein from human cerebrospinal fluid. *Brain Res*., 565, 337-40.

Lahiri S.D., Johnstone M.R., Ross P.L., McLaughlin R.E., Olivier N.B., Alm R.A., 2014. Avibactam and class C β-lactamases: mechanism of inhibition, conservation of the binding pocket, and implications for resistance. *Antimicrob Agents Chemother*., 58, 5704-13.

Lister P.D., Wolter D.J., Hanson N.D., 2009. Antibacterial-resistant *Pseudomonas aeruginosa*: clinical impact and complex regulation of chromosomally encoded resistance mechanisms. *Clin. Microbiol. Rev*. 22, 582-610.

Liu T., Lin Y., Wen X., Jorissen R.N., Gilson M.K., 2007. BindingDB: A Web-Accessible Database of Experimentally Determined Protein-ligand Binding Affinities. *Nucleic Acids Res*., 35, D198–D201.

Morris G.M., Goodsell D.S., Halliday R.S., Huey R., Hart W.E., Belew R.K., Olson A.J., 1998. Automated docking using a Lamarckian Genetic algorithm and an empirical binding free energy function. *J. of Comput. Chem.*, 19, 1639-1662.

Morris G.M., Huey R., Lindstrom W., Sanner M.F., Belew R. K., Goodsell D.S., Olson A.J., 2009. Autodock4 and AutoDockTools4: automated docking with selective receptor flexiblity. *J. Computational Chemistry*, 16, 2785-91.

Muegge I. and Rarey M., 2001. Small molecule docking and scoring. *Rev. Comput. Chem.*, 17, 1-60.

Nissink J.W., 2009. Simple size-independent measure of ligand efficiency. *J Chem Inf Model*., 49, 1617-22.

Nocedal J. and Wright S.J., 1999. Numerical Optimization. Springer Verlag Berlin, Springer Series in Operations Research.

O'Byrne S.M. and Blaner W.S., 2013. Retinol and retinyl esters: biochemistry and physiology. *J Lipid Res.*, 54, 1731-43.

Padhy B.M. and Gupta Y.K., 2011. Drug repositioning: re-investigating existing drugs for new therapeutic indications. *J Postgrad Med*., 57, 153-60.

Petitpas I., Petersen C.E., Ha C.E., Bhattacharya A.A., Zunszain P.A., Ghuman J., Bhagavan N.V., Curry S., 2003. Structural basis of albumin-thyroxine interactions and familial dysalbuminemic hyperthyroxinemia. *Proc Natl Acad Sci U S A*. 100, 6440-5.

Pettersen E.F., Goddard T.D., Huang C.C., Couch G.S., Greenblatt D.M., Meng E.C., Ferrin T.E., 2004. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem.* 25, 1605-12.

Rarey M., Kramer B., Lengauer T., Klebe G., 1996. A fast flexible docking method using an incremental construction algorithm. *J Mol Biol*., 261, 470-89.

Renaud J.P., Rochel N., Ruff M., Vivat V., Chambon P., Gronemeyer H., Moras D., 1995. Crystal structure of the RAR-gamma ligand-binding domain bound to all-trans retinoic acid. *Nature*, 378, 681-9.

Sliwoski G., Kothiwale S., Meiler J., Lowe E.W. Jr., 2014. Computational methods in drug discovery. *Pharmacol Rev*., 66, 334-95.

Sousa S.F., Fernandes P.A., Ramos M.J., 2006. Protein-ligand docking: current status and future challenges. *PROTEINS: Structure, Function, and Bioinformatics*, 65, 15-26.

Sousa S.F., Ribeiro A.J., Coimbra J.T., Neves R.P., Martins S.A., Moorthy N.S., Fernandes P.A., Ramos M.J., 2013. Protein-ligand docking in the new millennium--a retrospective of 10 years in the field. *Curr Med Chem*., 20, 2296-314.

Sugio S., Kashima A., Mochizuki S., Noda M., Kobayashi K., 1999. Crystal structure of human serum albumin at 2.5 Å resolution. *Protein Eng.*, 12, 439-46.

Talele T.T., Khedkar S.A., Rigby A.C., 2010. Successful applications of computer aided drug discovery: moving drugs from concept to the clinic. *Curr Top Med Chem.,* 10, 127–141.

Trott O. and Olson A.J., 2009. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* 31, 455-461.

Van Drie J.H., 2007. Computer-aided drug design: the next 20 years. *J Comput Aided Mol Des*., 21, 591–601.

Viet Hung L., Caprari S., Bizai M., Toti D., Polticelli F., 2015. LIBRA: LIgand Binding site Recognition Application. *Bioinformatics*. 31, 4020-2.

Vijayakrishnan R., 2009. Structure-based drug design and modern medicine. *J Postgrad Med*., 55, 301–304.

Wang R., Lai L., Wang S., 2002. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J-Comput. Aided Mol. Des.*, 16, 11-26.

Wang, R., Fang X., Lu Y., Wang S., 2004. The PDBbind Database: Collection of Binding Affinities for Protein-Ligand Complexes with Known Three-Dimensional Structures. J. *Med. Chem.*, 47, 2977-2980.

Wardell M., Wang Z., Ho J.X., Robert J., Ruker F., Ruble J., Carter D.C., 2002. The atomic structure of human methemalbumin at 1.9 Å. *Biochem Biophys Res Commun.*, 291, 813-9.

Wishart D.S., Knox C., Guo A.C, Shrivastava S., Hassanali M., Stothard P., Chang Z., Woolsey J., 2006. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res*. 34, D668-72.

Yamanishi Y., Araki M., Gutteridge A., Honda W., Kanehisa M., 2008. Prediction of drug-target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics*, 24, i232-40.

Yu S., Jensen V., Seeliger J., Feldmann I., Weber S., Schleicher E., Häussler S., Blankenfeldt W., 2009. Structure elucidation and preliminary assessment of hydrolase activity of PqsE, the *Pseudomonas* quinolone signal (PQS) response protein. Biochemistry 48, 10298 –10307.

Zahn M., Mäder M., Schmidt B., Bollensen E., Felgenhauer K., 1993. Purification and N-terminal sequence of beta-trace, a protein abundant in human cerebrospinal fluid. *Neurosci Lett*., 154, 93-5.

Zhang W.X., Zhou W., Zhang Z.M., Zhang Z.Q., He J.F., Shi B.Y., 2014. Decreased retinol-binding protein 4 in the sera of patients with end-stage renal disease after kidney transplantation. *Genet Mol Res.*, 13, 8126-34.

SUPPLEMENTARY MATERIALS

Reprints of papers published

In this chapter, the reprints of the three papers published during the Ph.D. course are reported. The first two papers concern the experimental and computational studies of interaction between human serum albumin (HSA) and the drug imatinib (Di Muzio *et al*., 2014) and retinoids (Di Muzio *et al*., 2015), respectively.

The third paper describes the biochemical and structural characterization of the invertebrate PAO enzyme from the Pacific oyster (*Crassostrea gigas*), named Cgi-PAO, in which docking simulations were used to study the molecular bases of the substrate specificity of this enzyme (Cervelli *et al*., 2015).

Archives of Biochemistry and Biophysics 560 (2014) 100–112

Contents lists available at ScienceDirect

Archives of Biochemistry and Biophysics

journal homepage: www.elsevier.com/locate/yabbi

Imatinib binding to human serum albumin modulates heme association and reactivity

Elena Di Muzio [a], Fabio Polticelli [a,b], Viviana Trezza [a], Gabriella Fanali [c], Mauro Fasano [c], Paolo Ascenzi [d,e,*]

[a] Department of Sciences, Roma Tre University, I-00146 Roma, Italy
[b] National Institute of Nuclear Physics, Roma Tre Section, I-00146 Roma, Italy
[c] Biomedical Research Division, Department of Theoretical and Applied Sciences and Center of Neuroscience, University of Insubria, I-21052 Busto Arsizio, VA, Italy
[d] Interdepartmental Laboratory for Electron Microscopy, Roma Tre University, I-00146 Roma, Italy
[e] National Institute of Biostructures and Biosystems, I-00136 Roma, Italy

ARTICLE INFO

Article history:
Received 22 May 2014
and in revised form 25 June 2014
Available online 21 July 2014

Keywords:
Human serum albumin
Human serum heme-albumin
Imatinib binding
Peroxynitrite detoxification
Allosteric effects
Molecular docking

ABSTRACT

Imatinib, an inhibitor of the Bcr-Abl tyrosine kinase, is approximately 95% bound to plasma proteins, α1-acid glycoprotein (AGP) being the primary carrier. However, human serum albumin (HSA) may represent the secondary carrier of imatinib in pathological states characterized by low AGP levels, such as pancreatic cancer, hepatic cirrhosis, hepatitis, hyperthyroidism, nephrotic syndrome, malnutrition, and cachexia. Here, thermodynamics of imatinib binding to full-length HSA and its recombinant Asp1-Glu382 truncated form (containing only the FA1, FA2, FA6, and FA7 binding sites; trHSA), in the absence and presence of ferric heme (heme-Fe(III)), and the thermodynamics of heme-Fe(III) binding to HSA and trHSA, in the absence and presence of imatinib, has been investigated. Moreover, the effect of imatinib on kinetics of peroxynitrite detoxification by ferric human serum heme-albumin (HSA-heme-Fe(III)) and ferric truncated human serum heme-albumin (trHSA-heme-Fe(III)) has been explored. All data were obtained at pH 7.0, and 20.0 °C and 37.0 °C. Imatinib binding to the FA7 site of HSA and trHSA inhibits allosterically heme-Fe(III) association to the FA1 site and *vice versa*, according to linked functions. Moreover, imatinib binding to the secondary FA2 site of HSA-heme-Fe(III) inhibits allosterically peroxynitrite detoxification. Docking simulations and local structural comparison with other imatinib-binding proteins support functional data indicating the preferential binding of imatinib to the FA1 and FA7 sites of HSA, and to the FA2 and FA7 sites of HSA-heme-Fe(III). Present results highlight the allosteric coupling of the FA1, FA2, and FA7 sites of HSA, and may be relevant in modulating ligand binding and reactivity properties of HSA *in vivo*.

Introduction

Imatinib is a potent and selective kinase inhibitor approved in the treatment of chronic myelogenous leukemia (CML)[1] and metastatic gastrointestinal stromal tumor (GIST). Imatinib acts by blocking the tyrosine kinase activity of key proteins involved in the pathogenesis of CML and GIST [1]. In particular, imatinib inhibits the oncogenic tyrosine kinase Bcr-Abl, that is found in 95% of patients with CML. Furthermore, imatinib inhibits the deregulated tyrosine kinase activity of the mutated receptor c-kit, that is found in up to 90% of metastatic GISTs [2,3]. The mechanism of action of imatinib also includes inhibition of the tyrosine kinase activity of c-kit, the receptor for kit ligand, and of the mutated platelet-derived growth factor receptors α and β that are observed in approximately 35% of metastatic GISTs [4,5].

Pharmacokinetic studies showed that, following oral administration, imatinib is well absorbed and has an absolute bioavailability of 98% irrespective of the pharmaceutical form (solution, capsule, and tablet) or dose used (100 mg or 400 mg) [2,3]. After repeated administration of imatinib, at the dose of 400 mg once a day, a mean peak plasma concentration of 2.6 ± 0.8 μg/mL (*i.e.*, *ca.* 5×10^{-6} M) was reached [3,6]. Imatinib is approximately 95% bound to plasma proteins. Although α1-acid glycoprotein (AGP) plasma concentration in healthy subjects ($1–2 \times 10^{-5}$ M) is much lower than that of human serum albumin (HSA; 7×10^{-4} M),

* Corresponding author at: Interdepartmental Laboratory for Electron Microscopy, Roma Tre University, Via della Vasca Navale 79, I-00146 Roma, Italy. Fax: +39 06 57336321.
E-mail address: ascenzi@uniroma3.it (P. Ascenzi).

[1] *Abbreviations used:* apoMb, heme-free recombinant sperm whale Mb; CML, chronic myelogenous leukemia; FA, fatty acid; GIST, gastrointestinal stromal tumor; HSA, human serum albumin; HSA-heme-Fe(III), ferric human serum heme-albumin; Mb, ferric recombinant sperm whale myoglobin; trHSA, truncated HSA; trHSA-heme-Fe(III), ferric truncated human serum heme-albumin.

http://dx.doi.org/10.1016/j.abb.2014.07.001
0003-9861/

AGP plays an important role in plasma binding of many basic drugs including imatinib [7].

In vitro studies showed that the dissociation equilibrium constant for imatinib binding to AGP (6×10^{-7} M) is lower than that for drug association to HSA (4×10^{-6} M to 3×10^{-5} M), indicating that AGP is the primary carrier of imatinib in plasma. Indeed, HSA does not affect the pharmacokinetic properties of imatinib [8–13]. Accordingly, erythromycin, an AGP but not a HSA specific ligand, has been reported to increase the pharmacologically-active free fraction of imatinib in plasma [14].

Although the plasma level of AGP is markedly elevated in malignant diseases, in contrast to the decline of HSA level [15–19], half of the normal plasma levels of AGP have been found in patients with hepatic cirrhosis and hepatitis [20]; therefore, in these pathological states, HSA may serve as a secondary carrier for imatinib.

Here, thermodynamics of imatinib binding to HSA and its recombinant Asp1-Glu382 truncated form (trHSA), in the absence and presence of ferric heme (heme-Fe(III)), and the thermodynamics of heme-Fe(III) binding to HSA and trHSA, in the absence and presence of imatinib, are reported. Moreover, the effect of imatinib on kinetics of peroxynitrite detoxification by ferric human serum heme-albumin (HSA-heme-Fe(III)) and ferric truncated human serum heme-albumin (trHSA-heme-Fe(III)) has been explored. Remarkably, imatinib binding to the FA7 site of HSA modulates allosterically heme-Fe(III) binding to the FA1 site and *vice versa*. Moreover, drug binding to the FA2 site impairs the heme-Fe(III)-based reactivity of HSA-heme-Fe(III). Experimental data have been interpreted in the light of docking simulations and analysis of local structural similarity between the available HSA three-dimensional structures and a set of 15 other imatinib-binding proteins structures. These analyses support the conclusions drawn from the functional data and provide an atomic level view of the imatinib-HSA(-heme-Fe(III)) complexes.

Materials

HSA (⩾96%, essentially FA-free) was obtained from Sigma–Aldrich (St. Louis, MO, USA). The HSA concentration was determined spectrophotometrically at 280 nm ($\varepsilon_{280\ nm}$ = 38.2×10^3 M^{-1} cm^{-1}) [21]. The HSA stock solution (2.0×10^{-4} M) was prepared by dissolving HSA in 2.0×10^{-2} M sodium phosphate buffer, at pH 7.0.

Recombinant FA-free trHSA was expressed and purified as previously reported [22]. The trHSA concentration was determined according to Bradford [23]. The trHSA stock solution (2.2×10^{-4} M) was prepared by dissolving trHSA in 2.0×10^{-2} M sodium phosphate buffer, at pH 7.0.

Ferric recombinant sperm whale myoglobin (Mb) and hemin (Fe(III)-protoporphyrin IX) chloride were purchased from Sigma–Aldrich (St. Louis, MO, USA). The heme stock solution (5.0×10^{-3} M) was prepared by dissolving heme-Fe(III) in 1.0×10^{-2} M NaOH [21]. The heme-Fe(III) concentration was determined spectrophotometrically at 535 nm, after converting heme-Fe(III) to the heme-Fe(III)-bis-imidazolate derivative by adding 1.0 M imidazole, in sodium dodecyl sulfate micelles ($\varepsilon_{535\ nm}$ = 1.45×10^4 M^{-1} cm^{-1}) [24].

Heme-free recombinant sperm whale Mb (apoMb) was prepared by the cold (−20.0 °C) acetone–HCl method. The apoMb concentration was determined by titration with heme-Fe(III) [25].

The HSA-heme-Fe(III) and trHSA-heme-Fe(III) stock solutions (1.6×10^{-4} M and 1.8×10^{-4} M, respectively) were prepared by adding a 0.8-M defect of the heme-Fe(III) stock solution to the HSA and trHSA solutions. The HSA-heme-Fe(III) concentration was determined spectrophotometrically at 403 nm ($\varepsilon_{403\ nm}$ = 1.1×10^5 M^{-1} cm^{-1}) [26]. The trHSA-heme-Fe(III) concentration was determined spectrophotometrically at 402 nm ($\varepsilon_{404\ nm}$ = 1.2×10^5 M^{-1} cm^{-1}).

Imatinib mesylate (4-(4-methyl-piperazin-1-yl-methyl)-*N*-[4-methyl-3-(4-pyridin-3-yl-pyrimidin-2-yl-amino)-phenyl]-benzamide methanesulfonate) was provided by Novartis Pharma AG (Basel, Switzerland). The imatinib stock solution (2.0×10^{-3} M) was prepared by dissolving the drug in water at pH 5.0. The final imatinib concentration ranged between 2.0×10^{-5} M and 2.0×10^{-3} M.

Dansyl-arginine and dansyl-sarcosine were obtained from Sigma–Aldrich (St. Louis, MO, USA). The dansyl-arginine stock solution (8.0×10^{-4} M) was prepared by dissolving the compound in 2.0×10^{-2} M phosphate buffer, at pH 7.0, in the presence of 10% (v/v) dimethylsulfoxide [27–31]. The final dansyl-arginine concentration ranged between 5.0×10^{-6} M and 2.0×10^{-4} M. The dansyl-sarcosine stock solution (2.0×10^{-4} M) was prepared by dissolving the compound in 2.0×10^{-2} M phosphate buffer, at pH 7.0, in the presence of 10% (v/v) dimethylsulfoxide [27–31]. The final dansyl-sarcosine concentration ranged between 5.0×10^{-6} M and 5.0×10^{-5} M.

Peroxynitrite was synthesized from KO_2 and NO and from HNO_2 and H_2O_2 and stored in small aliquots at −80.0 °C [32,33]. The peroxynitrite stock solution (2.0×10^{-3} M) was diluted immediately before use with degassed 5.0×10^{-2} M NaOH to reach the desired concentration [34–37]. The nitrate and nitrite contaminations were in the range of 0–8% and 7–18% of the peroxynitrite concentration, respectively. The nitrate and nitrite analysis was carried out spectrophotometrically at 543 nm by using the Griess reagent and VCl_3 to reduce NO_3^- to NO_2^-, as described previously [34,35, 38–42]. The concentration of peroxynitrite was determined spectrophotometrically prior to each experiment by measuring the absorbance at 302 nm ($\varepsilon_{302\ nm}$ = 1.705×10^3 M^{-1} cm^{-1}) [32,33].

All the other chemicals were obtained from Sigma–Aldrich and Merck AG (Darmstadt, Germany). All products were of analytical or reagent grade and were used without further purification.

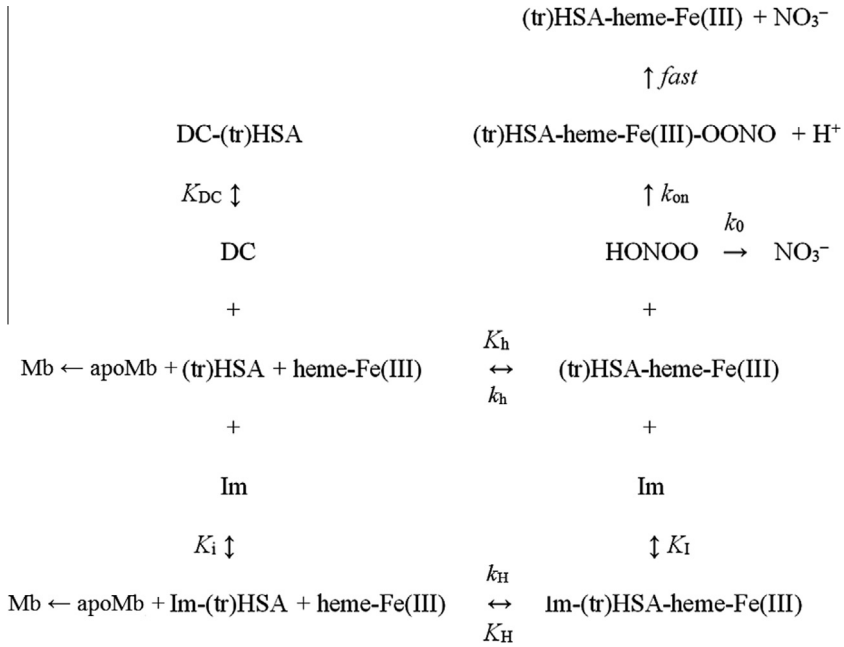
Methods

Thermodynamics of imatinib binding to HSA and trHSA, in the absence and presence of heme-Fe(III), thermodynamics and kinetics of heme-Fe(III) binding to HSA and trHSA, in the absence and presence of imatinib, and kinetics of peroxynitrite detoxification by HSA-heme-Fe(III) and trHSA-heme-Fe(III), in the absence and presence of imatinib, were analyzed in the framework of the reaction mechanism depicted by Scheme 1, where DC indicates dansyl-arginine or dansyl-sarcosine and Im indicates imatinib.

Imatinib binding to HSA and trHSA

Imatinib binding to HSA was followed by competitive inhibition of dansyl-arginine and dansyl-sarcosine association, at pH 7.0 (1.0×10^{-1} M phosphate buffer) and 20.0 °C and 37.0 °C. Imatinib binding to trHSA was followed by competitive inhibition of dansyl-arginine association, at pH 7.0 (1.0×10^{-1} M phosphate buffer) and 20.0 °C. The fluorophore of dansyl-arginine and dansyl-sarcosine was excited at 370 nm; the fluorescence emission intensities were measured at the maximum wavelengths (*i.e.*, 460 nm for dansyl-arginine, and at 475 nm for dansyl-sarcosine); the excitation and emission slits were 5 nm [27–31]. The HSA concentration was 3.2×10^{-6} M, the trHSA concentration was 2.9×10^{-6} M, the dansyl-arginine and dansyl-sarcosine concentration ranged between 5.0×10^{-6} M and 2.0×10^{-4} M, and the imatinib concentration ranged between 2.0×10^{-5} M and 2.0×10^{-3} M.

Values of the dissociation equilibrium constant for dansyl-arginine and/or dansyl-sarcosine binding to HSA and trHSA were



Scheme 1.

obtained from the dependence of the relative fluorescence intensity change (*i.e.*, $\Delta F/\Delta F_{max}$) of the HSA-dansyl-arginine, HSA-dansyl-sarcosine, and trHSA-dansyl-arginine complexes on the dansylated compound concentration (*i.e.*, [DC]) in the absence (*i.e.*, K_{DC}) and presence (*i.e.*, K_{DC}^{app}) of imatinib, according to Eq. (1) [27–31]:

$$\Delta F/\Delta F_{max} = Y = [DC]/(K_{DC} + [DC]) \quad (1)$$

where K_{DC} is K_{DCda} or K_{DCda}^{app} for dansyl-arginine binding and K_{DCds} or K_{DCds}^{app} for dansyl-sarcosine binding, ΔF is the fluorescence intensity change observed at each concentration of dansyl-arginine and dansyl-sarcosine, and ΔF_{max} is the maximum fluorescence intensity change.

The values of the dissociation equilibrium constant for imatinib binding to HSA and trHSA (*i.e.*, K_i) were obtained from the dependence of the K_{DC}^{app}/K_{DC} ratio on the imatinib concentration (*i.e.*, [imatinib], according to Eq. (2) [31,43]:

$$K_{DC}^{app}/K_{DC} = (1/K_i) \times [\text{imatinib}] + 1 \quad (2)$$

Test measurements performed after 2 h of imatinib/HSA/dansyl-arginine, imatinib/HSA/dansyl-sarcosine, and imatinib/trHSA/dansyl-arginine incubation excluded slow kinetic effects.

Spectrofluorimetric measurements were carried out with the Jasco FP-6500 spectrofluorimeter (Jasco International Co., Ltd., Tokyo, Japan).

Imatinib binding to HSA-heme-Fe(III) and trHSA-heme-Fe(III)

Imatinib binding to HSA-heme-Fe(III) was followed spectrophotometrically between 350 and 450 nm, at pH 7.0 (1.0×10^{-1} M phosphate buffer) and 20.0 °C and 37.0 °C. Moreover, the formation of the imatinib-trHSA-heme-Fe(III) complex was followed spectrophotometrically between 350 and 450 nm, at pH 7.0 (1.0×10^{-1} M phosphate buffer) and 20.0 °C. The HSA-heme-Fe(III) concentration was 3.0×10^{-6} M, the trHSA-heme-Fe(III) concentration was 2.7×10^{-6} M, and the imatinib concentration ranged between 2.0×10^{-5} M and 5.0×10^{-4} M.

Values of the dissociation equilibrium constant for imatinib binding to HSA-heme-Fe(III) and trHSA-heme-Fe(III) (*i.e.*, K_I; see Scheme 1) were obtained from the dependence of the relative absorbance change (*i.e.*, $\Delta A/\Delta A_{max}$) of the imatinib-HSA-heme-Fe(III) and imatinib-trHSA-heme-Fe(III) complexes on the imatinib concentration (*i.e.*, [imatinib]), according to Eq. (3) [26]:

$$\Delta A/\Delta A_{max} = Y = [\text{imatinib}]/(K_I + [\text{imatinib}]) \quad (3)$$

where ΔA is the absorbance change at a given drug concentration, and ΔA_{max} is the absorbance change at the saturating drug concentration.

Test measurements performed after 2 h of imatinib/HSA-heme-Fe(III) and imatinib/trHSA-heme-Fe(III) incubation excluded slow kinetic effects.

Spectrophotometric measurements were carried out with the Jasco V-560 spectrophotometer (Jasco International Co., Ltd., Tokyo, Japan).

Heme-Fe(III) binding to HSA and trHSA in the absence and presence of imatinib

Heme-Fe(III) binding to HSA and imatinib-HSA was followed spectrophotometrically between 350 and 450 nm, at pH 7.0 (1.0×10^{-1} M phosphate buffer) and 20.0 °C and 37.0 °C. Heme-Fe(III) binding to trHSA and imatinib-trHSA was followed spectrophotometrically between 350 and 450 nm, at pH 7.0 (1.0×10^{-1} M phosphate buffer) and 20.0 °C. The heme-Fe(III) concentration was 7.0×10^{-7} M, the imatinib concentration ranged between 1.0×10^{-6} M and 6.0×10^{-4} M, and the HSA and trHSA concentration ranged between 1.1×10^{-7} M and 1.2×10^{-6} M. Heme-Fe(III) binding to HSA in the absence and presence of imatinib was investigated at a fixed low heme-Fe(III) concentration by increasing the amount of HSA. This approach minimizes the problems with heme-Fe(III) dimerization and the absorbance changes are easy to interpret.

Values of the dissociation equilibrium constant for heme-Fe(III) binding to HSA and trHSA in the absence and presence of saturating imatinib concentration (*i.e.*, K_h and K_H, respectively; see Scheme 1) have been obtained from the dependence of the relative absorbance change (*i.e.*, $\Delta A/\Delta A_{max}$) of HSA-heme-Fe(III), imatinib-HSA-heme-Fe(III), trHSA-heme-Fe(III), and imatinib-trHSA-heme-Fe(III) on the

HSA or trHSA concentration (*i.e.*, [HSA] or [trHSA]), according to Eq. (4) [26]:

$$\Delta A/\Delta A_{max} = Y = [(tr)HSA]/(K + [(tr)HSA]) \quad (4)$$

where K is K_h or K_H.

Test measurements performed after 2 h of imatinib/HSA-heme-Fe(III) and imatinib/trHSA-heme-Fe(III) incubation excluded slow kinetic effects.

According to linked functions [44], the dependence of values of the dissociation equilibrium constants for heme-Fe(III) binding to HSA and trHSA (*i.e.*, K_h^{app}) on the imatinib concentration was analyzed according to Eq. (5) [44]:

$$\mathrm{Log}K_h^{app} = \mathrm{Log}K_h - \mathrm{Log}(([\mathrm{Imatinib}] + K_I)/([\mathrm{Imatinib}] + K_i)) + \mathrm{Log}(K_I/K_i) \quad (5)$$

where [Imatinib] indicates the imatinib concentration.

According to competitive inhibition [43], the dependence of K_h^{app} values on the imatinib concentration was analyzed according to Eq. (6) [43]:

$$\mathrm{Log}K_h^{app} = \mathrm{Log}(((K_h/K_i) \times [\mathrm{Imatinib}]) + K_h) \quad (6)$$

Values of the first-order rate constant for heme-Fe(III) dissociation from trHSA-heme-Fe(III) in the absence and presence of imatinib (*i.e.*, k_h and k_H, respectively; see Scheme 1) have been determined spectrophotometrically (between 370 nm and 420 nm), at pH 7.0 (1.0×10^{-1} M phosphate buffer) and 20.0 °C, by mixing trHSA-heme-Fe(III) with sperm whale apoMb according to Eq. (7) [25]:

$$[\mathrm{trHSA\text{-}heme\text{-}Fe(III)}]_t = [\mathrm{trHSA\text{-}heme\text{-}Fe(III)}]_i \times e^{-k \times t} \quad (7)$$

where k is k_h or k_H.

Spectrophotometric measurements were carried out with the Jasco V-560 spectrophotometer (Jasco International Co., Ltd., Tokyo, Japan) and the Bio-Logic SFM 2000 rapid-mixing stopped-flow apparatus (Bio-Logic Science Instruments, Claix, France).

Peroxynitrite isomerisation by HSA-heme-Fe(III) and trHSA-heme-Fe(III) in the absence and presence of imatinib

Kinetics of peroxynitrite isomerisation by HSA-heme-Fe(III), in the absence and presence of imatinib, was recorded at 302 nm ($\varepsilon = 1.705 \times 10^3$ M^{-1} cm^{-1}) by rapid-mixing the HSA-heme-Fe(III) solution with the peroxynitrite solution, at pH 7.0 (1.0×10^{-1} M phosphate buffer) and 20.0 °C and 37.0 °C. Kinetics of peroxynitrite isomerisation by trHSA-heme-Fe(III), in the absence and presence of imatinib, was recorded at 302 nm ($\varepsilon = 1.705 \times 10^3$ M^{-1} cm^{-1}) by rapid-mixing the trHSA-heme-Fe(III) solution with the peroxynitrite solution, at pH 7.0 (1.0×10^{-1} M phosphate buffer) and 20.0 °C. The HSA-heme-Fe(III) and trHSA-heme-Fe(III) concentration ranged between 5.0×10^{-6} M and 4.0×10^{-5} M, the peroxynitrite concentration was 2.5×10^{-4} M, and the imatinib concentration ranged between 2.0×10^{-4} M and 2.0×10^{-3} M.

Values of the pseudo-first-order rate constant for the HSA-heme-Fe(III)- and trHSA-heme-Fe(III)-mediated peroxynitrite isomerisation (*i.e.*, k^{obs}) were determined from the analysis of the time-dependent absorbance decrease at 302 nm, according to Eq. (8) [34,35,39–42,45]:

$$[\mathrm{peroxynitrite}]_t = [\mathrm{peroxynitrite}]_i \times e^{-kobs \times t} \quad (8)$$

Values of the second-order rate constant for the HSA-heme-Fe(III)- and trHSA-heme-Fe(III)-mediated peroxynitrite isomerisation (*i.e.*, k_{on}; see Scheme 1) and of the first-order rate constant for peroxynitrite isomerisation in the absence of HSA-heme-Fe(III) and trHSA-heme-Fe(III) (*i.e.*, k_0; see Scheme 1) were determined from the linear dependence of k^{obs} on the HSA-heme-Fe(III) and trHSA-heme-Fe(III) concentration, according to Eq. (9) [34,35,39–42,45]:

$$k^{obs} = k_{on} \times [(tr)\mathrm{HSA\text{-}heme\text{-}Fe(III)}] + k_0 \quad (9)$$

The value of the dissociation equilibrium constant for imatinib binding to HSA-heme-Fe(III) and trHSA-heme-Fe(III) (*i.e.*, K_I; see Scheme 1) was determined from the dependence of k_{on} values on the imatinib concentration (*i.e.*, [imatinib]), according to Eq. (10) [39–42,45,46]:

$$k_{on} = k_{on}^0 - (k_{on}^0 \times [\mathrm{imatinib}])/(K_I + [\mathrm{imatinib}]) \quad (10)$$

where k_{on}^0 indicates the k_{on} value obtained in the absence of imatinib.

Test measurements performed after 2 h of imatinib/HSA-heme-Fe(III)/peroxynitrite and imatinib/trHSA-heme-Fe(III)/peroxynitrite incubation excluded slow kinetic effects.

Kinetic measurements were carried out with the Bio-Logic SFM 2000 rapid-mixing stopped-flow apparatus (Bio-Logic Science Instruments, Claix, France).

Data analysis

Thermodynamics and kinetics were analyzed using the MatLab program (The Math Works Inc., Natick, MA, USA). The results are given as mean values of at least four experiments plus or minus the corresponding standard deviation.

Docking of imatinib to HSA and HSA-heme-Fe(III)

Docking simulations of imatinib binding to HSA were performed using the crystal structures of the ligand-free HSA (PDB ID: 1AO6) [47] and of HSA-heme-Fe(III) (PDB ID: 1N5U) [48].

Simulations were carried out using the docking program Autodock Vina [49] with a search space (docking grid) that included the whole protein, in order to carry out "blind" predictions of the imatinib binding site(s). Additional simulations were also carried out restricting the search space only to the FA2, FA6 and FA7 sites. The grid spacing was set to 1 Å per grid unit and the exhaustiveness parameter was increased from the default value of 8 to 24 as suggested by Autodock Vina developers for grid sizes larger than 27,000 $Å^3$, which is the case for HSA simulations.

Initially the number of solutions (docking poses) generated by Autodock Vina was set to the default value of 9, but simulations were also carried out with a number of solutions of 20 (see Results Section Docking analyses). The simulations were carried out both by keeping all protein residues rigid and by allowing flexibility of the residues building up the walls of the sites from FA1 to FA7 (see [26]). Dihedral angles involving single bonds of the flexible residues are, by default, varied by a 50 degrees increment during Autodock Vina execution. Residues for which flexibility was allowed are reported in Supplementary Materials (Table S1). Rotatable bonds of the imatinib lowest energy resonance structure were kept flexible in all the simulations, as suggested by Autodock Vina developers (see http://vina.scripps.edu/tutorial.html).

Analysis of local structural similarity between HSA and imatinib binding proteins

The analysis of the local structural similarities between HSA and a set of imatinib binding proteins was carried out using the in-house developed local structure comparison program ASSIST [50]. This program is based on a geometric hashing approach to find the largest subset of similar residues between an input protein structure and a set of structural motifs. A search of the PDB

retrieved a total of 9 different imatinib binding proteins with PDB ID: 1T46 [51], 1XBB [52], 2OIQ [53], 2PL0 [54], 3FW1 [55], 3GVU (to be published), 3HEC [56], 3MS9 [57] and 3PYY [58]. The structural motifs to be compared with HSA structures were composed of the residues involved in the binding of imatinib in the 9 proteins as determined by LigPlot+ [59].

Results

Imatinib binding to HSA

Imatinib binding to HSA was followed by competitive inhibition of dansyl-arginine and dansyl-sarcosine association, at 20.0 °C and 37.0 °C, whereas imatinib binding to trHSA was followed by competitive inhibition of dansyl-arginine association, at 20.0 °C. Competitive inhibition of dansyl-arginine and dansyl-sarcosine association reflects ligand binding to the FA7 site and the FA3–FA4 cleft, respectively [27–31]. In the absence and presence of imatinib, dansyl-arginine and/or dansyl-sarcosine binding to HSA and trHSA follows a simple equilibrium (according to Eq. (1); see Fig. 1, panels A–D), the average values of the Hill coefficient (*i.e.*, n) being 1.00 ± 0.02 and 0.99 ± 0.01, respectively. Notably, values of the dissociation equilibrium constant for dansyl-arginine and dansyl-sarcosine binding to HSA, at 20.0 °C, in the absence of imatinib ($K_{DCda} = (2.2 \pm 0.2) \times 10^{-5}$ M and $K_{DCds} = (5.1 \pm 0.5) \times 10^{-6}$ M) are in agreement with those reported in the literature [27,29,31].

Imatinib affects only dansyl-arginine binding to HSA and trHSA. In fact, values of the apparent equilibrium constant for dansyl-arginine binding to HSA and trHSA in the presence of imatinib (*i.e.*, K^{app}_{DAda}) increase linearly with the imatinib concentration, as expected for a simple competitive binding mechanism (according to Eq. (2); see Fig. 1, panels B–E). On the other hand, values of the apparent equilibrium constant for dansyl-sarcosine binding to HSA (*i.e.*, K^{app}_{DAds}) are unaffected by imatinib (see Fig. 1, panels A and E). The analysis of data according to Eq. (2) allows the

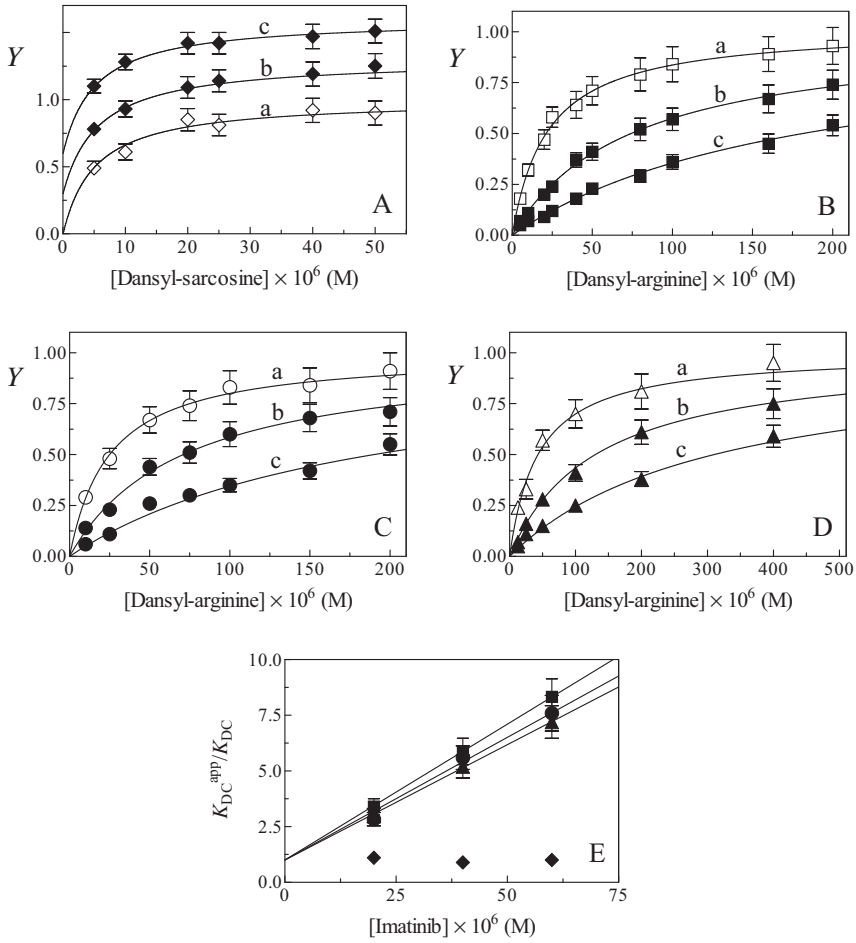


Fig. 1. Effect of imatinib on binding isotherms for dansyl-arginine and dansyl-sarcosine binding to HSA and trHSA, at pH 7.0. (A) Binding isotherms for dansyl-sarcosine association to HSA in the absence (a) and presence of 2.0×10^{-5} M (b) and 6.0×10^{-5} M (c) imatinib, at 20.0 °C. For clarity, values of $\Delta F/\Delta F_{max}$ (corresponding to Y) of the binding isotherms (b) and (c) have been arbitrarily up shifted of 0.3 and 0.6 units, respectively, with respect to those of the isotherm (a). The analysis of data according to Eq. (1) allowed the determination of the following values of $K_{DAds} = (5.1 \pm 0.5) \times 10^{-6}$ M (a), $K^{app}_{DAds} = (5.6 \pm 0.5) \times 10^{-6}$ M (b), and $K^{app}_{DAds} = (5.2 \pm 0.5) \times 10^{-6}$ M (c). (B) Binding isotherms for dansyl-arginine association to HSA in the absence (a) and presence of 2.0×10^{-5} M (b) and 6.0×10^{-5} M (c) imatinib, at 20.0 °C. The analysis of data according to Eq. (1) allowed the determination of the following values of $K_{DAda} = (2.2 \pm 0.2) \times 10^{-5}$ M (a), $K^{app}_{DAda} = (7.5 \pm 0.7) \times 10^{-5}$ M (b), and $K^{app}_{DAda} = (1.8 \pm 0.2) \times 10^{-4}$ M (c). (C) Binding isotherms for dansyl-arginine association to trHSA in the absence (a) and presence of 2.0×10^{-5} M (b) and 6.0×10^{-5} M (c) imatinib, at 20.0 °C. The analysis of data according to Eq. (1) allowed the determination of the following values of $K_{DAda} = (2.5 \pm 0.3) \times 10^{-5}$ M (a), $K^{app}_{DAda} = (7.1 \pm 0.7) \times 10^{-5}$ M (b), and $K^{app}_{DAda} = (1.9 \pm 0.2) \times 10^{-4}$ M (c). (D) Binding isotherms for dansyl-arginine association to HSA in the absence (a) and presence of 2.0×10^{-5} M (b) and 6.0×10^{-5} M (c) imatinib, at 37.0 °C. The analysis of data according to Eq. (1) allowed the determination of the following values of $K_{DAda} = (4.3 \pm 0.4) \times 10^{-5}$ M (a), $K^{app}_{DAda} = (1.3 \pm 0.1) \times 10^{-4}$ M (b), and $K^{app}_{DAda} = (3.1 \pm 0.3) \times 10^{-4}$ M (c). (E) Dependence of the K^{app}_{DAda}/K_{DAda} ratio for dansyl-arginine binding to HSA at 20.0 °C (squares), trHSA at 20.0 °C (circles), and HSA at 37.0 °C (triangles), and of the K^{app}_{DAds}/K_{DAds} ratio for dansyl-sarcosine binding to HSA at 20.0 °C (diamonds) on the imatinib concentration. The analysis of data according to Eq. (2) allowed the determination of the following values of K_i: HSA at 20.0 °C, $(8.3 \pm 0.8) \times 10^{-6}$ M; trHSA at 20.0 °C, $(9.1 \pm 0.9) \times 10^{-6}$ M; and HSA at 37.0 °C, $(9.6 \pm 1.0) \times 10^{-6}$ M. Values of K^{app}_{DAds}/K_{DAds} are unaffected by imatinib. Where not shown, the standard deviation is smaller than the symbol. For details, see text.

determination of values of the dissociation equilibrium constant for imatinib binding to HSA and trHSA in the absence of dansyl-arginine (*i.e.*, K_i), corresponding to the inverse of the slope of the linear plot (see Table 1). Remarkably, values of K_i for imatinib binding to HSA and trHSA, at 20.0 °C, match each other very well.

As a whole, data reported in Fig. 1 suggest that imatinib binds selectively to the FA7 site of HSA, which accommodates dansyl-arginine, but not to the FA3–FA4 cleft that binds dansyl-sarcosine [27–31].

Imatinib binding to HSA-heme-Fe(III) and trHSA-heme-Fe(III)

Imatinib binding to HSA-heme-Fe(III) was investigated spectrophotometrically, at 20.0 °C and 37.0 °C. Moreover, imatinib binding to trHSA-heme-Fe(III) was investigated spectrophotometrically, at 20.0 °C. The analysis of data reported in Fig. 2 according to Eq. (3) indicates that imatinib binds to HSA-heme-Fe(III) and trHSA-heme-Fe(III) with a simple equilibrium, the values of the Hill coefficient (*i.e.*, n) ranging between 0.99 ± 0.02 and 1.01 ± 0.02. Values of the dissociation equilibrium constant for imatinib binding to HSA-heme-Fe(III) and trHSA-heme-Fe(III) (*i.e.*, K_I) are reported in Table 1. Remarkably, values of K_I for imatinib binding to HSA-heme-Fe(III) and trHSA-heme-Fe(III), at 20.0 °C, match each other very well.

Heme-Fe(III) binding to HSA and trHSA in the absence and presence of imatinib

Heme-Fe(III) binding to the FA1 site of HSA [26,47,59,60] was investigated spectrophotometrically in the absence and presence of imatinib (ranging between 1.0×10^{-6} M and 6.0×10^{-4} M), at 20.0 °C and 37.0 °C. Moreover, the formation of the trHSA-heme-Fe(III) complex was investigated spectrophotometrically in the absence and presence of imatinib (ranging between 1.0×10^{-6} M and 6.0×10^{-4} M), at 20.0 °C.

The analysis of data (Fig. 3, panels A–C) according to Eq. (4) shows that heme-Fe(III) binds to HSA, trHSA, imatinib-HSA, and imatinib-trHSA with a simple equilibrium, values of the Hill coefficient (*i.e.*, n) ranging between 0.99 ± 0.02 and 1.02 ± 0.02. Values of the dissociation equilibrium constant for heme-Fe(III) binding to HSA, trHSA, imatinib-HSA, and imatinib-trHSA (K_h and K_H,

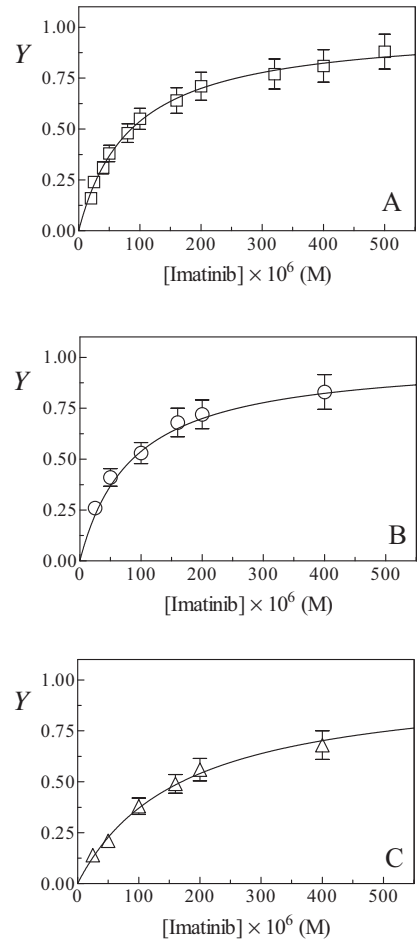


Fig. 2. Imatinib binding to HSA-heme-Fe(III) and trHSA-heme-Fe(III), at pH 7.0. (A) Binding isotherm for imatinib association to HSA-heme-Fe(III), at 20.0 °C. (B) Binding isotherm for imatinib association to trHSA-heme-Fe(III), at 20.0 °C. (C) Binding isotherm for imatinib association to HSA-heme-Fe(III), at 37.0 °C. The analysis of data according to Eq. (3) allowed the determination of the following values of $K_I = (8.6 \pm 0.9) \times 10^{-5}$ M (A), $(8.6 \pm 0.9) \times 10^{-5}$ M (B), and $(1.7 \pm 0.2) \times 10^{-4}$ M (C). Where not shown, the standard deviation is smaller than the symbol. For details, see text.

Table 1
Values of the dissociation equilibrium constants for imatinib binding to HSA, trHSA, HSA-heme-Fe(III), and trHSA-heme-Fe(III), and for heme-Fe(III) binding to HSA, trHSA, imatinib-HSA, and imatinib-trHSA (pH 7.0).

Ligand	HSA species	Equilibrium constant (M)	Temperature (°C)	FA site
Imatinib	HSA	$(8.3 \pm 0.8) \times 10^{-6}$	20.0	FA7
	trHSA	$(9.1 \pm 0.9) \times 10^{-6}$	20.0	FA7
	HSA	$(9.6 \pm 1.0) \times 10^{-6}$	37.0	FA7
Imatinib [a]	HSA-heme-Fe(III)	$(8.6 \pm 0.8) \times 10^{-5}$	20.0	FA7
	trHSA-heme-Fe(III)	$(8.6 \pm 0.8) \times 10^{-5}$	20.0	FA7
	HSA-heme-Fe(III)	$(1.7 \pm 0.2) \times 10^{-4}$	37.0	FA7
Heme-Fe(III)	HSA	$(3.4 \pm 0.4) \times 10^{-8}$	20.0	FA1
	trHSA	$(4.2 \pm 0.4) \times 10^{-8}$	20.0	FA1
	HSA	$(6.5 \pm 0.7) \times 10^{-8}$	37.0	FA1
Heme-Fe(III)	Imatinib-HSA	$(3.5 \pm 0.4) \times 10^{-7}$	20.0	FA1
	Imatinib-trHSA	$(3.3 \pm 0.3) \times 10^{-7}$	20.0	FA1
	Imatinib-HSA	$(9.1 \pm 1.0) \times 10^{-7}$	37.0	FA1
Imatinib [b]	HSA-heme-Fe(III)	$(7.1 \pm 0.7) \times 10^{-4}$	20.0	FA2
	trHSA-heme-Fe(III)	$(7.1 \pm 0.7) \times 10^{-4}$	20.0	FA2
	HSA-heme-Fe(III)	$(1.1 \pm 0.1) \times 10^{-3}$	37.0	FA2

[a] Values of the dissociation equilibrium constants for imatinib binding to HSA-heme-Fe(III) and trHSA-heme-Fe(III) were obtained spectrophotometrically from the dependence of $\Delta A/\Delta A_{max}$ on the drug concentration.
[b] Values of the dissociation equilibrium constants for imatinib binding to HSA-heme-Fe(III) and trHSA-heme-Fe(III) were determined from the dependence of k_{on} on the drug concentration.

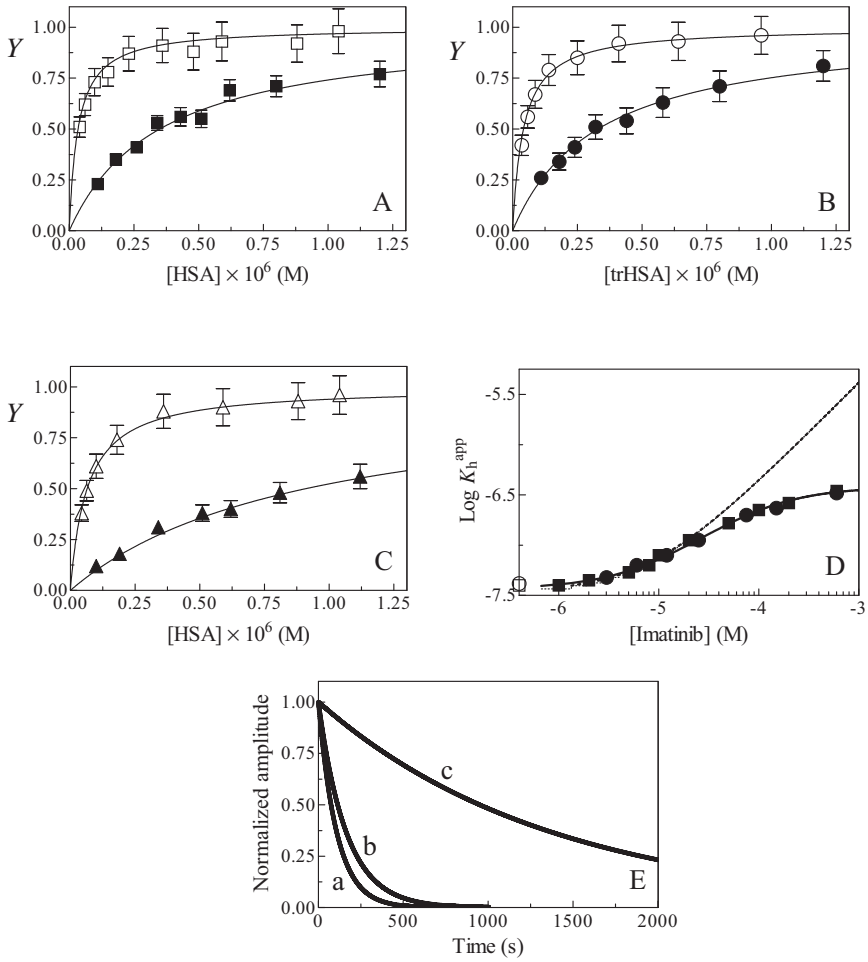


Fig. 3. Effect of imatinib on heme-Fe(III) binding to HSA and trHSA, at pH 7.0. (A) Binding isotherms for heme-Fe(III) association to HSA in the absence (open squares) and presence (filled squares) of imatinib, at 20.0 °C. The analysis of data according to Eq. (4) allowed the determination of $K_h = (3.4 \pm 0.4) \times 10^{-8}$ M and $K_H = (3.5 \pm 0.4) \times 10^{-7}$ M, respectively. (B) Binding isotherms for heme-Fe(III) association to trHSA in the absence (open circles) and presence (filled circles) of imatinib, at 20.0 °C. The analysis of data according to Eq. (4) allowed the determination of $K_h = (4.2 \pm 0.4) \times 10^{-8}$ M and $K_H = (3.3 \pm 0.3) \times 10^{-7}$ M, respectively. (C) Binding isotherms for heme-Fe(III) association to HSA in the absence (open triangles) and presence (filled triangles) of imatinib, at 37.0 °C. The analysis of data according to Eq. (4) allowed the determination of $K_h = (6.5 \pm 0.7) \times 10^{-8}$ M and $K_H = (9.1 \pm 1.0) \times 10^{-7}$ M, respectively. In panels A–C, the heme-Fe(III) concentration was 7.0×10^{-7} M, and the imatinib concentration was 6.0×10^{-4} M. (D) Dependence of Log K_h^{app} (M) for heme-Fe(III) association to HSA (filled squares) and trHSA (filled circles) on the imatinib concentration, at 20.0 °C. The continuous line was calculated according to Eq. (5) with the following parameters: $K_h = (3.4 \pm 0.4) \times 10^{-8}$ M, $K_I = (8.6 \pm 0.9) \times 10^{-5}$ M, and $K_i = (8.3 \pm 0.8) \times 10^{-6}$ M. The dashed line indicates the expected dependence of Log K_h^{app} on the imatinib concentration if the drug would have inhibited competitively heme-Fe(III) binding to HSA and trHSA, at 20.0 °C. The dashed line was calculated according to Eq. (6) with $K_h = (3.4 \pm 0.4) \times 10^{-8}$ M, and $K_I = (8.6 \pm 0.9) \times 10^{-5}$ M. The open square and the open circle indicate values of Log K_h for heme-Fe(III) binding to HSA and trHSA, respectively, in the absence of imatinib, at 20.0 °C. (E) Normalized averaged time courses of heme-Fe(III) dissociation from trHSA-heme-Fe(III), at 20.0 °C. The imatinib concentration was 0.0 M (a), 2.0×10^{-4} M (b), and 2.0×10^{-3} M (c). The time course analysis according to Eq. (7) allowed the determination of the following values of $k_h = (9.3 \pm 2.6) \times 10^{-3}$ s^{-1} (a), and of $k_H = (6.1 \pm 1.9) \times 10^{-3}$ s^{-1} (b) and $(7.3 \pm 2.1) \times 10^{-4}$ s^{-1} (c). The trHSA-heme-Fe(III) concentration was 2.6×10^{-6} M. The concentration of sperm whale apoMb was 8.0×10^{-6} M. Where not shown, the standard deviation is smaller than the symbol. For details, see text.

respectively) are reported in Table 1. Remarkably, values of K_h and K_H for heme-Fe(III) binding to HSA, trHSA, imatinib-HSA, and imatinib-trHSA, respectively, at 20.0 °C, match each other very well.

According to linked functions [44], values the dissociation equilibrium constant for heme-Fe(III) binding to imatinib-HSA and imatinib-trHSA at 20.0 °C, and to imatinib-HSA at 37.0 °C, calculated according to Eq. (5) ($K_H = K_h \times K_I/K_i = 3.5 \times 10^{-7}$ M, 3.9×10^{-7} M, and 1.2×10^{-6} M, respectively) agree with those obtained experimentally ($(3.5 \pm 0.4) \times 10^{-7}$ M, $(3.3 \pm 0.3) \times 10^{-7}$ M, and $(9.1 \pm 1.0) \times 10^{-7}$ M, respectively) at the highest imatinib concentration investigated (6.0×10^{-4} M), thus giving confidence that the assumptions underlying Eq. (5) are correct. The excellent fit of data reported in Fig. 3 (panel D) according to Eq. (5) excludes the possibility of competitive inhibition of heme-Fe(III) binding to HSA and trHSA by imatinib over the whole drug concentration range explored. Moreover, the analysis of data shown in Fig. 3 (panel D) according to Eq. (6), accounting for competitive inhibition of ligand binding, rules out the possibility that imatinib impairs competitively heme-Fe(III) binding to HSA and trHSA. By analogy with related drug-HSA-heme-Fe(III) systems (see [26,38,61–64]), present data indicate that imatinib binding to the FA7 site affects allosterically heme-Fe(III) binding to the FA1 site and *vice versa*.

As shown in Fig. 3 (panel E), kinetics of heme-Fe(III) transfer from trHSA-heme-Fe(III) to apoMb was fitted to a single-exponential decay for more than 85% of its course (according to Eq. (7)), indicating that no intermediate species (see Scheme 1) accumulate(s) in the course of the reaction. Imatinib affects significantly the value of the first-order rate constant for heme-Fe(III) dissociation from trHSA-heme-Fe(III) only at a concentration

higher than 2.0×10^{-4} M, which decreases from $k_h = (9.3 \pm 2.6) \times 10^{-3}$ s^{-1}, in the absence of imatinib, to $k_H = (7.3 \pm 2.1) \times 10^{-4}$ s^{-1}, in the presence of 2.0×10^{-4} M imatinib.

Peroxynitrite isomerisation by HSA-heme-Fe(III) and trHSA-heme-Fe(III) in the absence and presence of imatinib

Kinetics of peroxynitrite isomerization in the absence and presence of HSA-heme-Fe(III), trHSA-heme-Fe(III), and imatinib was recorded by using a single-wavelength rapid-mixing stopped-flow apparatus. Under all the experimental conditions, a decrease of the absorbance at 302 nm was observed, as previously reported [34,35,39,42,45]. Kinetics of peroxynitrite isomerization was fitted to a single-exponential decay for more than 90% of its course (according to Eq. (8); see Fig. 4, panel A). According to the literature (see [39–42,45]), this indicates that no intermediate species (*e.g.*, HSA-heme-Fe(III)-OONO and trHSA-heme-Fe(III)-OONO; see Scheme 1) accumulate(s) in the course of peroxynitrite isomerization. In particular, the formation of the transient HSA-heme-Fe(III)-OONO and trHSA-heme-Fe(III)-OONO species represents the

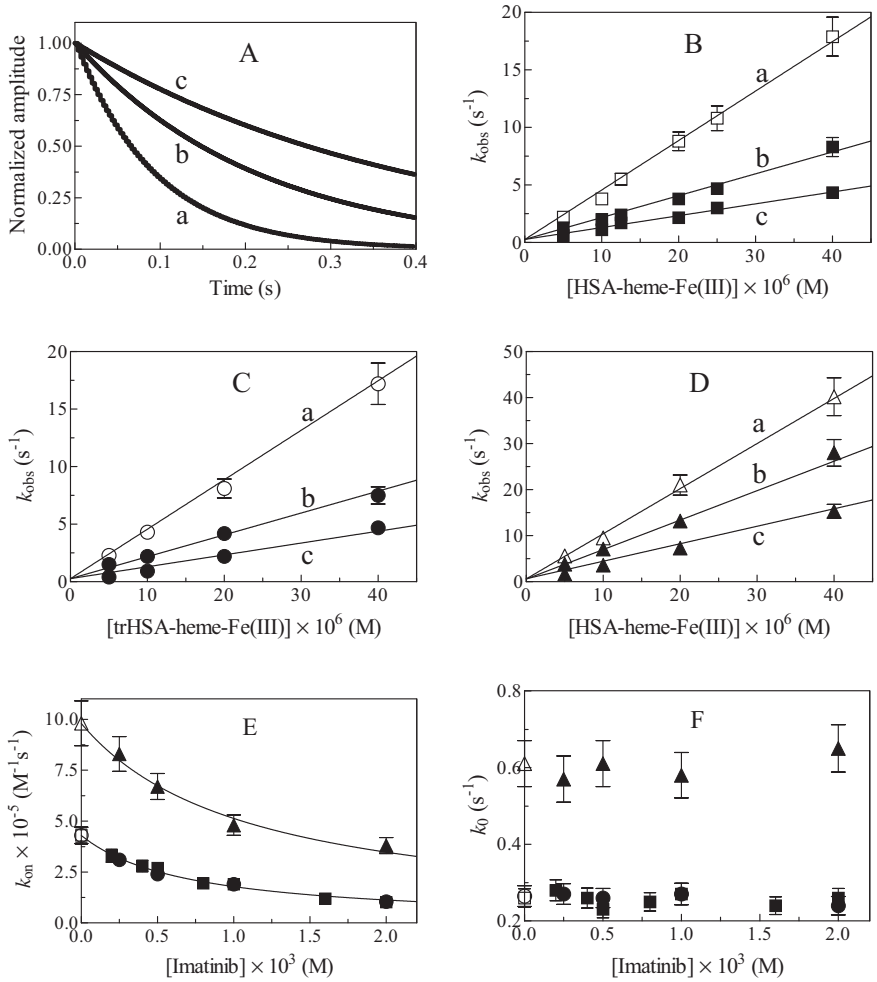


Fig. 4. Inhibition by imatinib of peroxynitrite isomerization by HSA-heme-Fe(III) and trHSA-heme-Fe(III), at pH 7.0. (A) Normalized averaged time courses of peroxynitrite isomerization by HSA-heme-Fe(III), at 20.0 °C. The imatinib concentration was 0.0 M (a), 1.0×10^{-3} M (b), and 2.0×10^{-3} M (c). The time course analysis according to Eq. (8) allowed the determination of the following values of $k_{obs} = 1.1 \times 10^1$ s^{-1} (a), 4.7 s^{-1} (b), and 2.5 s^{-1} (c). The HSA-heme-Fe(III) concentration was 2.5×10^{-5} M. The peroxynitrite concentration was 2.5×10^{-4} M. (B) Dependence of the pseudo-first-order rate constant for peroxynitrite isomerization (*i.e.*, k_{obs}) on the HSA-heme-Fe(III) concentration, at 20.0 °C. The imatinib concentration was 0.0 M (a), 1.0×10^{-3} M (b), and 2.0×10^{-3} M (c). The continuous lines were calculated according to Eq. (9) with the following parameters: (a) $k_{on} = (4.3 \pm 0.4) \times 10^5$ M^{-1} s^{-1}) and $k_0 = (2.6 \pm 0.3) \times 10^{-1}$; (b) $k_{on} = (1.9 \pm 0.2) \times 10^5$ M^{-1} s^{-1} and $k_0 = (2.7 \pm 0.3) \times 10^{-1}$ s^{-1}; and (c) $k_{on} = (1.0 \pm 0.1) \times 10^5$ M^{-1} s^{-1} and $k_0 = (2.7 \pm 0.3) \times 10^{-1}$ s^{-1}. (C) Dependence of the pseudo-first-order rate constant for peroxynitrite isomerization (*i.e.*, k_{obs}) on the trHSA-heme-Fe(III) concentration, at 20.0 °C. The imatinib concentration was 0.0 M (a), 1.0×10^{-3} M (b), and 2.0×10^{-3} M (c). The continuous lines were calculated according to Eq. (9) with the following parameters: (a) $k_{on} = (4.4 \pm 0.4) \times 10^5$ M^{-1} s^{-1}) and $k_0 = (2.7 \pm 0.3) \times 10^{-1}$; (b) $k_{on} = (1.8 \pm 0.2) \times 10^5$ M^{-1} s^{-1} and $k_0 = (2.6 \pm 0.3) \times 10^{-1}$ s^{-1}; and (c) $k_{on} = (1.1 \pm 0.1) \times 10^5$ M^{-1} s^{-1} and $k_0 = (2.7 \pm 0.3) \times 10^{-1}$ s^{-1}. (D) Dependence of the pseudo-first-order rate constant for peroxynitrite isomerization (*i.e.*, k_{obs}) on the HSA-heme-Fe(III) concentration, at 37.0 °C. The imatinib concentration was 0.0 M (a), 5.0×10^{-4} M (b), and 2.0×10^{-3} M (c). The continuous lines were calculated according to Eq. (9) with the following parameters: (a) $k_{on} = (9.8 \pm 1.0) \times 10^5$ M^{-1} s^{-1}) and $k_0 = (6.1 \pm 0.6) \times 10^{-1}$; (b) $k_{on} = (6.4 \pm 0.6) \times 10^5$ M^{-1} s^{-1} and $k_0 = (5.8 \pm 0.6) \times 10^{-1}$ s^{-1}; and (c) $k_{on} = (3.8 \pm 0.4) \times 10^5$ M^{-1} s^{-1} and $k_0 = (6.5 \pm 0.6) \times 10^{-1}$ s^{-1}. In panels B–D, the peroxynitrite concentration was 2.5×10^{-4} M. (E) Effect of the imatinib concentration on the second-order rate constant for HSA-heme-Fe(III)- and trHSA-heme-Fe(III)-catalyzed isomerization of peroxynitrite (*i.e.*, k_{on}). Squares and circles indicate values of k_{on} obtained at 20.0 °C. Triangles indicate values of k_{on} obtained at 37.0 °C. The open symbols on the ordinate indicates values of k_{on} obtained in the absence of imatinib ($k^0_{on} = (4.3 \pm 0.4) \times 10^5$ M^{-1} s^{-1}, at 20.0 °C, and $(9.8 \pm 1.0) \times 10^5$ M^{-1} s^{-1}), at 37.0 °C). The continuous lines were calculated according to Eq. (10) with the following parameters: 20.0 °C, $k^0_{on} = (4.3 \pm 0.4) \times 10^5$ M^{-1} s^{-1} and $K_I = (7.1 \pm 0.7) \times 10^{-4}$ M; and 37.0 °C, $k^0_{on} = (9.8 \pm 1.0) \times 10^5$ M^{-1} s^{-1} and $K_I = (1.1 \pm 0.1) \times 10^{-3}$ M. (F) Effect of the imatinib concentration on the first-order rate constant for the spontaneous isomerization of peroxynitrite (*i.e.*, k_0). Squares and circles indicate values of k_{on} obtained at 20.0 °C. Triangles indicate values of k_{on} obtained at 37.0 °C. The open symbols on the ordinate indicate values of k_0 obtained in the absence of imatinib ($k_0 = (2.7 \pm 0.3) \times 10^{-1}$ s^{-1}, at 20.0 °C; and $(6.1 \pm 0.6) \times 10^{-1}$, at 37.0 °C). Values of k_0 are independent of the imatinib concentration; the average values of k_0 are $(2.5 \pm 0.3) \times 10^{-1}$ s^{-1} and $(6.0 \pm 0.5) \times 10^{-1}$ s^{-1}, at 20.0 °C and 37.0 °C, respectively. Where not shown, the standard deviation is smaller than the symbol. For details, see text.

rate-limiting step in catalysis, the conversion of the HSA-heme-Fe(III)-OONO and trHSA-heme-Fe(III)-OONO complexes to HSA-heme-Fe(III) and trHSA-heme-Fe(III), respectively, and NO_3^- being faster by at least 1 order of magnitude.

In the absence and presence of imatinib, the observed rate constant for HSA-heme-Fe(III)- and trHSA-heme-Fe(III)-catalyzed isomerization of peroxynitrite (*i.e.*, k_{obs}) increases linearly with the HSA-heme-Fe(III) and trHSA-heme-Fe(III) concentration (Fig. 4, panels B–D). The analysis of the data reported in Fig. 4 (panels B–D), according to Eq. (9), allowed the determination of values of the second-order rate constant for peroxynitrite isomerization by HSA-heme-Fe(III) and trHSA-heme-Fe(III) (*i.e.*, k_{on}, corresponding to the slope of the linear plots) and of the first-order rate constant for peroxynitrite isomerization in the absence of HSA-heme-Fe(III) and trHSA-heme-Fe(III) (*i.e.*, k_0, corresponding to the *y* intercept of the linear plots). Of note, values of k_{on} and k_0 for peroxynitrite isomerization obtained according to Eq. (9) in the absence of imatinib are in excellent agreement with those reported in the literature (see [34,35,39–42,45,65]).

Imatinib affects dose-dependently k_{on} values for HSA-heme-Fe(III)- and trHSA-heme-Fe(III)-mediated isomerization of peroxynitrite (Fig. 4, panel E), decreasing from $(4.3 \pm 0.4) \times 10^5\ M^{-1}\ s^{-1}$ in the absence of the drug to $(1.0 \pm 0.1) \times 10^5\ M^{-1}\ s^{-1}$ at the imatinib concentration of 2.0×10^{-3} M, at 20.0 °C, and for HSA-heme-Fe(III)-mediated isomerization of peroxynitrite (Fig. 4, panel E), decreasing from $(9.8 \pm 1.1) \times 10^5\ M^{-1}\ s^{-1}$ in the absence of the drug to $(3.8 \pm 0.4) \times 10^5\ M^{-1}\ s^{-1}$ at the imatinib concentration of 2.0×10^{-3} M, at 37.0 °C. By contrast, values of k_0 are unaffected by imatinib, the average k_0 values for HSA-heme-Fe(III) and trHSA-heme-Fe(III) action being $(2.5 \pm 0.3) \times 10^{-1}\ s^{-1}$ at 20.0 °C, and for HSA-heme-Fe(III) action being $(6.0 \pm 0.5) \times 10^{-1}\ s^{-1}$, at 37.0 °C (Fig. 4, panel F).

According to Eq. (10) (Fig. 4, panel E), the analysis of the dependence of k_{on} for HSA-heme-Fe(III)- and trHSA-heme-Fe(III)-catalyzed isomerization of peroxynitrite on the imatinib concentration allowed the determination of the values of K_I for imatinib binding to HSA-heme-Fe(III) and trHSA-heme-Fe(III) at 20.0 °C ($(7.1 \pm 0.7) \times 10^{-4}$ M), and for drug binding to HSA-heme-Fe(III) at 37.0 °C ($(1.1 \pm 0.1) \times 10^{-3}$ M). The Hill coefficient (*n*) for imatinib binding to HSA-heme-Fe(III) and trHSA-heme-Fe(III) ranges between 0.99 ± 0.02 and 1.01 ± 0.01, indicating that drug binding to HSA-heme-Fe(III) and trHSA-heme-Fe(III) is a non-cooperative event. Under conditions where the imatinib concentration is much greater than K_I, HSA-heme-Fe(III) and trHSA-heme-Fe(III) do not catalyze the isomerization of peroxynitrite as observed in the presence of non-catalytic HSA and cyanide-bound HSA-heme-Fe(III) [45].

Docking analyses of imatinib binding to ligand-free HSA

Docking simulations of imatinib binding to HSA, with the search space extended to the whole protein (Fig. 5), indicated the preferential binding of the drug to the FA1 and FA7 sites with similar binding energy. The number of complexes observed in FA1 and FA7 in docking simulations with a maximum of 9 poses was 4 and 2, respectively. By increasing to 20 the number of possible poses, 9 and 3 complexes were observed in FA1 and FA7, respectively. The highest ranking complexes of imatinib in FA1 and FA7 are shown in Fig. 6. In simulations carried out restricting the docking search space only to FA2, FA6 and FA7 sites, imatinib was found to bind with the best binding energies and in the majority of the poses (7 out of 9) in FA7 rather than in FA6 site (2 out of 9), while binding in FA2 was not observed. These results suggest that in ligand free-HSA imatinib binds preferentially to FA1 and FA7 sites.

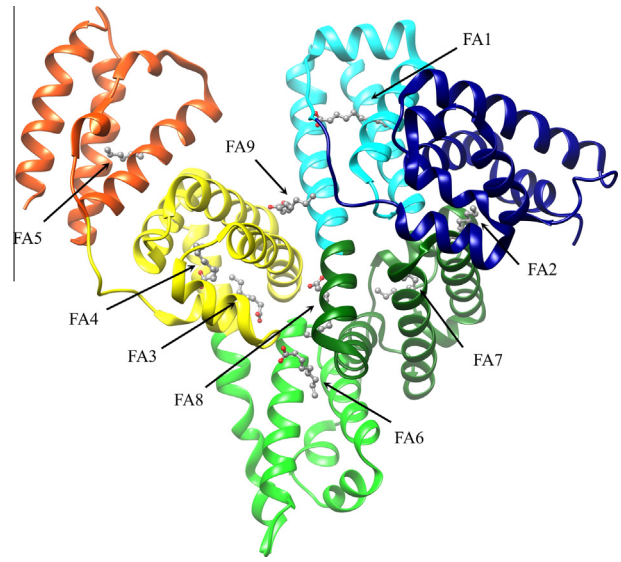


Fig. 5. Ribbon representation of the three-dimensional structure of HSA. The subdomains of HSA are rendered with different colors (domain IA, in blue; domain IB, in cyan; domain IIA, in dark green; domain IIB, in light green; domain IIIA, in yellow; and domain IIIB, in red). trHSA includes only domains I and II. The FA1–FA9 sites are occupied by capric acid (in ball-and-stick representation). Atomic coordinates were taken from PDB ID: 1E7E [82]. The FA binding sites are numbered according to literature [82]. For details, see text. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Docking analyses of imatinib binding to HSA-heme-Fe(III)

Docking simulations of imatinib binding to HSA-heme-Fe(III), with the search space extended to the whole protein (Fig. 5), indicated the most likely binding sites are FA2 and FA7, binding energy values being of the same order of magnitude. When the maximum number of docking poses was set to 9, the number of complexes observed in FA2 and FA7 was 2 and 6, respectively. Setting the maximum number of docking poses to 20 resulted in 1 and 9 complexes in FA6 and FA7, respectively. When the search space was restricted only to FA2 and FA7 sites, 1 complex was observed in FA2 and 8 in FA7. These results suggest that in HSA-heme-Fe(III), imatinib binds preferentially to the FA2 and FA7 sites.

Analysis of local structural similarity between HSA and imatinib binding proteins

The analysis of the local structural similarities between HSA and a set of imatinib binding proteins, carried out using the local structural comparison program ASSIST [50], identified four different sets of residues in HSA which display chemical and geometric similarity with the imatinib binding pocket of the spleen tyrosine kinase (PDB ID: 1XBB) [52]. The best match is obtained with HSA residues Ile264, Ser287, Leu260, Glu230, and Leu219 located on the walls of the FA7 site (Table 2 and Fig. 8), the calculated all-atoms RMSD value between these residues and the corresponding spleen tyrosine kinase residues being 1.78 Å.

Discussion

This study reports a detailed functional and structural characterization of imatinib binding to HSA, which may act as the secondary carrier of the drug in pathological states characterized by low plasma levels of AGP [20]; indeed AGP represents the first carrier of imatinib [7–11,13,66]. Moreover, the modulation of heme-Fe(III)

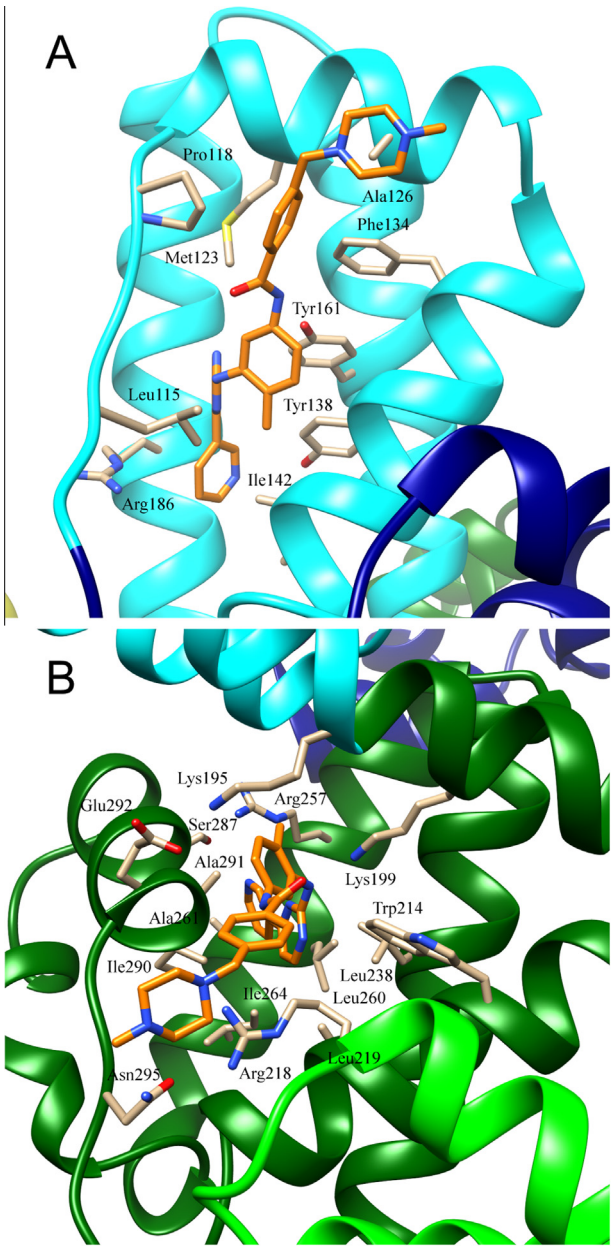


Fig. 6. Atomic details of the highest ranking complexes obtained by docking simulations between imatinib and ligand-free HSA. (A) Imatinib bound to FA1. (B) Imatinib bound to FA7. Atomic coordinates were taken from the PDB ID: 1AO6 [47]. For details, see text.

binding to HSA and the reactivity of HSA-heme-Fe(III) by imatinib has been deeply investigated. Furthermore, present results have been substantiated by docking simulations of imatinib binding to HSA and by the comparative analysis of three-dimensional structures of imatinib-HSA(-heme-Fe(III)) complexes with those of imatinib-binding proteins.

Previously, only imatinib binding to HSA was investigated [8,10,12]; values of the dissociation equilibrium constant for imatinib binding to HSA (*i.e.*, K_i) reported in the literature and obtained by different methods [8,10,12,67] range between 4×10^{-6} M and 3×10^{-5} M. The value of the dissociation equilibrium constant here determined by competitive inhibition (*i.e.*, K_i) is $(9.6 \pm 1.0) \times 10^{-6}$ M, at 37.0 °C and pH 7.0. Remarkably, the affinity of imatinib for HSA is lower than that for the Abl tyrosine kinases for which IC_{50} values ranging from 2.5×10^{-8} M to 2.0×10^{-7} M

Table 2
Structural matches between the imatinib binding sites of HSA and of the spleen tyrosine kinase.

HSA site	HSA residues	Matching residues	RMSD
FA7	Ile264	Leu377	1.78 Å
	Ser287	Ser379	
	Leu260	Val385	
	Glu230	Glu452	
	Leu219	Leu501	
FA6	Leu347	Leu377	1.85 Å
	Ser342	Ser379	
	Val343	Val385	
	Ala215	Ala400	
	Ala210	Ala451	
FA2	Ala143	Ala400	1.92 Å
	Leu154	Leu377	
	Ser192	Ser 79	
	Ala26	Ala451	
	Leu250	Leu501	
FA2	Asp13	Glu452	1.98 Å
	Leu251	Leu377	
	Leu22	Val385	
	Ala254	Ala400	
	Gly15	Gly454	

have been reported (see [67] and references therein). However, this is not surprising as imatinib has been specifically developed to target this specific class of kinases [68].

The allosteric modulation of heme-Fe(III) binding to HSA and the HSA-heme-Fe(III) reactivity by imatinib is highlighted from the overlapping functional properties of the full length HSA and its truncated form trHSA [26], containing only the FA1, FA2, FA6, and FA7 sites of HSA (Fig. 5).

As expected for ligand (*e.g.*, drugs) binding to the high affinity FA7 site of HSA [22,26,31,62,69–72], imatinib impairs allosterically heme-Fe(III) binding to the FA1 site of HSA. According to linked functions [44], heme-Fe(III) binding to the FA1 site impairs imatinib binding to the FA7 pocket of HSA by the same extent. The allosteric linkage between the FA1 and FA7 sites is mediated by the Phe149-Tyr150 dyad and by Arg257 [64]. Actually, drug binding to FA7 causes a shift of these residues towards the FA1 cavity impairing heme binding (*i.e.*, decreasing the heme affinity by one order of magnitude) [64]. Although docking simulations suggest imatinib binding to the FA1 site (Fig. 6, panel A), postulating direct competition between the drug and heme-Fe(III), data analysis shown in Fig. 3 (panel D) exclude this possibility over the imatinib concentration range here explored.

According to literature [62,63], the value of the first-order rate constant for heme-Fe(III) dissociation from trHSA-heme-Fe(III) is grossly unaffected by the imatinib concentration up to 2.0×10^{-4} M, *i.e.* over the drug concentration range where most of the allosteric modulation of the trHSA-heme-Fe(III) complex formation occurs (Fig. 3, panels D and E). However, on increasing the imatinib concentration from 2.0×10^{-4} M to 2.0×10^{-3} M, the value of the first-order rate constant for heme-Fe(III) dissociation from trHSA-heme-Fe(III) decreases by about one order of magnitude. According to literature [62,63,70,73], these data suggest that: (*i*) imatinib binding to the high affinity FA7 site modulates allosterically the formation of penta-coordinated trHSA-heme-Fe(III) without affecting kinetics of heme-Fe(III) release; and (*ii*) imatinib binding to the low affinity FA2 site induces the penta-to-hexa-coordination transition of the heme-Fe(III) atom, which in turn becomes unreactive and dissociates slowly.

At present, the available data do not allow to identify unambiguously the low affinity imatinib binding site(s) modulating allosterically the HSA-heme-Fe(III) reactivity, *i.e.* peroxynitrite

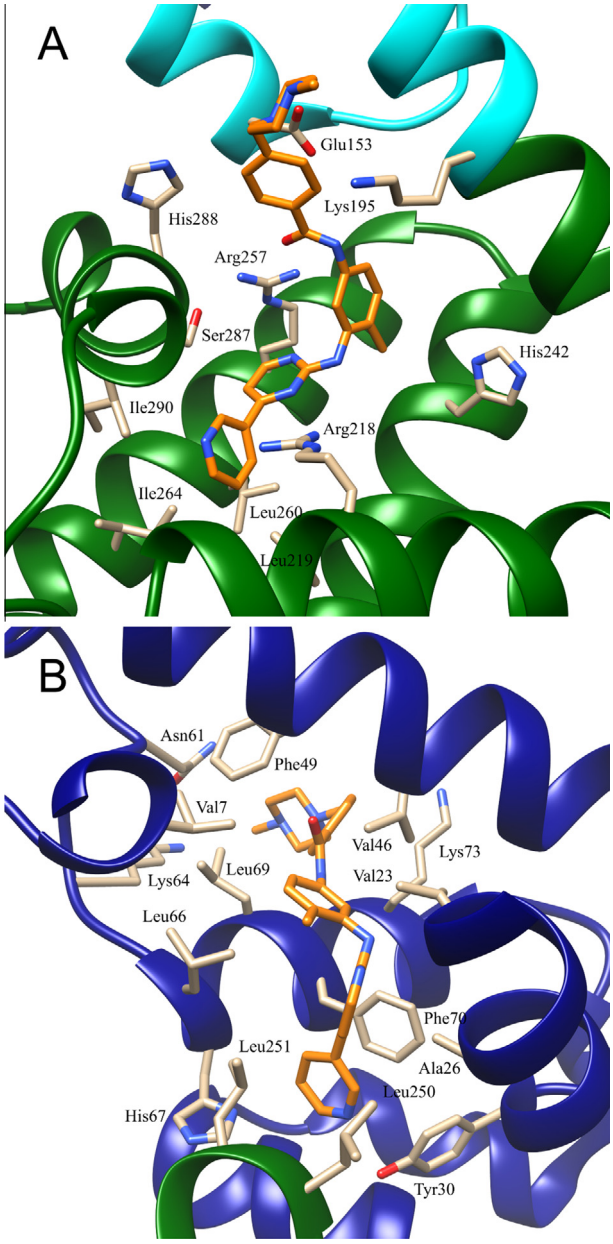


Fig. 7. Atomic details of the highest ranking complexes obtained by docking simulations between imatinib and HSA-heme-Fe(III). (A) Imatinib bound to FA7. (B) Imatinib bound to FA2. Atomic coordinates were taken from the PDB ID: 1N5U [48]. For details, see text.

isomerization, possible candidates being FA2 and/or FA6 sites. This is because no single chromophore is available to probe both sites at the same time. However, docking simulations (Fig. 7, panel B) suggest that FA2 is the most probable imatinib low affinity binding site.

The involvement of FA6 as the binding region responsible for the observed change in heme-based reactivity may be excluded on the ground of structural and functional studies performed on the ibuprofen-(tr)HSA and ibuprofen-(tr)HSA-heme complexes. Based on structural evidence, ibuprofen binds to both FA3–FA4 and FA6 regions of HSA [69]. However, from the functional viewpoint, ibuprofen shows three distinct binding constants for (tr)HSA(-heme-Fe) (*ca.* 10^{-7} M, 10^{-5} M, and 10^{-3} M), the highest and medium affinity binding sites of ibuprofen being the FA3–FA4 cleft and the FA6 site [26]. Since only a drug concentration higher than 10^{-3} M induces the hexa-coordination of HSA-heme-Fe(II)-NO and triggers the heme-Fe-ligand dissociation kinetics [45,70,71], it has been postulated that the allosteric linkage occurs between the FA1 and FA2 sites hosting the heme and the drug, respectively. Moreover, the X-ray crystal structure of HSA binding two ibuprofen molecules at the FA3–FA4 and FA6 regions is almost superimposable to that of the drug-free protein [69]. These findings suggest that drug binding to FA3–FA4 region and the FA6 site does not appear to induce allosteric conformational changes of the FA1 site.

Imatinib-based modulation of HSA-heme-Fe(III) reactivity resembles that induced by ibuprofen and warfarin upon binding to the low affinity site [26,38,45,64]. On the basis of resonance Raman spectroscopic properties of HSA-heme-Fe(III) complexes binding ibuprofen and warfarin at the low affinity site, the His146 residue has been suggested as the putative ligand able to coordinate the heme-Fe(III) atom at the sixth position, the fifth coordination position of the heme-Fe(III) atom being occupied by the oxygen phenoxy atom of Tyr161 [70]. Therefore, the presence of the His146 residue at the sixth coordination position of the heme-Fe(III) atom inhibits peroxynitrite isomerization, thus rendering the imatinib-HSA-heme-Fe(III) complex non-reactive.

Present data highlight ligand binding modulation by the allosterically linked FA1, FA2 and FA7 sites present in domains I and II of (tr)HSA (Fig. 5). In fact, the removal of the C-terminal domain III of HSA, containing the FA3–FA4 cleft and the FA5 site, does not affect the allosteric linkage between the FA1, FA2 and FA7 sites [22,31,73]. Moreover, the interface of subdomains IA and IIA has been reported to play a pivotal role in the allosteric ligand interactions of HSA [74]. However, different ligands may display different allosteric binding interactions. Indeed, (*i*) long-chain fatty acids regulate the binding properties of HSA in a complex competitive and allosteric manner mainly involving domain I [75], and (*ii*) the allosteric linkage between the FA3–FA4 cleft (in domain III) and the FA7 site (in domain II) has been postulated to modulate benzodiazepine and warfarin binding [76] as well as phenprocoumon enantiomers and 3-substituted 1,4-benzodiazepines association [77].

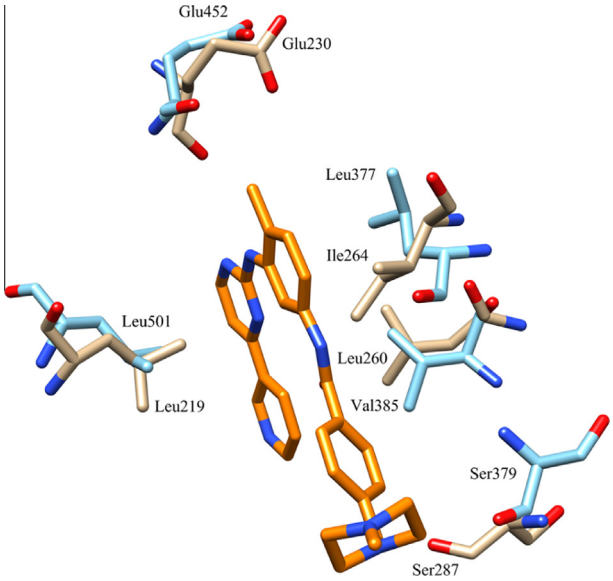


Fig. 8. Comparison of HSA residues located in the FA7 site and the equivalent residues located in the imatinib binding pocket of the spleen tyrosine kinase as determined by local structural comparison using the program ASSIST [50]. HSA residues are in light brown and spleen tyrosine kinase residues are in light blue. For details see text. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Docking studies undertaken to complement the experimental characterization of imatinib binding to HSA indicate a preferential binding of imatinib in FA1 and FA7 sites in ligand-free HSA, and in FA2 and FA7 sites in HSA-heme-Fe(III) (Figs 6 and 7). Furthermore, analysis of the local structural similarities between HSA and a set of imatinib binding proteins identifies a set of residues in the FA7 pocket of HSA which display chemical and geometric similarity with the imatinib binding pocket of the spleen tyrosine kinase (Fig. 8 and Table 2).

Imatinib binding to HSA affects not only the protein binding capacity (*i.e.*, heme-Fe(III) association), but also the heme-Fe(III)-based reactivity. Because it is compelling evidence that *in vivo* binding of endogenous and exogenous ligands to HSA is a transient event, it is plausible that HSA-ligand complexes display time-dependent functional properties. Thus, HSA should be considered as a "chronosteric" protein [26,78–81].

In conclusion, since the AGP levels in patients with hepatic cirrhosis and hepatitis are half of the normal plasma levels [20] and HSA plasma levels are at least 40-times higher [81], HSA may serve as a secondary carrier for imatinib.

Acknowledgment

This study was fully supported by a grant from Roma Tre University, Roma, Italy (CLA 2013 to P.A.).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.abb.2014.07.001.

References

[1] J.T. Hartmann, M. Haap, H.G. Kopp, H.P. Lipp, Curr. Drug Metab. 10 (2009) 470–481.
[2] D. Levêque, F. Maloisel, In Vivo 19 (2005) 77–84.
[3] B. Peng, P. Lloyd, H. Schran, Clin. Pharmacokinet. 44 (2005) 879–894.
[4] M.W. Deininger, B.J. Druker, Pharmacol. Rev. 55 (2003) 401–423.
[5] S. Basciani, M. Brama, S. Mariani, G. De Luca, M. Arizzi, L. Vesci, C. Pisano, S. Dolci, G. Spera, L. Gnessi, Cancer Res. 65 (2005) 1897–1903.
[6] N.A. Lankheet, L.M. Knapen, J.H. Schellens, J.H. Beijnen, N. Steeghs, A.D. Huitema, Ther. Drug Monit. 36 (2014) 326–334.
[7] Z. Huang, T. Ung, Curr. Drug Metab. 14 (2013) 226–238.
[8] C. Gambacorti-Passerini, R. Barni, P. le Coutre, M. Zucchetti, G. Cabrita, L. Cleris, F. Rossi, E. Gianazza, J. Brueggen, R. Cozens, P. Pioltelli, E. Pogliani, G. Corneo, F. Formelli, M. D'Incalci, J. Natl Cancer Inst. 92 (2000) 1641–1650.
[9] C. Gambacorti-Passerini, M. Zucchetti, D. Russo, R. Frapolli, M. Verga, S. Bungaro, L. Tornaghi, F. Rossi, P. Pioltelli, E. Pogliani, D. Alberti, G. Corneo, M. D'Incalci, Clin. Cancer Res. 9 (2003) 625–632.
[10] I. Fitos, J. Visy, F. Zsila, Biochim. Biophys. Acta 1760 (2006) 1704–1712.
[11] N. Widmer, L.A. Decosterd, C. Csajka, S. Leyvraz, M.A. Duchosal, A. Rosselet, B. Rochat, C.B. Eap, H. Henry, J. Biollaz, T. Buclin, Br. J. Clin. Pharmacol. 62 (2006) 97–112.
[12] A.H. Hegde, R. Punith, J. Seetharamappa, J. Fluoresc. 22 (2012) 521–528.
[13] A. Haouala, N. Widmer, M. Guidi, M. Montemurro, S. Leyvraz, T. Buclin, C.B. Eap, L.A. Decosterd, C. Csajka, Br. J. Clin. Pharmacol. 75 (2013) 1007–1018.
[14] C.B. Gambacorti-Passerini, F. Rossi, M. Verga, H. Ruchatz, R. Gunby, R. Frapolli, M. Zucchetti, L. Scapozza, S. Bungaro, L. Tornaghi, F. Rossi, P. Pioltelli, E. Pogliani, M. D'Incalci, G. Corneo, Blood Cells Mol. Dis. 28 (2002) 361–372.
[15] M.D. Barber, J.A. Ross, K.C. Fearon, Nutr. Cancer 35 (1999) 106–110.
[16] D.C. McMillan, W.S. Watson, P. O'Gorman, T. Preston, H.R. Scott, C.S. McArdle, Nutr. Cancer 39 (2001) 210–213.
[17] M.R. Dixon, J.S. Haukoos, S.M. Udani, J.J. Naghi, T.D. Arnell, R.R. Kumar, M.J. Stamos, Arch. Surg. 138 (2003) 962–966.
[18] L.C. Sun, K.S. Chu, S.C. Cheng, C.Y. Lu, C.H. Kuo, J.S. Hsieh, Y.L. Shih, S.J. Chang, J.Y. Wang, BMC Cancer 9 (2009) 288.
[19] D. Gupta, C.G. Lis, Nutrition 9 (2010) 69.
[20] J.M. Kremer, J. Wilting, L.H. Janssen, Pharmacol. Rev. 40 (1988) 1–47.
[21] V.G. Kharitonov, V.S. Sharma, D. Magde, D. Koesling, Biochemistry 36 (1997) 6814–6818.
[22] G. Fanali, G. Pariani, P. Ascenzi, M. Fasano, FEBS J. 276 (2009) 2241–2250.
[23] M.M. Bradford, Anal. Biochem. 72 (1976) 248–254.
[24] A. Boffi, T.K. Das, S. Della, Biophys. J. 77 (1999) 1143–1149.
[25] E. Antonini, M. Brunori, Hemoglobin and Myoglobin in their Reactions with Ligands, North-Holland Publishing Co., Amsterdam, London, 1971, pp. 123–130.
[26] G. Fanali, P. Ascenzi, G. Bernardi, M. Fasano, J. Biomol. Struct. Dyn. 29 (2012) 691–701.
[27] G. Sudlow, D.J. Birkett, D.N. Wade, Mol. Pharmacol. 11 (1975) 824–832.
[28] G. Sudlow, D.J. Birkett, D.N. Wade, Mol. Pharmacol. 12 (1976) 1052–1061.
[29] K. Nakajou, H. Watanabe, U. Kragh-Hansen, T. Maruyama, M. Otagiri, Biochim. Biophys. Acta 1623 (2003) 88–97.
[30] A.J. Ryan, J. Ghuman, P.A. Zunszain, C.W. Chung, S. Curry, J. Struct. Biol. 174 (2011) 84–91.
[31] G. Fanali, M. Fasano, P. Ascenzi, J.-M. Zingg, A. Azzi, BioFactors 39 (2013) 294–303.
[32] D.S. Bohle, P.A. Glassbrenner, B. Hansert, Methods Enzymol. 269 (1996) 302–311.
[33] W.H. Koppenol, R. Kissner, J.S. Beckman, Methods Enzymol. 269 (1996) 296–302.
[34] S. Herold, K. Shivashankar, Biochemistry 42 (2003) 14036–14046.
[35] S. Herold, S. Kalinga, T. Matsui, Y. Watanabe, J. Am. Chem. Soc. 126 (2004) 6945–6955.
[36] P. Ascenzi, M. Fasano, Biochem. Biophys. Res. Commun. 353 (2007) 469–474.
[37] S. Goldstein, G. Merényi, Methods Enzymol. 436 (2008) 49–61.
[38] P. Ascenzi, M. Fasano, IUBMB Life 61 (2009) 1118–1122.
[39] P. Ascenzi, A. Bolli, F. Gullotta, G. Fanali, M. Fasano, IUBMB Life 62 (2010) 776–780.
[40] P. Ascenzi, C. Ciaccio, F. Sinibaldi, R. Santucci, M. Coletta, Biochem. Biophys. Res. Commun. 404 (2011) 190–194.
[41] P. Ascenzi, C. Ciaccio, F. Sinibaldi, R. Santucci, M. Coletta, Biochem. Biophys. Res. Commun. 415 (2011) 463–467.
[42] P. Ascenzi, A. Bolli, A. di Masi, G.R. Tundo, G. Fanali, M. Coletta, M. Fasano, J. Biol. Inorg. Chem. 16 (2011) 97–108.
[43] P. Ascenzi, M.G. Ascenzi, G. Amiconi, Biochem. Mol. Biol. Edu. 15 (1987) 134–135.
[44] J. Wyman Jr., Adv. Protein Chem. 19 (1964) 223–286.
[45] P. Ascenzi, A. di Masi, M. Coletta, C. Ciaccio, G. Fanali, F.P. Nicoletti, G. Smulevich, M. Fasano, J. Biol. Chem. 284 (2009) 31006–31017.
[46] P. Ascenzi, Y. Cao, G.R. Tundo, M. Coletta, G. Fanali, M. Fasano, Biochem. Biophys. Res. Commun. 411 (2011) 185–189.
[47] S. Sugio, A. Kashima, S. Mochizuki, M. Noda, K. Kobayashi, Protein Eng. 12 (1999) 439–446.
[48] M. Wardell, Z. Wang, J.X. Ho, J. Robert, F. Ruker, J. Ruble, D.C. Carter, Biochem. Biophys. Res. Commun. 291 (2002) 813–819.
[49] O. Trott, A.J. Olson, J. Comp. Chem. 31 (2010) 455–461.
[50] S. Caprari, D. Toti, L. Viet Hung, M. Di Stefano, F. Polticelli, Bioinformatics 30 (2014) 1022–1024.
[51] C.D. Mol, D.R. Dougan, T.R. Schneider, R.J. Skene, M.L. Kraus, D.N. Scheibe, G.P. Snell, H. Zou, B.C. Sang, K.P. Wilson, J. Biol. Chem. 279 (2004) 31655–31663.
[52] S. Atwell, J.M. Adams, J. Badger, M.D. Buchanan, I.K. Feil, K.J. Froning, X. Gao, J. Hendle, K. Keegan, B.C. Leon, H.J. Müller-Dieckmann, V.L. Nienaber, B.W. Noland, K. Post, K.R. Rajashankar, A. Ramos, M. Russell, S.K. Burley, S.G. Buchanan, J. Biol. Chem. 279 (2004) 55827–55832.
[53] M.A. Seeliger, B. Nagar, F. Frank, X. Cao, M.N. Henderson, J. Struct. 15 (2007) 299–311.
[54] M.D. Jacobs, P.R. Caron, B.J. Hare, Proteins 70 (2008) 1451–1460.
[55] J.A. Winger, O. Hantschel, G. Superti-Furga, J. Kuriyan, BMC Struct. Biol. 9 (2009) 7.
[56] H.V. Namboodiri, M. Bukhtiyarova, J. Ramcharan, M. Karpusas, Y. Lee, E.B. Springman, Biochemistry 49 (2010) 3611–3618.
[57] W. Jahnke, R.M. Grotzfeld, X. Pellé, A. Strauss, G. Fendrich, S.W. Cowan-Jacob, S. Cotesta, D. Fabbro, P. Furet, J. Mestan, A.L. Marzinzik, J. Am. Chem. Soc. 132 (2010) 7043–7048.
[58] J. Yang, N. Campobasso, M.P. Biju, K. Fisher, X.Q. Pan, J. Cottom, S. Galbraith, T. Ho, H. Zhang, X. Hong, P. Ward, G. Hofmann, B. Siegfried, F. Zappacosta, Y. Washio, P. Cao, J. Qu, S. Bertrand, D.Y. Wang, M.S. Head, H. Li, S. Moores, Z. Lai, K. Johanson, G. Burton, C. Erickson-Miller, G. Simpson, P. Tummino, R.A. Copeland, A. Oliff, Chem. Biol. 18 (2011) 177–186.
[59] R.A. Laskowski, M.B. Swindells, J. Chem. Inf. Model. 51 (2011) 2778–2786.
[60] P.A. Zunszain, J. Ghuman, T. Komatsu, E. Tsuchida, S. Curry, BMC Struct. Biol. 3 (2003) 6.
[61] P. Ascenzi, A. Bocedi, S. Notari, E. Menegatti, M. Fasano, Biochem. Biophys. Res. Commun. 334 (2005) 481–486.
[62] A. Bocedi, S. Notari, E. Menegatti, G. Fanali, M. Fasano, P. Ascenzi, FEBS J. 272 (2005) 6287–6296.
[63] P. Ascenzi, A. Bocedi, S. Notari, G. Fanali, R. Fesce, M. Fasano, Med. Chem. 6 (2006) 483–489.
[64] P. Ascenzi, M. Fasano, Biophys. Chem. 148 (2010) 16–22.
[65] S. Goldstein, J. Lind, G. Merényi, Chem. Rev. 105 (2005) 2457–2470.
[66] I. Fitos, Á. Simon, F. Zsila, G. Mády, Á. Bencsura, Z. Varga, L. Orfi, G. Kéri, J. Visy, Int. J. Biol. Macromol. 50 (2012) 788–795.
[67] O. Kretz, H.M. Weiss, M.M. Schumacher, G. Gross, Br. J. Clin. Pharmacol. 58 (2004) 212–216.
[68] M. Deininger, E. Buchdunger, B.J. Druker, Blood 105 (2005) 2640–2653.
[69] J. Ghuman, P.A. Zunszain, I. Petitpas, A.A. Bhattacharya, M. Otagiri, S. Curry, J. Mol. Biol. 353 (2005) 38–52.
[70] F.P. Nicoletti, B.D. Howes, M. Fittipaldi, G. Fanali, M. Fasano, P. Ascenzi, G. Smulevich, J. Am. Chem. Soc. 130 (2008) 11677–11688.
[71] A. Bocedi, G. De Sanctis, C. Ciaccio, G.R. Tundo, A. Di Masi, G. Fanali, F.P. Nicoletti, M. Fasano, G. Smulevich, P. Ascenzi, M. Coletta, PLoS ONE 8 (2013) e58842.

[72] A. Bolli, M. Marino, G. Rimbach, G. Fanali, M. Fasano, P. Ascenzi, Biochem. Biophys. Res. Commun. 398 (2010) 444–449.
[73] A. di Masi, F. Gullotta, A. Bolli, G. Fanali, M. Fasano, P. Ascenzi, FEBS J. 278 (2011) 654–662.
[74] K. Kaneko, V.T. Chuang, A. Minomo, K. Yamasaki, N.V. Bhagavan, T. Maruyama, M. Otagiri, IUBMB Life 63 (2011) 277–285.
[75] V.T. Chuang, M. Otagiri, Pharm. Res. 19 (2002) 1458–1464.
[76] E. Domenici, C. Bertucci, P. Salvadori, I.W. Wainer, J. Pharm. Sci. 80 (1991) 164–166.
[77] I. Fitos, M. Simonyi, Chirality 4 (1992) 21–23.
[78] U. Muller-Eberhard, J. Javid, H.H. Liem, A. Hanstein, M. Hanna, Blood 32 (1968) 811–815.
[79] Y.I. Miller, N. Shaklai, Biochim. Biophys. Acta 1454 (1999) 153–164.
[80] P. Ascenzi, S. Gianni, IUBMB Life 65 (2013) 836–844.
[81] G. Fanali, A. di Masi, V. Trezza, M. Marino, M. Fasano, P. Ascenzi, Mol. Aspects Med. 33 (2012) 209–290.
[82] A.A. Bhattacharya, T. Grüne, S. Curry, J. Mol. Biol. 303 (2000) 721–732.

Archives of Biochemistry and Biophysics 590 (2016) 56–63

Contents lists available at ScienceDirect

Archives of Biochemistry and Biophysics

journal homepage: www.elsevier.com/locate/yabbi

All-*trans*-retinoic acid and retinol binding to the FA1 site of human serum albumin competitively inhibits heme-Fe(III) association

Elena Di Muzio [a], Fabio Polticelli [a, b], Alessandra di Masi [a], Gabriella Fanali [c], Mauro Fasano [c, d], Paolo Ascenzi [e, *]

[a] Department of Sciences, Roma Tre University, I-00146 Roma, Italy
[b] National Institute of Nuclear Physics, Roma Tre Section, I-00146 Roma, Italy
[c] Division of Biomedical Sciences, Department of Theoretical and Applied Sciences, University of Insubria, I-21052 Busto Arsizio (VA), Italy
[d] Center of Neuroscience, University of Insubria, I-21052 Busto Arsizio (VA), Italy
[e] Interdepartmental Laboratory for Electron Microscopy, Roma Tre University, I-00146 Roma, Italy

ARTICLE INFO

Article history:
Received 31 August 2015
Received in revised form
19 October 2015
Accepted 22 October 2015
Available online 27 October 2015

Keywords:
Human serum albumin
All-*trans*-retinoic acid
Retinol
Heme binding
Competitive inhibition
Molecular docking

ABSTRACT

Retinoids are a class of chemicals derived from vitamin A metabolism, playing important and diverse functions. Vitamin A, also named all-*trans*-retinol (all-*trans*-ROL), is coverted into two classes of biologically active retinoids, *i.e.* 11-*cis*-retinoids and acidic retinoids. Among acidic retinoids, all-*trans*-retinoic acid (all-*trans*-RA) and 9-*cis*-retinoic acid (9-*cis*-RA) represent the main metabolic products. Specific and aspecific proteins solubilize, protect, and detoxify retinoids in the extracellular environment. The retinoid binding protein 4 (RBP4), the epididymal retinoid-binding protein (ERBP), and the interphotoreceptor matrix retinoid-binding protein (IRBP) play a central role in ROL transport, whereas lipocalin-type prostaglandin D synthase (also named β-trace) and human serum albumin (HSA) transport preferentially all-*trans*-RA. Here, the modulatory effect of all-*trans*-RA and all-*trans*-ROL on ferric heme (heme-Fe(III)) binding to HSA is reported. All-*trans*-RA and all-*trans*-ROL binding to the FA1 site of HSA competitively inhibit heme-Fe(III) association. Docking simulations and local structural comparison of HSA with all-*trans*-RA- and all-*trans*-ROL-binding proteins support functional data indicating the preferential binding of all-*trans*-RA and all-*trans*-ROL to the FA1 site of HSA. Present results may be relevant *in vivo*, in fact HSA could act as a secondary carrier of retinoids in human diseases associated with reduced levels of RBP4 and IRBP.

1. Introduction

Dietary vitamin A and its natural and synthetic analogs (named retinoids) are naturally-occurring, fat-soluble, unsaturated isoprenoids present in all living organisms. Retinoids mainly influence cell growth, differentiation and death, the deregulation of retinoid signaling pathways being linked to tumorigenesis. For these reasons, they soon emerged as potential therapeutic agents for several diseases, including cancer [1].

Vitamin A, also named all-*trans*-retinol (all-*trans*-ROL), is stored in the intestinal mucosa as retinyl esters, and is metabolized within the intestinal lumen by a sequential oxidation process. All-*trans*-ROL metabolism produces two classes of biologically active retinoids, *i.e.* 11-*cis*-retinoids and acidic retinoids (RA). During the conversion of ROL to RA, all-*trans*-retinal is produced. Among acidic retinoids, all-*trans*-retinoic acid (all-*trans*-RA) and 9-*cis*-retinoic acid (9-*cis*-RA) represent the main metabolic products [1,2]. While 11-*cis*-retinoids act mainly in vision, all-*trans*-RA (that binds to the retinoic acid receptors (RAR) α, β, and γ) and 9-*cis*-RA (that binds both RARs and members of the retinoid X receptors (RXR) α, β, and γ) regulate several biological processes, including cell growth, differentiation and death [1–3].

RA, ROL, and all-*trans*-retinal bind to specific and aspecific proteins, which solubilize, protect, and detoxify retinoids in the

Abbreviations: 9-*cis*-RA, 9-*cis*-retinoic acid; all-*trans*-RA, all-*trans*-retinoic acid; all-*trans*-ROL, all-*trans*-retinol; FA, fatty acid; FA1 site, fatty acid binding site 1; HSA, human serum albumin; HSA-heme-Fe(III), ferric human serum heme-albumin; heme-Fe(III), ferric heme (Fe(III)-protoporphyrin IX); RA, retinoic acid; RAR, retinoic acid receptor; RBP, retinoid-binding proteins; RBP4, retinol-binding protein 4; ROL, retinol; RXR, retinoid X receptor.

* Corresponding author. Interdepartmental Laboratory for Electron Microscopy, Roma Tre University, Via della Vasca Navale 79, I-00146 Roma, Italy.
E-mail address: ascenzi@uniroma3.it (P. Ascenzi).

http://dx.doi.org/10.1016/j.abb.2015.10.014
0003-9861/

extracellular environment [1,2,4–6]. Among the extracellular retinoid-binding proteins (RBP), the epididymal retinoid-binding protein (ERBP), the interphotoreceptor matrix retinoid-binding protein (IRBP), and the retinoid binding protein 4 (RBP4) play a central role in ROL transport, whereas lipocalin-type prostaglandin D synthase (also known as β-trace) and HSA transport preferentially all-*trans*-RA [1,2,7,8]. This suggests that RBPs may function as alterative carriers of ROL and RA in health and disease. Indeed, HSA may act as a secondary carrier in human diseases associated with reduced levels of IRBP (*e.g.*, in the early stages of diabetic retinopathy) [9] or RBP4 (*e.g.*, in prolonged dietary vitamin A deficiency, in migraine, in HIV, in diabetic patients treated with anti-TNFα-therapy, and in end-stage renal disease after kidney transplantation) [10–13].

Here, the effect of all-*trans*-RA and all-*trans*-ROL on ferric heme (heme-Fe(III)) binding to HSA is reported. All-*trans*-RA and all-*trans*-ROL inhibit competitively heme-Fe(III) association to the FA binding site 1 (FA1 site) of HSA. Experimental data have been interpreted in the light of docking simulations and analysis of local structural similarity between HSA and a set of 25 RA-binding proteins and 18 all-*trans*-ROL-binding proteins. Functional and molecular modeling data match each other very well and provide an atomic level view of the all-*trans*-RA- and all-*trans*-ROL-HSA complexes.

2. Materials

HSA (essentially FA-free), hemin (Fe(III)-protoporphyrin IX) chloride, all-*trans*-RA, all-*trans*-ROL, dansyl-arginine, and dansyl-sarcosine were obtained from Sigma-Aldrich (St. Louis, MO, USA). All the other chemicals were obtained from Merck AG (Darmstadt, Germany). All products were of analytical or reagent grade and were used without further purification. The HSA concentration was determined spectrophotometrically at 280 nm ($\varepsilon_{280\ nm} = 3.82 \times 10^4\ M^{-1}\ cm^{-1}$) [14]. The HSA stock solution (2.0×10^{-4} M) was prepared by dissolving HSA in 2.0×10^{-2} M sodium phosphate buffer, at pH 7.1. The heme-Fe(III) stock solution (5.0×10^{-3} M) was prepared by dissolving heme-Fe(III) in 1.0×10^{-2} M NaOH [14]. The heme-Fe(III) concentration was determined spectrophotometrically at 535 nm, after converting heme-Fe(III) to the heme-Fe(III)-bis-imidazolate derivative by adding 1.0 M imidazole, in sodium dodecylsulfate micelles ($\varepsilon_{535\ nm} = 1.45 \times 10^4\ M^{-1}\ cm^{-1}$) [15]. The all-*trans*-RA and all-*trans*-ROL stock solutions (2.0×10^{-4} M) were prepared by dissolving both retinoids in dimethylsulfoxide. The final all-*trans*-RA and all-*trans*-ROL concentration ranged between 3.0×10^{-6} M and 4.0×10^{-5} M.

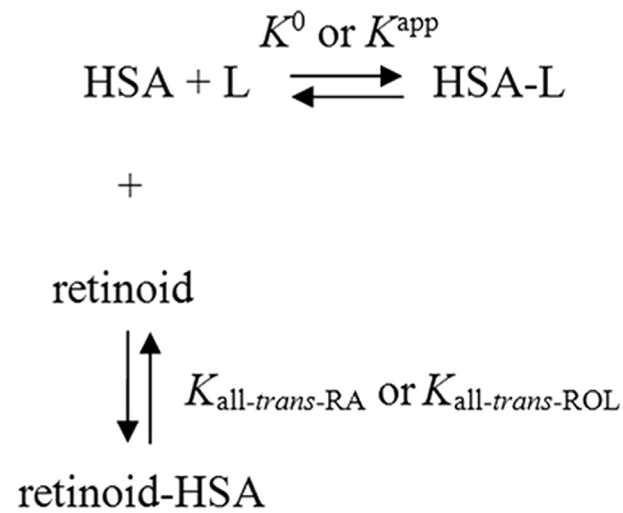
3. Methods

3.1. All-trans-RA and all-trans-ROL binding to HSA

Thermodynamics of all-*trans*-RA and all-*trans*-ROL (*i.e.*, retinoid) binding to the FA1 site, the FA3-FA4 cleft, and the FA7 site of HSA was followed by competitive inhibition of heme-Fe(III), dansyl-sarcosine, and dansyl-arginine (*i.e.*, L) association, respectively, according to Scheme 1.

In Scheme 1, K^0 and K^{app} indicate values of the dissociation equilibrium constant for heme-Fe(III) (*i.e.*, K^0_{heme} and K^{app}_{heme}), dansyl-arginine (*i.e.*, K^0_{da} and K^{app}_{da}), and dansyl-sarcosine (*i.e.*, K^0_{ds} and K^{app}_{ds}) binding to HSA in the absence (*i.e.*, K^0) and presence (*i.e.*, K^{app}) of all-*trans*-RA or all-*trans*-ROL, respectively. $K_{all\text{-}trans\text{-}RA}$ or $K_{all\text{-}trans\text{-}ROL}$ indicate values of the dissociation equilibrium constant for all-*trans*-RA or all-*trans*-ROL binding to HSA, respectively.

Heme-Fe(III) binding to HSA, in the absence and presence of all-*trans*-RA and all-*trans*-ROL, was followed spectrophotometrically between 350 and 450 nm, at pH 7.1 (2.0×10^{-2} phosphate buffer) and 20.0 °C [16]. The heme-Fe(III) concentration was 7.0×10^{-7} M, the all-*trans*-RA and all-*trans*-ROL concentration ranged between 3.0×10^{-6} M and 4.0×10^{-5} M, and the HSA concentration ranged between 1.1×10^{-7} M and 1.3×10^{-6} M. Heme-Fe(III) binding to HSA, in the absence and presence of all-*trans*-RA and all-*trans*-ROL, was investigated at a fixed low heme-Fe(III) concentration by increasing the amount of HSA. This approach minimizes the problems with heme-Fe(III) dimerization and the absorbance changes are easy to interpret [16].



Scheme 1. Binding of retinoids, heme-Fe(III), dansyl-arginine and dansyl-sarcosine (*i.e.*, L) to HSA.

Values of the dissociation equilibrium constant for heme-Fe(III) binding to HSA, in the absence and presence of all-*trans*-RA and all-*trans*-ROL, (*i.e.*, K^0_{heme} and K^{app}_{heme}, respectively; see Scheme 1) have been obtained from the dependence of the relative absorbance change (*i.e.*, $\Delta A/\Delta A_{max}$) of HSA-heme-Fe(III) formation on the free HSA concentration (*i.e.*, [HSA]), according to Eqn. (1) [16]:

$$\Delta A/\Delta A_{max} = [HSA]/(K_{heme} + [HSA]) \quad (1)$$

where K_{heme} is K^0_{heme} or K^{app}_{heme}, ΔA is the absorbance intensity change observed at each HSA concentration, and ΔA_{max} is the maximum absorbance intensity change.

Dansyl-arginine and dansyl-sarcosine binding to HSA, in the absence and presence of all-*trans*-RA and all-*trans*-ROL, was followed spectrofluorimetrically at pH 7.1 (2.0×10^{-2} phosphate buffer) and 20.0 °C. The fluorophore of dansyl-arginine and dansyl-sarcosine was excited at 370 nm; the fluorescence emission intensities were measured at the maximum wavelengths (*i.e.*, 460 nm for dansyl-arginine, and at 475 nm for dansyl-sarcosine); the excitation and emission slits were 5 nm [16–21]. The HSA concentration was 2.8×10^{-6} M, the dansyl-arginine and dansyl-sarcosine concentration ranged between 5.0×10^{-6} M and 1.0×10^{-4} M, and the all-*trans*-RA and all-*trans*-ROL concentration ranged between 5.0×10^{-6} M and 4.0×10^{-5} M.

Values of the dissociation equilibrium constants for dansyl-arginine and dansyl-sarcosine binding to HSA, in the absence and presence of all-*trans*-RA and all-*trans*-ROL, (*i.e.*, K^0_{da}, K^{app}_{da}, K^0_{ds}, and K^{app}_{ds}, respectively; see Scheme 1) were obtained from the dependence of the relative fluorescence intensity change (*i.e.*, $\Delta F/\Delta F_{max}$) of the HSA-dansyl-arginine and HSA-dansyl-sarcosine complexes on the dansylated compound concentration (*i.e.*, [dc])

in the absence (*i.e.*, K^0) and presence (*i.e.*, K^{app}) of all-*trans*-RA and all-*trans*-ROL, according to eqn. (2) [16–21]:

$$\Delta F/\Delta F_{max} = [dc]/(K + [dc]) \quad (2)$$

where K is K^0_{da} or K^{app}_{da} for dansyl-arginine binding and K^0_{ds} or K^{app}_{ds} for dansyl-sarcosine binding, ΔF is the fluorescence intensity change observed at each concentration of dansyl-arginine and dansyl-sarcosine, and ΔF_{max} is the maximum fluorescence intensity change.

Test measurements performed after 2 h of all-*trans*-RA/HSA/heme-Fe(III), all-*trans*-ROL/HSA/heme-Fe(III), all-*trans*-RA/HSA/dansyl-arginine, all-*trans*-ROL/HSA/dansyl-arginine, all-*trans*-RA/HSA/dansyl-sarcosine, and all-*trans*-ROL/HSA/dansyl-sarcosine incubation excluded slow kinetic effects.

Spectrophotometric and spectrofluorimetric measurements were carried out with a Jasco V-560 spectrophotometer and a Jasco FP-6500 spectrofluorimeter (Jasco International Co., Ltd., Tokyo, Japan), respectively.

According to competitive inhibition, the dependence of K^{app} values on the all-*trans*-RA and all-*trans*-ROL concentration was analyzed according to Eqn. (3) [22]:

$$K^{app}/K^0 = ([\text{retinoid}]/K_R) + 1 \quad (3)$$

where K^{app} is K^{app}_{heme}, K^{app}_{da} or K^{app}_{ds}, K^0 is K^0_{heme}, K^0_{da} or K^0_{ds}, retinoid is either all-*trans*-RA or all-*trans*-ROL, and K_R is either $K_{\text{all-}trans\text{-RA}}$ or $K_{\text{all-}trans\text{-ROL}}$.

Thermodynamics were analyzed using the MatLab program (The Math Works Inc., Natick, MA, USA). The results are given as mean values of at least four experiments plus or minus the corresponding standard deviation.

3.2. Docking of all-trans-RA and all-trans-ROL to HSA

Docking simulations of all-*trans*-RA and all-*trans*-ROL binding to HSA were performed using the crystal structure of ligand-free HSA (PDB ID: 1AO6) [23] and of FA-bound HSA (PDB ID: 3SQJ) [24]. All-*trans*-RA and all-*trans*-ROL three-dimensional structures were obtained respectively from the crystal structures of the all-*trans*-RA-bound ligand binding domain of RAR-γ (PDB ID: 2LBD) [25] and of the all-*trans*-ROL-bound retinol-binding protein (PDB ID: 1KT6) [26].

Simulations were carried out using the docking program Autodock Vina [27] with a search space (docking grid) that included the whole protein, in order to carry out "blind" predictions of the all-*trans*-RA and all-*trans*-ROL binding sites. Additional simulations were also carried out restricting the search space only to the FA1 site, in order to obtain more accurate details on all-*trans*-RA and all-*trans*-ROL interaction with HSA. The grid spacing was set to 1 Å per grid unit and the exhaustiveness parameter was increased from the default value of 8 to 24, as suggested by Autodock Vina developers for grid sizes larger than 27,000 Å^3 (see http://vina.scripps.edu/tutorial.html), which is the case for HSA simulations.

The simulations were carried out both by keeping all protein residues rigid and by allowing flexibility only of the residues building up the walls of the FA sites (FA1 to FA9) (see [28]). Residues for which flexibility was allowed are reported in Supplementary Table S1. Rotatable bonds of the all-*trans*-RA ad ROL lowest energy resonance structure were kept flexible in all the simulations, as suggested by Autodock Vina developers (see http://vina.scripps.edu/tutorial.html).

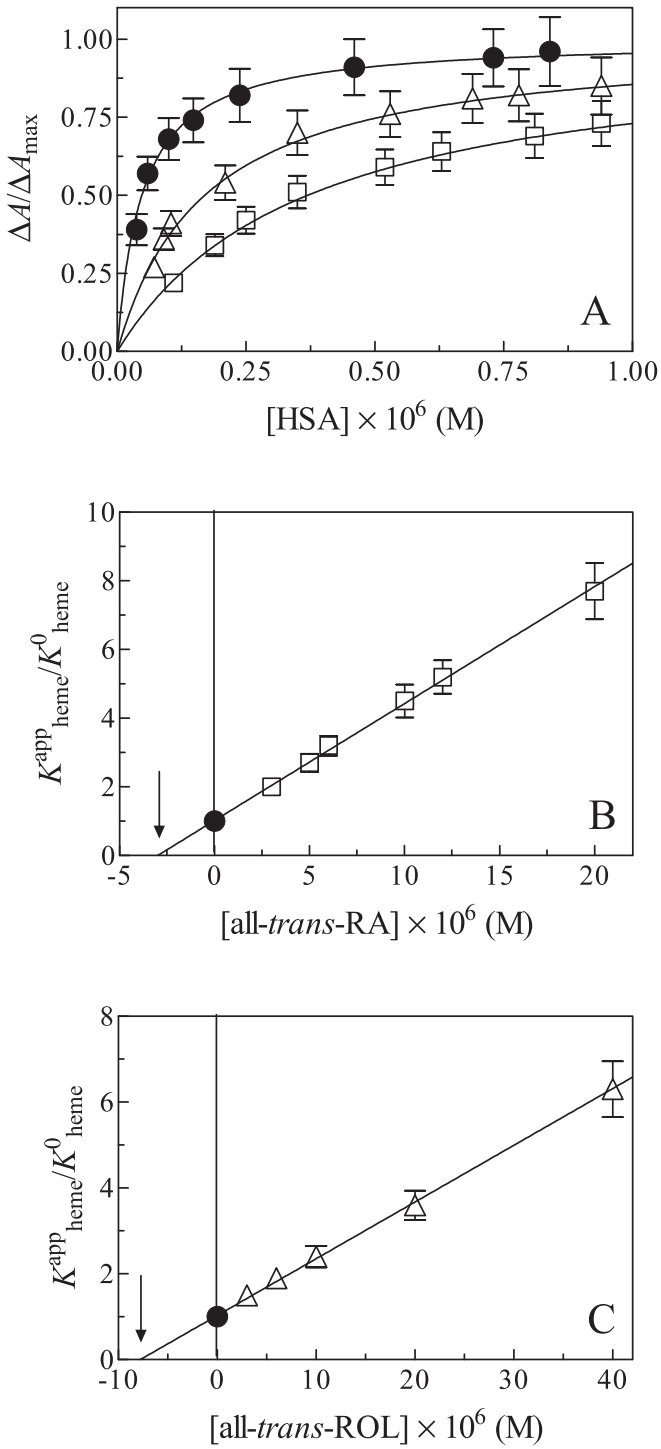


Fig. 1. Competitive inhibition of heme-Fe(III) binding to HSA by all-*trans*-RA and all-*trans*-ROL, at pH 7.1 and 20.0 °C. Binding isotherms for heme-Fe(III) association to HSA in the absence (filled circles) and presence of 2.0×10^{-5} M all-*trans*-RA (open squares) and 2.0×10^{-5} M all-*trans*-ROL (open triangles) (panel A). The HSA concentration refers to that of the free protein. The analysis of data according to Eqn. (1) allowed the determination of the following parameters: $K^0_{heme} = (4.8 \pm 0.5) \times 10^{-8}$ M (filled circles), $K^{app}_{heme} = (3.6 \pm 0.4) \times 10^{-7}$ M (open squares), and $K^{app}_{heme} = (1.7 \pm 0.2) \times 10^{-7}$ M (open triangles). Dependence of the $K^{app}_{heme}/K^0_{heme}$ ratio on the all-*trans*-RA (panel B) and all-*trans*-ROL (panel C) concentration. The analysis of data according to eqn. (2) allowed the determination of the following parameters: $K_{\text{all-}trans\text{-RA}} = (3.0 \pm 0.3) \times 10^{-6}$ M and $K_{\text{all-}trans\text{-ROL}} = (7.8 \pm 0.8) \times 10^{-6}$ M. Where not shown, the standard deviation is smaller than the symbol. For details, see text.

3.3. Analysis of local structural similarity between HSA and RA and all-trans-ROL binding proteins

The analysis of the local structural similarities between HSA and a set of RA and all-*trans*-ROL-binding proteins was carried out using the in-house developed local structure comparison programs ASSIST [29] and LIBRA [30]. These programs are based on a geometric hashing and graph theory approach, respectively, to find the largest subset of similar residues between an input protein structure and a set of structural motifs. The analysis was carried out through a structural comparison between HSA and a total of 25 different RA-binding proteins and 18 all-trans-ROL-binding proteins (see Supplementary Tables S2 and S3, respectively). The number of binding sites analyzed was 26 and 18 for RA and all-*trans*-ROL respectively.

4. Results

4.1. All-trans-RA and all-trans-ROL binding to FA-free HSA

All-*trans*-RA and all-*trans*-ROL binding to the FA1 site of HSA was followed by competitive inhibition of heme-Fe(III) association, at pH 7.1 and 20.0 °C. Heme-Fe(III) binding to the FA1 site of HSA [23,28,31,32] was investigated spectrophotometrically in the absence and presence of all-*trans*-RA and all-*trans*-ROL (the concentration of all-*trans*-RA and all-*trans*-ROL ranged between 3.0×10^{-6} M and 2.0×10^{-5} M, and between 3.0×10^{-6} M and 4.0×10^{-5} M, respectively).

The analysis of data (Fig. 1, panel A) according to eqn. (1) [16] showed that heme-Fe(III) binds to HSA, all-*trans*-RA-HSA, and all-*trans*-ROL-HSA with a simple equilibrium, values of the Hill coefficient (*i.e.*, *n*) ranging between 0.98 ± 0.02 and 1.01 ± 0.02.

As shown in Fig. 1 (panels B and C), all-*trans*-RA and all-*trans*-ROL inhibited competitively heme-Fe(III) binding to HSA, over the whole all-*trans*-RA and all-*trans*-ROL concentration range explored. The analysis of data according to Eqn. (1) [22] allowed the determination of values of the dissociation equilibrium constants for heme-Fe(III) binding to HSA in the absence of all-*trans*-RA and all-*trans*-ROL (*i.e.*, $K^0_{heme} = (4.8 \pm 0.5) \times 10^{-8}$ M) and of all-*trans*-RA and all-*trans*-ROL binding to HSA (*i.e.*, $K_{all\text{-}trans\text{-}RA} = (3.0 \pm 0.3) \times 10^{-6}$ M and $K_{all\text{-}trans\text{-}ROL} = (7.8 \pm 0.8) \times 10^{-6}$ M). Values of K^0_{heme}, $K_{all\text{-}trans\text{-}RA}$, and $K_{all\text{-}trans\text{-}ROL}$ matched very well those reported in the literature [33–36].

All-*trans*-RA and all-*trans*-ROL binding to the FA3-FA4 cleft and the FA7 site of HSA was followed by competitive inhibition of dansyl-sarcosine and dansyl-arginine association, respectively, at pH 7.1 and 20.0 °C. Dansyl-sarcosine and dansyl-arginine binding to the FA3-FA4 cleft and the FA7 site of HSA, respectively [16–21], was investigated spectrofluorimetrically in the absence and presence of all-*trans*-RA and all-*trans*-ROL (the concentration ranged between 3.0×10^{-6} M and 2.0×10^{-5} M, and between 3.0×10^{-6} M and 4.0×10^{-5} M, respectively). In the absence and presence of all-*trans*-RA and all-*trans*-ROL, dansyl-sarcosine and dansyl-arginine binding to HSA followed a simple equilibrium (according to Eqn.

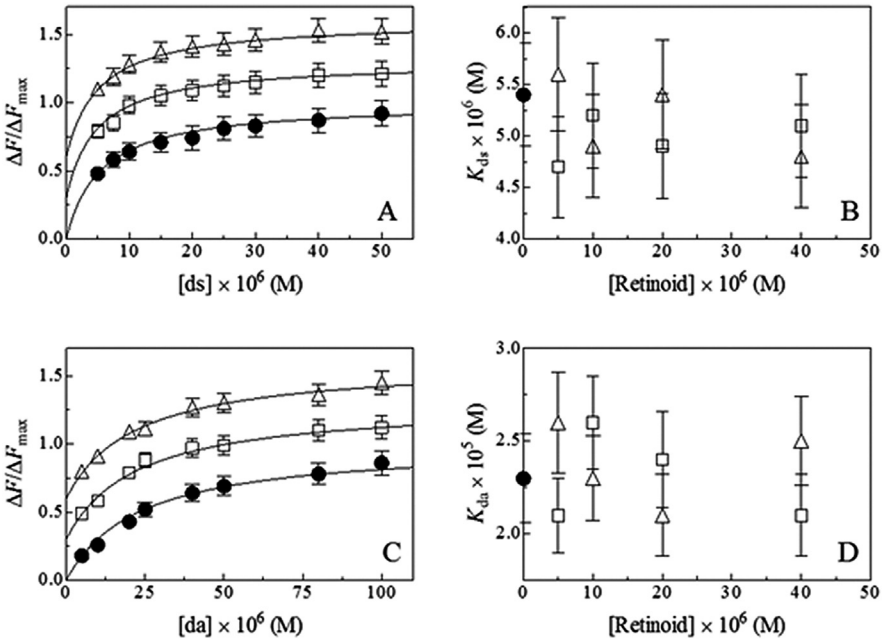


Fig. 2. Effect of all-*trans*-RA and all-*trans*-ROL concentration on dansyl-sarcosine and dansyl-arginine binding to HSA, at pH 7.1 and 20.0 °C. Binding isotherms for dansyl-sarcosine association to HSA in the absence (filled circles) and presence of 4.0×10^{-5} M all-*trans*-RA (open squares) and 4.0×10^{-5} M all-*trans*-ROL (open triangles) (panel A). For clarity, values of $\Delta F/\Delta F_{max}$ (corresponding to *Y*) of the binding isotherms for dansyl-sarcosine association to HSA in the presence of all-*trans*-RA (open squares) and all-*trans*-ROL (open triangles) have been arbitrarily up shifted of 0.3 and 0.6 units, respectively, with respect to those of the binding isotherm for dansyl-sarcosine association to HSA in the absence of all-*trans*-RA and all-*trans*-ROL (filled circles). The analysis of data according to eqn. (2) allowed the determination of the following parameters: $K^0_{ds} = (5.4 \pm 0.5) \times 10^{-6}$ M (filled circles), $K^{app}_{ds} = (4.8 \pm 0.5) \times 10^{-6}$ M (open squares), and $K^{app}_{ds} = (5.1 \pm 0.5) \times 10^{-6}$ M (open triangles). Dependence of K_{ds} on the all-*trans*-RA (open squares) and ROL (open triangles) concentration (panel B). The filled circle on the ordinate indicates the K_{ds} value obtained in the absence of all-*trans*-RA and all-*trans*-ROL. Values of K_{ds} are unaffected by all-*trans*-RA and all-*trans*-ROL. Binding isotherms for dansyl-arginine association to HSA in the absence (filled circles) and presence of 4.0×10^{-5} M all-*trans*-RA (open squares) and 4.0×10^{-5} M all-*trans*-ROL (open triangles) (panel C). For clarity, values of $\Delta F/\Delta F_{max}$ (corresponding to *Y*) of the binding isotherms for dansyl-arginine association to HSA in the presence of all-*trans*-RA (open squares) and all-*trans*-ROL (open triangles) have been arbitrarily up shifted of 0.3 and 0.6 units, respectively, with respect to those of the binding isotherm for dansyl-arginine association to HSA in the absence of all-*trans*-RA and all-*trans*-ROL (filled circles). The analysis of data according to eqn. (2) allowed the determination of the following parameters: $K^0_{da} = (2.3 \pm 0.2) \times 10^{-5}$ M (filled circles), $K^{app}_{da} = (2.2 \pm 0.2) \times 10^{-5}$ M (open squares), and $K^{app}_{da} = (2.2 \pm 0.2) \times 10^{-5}$ M (open triangles). Dependence of K_{da} on the all-*trans*-RA (open squares) and all-*trans*-ROL (open triangles) concentration (panel D). The filled circle on the ordinate indicates the K_{da} value obtained in the absence of all-*trans*-RA and all-*trans*-ROL. Values of K_{da} are unaffected by all-*trans*-RA and all-*trans*-ROL. Where not shown, the standard deviation is smaller than the symbol. For details, see text.

(2); see Fig. 2), values of the Hill coefficient (*i.e.*, n) ranging between 0.98 ± 0.02 and 1.01 ± 0.02. Values of the dissociation equilibrium constant for dansyl-sarcosine and dansyl-arginine binding to HSA (*i.e.*, K_{ds} and K_{da}, respectively) were independent of the all-*trans*-RA and all-*trans*-ROL concentration (Fig. 2). Moreover, values of K_{ds} and K_{da} obtained in the absence and presence of all-*trans*-RA and all-*trans*-ROL were in agreement with those reported in the literature [16,17,19,21]. This indicates that neither all-*trans*-RA nor all-*trans*-ROL bind to the FA3-FA4 cleft and the FA7 site of HSA over the whole retinoid concentration range explored.

4.2. Docking analysis of all-trans-RA and all-trans-ROL binding to FA-free HSA

Docking simulations of all-*trans*-RA and all-*trans*-ROL binding to HSA, with the search space extended to the whole protein, indicated the preferential binding of both compounds to the FA1 site (Fig. 3). The number of complexes observed in FA1 in docking simulations with a maximum of 9 poses was 5 for all-*trans*-RA and 6 for all-*trans*-ROL. The binding affinity of the best all-*trans*-RA and all-*trans*-ROL poses was −8.1 kcal/mol in both cases. In other poses of this simulation, all-*trans*-RA was found in the FA6 site (3 poses) and in the FA9 site (1 pose), and all-*trans*-ROL in the FA9 site (3 poses).

The simulation with the search space reduced to FA1 site provided more reliable details of the interaction: in this case, the binding affinity of all-*trans*-RA and all-*trans*-ROL for the FA1 site of HSA was −8.8 kcal/mol and −8.0 kcal/mol, respectively. The highest ranking complexes of all-*trans*-RA and all-*trans*-ROL bound to the FA1 site of HSA are shown in Fig. 3.

4.3. Docking analysis of all-trans-RA and all-trans-ROL binding to FA-bound HSA

Docking simulations of all-*trans*-RA and all-*trans*-ROL binding to FA-bound HSA, with the search space extended to the whole protein, indicated the binding of all-*trans*-RA and all-*trans*-ROL in three different sites (Fig. 4). It should be noted that in the FA-HSA structure used in this simulations the FA1 to FA7 sites are occupied by myristic acid [24]. In the first-ranking solution both ligands were placed in the FA8 site, located at the base of the crevice between subdomains IA-IB-IIA on one side and subdomains IIB-IIIA-IIIB on the other side. The FA8 site is relevant for ligand recognition only when the FA1 to FA7 sites of HSA are occupied by long-chain FAs such as myristic acid [28,37]. The binding affinity of the best all-*trans*-RA and all-*trans*-ROL poses in the FA8 site was −8.5 kcal/mol and −8.6 kcal/mol, respectively. In other poses, all-*trans*-RA and all-*trans*-ROL bind near the FA1 site occupied by myristate with an apparent free energy of −8.4 kcal/mol and −8.3 kcal/mol, respectively. Similar co-binding of aromatic

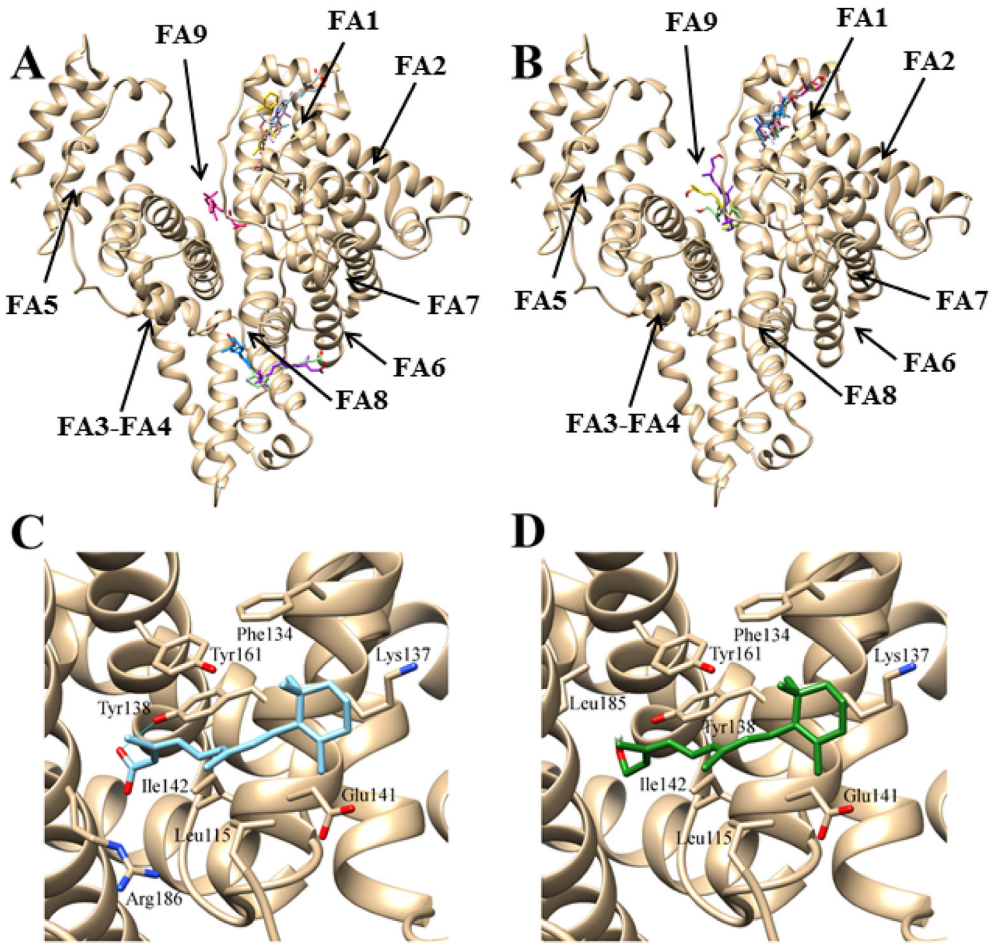


Fig. 3. All-*trans*-RA and all-*trans*-ROL recognition by FA-free HSA as obtained by docking simulations. Overall view of the nine lowest energy all-*trans*-RA-HSA complexes (panel A). Overall view of the nine lowest energy all-*trans*-ROL-HSA complexes (panel B). Atomic details of the lowest energy all-*trans*-RA-HSA complex (panel C). Atomic details of the lowest energy all-*trans*-ROL-HSA complex (panel D). Both all-*trans*-RA and all-*trans*-ROL bind in the vicinity of the FA1 site residue Tyr161 coordinating the heme-Fe(III) atom. For details, see text. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

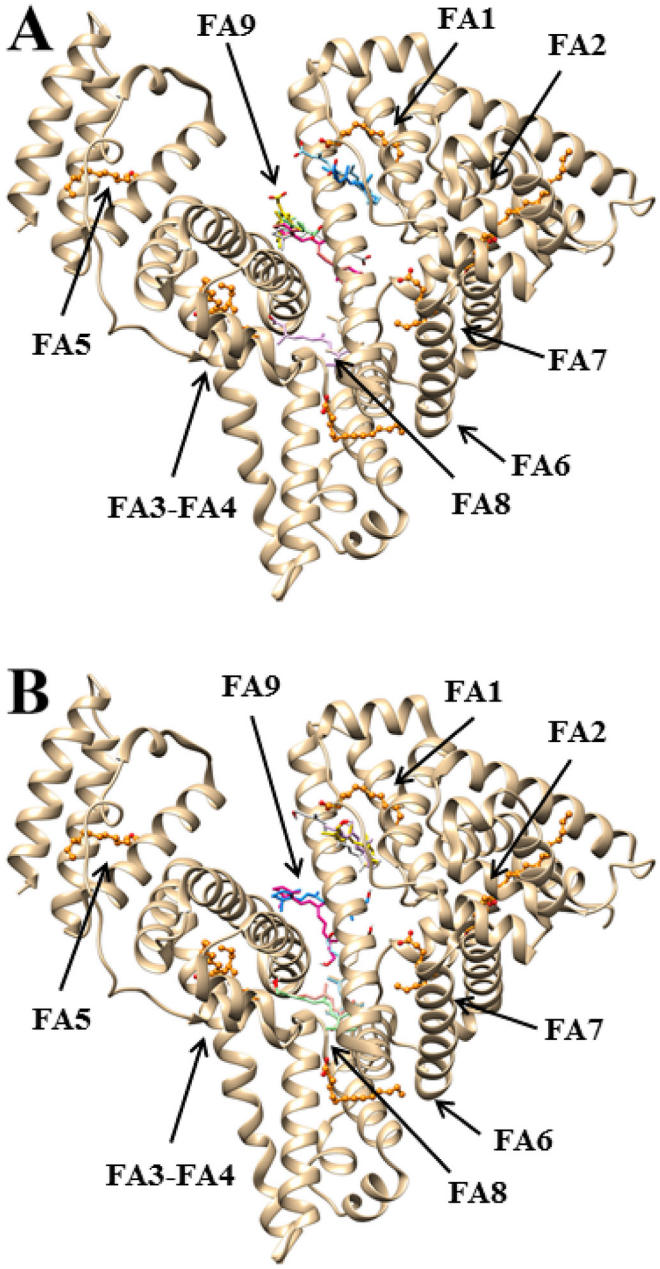


Fig. 4. All-*trans*-RA and all-*trans*-ROL recognition by FA-bound HSA as obtained by docking simulations. Overall view of the nine lowest energy all-*trans*-RA-HSA complexes (panel A). Overall view of the nine lowest energy all-*trans*-ROL-HSA complexes (panel B). Myristate molecules occupying the FA1 to FA7 sites are shown in ball-and-stick representation and colored in orange. For details, see text. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

drugs or drug-like molecules such as indomethacin and triiodobenzoic acid in the FA1 site was already reported [38,39]. Lastly, all-*trans*-RA and all-*trans*-ROL bind to the FA9 site of HSA, located in the upper region of the cleft built up by subdomains IA-IB-IIA on one side and subdomains IIB-IIIA-IIIB on the other side. Of note, the FA9 site is an additional binding pocket of HSA relevant for ligand recognition only when long-chain FAs, such as myristic acid, occupy the FA1 to FA7 sites of HSA [37]. All-*trans*-RA and all-*trans*-ROL bind to the FA9 site of HSA in a fashion similar to that observed for thyroxine recognition when long-chain FAs, such as myristic acid, occupy the FA1 to FA7 sites [40]. The binding affinity of the best RA and ROL poses in this cleft was −8.0 kcal/mol and −7.6 kcal/mol.

4.4. Analysis of local structural similarity between HSA and RA and ROL binding proteins

The analysis of the local structural similarities between HSA and the set of 25 all-*trans*-RA binding proteins (see Supplementary Table S2), carried out using the local structural comparison program LIBRA [30], identified one set of residues in HSA, which displays chemical and geometric similarity with the ligand binding domain of human RAR-γ (PDB ID: 2LBD) [25]. The match was obtained with HSA residues Phe134, Leu178, Phe165, Ala158, Val122, Arg144, and Ile142 located on the walls of the FA1 site (Fig. 5). The all-atoms rmsd value calculated between these residues of HSA and the corresponding side chains of human RAR-γ was 1.7 Å. Noteworthy, amino acid residues coordinating all-*trans*-RA in human RAR-γ (Fig. 5) are conserved also in human RAR-α and -β [1].

The same analysis carried out between three-dimensional structures of HSA and the set of 18 all-*trans*-ROL binding proteins (see Supplementary Table S3) gave no significant results. This may be due to the fact that, with the only exception of retinol dehydratase, all the ROL binding proteins present in the dataset belong to the lipocalin family and display a very similar ROL binding environment. Thus, this dataset is less rich and much more redundant with respect to that of the RA binding proteins, which, in addition to lipocalins, comprises also nuclear receptors and CYP450 variants.

5. Discussion

Functional studies here reported demonstrate that, in the absence of FAs, all-*trans*-RA and all-*trans*-ROL bind to the FA1 site of HSA, impairing competitively heme-Fe(III) association (Fig. 1). On the other hand, all-*trans*-RA and all-*trans*-ROL do not bind to the FA3-FA4 cleft and the FA7 site, respectively; in fact, retinoids do not affect the association of dansyl-sarcosine and dansyl-arginine (Fig. 2). Noteworthy, the FA7 site, the FA3-FA4 cleft, and the FA1 pocket (located in subdomains IIA, IIIA, and IB, respectively) represent the first, the second, and the third major ligand binding region of HSA, respectively [41].

Docking studies undertaken to complement the experimental characterization of RA and ROL binding to HSA indicate a preferential binding of these compounds in the FA1 site in FA-free HSA, and in FA8 and FA9 sites in the myristate-HSA complex (Figs. 3 and 4). Furthermore, the analysis of the local structural similarities between HSA and a set of RA binding proteins identifies a set of residues in the FA1 pocket of HSA which display chemical and geometric similarity with those present in the ligand (*e.g.*, RA) binding domain of the RAR-γ (Fig. 5). Noteworthy, the amino acid sequence analysis indicates that residues coordinating all-*trans*-RA are conserved also in the ligand binding domain of human RAR-α and -β [1]. Both all-*trans*-RA and ROL bind HSA in the vicinity of Tyr161, which in HSA-heme-Fe(III) coordinates the heme iron atom [31,42–44], and establish hydrophobic interactions with Tyr138 (Fig. 3). Sitting of Tyr138 outside the FA1 site is a pre-requisite for heme-Fe(III) binding, while, according to docking simulations, retinoids relocate Tyr138 within the FA1 pocket, thus inhibiting heme-Fe(III) association [45]. In this respect, it must be stressed that binding of heme-Fe(III) and retinoids to the FA1 site of HSA must be regarded as alternative, as the FA1 pocket dimensions would not allow simultaneous binding of heme-Fe(III) and retinoids. On the other hand, both functional and structural evidences exclude RA and ROL binding to the FA3-FA4 cleft and the FA7 site (Figs. 1–4). Interestingly, docking simulations carried out on the FA-bound HSA predict binding of retinoids in alternative sites (FA8 site, nearby FA1 site and FA9 site) as it was observed for aromatic drugs and drug-like compounds as well as for thyroxine [40]. RA and ROL-

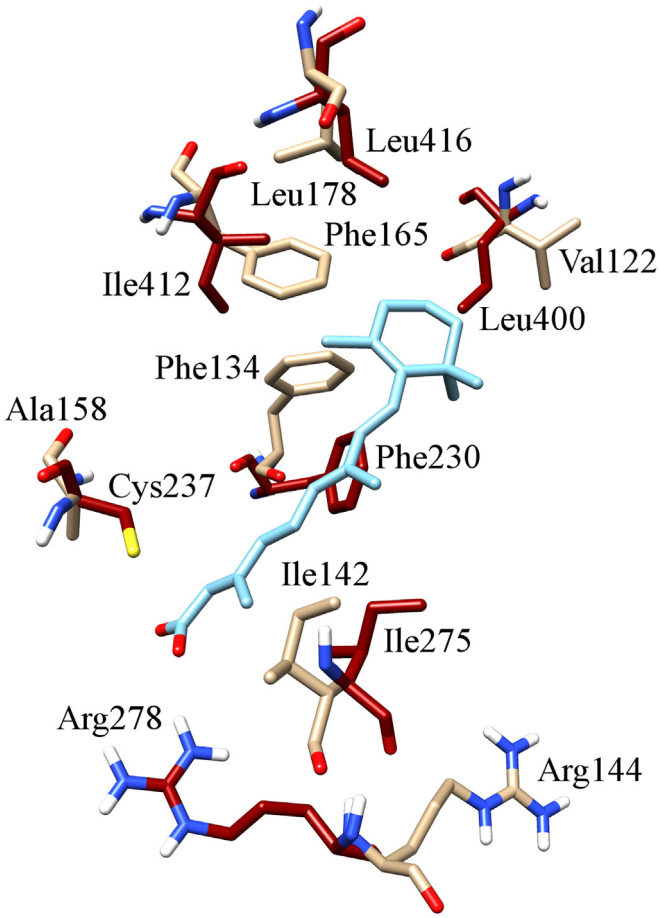


Fig. 5. Schematic representation of residues forming the walls of the FA1 site of HSA, in light brown (PDB ID: 1AO6) [23], displaying chemical and geometric similarity with those of the all-*trans*-RA binding pocket of human RAR-γ, in dark red (PDB ID: 2LBD) [25]. For details, see text. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

based modulation of HSA-heme-Fe(III) binding resembles that induced by FAs that inhibit both competitively and allosterically heme-Fe(III) binding and heme-Fe-based reactivity of HSA [5,46].

The low affinity of RA and ROL for FA-free HSA with respect to retinoid binding to extracellular RBPs [47–49], reflects first of all a lower ligand-receptor shape complementarity. Indeed, RBPs are characterized by a rigid hydrophobic binding pocket that displays a high surface complementarity with the ligand [50]. This increases the contact surface with the ligand and thus maximizes the contribution of hydrophobic interactions to the binding energy, which is proportional to the contact surface area [51]. An additional factor is the desolvation penalty due to partial burial of the alcohol/carboxylate group of all-*trans*-ROL and all-*trans*-RA, respectively, in retinoid-bound HSA (Fig. 3). In fact in RBPs the polar group of the ligand is always solvent accessible, which is not the case of retinoids-HSA complexes, according to docking simulations. This latter aspect is particularly relevant, in that it has been shown that the energetic penalty of burying polar groups in protein–protein and protein-ligand complexes is usually not compensated by the established coulombic interactions [52]. These unfavorable contributions, though partially balanced by the hydrophobic interactions and the hydrogen bond/salt bridge (in the case of ROL and RA, respectively) formed between retinoids and HSA (Fig. 3), are probably at the basis of the lower affinity of HSA for retinoids as compared to extracellular RBPs.

Although values of the dissociation equilibrium constant for all-*trans*-RA and all-*trans*-ROL binding to HSA (ranging between ~ 3×10^{-6} M and ~ 8×10^{-6} M) (present data and refs [33,34]) are lower than those for retinoid binding to extracellular RBPs (ranging between ~ 1×10^{-10} M and ~1×10^{-8} M) [47–49], HSA may serve as a secondary carrier for retinoids [1,2,53] in some pathological conditions characterized by low levels of IRBP and RBP4 [9–13].

Competitive inhibition of heme-Fe(III) binding to HSA by retinoids may affect the heme-Fe(III)-based catalytic properties and the maintenance of cellular homeostasis. Indeed, heme-Fe(III) not incorporated into proteins as well as released from heme-proteins: (*i*) affects NO homeostasis, (*ii*) does not contribute to the detoxification of nitrogen and oxygen reactive species, (*iii*) does not participate to the conversion of toxic peroxynitrite to harmless nitrate, (*iv*) catalyzes the synthesis of toxic free hydroxyl radicals, (*v*) induces hemolysis affecting the erythrocyte membrane stability, and (*vi*) is strongly pro-inflammatory. As a whole, the amount of free heme-Fe must be therefore tightly controlled to maintain cellular homeostasis and avoid pathological conditions [28,54–60].

Acknowledgments

Authors wish to thank Prof. Clara Nervi (University of Roma "La Sapienza", Roma, Italy) for helpful discussions. This study was partially supported by CUIA 2014 to PA.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.abb.2015.10.014.

References

[1] A. di Masi, L. Leboffe, E. De Marinis, F. Pagano, L. Cicconi, C. Rochette-Egly, F. Lo-Coco, P. Ascenzi, C. Nervi, Mol. Asp. Med. 41 (2015) 1–115.
[2] S.M. O'Byrne, W.S. Blaner, J. Lipid Res. 54 (2013) 1731–1743.
[3] A.C. Ross, J. Nutr. 133 (2003) 291S–296S.
[4] S. Vogel, M. Gamble, W.S. Blaner, in: H. Nau, W.S. Blaner (Eds.), The Handbook of Experimental Pharmacology: Retinoids, vol. 139, Springer Verlag, Berlin, 1999, pp. 31–95.
[5] N. Noy, Biochem. J. 348 (2000) 481–495.
[6] C. Folli, V. Calderone, I. Ramazzina, G. Zanotti, R. Berni, J. Biol. Chem. 277 (2002) 41970–41977.
[7] A.P. Kuruvilla, G.M. Hochwald, J. Ghiso, E.M. Castano, M. Pizzolato, B. Brain Res. 565 (1991) 337–340.
[8] M. Zahn, M. Mäder, B. Schmidt, E. Bollensen, K. Felgenhauer, Neurosci. Lett. 154 (1993) 93–95.
[9] M. Garcia-Ramírez, C. Hernández, M. Villarroel, F. Canals, M.A. Alons, R. Fortuny, L. Masmiquel, A. Navarro, J. García-Arumí, R. Simó, Diabetologia 52 (2009) 2633–2641.
[10] W.X. Zhang, W. Zhou, Z.M. Zhang, Z.Q. Zhang, J.F. He, B.Y. Shi, Genet. Mol. Res. 13 (2014) 8126–8134.
[11] S.R. Kotzé, R. Zinyama-Gutsire, P. Kallestrup, C.S. Benn, E. Gomo, J. Gerstoft, G. van Dam, O.H. Mortensen, H. Ullum, C. Erikstrup, Int. J. Infect. Dis. 33 (2015) 159–164.
[12] T. Pina, F. Genre, R. Lopez-Mejias, S. Armesto, B. Ubilla, V. Mijares, T. Dierssen-Sotos, A. Corrales, M.A. Gonzalez-Lopez, M.C. Gonzalez-Vela, R. Blanco, J. Llorca, M.A. Gonzalez-Gay, J. Eur. Acad. Dermatol. Venereol. (2015), http://dx.doi.org/10.1111/jdv.13005.
[13] N. Tanik, A. Celikbilek, A. Metin, A.Y. Gocmen, L.E. Inan, Neurol. Sci. (2015), http://dx.doi.org/10.1007/s10072-015-2262-6.
[14] V.G. Kharitonov, V.S. Sharma, D. Magde, D. Koesling, Biochemistry 36 (1997) 6814–6818.
[15] A. Boffi, T.K. Das, S. Della Longa, C. Spagnuolo, D.L. Rousseau, Biophys. J. 77 (1999) 1143–1149.
[16] E. Di Muzio, F. Polticelli, V. Trezza, G. Fanali, M. Fasano, P. Ascenzi, Arch. Biochem. Biophys. 560 (2014) 100–112.
[17] G. Sudlow, D.J. Birkett, D.N. Wade, Mol. Pharmacol. 11 (1975) 824–832.
[18] G. Sudlow, D.J. Birkett, D.N. Wade, Mol. Pharmacol. 12 (1976) 1052–1061.
[19] K. Nakajou, H. Watanabe, U. Kragh-Hansen, T. Maruyama, M. Otagiri, Biochim. Biophys. Acta 1623 (2003) 88–97.
[20] A.J. Ryan, J. Ghuman, P.A. Zunszain, C.W. Chung, S. Curry, J. Struct. Biol. 174 (2011) 84–91.
[21] G. Fanali, M. Fasano, P. Ascenzi, J.-M. Zingg, A. Azzi, BioFactors 39 (2013) 294–303.

[22] P. Ascenzi, M.G. Ascenzi, G. Amiconi, Biochem. Mol. Biol. Edu 15 (1987) 134–135.
[23] S. Sugio, A. Kashima, S. Mochizuki, M. Noda, K. Kobayashi, Protein Eng. 12 (1999) 439–446.
[24] Y. He, T. Ning, T. Xie, Q. Qiu, L. Zhang, Y. Sun, D. Jiang, K. Fu, F. Yin, W. Zhang, L. Shen, H. Wang, J. Li, Q. Lin, Y. Sun, H. Li, Y. Zhu, D. Yang, Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 19078–19083.
[25] J.P. Renaud, N. Rochel, M. Ruff, V. Vivat, P. Chambon, H. Gronemeyer, D. Moras, Nature 378 (1995) 681–689.
[26] V. Calderone, R. Berni, G. Zanotti, J. Mol. Biol. 329 (2003) 841–850.
[27] O. Trott, A.J. Olson, J. Comp. Chem. 31 (2010) 455–461.
[28] G. Fanali, A. di Masi, V. Trezza, M. Marino, M. Fasano, P. Ascenzi, Mol. Asp. Med. 33 (2012) 209–290.
[29] S. Caprari, D. Toti, L. Viet Hung, M. Di Stefano, F. Polticelli, Bioinformatics 30 (2014) 1022–1024.
[30] L. Viet Hung, S. Caprari, M. Bizai, D. Toti, F. Polticelli, Bioinformatics (2015), http://dx.doi.org/10.1093/bioinformatics/btv489.
[31] P.A. Zunszain, J. Ghuman, T. Komatsu, E. Tsuchida, S. Curry, BMC Struct. Biol. 3 (2003) 6.
[32] R.A. Laskowski, M.B. Swindells, J. Chem. Inf. Model 51 (2011) 2778–2786.
[33] T.K. Maiti, K.S. Ghosh, J. Debnath, S. Dasgupta, Int. J. Biol. Macromol. 38 (2006) 197–202.
[34] C.N. N'soukpoé-Kossi, R. Sedaghat-Herati, C. Ragi, S. Hotchandani, H.A. Tajmir-Riahi, Int. J. Biol. Macromol. 40 (2007) 484–490.
[35] D.A. El-Hady, H.M. Albishri, J. Chromatogr, B. Analyt, Technol. Biomed. Life Sci. 911 (2012) 180–185.
[36] A. Belatik, S. Hotchandani, J. Bariyanga, H.A. Tajmir-Riahi, Eur. J. Med. Chem. 48 (2012) 114–123.
[37] A.A. Bhattacharya, T. Grüne, S. Curry, J. Mol. Biol. 303 (2000) 721–732.
[38] S. Curry, H. Mandelkow, P. Brick, N. Franks, Nat. Struct. Biol. 5 (1998) 827–835.
[39] J. Ghuman, P.A. Zunszain, I. Petitpas, A.A. Bhattacharya, M. Otagiri, S. Curry, J. Mol. Biol. 353 (2005) 38–52.
[40] I. Petitpas, C.E. Petersen, C.E. Ha, A.A. Bhattacharya, P.A. Zunszain, J. Ghuman, N.V. Bhagavan, S. Curry, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 6440–6445.
[41] F. Zsila, Mol. Pharm. 10 (2013) 1668–1682.
[42] J.K. Kamal, D.V. Behere, Indian. J. Biochem. Biophys. 42 (2005) 7–12.
[43] F.P. Nicoletti, B.D. Howes, M. Fittipaldi, G. Fanali, M. Fasano, P. Ascenzi, G. Smulevich, J. Am. Chem. Soc. 130 (2008) 11677–11688.
[44] C. Meneghini, L. Leboffe, M. Bionducci, G. Fanali, M. Meli, G. Colombo, M. Fasano, P. Ascenzi, S. Mobilio, PLoS One 9 (2014) e104231.
[45] F. Zsila, J. Phys. Chem. B 117 (2013) 10798–10806.
[46] G. Fanali, R. Fesce, C. Agrati, P. Ascenzi, M. Fasano, FEBS J. 272 (2005) 4672–4683.
[47] M.S. Levin, B. Locke, N.C. Yang, E. Li, J.I. Gordon, J. Biol. Chem. 263 (1988) 17715–17723.
[48] Y. Urade, O. Hayaishi, Biochim. Biophys. Acta 1482 (2000) 259–271.
[49] F. Gonzalez-Fernandez, T. Bevilacqua, K.I. Lee, R. Chandrashekar, L. Hsu, M.A. Garlipp, J.B. Griswold, R.K. Crouch, D. Ghosh, Invest. Ophthalmol. Vis. Sci. 50 (2009) 5577–5586.
[50] C.I. Branden, J. Tooze, Introduction to Protein Structure, Garland Publishing, New York, 1991.
[51] K.A. Sharp, A. Nicholls, R.F. Fine, B. Honig, Science 252 (1991) 106–109.
[52] F. Polticelli, P. Ascenzi, M. Bolognesi, B. Honig, Protein Sci. 8 (1999) 2621–2629.
[53] P. Ascenzi, G. Fanali, M. Fasano, V. Pallottini, V. Trezza, J. Mol. Struct. 1077 (2014) 4–13.
[54] S.W. Ryter, R.M. Tyrrell, Free Radic. Biol. Med. 228 (2000) 289–309.
[55] V. Jeney, J. Balla, A. Yachie, Z. Varga, G.M. Vercellotti, J.W. Eaton, G. Balla, Blood 100 (2002) 879–887.
[56] S. Kumar, U. Bandyopadhyay, Toxicol. Lett. 157 (2005) 175–188.
[57] J. Balla, G.M. Vercellotti, V. Jeney, A. Yachie, Z. Varga, H.S. Jacob, J.W. Eaton, G. Balla, Antioxid. Redox Signal 9 (2007) 2119–2137.
[58] E. Tolosano, S. Fagoone, N. Morello, F. Vinchi, V. Fiorito, Antioxid. Redox Signal 12 (2010) 305–320.
[59] P. Ascenzi, S. Gianni, IUBMB Life 65 (2013) 836–844.
[60] A. di Masi, L. Leboffe, V. Trezza, G. Fanali, M. Coletta, M. Fasano, P. Ascenzi, Curr. Pharm. Des. 21 (2015) 1837–1847.

Amino Acids (2015) 47:949–961
DOI 10.1007/s00726-015-1924-2

ORIGINAL ARTICLE

Pacific oyster polyamine oxidase: a protein missing link in invertebrate evolution

Manuela Cervelli · Fabio Polticelli · Emanuela Angelucci · Elena Di Muzio · Pasquale Stano · Paolo Mariottini

Received: 12 November 2014 / Accepted: 15 January 2015 / Published online: 6 February 2015

Abstract Polyamine oxidases catalyse the oxidation of polyamines and acetylpolyamines and are responsible for the polyamine interconversion metabolism in animal cells. Polyamine oxidases from yeast can oxidize spermine, N^1-acetylspermine, and N^1-acetylspermidine, while in vertebrates two different enzymes, namely spermine oxidase and acetylpolyamine oxidase, specifically catalyse the oxidation of spermine, and N^1-acetylspermine/N^1-acetylspermidine, respectively. In this work we proved that the specialized vertebrate spermine and acetylpolyamine oxidases have arisen from an ancestor invertebrate polyamine oxidase with lower specificity for polyamine substrates, as demonstrated by the enzymatic activity of the mollusc polyamine oxidase characterized here. This is the first report of an invertebrate polyamine oxidase, the Pacific oyster *Crassostrea gigas* (CgiPAO), overexpressed as a recombinant protein. This enzyme was biochemically characterized and demonstrated to be able to oxidase both N^1-acetylspermine and spermine, albeit with different efficiency. Circular dichroism analysis gave an estimation of the secondary structure content and modelling of the three-dimensional structure of this protein and docking studies highlighted active site features. The availability of this pluripotent enzyme can have applications in crystallographic studies and pharmaceutical biotechnologies, including anticancer therapy as a source of hydrogen peroxide able to induce cancer cell death.

Keywords Polyamine oxidase · Pacific oyster · Enzyme activity · Tertiary structure

Abbreviations

AcSpd	N^1-acetylspermidine
AcSpm	N^1-acetylspermine
APAO	N^1-acetylpolyamine oxidase
CgiPAO	*Crassostrea gigas* polyamine oxidase
FAD	Flavin adenine dinucleotide

Handling Editor: S. Beninati.

Electronic supplementary material The online version of this article (doi:10.1007/s00726-015-1924-2) contains supplementary material, which is available to authorized users.

M. Cervelli (✉) · F. Polticelli · E. Angelucci · E. Di Muzio · P. Stano · P. Mariottini
Department of Sciences, University of Roma Tre, 00146 Rome, Italy
e-mail: manuela.cervelli@uniroma3.it

F. Polticelli
e-mail: polticel@uniroma3.it

E. Angelucci
e-mail: emanuela.angelucci@uniroma3.it

E. Di Muzio
e-mail: elena.dimuzio@uniroma3.it

P. Stano
e-mail: pasquale.stano@uniroma3.it

P. Mariottini
e-mail: paolo.mariottini@uniroma3.it

M. Cervelli · P. Mariottini
Interuniversity Consortium of Structural and Systems Biology, 00136 Rome, Italy

F. Polticelli
National Institute of Nuclear Physics, Roma Tre Section, 00146 Rome, Italy

FMN	Flavin mononucleotide
H_2O_2	Hydrogen peroxide
PAO	Polyamine oxidase
PAs	Polyamines
Put	Putrescine
Spd	Spermidine
Spm	Spermine
SMO	Spermine oxidase
ZmPAO	Maize polyamine oxidase

Introduction

Polyamine oxidases (PAOs) are flavin adenine dinucleotide (FAD)-containing enzymes that catalyse the oxidation of polyamines (PAs) at the secondary amino group, thereby giving different products according to the organism considered. Vertebrate acetylpolyamine oxidases (APAOs, EC 1.5.3.11) participate in the interconversion metabolism of PAs, converting N^1-acetyl derivatives of spermine (Spm) and spermidine (Spd) into Spd and putrescine (Put), respectively, plus 3-acetamidopropanal and hydrogen peroxide (H_2O_2) (McIntire and Hartman 1993; Van den Munckhof et al. 1995; Cervelli et al. 2003, 2004; Bellelli et al. 2004). On the other hand, in vertebrates, Spm is directly oxidized by the enzyme spermine oxidase (SMO, EC 1.5.3.16), a well-characterized flavoprotein (Cervelli et al. 2003, 2013a), and Spm oxidation leads to the production of Spd, 3-aminopropanal and H_2O_2. In the last decades, these PA catabolic enzymes have been extensively characterized and it is well documented that both oxidases play an essential role in maintaining vertebrate PA homeostasis, which is mandatory for cellular life, and their dysregulation is associated with pathological diseases (Thomas and Thomas 2001; Agostinelli et al. 2004; Casero and Marton 2007; Amendola et al. 2009, 2014; Cervelli et al. 2010, 2013a, b, 2014a; Capone et al. 2013). Modelling studies of mouse APAO (Bianchi et al. 2006) and SMO (Bianchi et al. 2005; Cervelli et al. 2003; Tavladoraki et al. 2011) enzymes have indicated a very similar active site environment for the two proteins, making it difficult to rationalize their different substrate specificity (Cervelli et al. 2012, 2014b). The biochemical characterization and the crystal structure analysis of the *Saccharomyces cerevisiae* enzyme Fms1 have demonstrated that this PAO is able to oxidize Spm, N^1-acetylSpm (AcSpm) and N^1-acetylSpd (AcSpd) (Landry and Sternglanz 2003; Huang et al. 2005). The molecular evolutionary history of this gene family has been described in Metazoa (Polticelli et al. 2012). It was proposed that the biochemical function of invertebrate enzymes, able to catalyse the oxidation of both acetylated and non-acetylated polyamines, is accomplished in vertebrates by the two specialized polyamine oxidases APAO and SMO, which are derived from the ancestral invertebrate enzyme (Polticelli et al. 2012). In this study we report the cloning and biochemical characterization of the invertebrate PAO enzyme from the Pacific oyster (*Crassostrea gigas*), named CgiPAO, capable of oxidizing AcSpm and Spm. In particular, the recombinant enzyme was expressed in a tagged form into *Escherichia coli* BL21 DE3 cells. The purified recombinant enzyme displays K_m and k_{cat} values that corroborate this view. Molecular modelling of the CgiPAO protein highlights active site features shared with other structurally characterized PAOS and support the hypothesis that the specialized vertebrate APAOs and SMOs have arisen from less substrate-specific invertebrate PAOs.

Materials and methods

Protein sequence homology search and phylogenetic analysis

Full-length PAO protein sequences were retrieved from public databases (UniProt, http://www.ebi.ac.uk/uniprot/ PFAM, http://www.sanger.ac.uk/Sofware/Pfam/ and GenBank http://www.ncbi.nlm.nih.gov/). Whenever possible, sequences whose enzymatic activities had been previously verified have been used as query sequences in BLAST searches. Blastp scoring parameters were the following: max target sequences: 100; short queries: automatically adjust parameters for short input sequences; expected threshold: 10; word size: 3; max matches in a query range: 0; matrix: BLOSUM62; gap costs: existence: 11, extension: 1; compositional adjustments: conditional compositional score matrix adjustment. Additional predicted invertebrate PAO and plant protein sequences were retrieved from currently sequenced and unfinished genomes at the Ensembl database (http://www.ensembl.org/index.html). Multiple amino acidic sequences alignment was performed using CLUSTALW (Larkin et al. 2007). Phylogenetic analyses were performed by the Maximum Likelihood and the Bayesian inference methods using plant PAOs as outgroup. Maximum Likelihood analysis was carried out in PALM (Chen et al. 2009). The JTT model with a proportion of invariable sites (I) and gamma-distributed rates across sites (G) was selected for the phylogenetic inference. Nodes support for the resulting phylogenetic tree was evaluated by 1,000 mL bootstrap replications. Bayesian inference analysis was performed with MrBayes 3.2 (Ronquist et al. 2012) under the same model of protein evolution [two independent runs of four Markov chain Monte Carlo (MCMC) chains for 1 million generations each, sampling every 100 generations, and posterior probabilities for nodes derived from a majority rule consensus of the trees sampled after convergence (25 % was sett as burn-in for sumt and sump).

Chemicals and DNA methodology

Spm and Spd were purchased from Sigma; **AcSpm** originally purchased from Sigma is no longer commercially available. The methods described by Sambrook et al. (1989) were used for the extraction and manipulation of plasmid DNA and general DNA in vitro methods. Nucleotide sequencing was carried out for both strands using the automated fluorescent dye terminator technique (Perkin Elmer ABI model 373A).

Construction of pCgiSMO-HT expression plasmid

The Pacific oyster CgiPAO cDNA sequence was deduced from the corresponding gene (GenBank accession number EKC38899) (Zhang et al. 2012). For the CgiPAO cDNA synthesis, total RNA was purified, using TRIzol reagent (Gibco BRL) according to the manufacturer's instructions, from the mantle membrane of a *C. gigas* specimen collected from the Tyrrhenian Sea and then used for RT-PCR. Synthesis of the specific cDNA was obtained using the primer CgiPAO-REV (5′-ACGTTTAGTCTGAGTACAGTTTG-3′) and the SuperScript III First-Strand (Invitrogen) according to the manufacturer's instructions. Full-length CgiPAO cDNA synthesis was obtained from the first-strand cDNA using as primers two oligonucleotides designed to introduce *Nde*I and *Xho*I restriction sites at the 5′ and 3′ ends: CgiPAO-FOR (5′-CATATGGGCGAAAAAGTGATAGATG-3′) and CgiPAO-HT-REV (5′-AAATATCTCGAGTCAATGAT GATGATGATGATGGTCTGAGTACAGTTTG-3′), respectively. The amplified PCR product was restricted with *Nde*I and *Xho*I and ligated with the restricted *Nde*I/*Xho*I pET17b vector (Novagen), to obtain the genetic construct named pCgiPAO-HT containing the mature form of CgiPAO protein joined to a C-terminal 6xHisTAG. The recombinant cDNA construct was sequenced to check the accuracy of the nucleotide sequence and then used to transform *E. coli* BL21 DE3 (Novagen) competent cells.

Expression of CgiPAO in *E. coli* cells

Escherichia coli BL21 DE3 cells transformed with the pCgiPAO-HT plasmid were grown at 30 °C in LB medium containing 100 μg/mL ampicillin to $A_{660} = 0.6$ and then induced with isopropyl-β-D-thiogalactopyranoside (1 mM final concentration), followed by further cultivation for 5 h at 28 °C. The *E. coli* BL21 DE3 cells were harvested by centrifugation (4 °C, 10 min, 10,000*g*), washed with 0.4 volume/culture weight of 30 mM Tris–HCl (pH 8.0), containing 20 % sucrose and 1 mM EDTA, and incubated 5–10 min at room temperature. The suspension was then centrifuged (10,000*g* for 10 min at 4 °C) and the cell pellet stored at −20 °C until use.

Rapid affinity purification with pET His Tag systems

Cell pellet was resuspended in 5 mM imidazole, 0.5 M NaCl, 10 % glycerol, 20 mM Tris–HCl (pH 7.9) and then sonicated. After centrifugation (12,000*g* for 30 min at 4 °C) the soluble cell extract was purified by the His-Bind chromatography kit (Novagen). In particular, the supernatant was applied to a column (3 mL) with Ni^{2+} cations immobilized on the His-Bind resin (Novagen), equilibrated with the binding buffer (5 mM imidazole, 0.5 M NaCl, 10 % glycerol, 20 mM Tris–HCl, pH 7.9). The column was washed with binding buffer and then elution was performed with 500 mM imidazole, 0.5 M NaCl, 10 % glycerol, 20 mM Tris–HCl (pH 7.9). SDS/PAGE electrophoretic analysis was performed on the purified recombinant CgiPAO tagged protein to assess the purity grade.

Absorption and fluorescence spectroscopy

Absorption spectra were measured by an Agilent 8453 diode-array spectrophotometer using a 1 cm quartz microcuvette (sample volume 80 μL). Excitation and emission spectra were recorded by a Jasco FP-6200 fluorimeter, using a 0.3 cm quartz microcuvette (sample volume 60 μL), with the following instrumental settings: excitation wavelength: 450 nm, emission wavelength: 530 nm, excitation and emission slits: 5 nm, scan rate: 250 nm/min, sensitivity: high. Fluorescence spectra were recorded at a protein concentration of 2.5 μM. All measurements were carried out in 0.1 M Tris–HCl (pH 8.5) at 25.0 °C.

Cofactor analysis

Purified recombinant CgiPAO was denatured by boiling for 15 min in the dark. The denatured protein was removed by centrifugation and the excitation/emission spectra of the supernatant were recorded before and after 15×10^{-3} U/mL of phosphodiesterase (PDE) treatment [which converts FAD to flavin mononucleotide (FMN)]. Once determined the chemical nature of the cofactor (FAD), the protein concentration was evaluated spectrophotometrically using a molar extinction coefficient ε_{450} of 10,000 cm^{-1} M^{-1}.

Circular dichroism (CD)

CD spectra of CgiPAO in 5 mM Tris–HCl (pH 8.5) were recorded at 25.0 °C using a Jasco J-600 spectropolarimeter and cylindrical quartz cells having 0.05 cm (far UV) or 0.5 cm (near UV/Vis) path length. Protein concentration was calculated spectrophotometrically, and resulted to be 1 and 100 μM (i.e. 55 and 5,500 μg/mL) for far-UV and near-UV/Vis analyses, respectively. Instrumental

ellipticity θ (mdeg) was converted to mean molar ellipticity $[\theta]$ (deg cm^2/dmol) and $\Delta\varepsilon$ (cm^{-1} M^{-1}) by taking into account the cell path length and the protein concentration. For each sample, sixteen CD spectra in the far-UV region (190–240 nm) and four in the near-UV/Vis region (300–550 nm) were recorded, averaged and smoothed. Spectra are shown in terms of $\Delta\varepsilon$ per residue in the far-UV region and $[\theta]_{FAD}$ in the near-UV/Vis region. Estimation of the secondary structure content was carried out using Dicroprot2000 (Deleage and Geourjon 1993) and CD-PRO (Sreerama and Woody 2000) software packages, according to the K2D neural network algorithm (Andrade et al. 1993).

CgiPAO catalytic activity assay and determination of the kinetic constants

Enzyme activity was measured spectrophotometrically by following the formation of a pink adduct ($\varepsilon_{515} = 2.6 \times 10^4$ M^{-1} cm^{-1}) as a result of the oxidation and following condensation of aminoantipyrine and 3,5-dichloro-2-hydroxybenzenesulfonic acid, catalysed by horseradish peroxidase (Polticelli et al. 2005). Measurements were performed in 0.1 M Tris–HCl buffer (pH 8.5), with substrates at various concentrations. In the enzyme assays the CgiPAO concentration was in the range of 1.0–5.0 μM. k_{cat} values were calculated using saturating concentrations of AcSpm (2 mM) and Spm (4 mM) and keeping the O_2 concentration constant at the air-saturated level (ca. 300 μM). The K_m and V_{max} values for AcSpm and Spm were determined by Lineweaver–Burk plots; fitting of the data was performed using the GraphPad Prism 4.0 software. Studies of the pH dependence of CgiPAO activity were conducted in 0.1 M Tris–HCl buffer (pH range values from 6.5 to and 9.0), and in 0.1 M sodium borate buffer (pH range values from 8.5 to 10.0) at 25 °C using saturating concentrations of both substrates. Data reported in the Results section are the average of at least three independent experiments carried out in duplicates. Enzyme activities were expressed in International Units (IU: the enzyme amount catalysing the oxidation of 1 μmol of substrate/min) per litre of culture broth. Catalytic activity of CgiPAO was negligible using Spd as substrate. When needed, protein content was estimated by the method of Bradford (1976) with bovine serum albumin as a standard. Purity was assayed by SDS-PAGE, according to the method of Laemmli (1970).

Modelling of the three-dimensional structure of CgiPAO and docking studies

The molecular model of CgiPAO was built using the pipeline implemented in the ab initio molecular modelling package I-TASSER which has been shown to yield reliable results in a number of different studies (see Xu et al. 2011 and references therein). I-TASSER first retrieves template proteins which are predicted to display a fold similar to the protein of interest, using several different threading approaches implemented in LOMETS (FUGUE, HHsearch, MUSTER, PPA, PROSPECT2, SAM-T02, SPARKS, SP3, FFAS and PRC). Subsequently, structure fragments excised from the templates are reassembled by replica-exchange Monte Carlo simulations and models are refined in an iterative procedure that optimizes the free energy and the global topology of the models (Roy et al. 2010). The best templates identified by I-TASSER were a FAD-dependent oxidoreductase from *Arthrobacter aurescens* (PDB code 3RHA; to be published) and *Zea mays* PAO (PDB code 1B5Q; Binda et al. 1999). Docking simulations of AcSpm and Spm into the active site of CgiPAO were performed using Autodock Vina (Trott and Olson 2010), allowing flexibility of the active site residues selected by comparison of CgiPAO structural model with the three-dimensional structure of the Fms1–Spm complex (PDB code 1XPQ; Huang et al. 2005). The water molecule bridging the FAD cofactor to the catalytically essential Lys residue (Polticelli et al. 2005; Tavladoraki et al. 2011), Lys315 in CgiPAO, was explicitly included in the docking simulations.

Results

Phylogeny of the invertebrate polyamine oxidase family

A BLASTP search of invertebrate PAO homologs carried out in the available databases (see "Materials and methods" section for details), retrieved 29 sequences annotated as PAO proteins from plants, baker's yeast (*Saccharomyces cerevisiae*) and invertebrate taxa, including Placozoa (*Trichoplax adhaerens*), Cnidaria, Echinodermata, Mollusca, Annelida, Arthropoda, Nematoda and Chordata (Table 1). The PAO protein sequence from the gastropod *Colubraria reticulata* was derived from the unpublished transcriptome of this species, currently under study (Oliverio and Modica, personal communication). Multiple amino acid sequence alignment of PAO proteins highlighted a fairly good degree of sequence similarity (Fig. 1). The evolutionary tree (Fig. 2) was coherent with the phylogenetic reconstruction of the invertebrate PAOs proposed by Polticelli et al. (2012). In particular, the oyster *C. gigas* PAO which belongs to a lophotrochozoan clade was nested in a molluscan lineage and shared with the mollusc PAOs an identity of 66 % (*Lottia gigantean*, ESO98426), 58 % (*Colubraria reticulata* Oliverio and Modica) and 55 % (*Aplysia californica*, 005092230), while shared with the Anellida

Table 1 Polyamine oxidase proteins sequences included in the phylogenetic analyses and their corresponding accession numbers

Organism and acronym	Accession number
Metazoa	
Acromyrmex echinatior Panamanian leafcutter ant (Aec)	[EGI59478]
Anopheles gambia malaria mosquito (Aga)	[XP_312316.3]
Apis florea little honeybee (Afl)	[XP_003691813]
Apis mellifera honeybee (Ame)	[XP_001122522]
Aplysia californica Califonia sea hare (Aca)	[XP_005092230]
Camponotus floridanus Florida carpenter ant (Cfl)	[EFN70585]
Capitella teneta marine polychaete worm (Cte)	[ELT94559]
Ciona intestinalis sea squirt (Cin)	[XP_002132119]
Crassostrea gigas Pacific oyster (Cgi)	[EKC38899]
Colubraria reticulata sea gastropod (Cre)	[Oliverio and Modica, pers. com]
Dendroctonus ponderosae mountain pine beetle (Dpo)	[478251226]
Drosophila melanogaster fruit fly (Dme)	[Q9VHN8]
Harpegnathos saltator Jerdon's jumping ant (Hsa)	[EFN82405]
Helobdella robusta freshwater leech (Hro)	[ESO00558]
Lottia gigantea owl limpet (Lgi)	[ESO98426]
Megachile rotundata alfalfa leafcutting bee (Mro)	[XP_003699937]
Nasonia vitripennis jewel wasp (Nvi)	[XP_001599761]
Nematostella vectensis starlet sea anemone (Nve)	[XP_001637328]
Pediculus humanus corporis human body louse (Phu)	[EEB13427]
Solenopsis invicta red fire ant (Sin)	[EFZ20860]
Strongylocentrotus purpuratus purple urchin (Spu)	[XP_001195328]
Tribolium castaneum red flour beetle (Tca)	[XP_971067]
Trichoplax adhaerens tablet animal-Placozoa (Tad)	[XP_002107802]
Plants	
Arabidopsis thaliana (Ath1)	[NP_196874]
Hordeum vulgare	
(Hvu1)	[Q93WM8]
(Hvu2)	[Q93WC0]
Zea mays (Zma)	[NP_001105106]
Fungi	
Saccharomyces cerevisiae baker's yeast (FMS1)	[YDL174C]

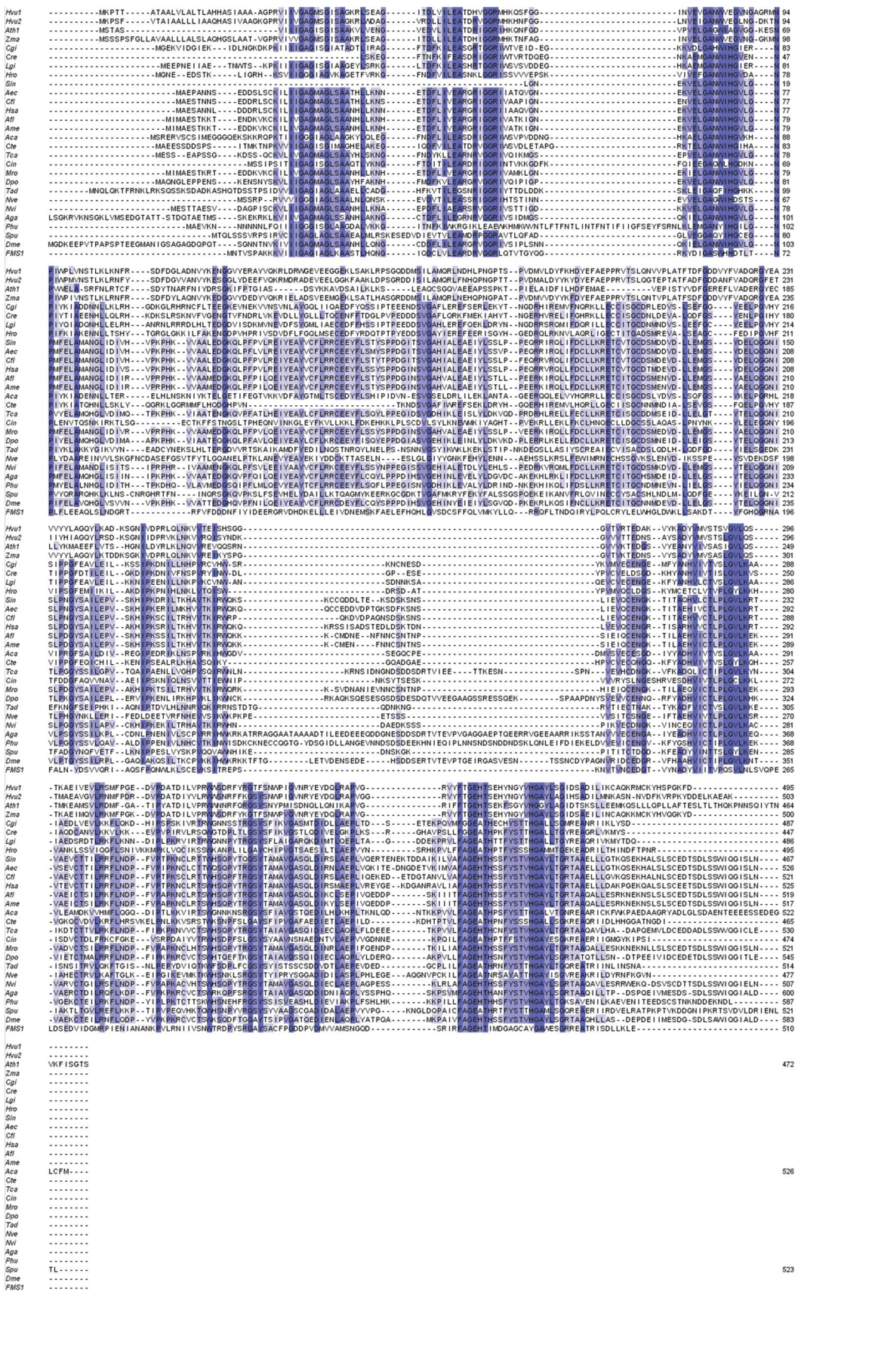
PAOs an identity of 48 % (*Capitella tenella*, ELT94559) and 43 % (*Helobdella robusta*, ESO00558).

Heterologous expression and protein purification of CgiPAO

The CgiPAO cDNA clone was obtained as described in the Methods section. The recombinant cDNA construct pCgi-PAO-HT (secreted-tagged form) was used to transform *E. coli* BL21 DE3 cells. After induction with IPTG, under the control of the T7 promoter, the catalytically active protein was expressed at a level of about 1.0 IU/litre of culture broth. The recombinant CgiPAO protein was purified (see "Materials and methods" section) and assayed by SDS/PAGE electrophoretic analysis, resulting in a single band of 55 kDa (Fig. 3).

Spectroscopic characterization of CgiPAO

The absorption spectrum of the native enzyme CgiPAO showed two peaks in the near-UV/Vis region, typical of oxidized flavoproteins, as previously described for animal SMO by Cervelli et al. (2003), with maxima at 365 and 450 nm (Fig. 4a). Being the first time, to the best of our knowledge, that a polyamine oxidase from *C. gigas* has been purified, the chemical nature of the flavin cofactor was first verified. The identity of the cofactor was determined by fluorescence analysis, according to the method described by Aliverti et al. (1999). The method exploits the fact that the fluorescence of a FMN solution is about tenfold higher than that of a FAD solution. The excitation and the emission spectra of (a) native recombinant CgiPAO, was obtained after purification; (b) supernatant


Hvu1
Hvu2
Atb1
Zma
Cgi
Cre
Lgi
Hro
Sln
Aec
Ctf
Hsa
Atf
Ame
Aca
Cte
Tca
Cin
Mro
Dpo
Tad
Nve
Nvi
Aga
Phu
Spu
Dme
FMS1

◀**Fig. 1** Multiple amino acid sequence alignment of polyamine oxidase proteins. The analysis involved 29 amino acidic sequences. Sequence blocks are coloured from *light* to *dark blue* according to increasing conservation score. Sequence-specific numbering is given on the *right end side* of the figure. The alignment was obtained using CLUSTALW2 (Larkin et al. 2007) and visualized using JalView (Waterhouse et al. 2009) (colour figure online)

was obtained after heat denaturation; and (c) supernatant after treatment with phosphodiesterase (15×10^{-3} U/mL) are shown in Fig. 4b. The very low fluorescence of the CgiPAO-bound cofactor (F_{530} ~6 a.u.) increased when the free cofactor was released after protein denaturation (F_{530} ~32 a.u.). Next, the addition of phosphodiesterase led to a strong fluorescence increase following FAD hydrolysis, to give FMN (F_{530} ~350 a.u.). The latter increase demonstrated that FAD was the CgiPAO flavin cofactor. Circular dichroism (CD) was employed to characterize the structure of CgiPAO in the near-UV/Vis and in the far-UV regions. The near-UV/Vis CD spectrum in the FAD region was analysed and compared with the spectrum of the well-known mouse SMO, for which a structural model is available (Tavladoraki et al. 2011; Cervelli et al. 2013a, b, 2014b), evidencing several similarities. In particular, the spectrum (Fig. 4c) was characterized by several valleys and peaks, which revealed underlying individual transitions (Edmonson and Tollin 1971). Two valleys at 308 and 353 nm were evident, but also two peaks at 324 and 447 nm were present. However, the peaks displayed several shoulders. By comparison, the values for valley and peaks in the mouse SMO FAD spectrum were 304 and 352 nm, and 320 and 461 nm, respectively. This very similar spectral pattern suggested that FAD molecular microenvironment in CgiPAO and in mouse SMO was similar. The far-UV CD spectrum showed the typical shape of a well-folded protein, with a broad negative band between 205 and 225 nm and a positive peak centred around 192 nm (Fig. 4d). The deconvolution of the CD spectrum gave an estimate of α-helix and β-sheet content of about 38 and 15 %, which was in a quite satisfactory agreement with the values obtained by structural modelling (28 and 13 %, respectively), see below.

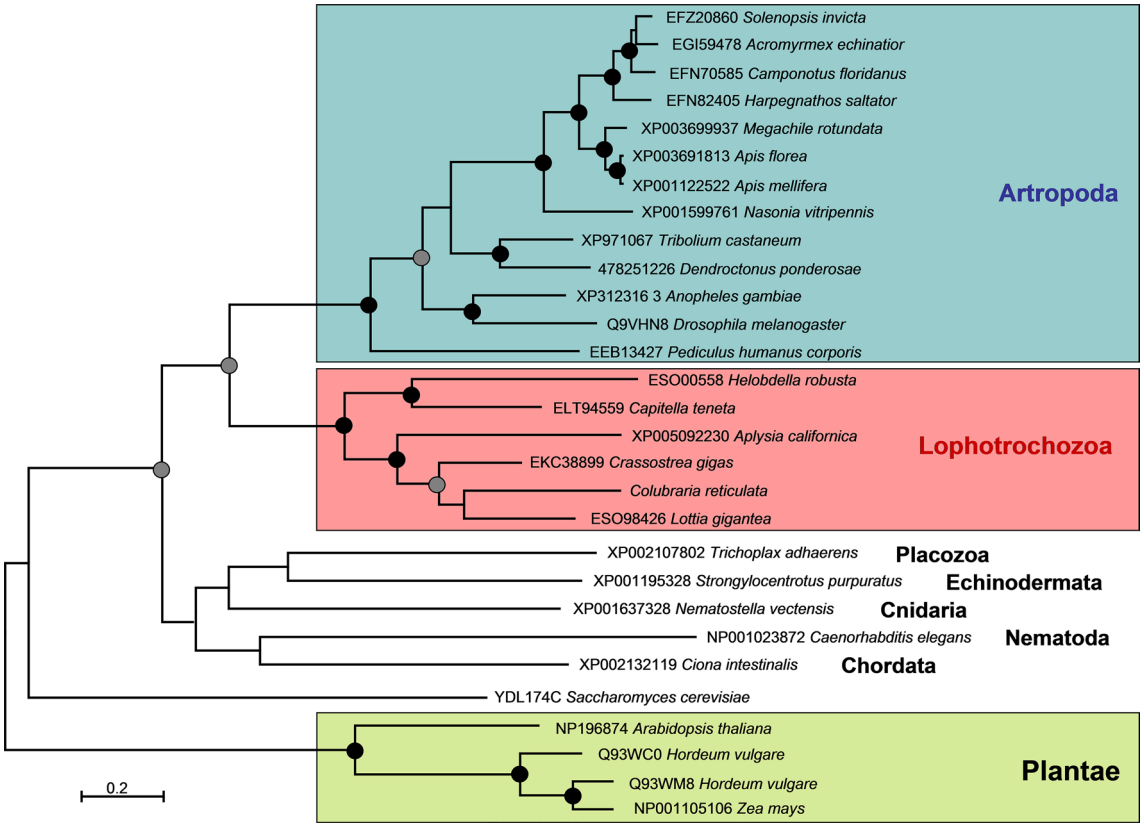


Fig. 2 The evolutionary tree of the invertebrate polyamine oxidase proteins. The evolutionary history tree of the polyamine oxidase proteins was inferred using the Maximum Likelihood method (log likelihood = −21985.9337) based on the JTT model with a proportion of invariable sites (I) and gamma-distributed rates across sites (G) (Bayesian tree showed identical topology at the main nodes; data not shown). The analysis involved 29 amino acidic sequences, and the tree was rooted with the plant sequences as outgroup. In correspondence of the nodes the bootstrap support (BS) and Bayesian posterior probabilities (BPP) are indicated by *black circles* (BS ranging from 95 to 100; BPP = 1.00) and *grey circles* (BS ranging from 90 to 94; BPP ranging from 0.95 to 0.99). The scale for branch lengths was measured by the number of substitutions per site

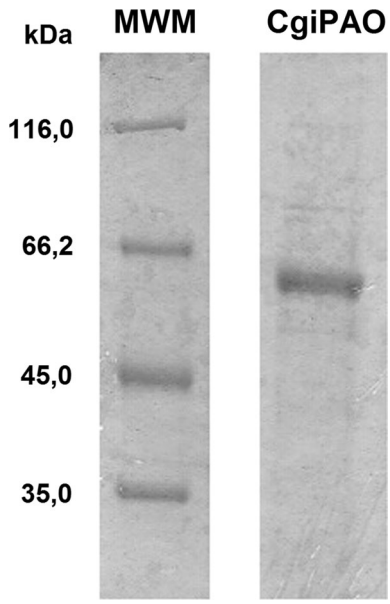


Fig. 3 Purified recombinant CgiPAO. SDS-PAGE analysis of the recombinant protein (3 μg of the purified enzyme), after Coomassie blue-brilliant staining. *MWM* molecular weight marker (Protein Molecular Weight Thermo)

Kinetic properties of CgiPAO

The purified CgiPAO exhibited a pH optimum of 8.5 in 0.1 M Tris–HCl buffer (see "Materials and methods" section). The enzyme activity at pH 7.5 and 9.5 resulted to be 40 and 45 % of the maximum for AcSpm, while for Spm resulted to be 35 and 40 % of the maximum (data not shown). The substrate specificity of CgiPAO for AcSpm and Spm was then investigated under standard conditions at pH 8.5. The purified enzyme oxidizes rapidly AcSpm and less efficiently Spm. The kinetic parameters of the tagged recombinant enzyme for AcSpm and Spm oxidation were also determined at pH 8.5. The values of K_m and k_{cat} resulted to be 33 (±1.0) μM and 19.0 (±0.6) × 10^{-3}/s, respectively, using AcSpm as a substrate, and 4.4 (±1.1) × 10^3 μM and 4.6 (±1.2) × $10^{-3}/s^{-1}$, respectively, using Spm (Table 2; Fig. S1). The catalytic efficiency k_{cat}/K_m values was 576 and 1.0 s^{-1} M^{-1} for AcSpm and Spm, respectively, indicating that AcSpm was the preferred substrate for CgiPAO, although this difference is mainly due to K_m rather than k_{cat} values (Table 2).

Structural features of CgiPAO structural model

The primary structure of CgiPAO was deduced from the cDNA sequence, and its open reading frame predicts a 487-aminoacid protein with a theoretical pI of 5.21 and a MW of 55,148 Da. A structural model of CgiPAO was built using the pipeline implemented in the ab initio molecular modelling package I-TASSER (Roy et al. 2010; Xu et al. 2011; see "Materials and methods" section for details). The best templates identified by I-TASSER were a FAD-dependent oxidoreductase from *Arthrobacter aurescens* (PDB code 3RHA) and *Zea mays* PAO (PDB code 1B5Q). The top panel of Fig. 5 shows a schematic representation of CgiPAO model compared to the three-dimensional structures of yeast PAO Fms1 (PDB code 1XPQ; Huang et al. 2005) and *Zea mays* PAO (ZmaPAO; PDB code 1B5Q; Binda et al. 1999). It must be remarked that these crystal structures are the only structures available for members of the PAO family. As it can be seen, the overall fold of CgiPAO was typical of PAOs, with a FAD-binding domain and a substrate-binding domain. Analysis of the secondary structure content of the CgiPAO molecular model revealed that, out of 487 residues, 138 (28 %) are in α-helix conformation and 61 (13 %) in β-sheet conformation, percentage values which were in good agreement with those derived from the fitting of circular dichroism spectroscopy data. The FAD molecular environment was well conserved with respect to both Fms1 and ZmPAO. However, as it can be seen in the bottom panel of Fig. 5, CgiPAO active site composition is more similar to that of Fms1 than that of ZmPAO. In particular CgiPAO displayed the presence of His78 and Tyr423, conserved in Fms1 as His67 and Tyr450 while substituted in ZmPAO as Glu62 and Phe403.

Molecular docking of AcSpm and Spm into CgiPAO active site

In Fig. 6 is shown a schematic representation of the complexes formed by CgiPAO with AcSpm and Spm obtained by docking simulations. The two substrates bind in a similar fashion within the active site. In particular, in the CgiPAO–AcSpm complex the substrate forms hydrophobic interactions with Tyr423 and establishes hydrogen bonds with the FAD N5 atom, with the carbonyl oxygen atoms of Cys189 and Ser465, and with the side chain hydroxyl group of Ser465. In the CgiPAO–Spm complex the substrate forms hydrophobic interactions with Tyr423 and establishes hydrogen bonds with the FAD-bound water molecule, with the carbonyl oxygen atoms of Cys189, Ser465 and Thr466, and with the side chain hydroxyl group of Ser465. The substrate C4 atom which is oxidized during the catalytic reaction (the carbon atom in *eso* with respect to the secondary amino group of the substrate) is located at a 3.7 and 5.9 Å distance from the redox active N5 atom of FAD in the CgiPAO–AcSpm and the CgiPAO–Spm complex, respectively. Docking simulations correctly rank the two substrates and the binding energy values, calculated using the semi-empirical AutoDock force field (Trott and Olson 2010), are −6.6 and −5.7 kcal/mol for AcSpm and Spm complex, respectively, therefore, being in agreement with the lower K_m values determined for AcSpm.

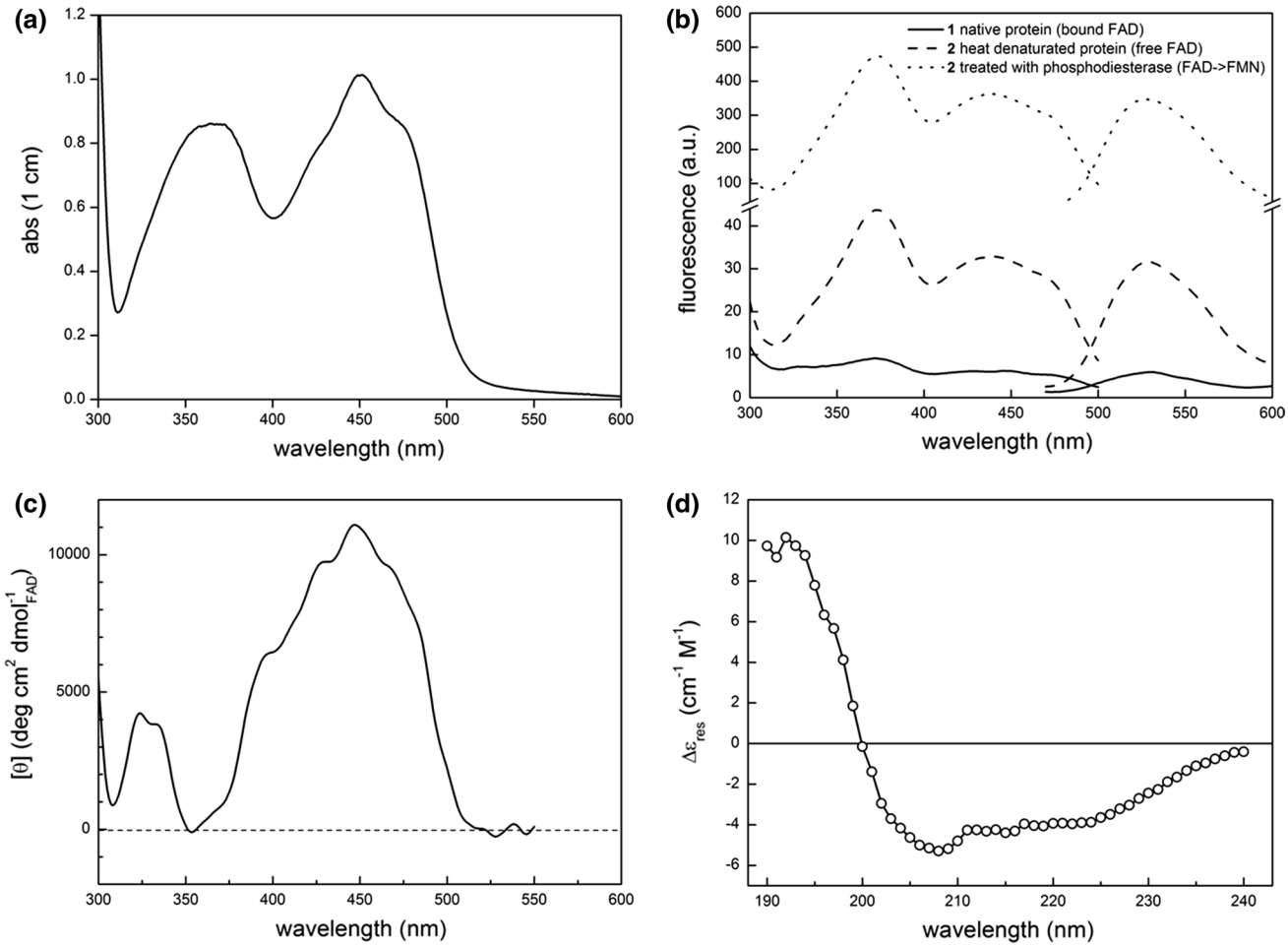


Fig. 4 Spectroscopic characterization of CgiPAO. **a** Absorption spectrum of CgiPAO (100 μM) in the near-UV/Vis region. **b** Fluorescence excitation (*curves* on the *left*) and emission spectra (*curves* on the *right*) of recombinant CgiPAO. Spectra 1 (*continuous lines*) were recorded under native conditions (whereby FAD is bound to the protein, therefore, quenched). Spectra 2 (*dashed lines*) refers to supernatant obtained after denaturation by boiling and centrifugation (released FAD, unquenched). Spectra 3 (*dotted lines*) were recorded after the treatment of supernatants with phosphodiesterase, that converts FAD in FMN, with further fluorescence enhancement (Aliverti et al. 1999). **c** Circular dichroism spectrum of CgiPAO (100 μM) in the near-UV/Vis region, expressed in FAD molar ellipticity. **d** Circular dichroism spectrum of GgiPAO (1 μM) in the far-UV region expressed as mean $\Delta\varepsilon$ per residue

Table 2 Kinetic parameters (apparent values) of N^1-acetylSpm and Spm oxidation by CgiPAO

Substrate	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m ($s^{-1}\,M^{-1}$)
N^1-acetylSpm	$19.0 \pm 0.6 \times 10^{-3}$	$3.3 \pm 0.1 \times 10^{1}$	576 ± 23
Spm	$4.6 \pm 1.2 \times 10^{-3}$	$4.4 \pm 1.1 \times 10^{3}$	1.0 ± 0.3

Discussion

Members of the polyamine oxidase gene family have been identified in a wide variety of animals, including vertebrates and invertebrates, as well as in plants and fungi. The CgiPAO displayed a high homology with the molluscan and annelida PAOs retrieved from GenBank and represents a bona fide mollusc polyamine oxidase. In fact, the phylogenetic position of CgiPAO within the lophotrochozoan clade was in agreement with the full-genome evidence proposing the monophyly of lophotrochozoans (Simakov et al. 2013). In addition, multiple sequence alignment revealed a fairly good conservation also with the other invertebrate and plant PAOs, widespread all over the sequence length. Invertebrate PAOs are proposed to be pluripotent enzymes able to oxidase both polyamines and their acetyl derivatives (Polticelli et al. 2012). In this regard, APAO and SMO enzymes are considered to be the result of the vertebrates'

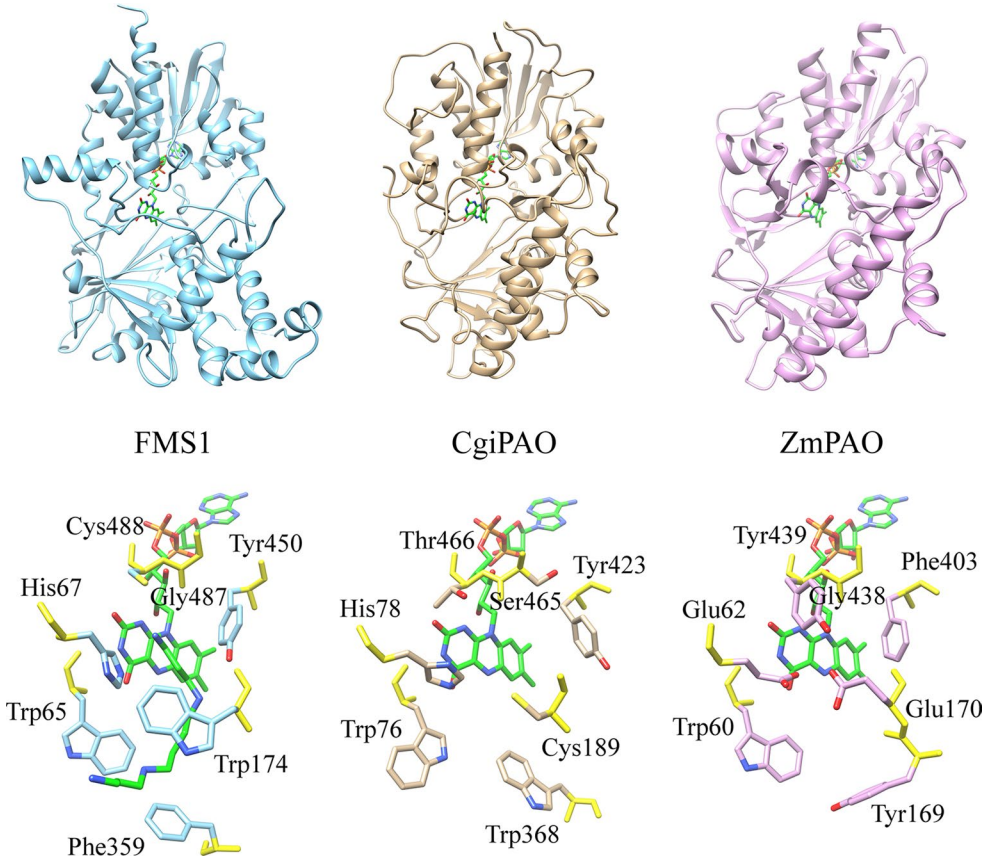


Fig. 5 Comparison of the CgiPAO structural model with Fms1 and ZmPAO three-dimensional structure. *Top* schematic representation of CgiPAO structural model as compared to the three-dimensional structures of Fms1 and ZmaPAO. *Bottom* Structural details of the active site of the three proteins. In the Fms1 active site the substrate Spm is also shown. For clarity FAD and Spm carbon atoms are coloured in *green* (colour figure online)

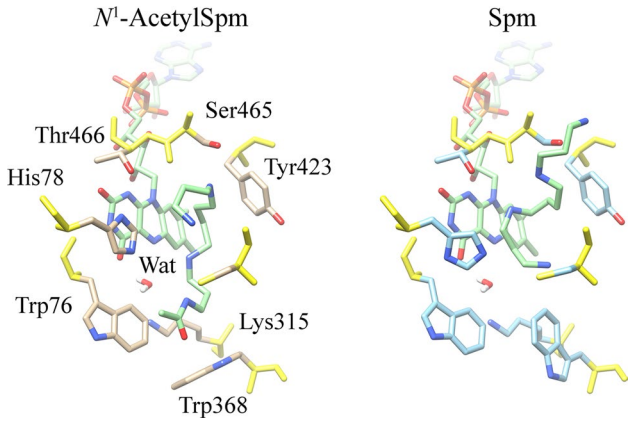


Fig. 6 CgiPAO–substrates complexes obtained by docking simulations. Structural details of the CgiPAO–AcSpm and –Spm complexes obtained by docking simulations. FAD and Spm carbon atoms are coloured in *light green*. See text for details (colour figure online)

specialization of an ancestor PAO protein, following a gene duplication event (Polticelli et al. 2012). To test this hypothesis, the Pacific oyster CgiPAO was cloned, heterologously expressed and biochemical characterized. To the best of our knowledge, no invertebrate PAO has ever been expressed in recombinant form in *E. coli* cells and this is the first report of a molluscan polyamine oxidase overexpressed as a recombinant protein. The analysis of the catalytically active CgiPAO protein reported in this work has shed light on the biochemical features of this enzyme. The purified recombinant enzyme displayed spectroscopic properties typical of PAOs as far as both FAD optical spectrum features and secondary structure content determined by circular dichroism. CgiPAO was able to oxidize both AcSpm and Spm, although the first one has shown to be a better substrate. Docking simulations were in agreement with this finding as both the value of the binding energy and the distance between the substrate C4 atom and the FAD N5 atom were lower in the case of AcSpm, suggesting that the enzyme displayed a higher binding affinity and electron transfer rate for this substrate. From a structural viewpoint, CgiPAO active site showed several features in common with the yeast PAO Fms1. In particular, the presence of Trp76, His78 and Tyr423 was conserved in Fms1 as Trp65, His67 and Tyr450. Worthwhile of note, all animal PAOs/APAOs/SMOs characterized to date, which oxidize the *eso* carbon with respect to the secondary amino group of the substrate, displayed the conservation of a His and a Tyr residue in position orthologous to CgiPAO His78 and Tyr423 (Tavladoraki et al. 2011; Polticelli et al. 2012). Extensive gene duplications have occurred early in

Table 3 Kinetic parameters of selected examples of polyamine oxidases from different organisms

Species	Enzyme	Spm			N^1-acetylSpm			R^a	Ref.
		k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m (s^{-1} M^{-1})	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m (s^{-1} M^{-1})		
C. gigas	CgiPAO	4.6×10^{-3}	4,400	1.0	1.9×10^{-2}	33	576	576	b
S. cerevisiae	FMS1	165	10	1.65×10^7	47	1.58	2.97×10^7	1.8	c
M. musculus	APAO	0.175	716	244	4.5	1.78	2.5×10^6	1.0×10^4	d
M. musculus	SMO	4.5	90	5×10^4	6.8×10^{-2}	800	85	1.7×10^{-3}	e
H. sapiens	SMO	6.6	190	3.47×10^4	0.4	492	813	2.3×10^{-2}	f

[a] R is defined as the ratio $[(k_{cat}/K_m)_{AcSpm}]/[(k_{cat}/K_m)_{Spm}]$

[b] This work; 100 mM Tris–HCl buffer (pH 8.5), 25 °C

[c] Landry and Sternglanz 2003; 62.5 mM Tris–HCl buffer (pH 9.0), room temperature

[d] Wu et al. 2003; 50 mM phosphate buffer (pH 7.6), 30 °C

[e] Cervelli et al. 2014b; 100 mM Tris–HCl buffer (pH 8.3), 25 °C

[f] Adachi et al. 2010; 200 mM Tris–HCl buffer (pH 8.3), 25 °C

chordate evolution and whether these duplications were due to complete genome duplication(s) as opposed to segmental or even sequential single gene duplications (Panopoulou et al. 2003) is still matter of debate. In vertebrates' cells, PA catabolism is mediated by the activity of two enzymes, i.e. APAO and SMO, derived by gene duplication, while in invertebrates this biochemical pathway has been proposed to be fulfilled by an ancestor single PAO enzyme (Polticelli et al. 2012). From this viewpoint, comparative analysis of the kinetic properties of Fms1, CgiPAO, APAO and SMO (Table 3), indicates that CgiPAO may represent a first step towards the evolution of specialized PAO enzyme. In fact, *S. cerevisiae* Fms1 shows similar values of k_{cat}/K_m for acetylated and non-acetylated polyamines. The ratio R between the k_{cat}/K_m values (acetylated/non-acetylated) is 1.8. On the contrary, both APAO and SMO from vertebrate (*M. musculus*) are very selective for acetylated polyamines and spermine, respectively (R being about 1.0×10^4 for APAO and 1.7×10^{-3} for SMO; and 2.3×10^{-2} in the case of human SMO). This study showed that CgiPAO was able to catalyse the oxidation of both acetylated and non-acetylated substrates, but the k_{cat}/K_m value was much higher for AcSpm than for Spm. As commented above, this was due to a large difference in K_m values (33 vs 4,400 μM), whereas the k_{cat} values are comparable. The ratio R for *C. gigas* PAO was 576, an intermediate value between 1.8 (Fms1) and 1.0×10^4 (APAO). On the basis of these considerations, it can be postulated that from an ancestral unspecific enzyme, an APAO-like enzyme gradually evolved which, upon gene duplication and further divergence, gave rise in vertebrates to the highly specific APAOs and SMOs. However, further studies on other invertebrate PAOs are needed to confirm this hypothesis. To summarize, the biochemical characterization of the Pacific oyster CgiPAO presented in this work strongly suggests that from yeast to arthropods a single enzyme can satisfy PAs oxidation metabolic need and CgiPAO may represent a typical member of the invertebrate polyamine oxidase family.

Perspectives

The CgiPAO enzyme is a new characterized member of the animal polyamine oxidase gene family and it was expressed for the first time in the *E. coli*. Circular dichroism analysis gave an estimation of the secondary structure content and modelling of the three-dimensional structure of this protein and docking studies highlighted active site features. Nowadays only the crystal structure from yeast Fms1 and maize *Zma* PAOs is available. On the contrary, no tridimensional structure of invertebrate and vertebrate PAOs is known. Despite the huge amount of data on the role played by the vertebrate SMO in PA metabolism, cancer and brain pathologies, many attempts from independent laboratories to obtain SMO protein crystals suitable for X-ray diffraction studies have been made, but were unsuccessful (Tavladoraki et al. 2011). The CgiPAO could represent a novel enzyme to challenge the PAO crystal structure, considering that the smaller size of this protein with respect to the vertebrate SMO could be very promising. The disposal of this pluripotent enzyme can have as well multiple applications in biotechnology and pharmaceutical fields, including anticancer therapy. In fact, in the last decade an increasing interest has been focused on the exploitation of the human SMO and BSAO enzyme activities in antitumoral treatment, since Spm catabolic degradation is closely related to DNA oxidation and apoptosis, mainly via H_2O_2 production (Agostinelli et al. 2007, 2009; Amendola et al. 2009, 2013, 2014; Cervelli et al. 2014a). In particular, it has been demonstrated that the release of toxic Spm metabolites (H_2O_2

and aldehydes) into melanoma, wild-type and multidrug resistant cells, after treatment with BSAO-catalysed oxidation of Spm, can induce inhibition of tumour growth (Agostinelli et al. 2009). Analogously, it might be conceived that CgiPAO, immobilized in a biocompatible polymer, could be utilized for anticancer therapeutic setting as a new and broader enzyme tool when combined with both Spm and acetyl derivatives, to produce the slow release of H_2O_2 and aldehydes at such concentration that induces tumour cell death by apoptosis rather than necrosis. In conclusion, the toxic enzymatic oxidation products generated by CgiPAO could be utilized in a combined treatment with drugs currently used in therapy and/or compounds with lysosomotropic properties, and hyperthermia associated with irradiation (Agostinelli et al. 2009; Amendola et al. 2013, 2014), to selectively generate reactive oxygen species in cancer cells.

Acknowledgments We wish to thank the University of Roma Tre for financial support and Prof. Marco Oliverio (Sapienza University of Rome, Italy) for useful discussion.

Conflict of interest The authors declare that they have no conflict of interest.

References

Adachi MS, Juarez PR, Fitzpatrick PF (2010) Mechanistic studies of human spermine oxidase: kinetic mechanism and pH effects. Biochemistry 49:386–392

Agostinelli E, Arancia G, Dalla Vedova L, Belli F, Marra M, Salvi M, Toninello A (2004) The biological functions of polyamine oxidation products by amine oxidises: perspectives of clinical applications. Amino Acids 27:347–358

Agostinelli E, Tempera G, Molinari A, Salvi M, Battaglia V, Toninello A, Arancia G (2007) The physiological role of biogenic amines redox reactions in mitochondria. New perspectives in cancer therapy. Amino Acids 33:175–187

Agostinelli E, Condello M, Tempera G, Macone A, Bozzuto G, Ohkubo S, Calcabrini A, Arancia G, Molinari A (2009) The combined treatment with chloroquine and the enzymatic oxidation products of spermine overcomes multidrug resistance of melanoma M14 ADR2 cells: a new therapeutic approach. Int J Oncol 45:1109–1122

Aliverti A, Curti B, Vanoni MA (1999) Identifying and quantitating FAD and FMN in simple and iron–sulfur-containing flavoproteins. Met Mol Biol 131:9–23

Amendola R, Cervelli M, Fratini E, Polticelli F, Sallustio DE, Mariottini P (2009) Spermine metabolism and anticancer therapy. Curr Cancer Drug Targets 9:118–130

Amendola R, Cervelli M, Fratini E, Sallustio DE, Tempera G, Ueshima T, Mariottini P, Agostinelli E (2013) Reactive oxygen species spermine metabolites generated from amine oxidases and radiation represent a therapeutic gain in cancer treatments. Int J Oncol 43:813–820

Amendola R, Cervelli M, Tempera G, Fratini E, Varesio L, Mariottini P, Agostinelli E (2014) Spermine metabolism and radiation-derived reactive oxygen species for future therapeutic implications in cancer: an additive or adaptive response. Amino Acids 46:487–498

Andrade MA, Chacón P, Merlo JJ, Morán F (1993) Evaluation of secondary structure of proteins from UV circular dichroism using an unsupervised learning neural network. Prot Eng 6:383–390

Bellelli A, Cavallo S, Nicolini L, Cervelli M, Bianchi M, Mariottini P, Zelli M, Federico R (2004) Mouse spermine oxidase: a model of the catalytic cycle and its inhibition by N, N1-bis(2,3-butadienyl)-1,4-butanediamine. Biochem Biophys Res Commun 322:1–8

Bianchi M, Amendola R, Federico R, Polticelli F, Mariottini P (2005) Two short protein domains are responsible for the nuclear localization of mouse spermine oxidase (mSMO)μ isoform. FEBS J 272:3052–3059

Bianchi M, Polticelli F, Ascenzi P, Botta M, Federico R, Mariottini P, Cona A (2006) Inhibition of polyamine and spermine oxidases by polyamine analogues. FEBS J 273:1115–1123

Binda C, Coda A, Angelini R, Federico R, Ascenzi P, Mattevi A (1999) A 30-angstrom-long U-shaped catalytic tunnel in the crystal structure of polyamine oxidase. Structure 7:265–276

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilising the principle of protein-dye binding. Anal Biochem 72:248–254

Capone C, Cervelli M, Angelucci E, Colasanti M, Macone A, Mariottini P, Persichini T (2013) A role for spermine oxidase as a mediator of reactive oxygen species production in HIV-Tat-induced neuronal toxicity. Free Radic Biol Med 63:99–107

Casero RA Jr, Marton LJ (2007) Targeting polyamine metabolism and function in cancer and other hyperproliferative diseases. Nat Rev Drug Disc 6:373–390

Cervelli M, Polticelli F, Federico F, Mariottini P (2003) Heterologous expression and characterization of mouse spermine oxidase. J Biol Chem 278:5271–5276

Cervelli M, Bellini A, Bianchi M, Marcocci L, Nocera S, Polticelli F, Federico R, Amendola R, Mariottini P (2004) Mouse spermine oxidase gene splice variants. Nuclear subcellular localization of a novel active isoform. Eur J Biochem 271:760–770

Cervelli M, Bellavia G, Fratini E, Amendola R, Polticelli F, Barba M, Federico R, Signore F, Gucciardo G, Grillo R, Woster PM, Casero RA Jr, Mariottini P (2010) Spermine oxidase (SMO) activity in breast tumor tissues and biochemical analysis of the anticancer spermine analogues BENSpm and CPENSpm. BMC Cancer 10:555

Cervelli M, Amendola R, Polticelli F, Mariottini P (2012) Spermine oxidase: ten years after. Amino Acids 42:441–450

Cervelli M, Salvi D, Polticelli F, Amendola R, Mariottini P (2013a) Structure-function relationships in the evolutionary framework of spermine oxidase. J Mol Evol 76:365–370

Cervelli M, Bellavia G, D'Amelio M, Cavallucci V, Moreno S, Berger J, Nardacci R, Marcoli M, Maura G, Piacentini M, Amendola R, Cecconi F, Mariottini P (2013b) A new transgenic mouse model for studying the neurotoxicity of spermine oxidase dosage in the response to excitotoxic injury. PLoS One 8:e64810

Cervelli M, Angelucci E, Germani F, Amendola R, Mariottini P (2014a) Inflammation, carcinogenesis and neurodegeneration studies in transgenic animal models for polyamine research. Amino Acids 46:521–530

Cervelli M, Angelucci E, Stano P, Leboffe L, Federico R, Antonini G, Mariottini P, Polticelli F (2014b) The Glu216/Ser218 pocket is a major determinant of spermine oxidase substrate specificity. Biochem J 461:453–459

Chen SH, Su SY, Lo CZ, Chen KH, Huang TJ, Kuo BH, Lin CY (2009) PALM: a paralleled and integrated framework for phylogenetic inference with automatic likelihood model selectors. PLoS One 4:e8116

Deleage G, Geourjon C (1993) An interactive graphic program for calculating the secondary structure-content of proteins from circular-dichroism spectrum. Comput Appl Biosci 9:197–199

Edmonson DE, Tollin G (1971) Circular dichroism studies of the flavin chromophore and of the relation between redox properties and flavin environment in oxidases and dehydrogenases. Biochemistry 10:113–124

Huang Q, Liu Q, Hao Q (2005) Crystal structures of Fms1 and its complex with spermine reveal substrate specificity. J Mol Biol 348:951–959

Laemmli UK (1970) Cleavage of structural proteins during assembly of the head of bacteriophage T4. Nature 277:680–685

Landry J, Sternglanz R (2003) Yeast Fms1 is a FAD-utilizing polyamine oxidase. Biochem Biophys Res Commun 303:771–776

Larkin M, Blackshields G, Brown N, Chenna R, McGettigan P, McWilliam H, Valentin F, Wallace I, Wilm A, Lopez R, Thompson J, Gibson T, Higgins D (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23:2947–2948

McIntire WS, Hartman C (1993) In: Davison VL (ed) Principle and application of quinoproteins. Marcel Dekker Inc., New York, pp 97–171

Panopoulou G, Hennig S, Groth D, Krause A, Poustka AJ, Herwig R, Vingron M, Lehrach H (2003) New evidence for genome-wide duplications at the origin of vertebrates using an amphioxus gene set and completed animal genomes. Genome Res 13:1056–1066

Polticelli F, Basran J, Faso C, Cona A, Minervini G, Angelini R, Federico R, Scrutton N, Tavladoraki P (2005) Biochemistry 44:16108–16120

Polticelli F, Salvi D, Mariottini P, Amendola R, Cervelli M (2012) Molecular evolution of the polyamine oxidase gene family in Metazoa. BMC Evol Biol 12:90

Ronquist F, Teslenko M, van der Mark P, Ayres D, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol 61:539–542

Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 5:725–738

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor

Simakov O, Marletaz F, Cho SJ, Edsinger-Gonzales E, Havlak P, Hellsten U, Kuo DH, Larsson T, Lv J, Arendt D, Savage R, Osoegawa K, de Jong P, Grimwood J, Chapman JA, Shapiro H, Aerts A, Otillar RP, Terry AY, Boore JL, Grigoriev IV, Lindberg DR, Seaver EC, Weisblat DA, Putnam NH, Rokhsar DS (2013) Insights into bilaterian evolution from three spiralian genomes. Nature 493:526–531

Sreerama N, Woody RW (2000) Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON and CDSSTR methods with an expand reference set. Anal Biochem 287:252–260

Tavladoraki P, Cervelli M, Antonangeli F, Minervini G, Stano P, Federico R, Mariottini P, Polticelli F (2011) Probing mammalian spermine oxidase enzyme–substrate complex through molecular modeling, site-directed mutagenesis and biochemical characterization. Amino Acids 840:1115–1126

Thomas T, Thomas TJ (2001) Polyamines in cell growth and cell death: molecular mechanisms and therapeutic applications. Cell Mol Life Sci 2:244–258

Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. J Comput Chem 31:455–461

Van den Munckhof RJ, Denyn M, Tigchelaar-Gutter W, Schipper RG, Verhofstad AAJ, Van Noorden CJF, Frederiks WM (1995) *In situ* substrate specificity and ultrastructural localization of polyamine oxidase activity in unfixed rat tissues. J Histochem Cytochem 43:1155–1162

Wallace HM, Fraser AV, Hughes A (2003) A perspective of polyamine metabolism. Biochem J 376:1–14

Waterhouse A, Procter J, Martin D, Clamp M, Barton G (2009) Jalview version 2–a multiple sequence alignment editor and analysis workbench. Bioinformatics 25:1189–1191

Wu T, Yankovskaya V, William S, McIntire W (2003) Cloning, sequencing, and heterologous expression of the murine peroxisomal flavoprotein, *N*1-acetylated polyamine oxidase. J Biol Chem 278:20514–20525

Xu D, Zhang J, Roy A, Zhang Y (2011) Automated protein structure modeling in CASP9 by I-TASSER pipeline combined with QUARK-based ab initio folding and FG-MD-based structure refinement. Proteins 79(S10):147–160

Zhang G, Fang X, Guo X, Li L, Luo R, Xu F, Yang P, Zhang L, Wang X, Qi H, Xiong Z, Que H, Xie Y, Holland PW, Paps J, Zhu Y, Wu F, Chen Y, Wang J, Peng C, Meng J, Yang L, Liu J, Wen B, Zhang N, Huang Z, Zhu Q, Feng Y, Mount A, Hedgecock D, Xu Z, Liu Y, Domazet-Loso T, Du Y, Sun X, Zhang S, Liu B, Cheng P, Jiang X, Li J, Fan D, Wang W, Fu W, Wang T, Wang B, Zhang J, Peng Z, Li Y, Li N, Wang J, Chen M, He Y, Tan F, Song X, Zheng Q, Huang R, Yang H, Du X, Chen L, Yang M, Gaffney PM, Wang S, Luo L, She Z, Ming Y, Huang W, Zhang S, Huang B, Zhang Y, Qu T, Ni P, Miao G, Wang J, Wang Q, Steinberg CE, Wang H, Li N, Qian L, Zhang G, Li Y, Yang H, Liu X, Wang J, Yin Y, Wang J (2012) The oyster genome reveals stress adaptation and complexity of shell formation. Nature 490:49–54

ACKNOWLEDGEMENTS

First of all, I wish to thank everyone at Computational Biology Lab in Rome, and especially Prof. Fabio Polticelli for giving me the opportunity to work at this project and for his support, both professional and moral, during these years, for his teaching and availability.

I'm also grateful to Prof. Paolo Ascenzi who involved me in such interesting projects.

I wish to acknowledge also Daniele who was so helpful with me during the last months, giving me many useful suggestion and for his 'contagious' diligence.

A sincere thanks to Silvia who shared with me the beginning of this experience, for her optimism and support: even if she moved abroad, I know that I can always count on her.

I would like to thanks the "Valentinas", my lab collegues, these years would't have been so pleasant and exciting without them. To all the people orbiting around the Lab: thank you for all the ideas, the teachings and the laughs.

Finally, some personal thanks: to my "super-cousins", Elena, Marco, Antonio, Ilaria, Luisa, Antonio and their families, for their kindness and sincere interest in my work; to Flavia and Ida, my adorable nieces, for the enthusiasm and love that only children know how to give us; to my uncles, especially who watch over me from on high, you will be always in my heart.

I wish to express my deep gratitude to Enrico, for support and love he gives me, day by day, and to his family, Michela, Emanuela and Francesco: they treated me as a daughter since I moved to Rome.

A special and loving thanks goes to my family, my mother, my dad and my sister, for always being there, providing me encouragement, moral support, love and everything I needed. Without them, I would not be able to do anything.