



# Modeling protein dynamics and protein-drug interactions with Monte Carlo based techniques

Seyed Ali Hosseini

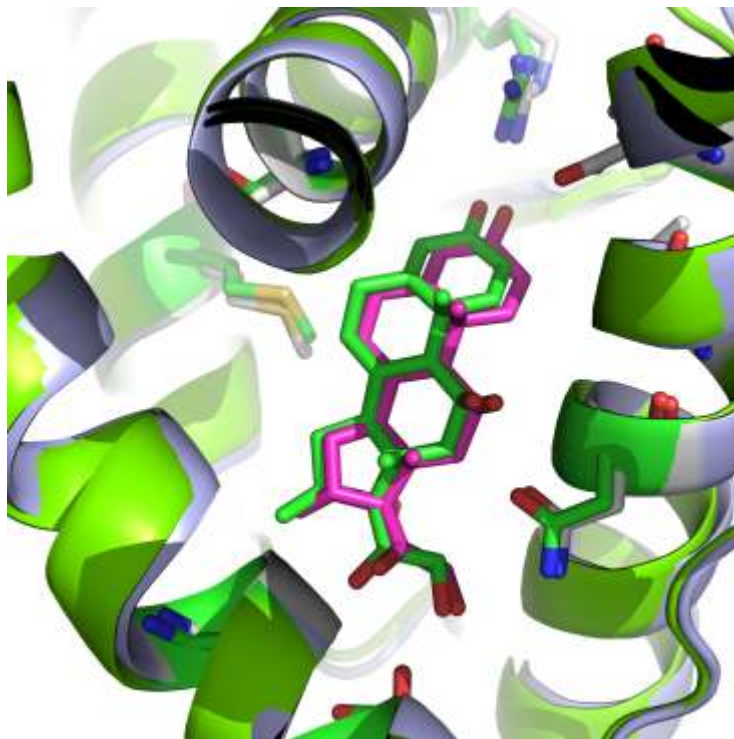
**ADVERTIMENT.** La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX ([www.tdx.cat](http://www.tdx.cat)) i a través del Dipòsit Digital de la UB ([diposit.ub.edu](http://diposit.ub.edu)) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

**ADVERTENCIA.** La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR ([www.tdx.cat](http://www.tdx.cat)) y a través del Repositorio Digital de la UB ([diposit.ub.edu](http://diposit.ub.edu)) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

**WARNING.** On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX ([www.tdx.cat](http://www.tdx.cat)) service and by the UB Digital Repository ([diposit.ub.edu](http://diposit.ub.edu)) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

**University of Barcelona  
Faculty of Pharmacy  
Barcelona Supercomputing Center**

**Modeling protein dynamics and protein-drug  
interactions with Monte Carlo based techniques**



**Seyed Ali Hosseini  
Doctoral thesis, 2015**



Universitat de Barcelona  
Facultat de Farmacia  
PhD program in Biomedicine  
Bieni 2008/2009

**Modeling protein dynamics and protein-drug  
interactions with Monte Carlo based techniques**

Thesis submitted by - Memoria presentada por

Seyed Ali Hosseini

Director: Dr. Victor Guallar

Seyed Ali Hosseini  
Barcelona Supercomputing Center

Barcelona 2015





*To my parents, my brother, my wife and my family,*

*For their love and support.*



## **Acknowledgment**

*Here is where I would acknowledge all those who helped me with my academic research.*

First of all, I would like to express my sincere gratitude to my advisor Dr. Victor Guallar for the continuous support of my Ph.D study and research, for his patience, enthusiasm and motivation. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my Ph.D study. He trained me how to analyze the results, how to project scientific questions, how to present the research and how to publish the job in good journals.

Great thanks go to the University of Barcelona for giving me this opportunity. I would like to thank Prof. Ricardo Perez, with whom I had a very successful collaboration.

A fellowship at the EAPM group, through an ERC-Adg2009 grant, as well as support by the Barcelona Supercomputing Center is greatly acknowledged.

I thank my colleagues in Barcelona Supercomputing Center, Life Science Department, especially people from our group: Fatima, Diego, Frank, Ben, Ken, Israel, Dani, Swuipa, Martin, Ryoji, Victor, Armin, Manuel, Sandra, Marina, Emanuele, James, and Marcelo for all their help and support and for all the fun we have had in the last five years. Special thanks go to Fatima who was always available for my tricky questions and also for her scientific advice and knowledge and many insightful discussions and suggestions.

Last but not the least; I would like to thank my whole family: my lovely mom: Akram, my brother: Majid, my sister: Mahnaz, my wife: Elham and her family (Abbas, Simin and Alireza) and my fabulous nephews: Farzad and Farshad for supporting me spiritually throughout my PhD.

Special thanks go to my parents (Amir and Akram) giving birth to me at the first place and support me through my life.

Very special thanks go to my wife Elham and my brother Majid, thank you for your great support and trust in me.





# Table of Contents

1	Introduction .....	1
1.1	Biomolecules: Proteins.....	3
1.2	Proteins targets for drug action .....	4
1.3	Characterization of protein-ligand interactions for drug discovery.....	7
1.4	Experimental characterization of protein-drug interactions .....	8
1.4.1	Structural characterization of proteins by experimental techniques.....	9
1.4.1.1	X-ray crystallography .....	9
1.4.1.2	Nuclear magnetic resonance (NMR).....	11
1.4.2	Measuring protein-ligand affinities .....	12
1.4.2.1	Surface Plasmon Resonance .....	12
1.4.2.2	Isothermal titration calorimetry .....	13
1.5	Disadvantage of experimental techniques .....	14
1.6	Protein dynamics and their effect on molecular interactions .	14
1.7	Computational methods to investigate protein structures and protein-ligand interactions.....	16
1.7.1	Pharmacophores .....	16
1.7.2	Homology modeling.....	17
1.7.3	Molecular docking.....	18
1.7.4	Introduction to Molecular Mechanics .....	21
1.7.5	Force filed .....	22
1.7.6	Molecular dynamics .....	24
1.7.7	Monte Carlo simulations .....	25
1.7.7.1	Metropolis Monte Carlo.....	26
1.7.8	Normal modes .....	27
1.8	Computational methods and drug discovery process .....	28
1.9	Successful application of in silico methods .....	29

1.9.1	Emerging computational techniques in drug discovery .	31
1.10	The Protein Energy Landscape Exploration (PELE) program.....	32
1.10.1	Parallel implementation and spawning .....	34
1.10.2	PELE's Application .....	35
1.11	Remaining challenges for computational chemists .....	36
2	Objectives.....	37
3	Publications .....	41
3.1	Identification of dual mTORC1 and mTORC2 inhibitors in melanoma cells: prodigiosin vs. obatoclax.....	43
3.2	Molecular Interactions of Prodiginines with the BH3 Domain of Anti-Apoptotic Bcl-2 Family Members.....	61
3.3	Atomic Picture of Ligand Migration in Toluene 4-Monooxygenase .....	75
3.4	Exploration of Protein Conformational Change with PELE and Meta-Dynamics .....	87
3.5	Computational Prediction of HIV-1 Resistance to Protease Inhibitors.....	99
3.6	Ligand recognition in steroid hormone receptors: from conserved plasticity to binding mechanism .....	123
4	Summary of the results and discussion .....	167
4.1	Validate our in-house technology PELE (Protein Energy Landscape Exploration) on sampling protein-ligand interactions and induce fit procedure. ....	169
4.2	Besides protocols and software validation, we aim to develop specific application on biomedical and biotechnological relevant systems. Thus, we aim at adding information for contributing to the mechanistic knowledge of important protein-ligand interactions. ....	172
4.3	Following the previous goal, we aim at the implementation of the atomic detailed knowledge into the rational design of new inhibitors, aiming to enhance specificity and binding strength. ....	175
4.4	An added value of (accurately) describing protein-ligand and protein/substrate interactions at a dynamical level, is being able to map possible changes in ligand/substrate affinities derived from mutations. We	

aim to develop protocols in PELE for the description of mutational effects in ligand binding. We tested this part on one of the most well studied systems with important mutational effects: HIV-1 protease..... 176

4.5 Beside these main objectives based on methods application, we aim to add methodological improvements derived from the application and validation studies..... 180

5 Conclusion..... 183

6 List of publications..... 187

7 References ..... 191



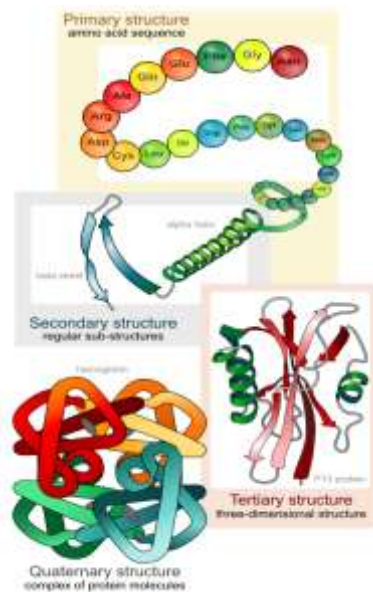
# 1 Introduction

## Introduction

## 1.1 Biomolecules: Proteins

Biomolecules are produced by living organisms and include macromolecules such as carbohydrates, (Queneau, Rauter et al. 2014) proteins (Brocchieri and Karlin 2005; Zhang 2008), lipids (Fahy, Subramaniam et al. 2009), and nucleic acids (Bloomfield, Crothers et al. 2000). Bacteria, algae, plants, and animals produce similar sets of macromolecules and these are responsible for maintaining life. Of these biomolecules, proteins have essential roles in most cellular functions. Proteins mediate these functions via direct or indirect interactions with: 1) different proteins (e.g., protein-protein interactions), 2) small molecules (e.g., protein-ligand interactions), and 3) ribonucleic acid polymers (e.g., protein-DNA/RNA interactions).

Four levels of protein structure are commonly defined (see Figure 1). Primary structure refers to all of covalent bonds linking amino acid residues in a polypeptide chain (Brocchieri and Karlin 2005). The most important element of primary structure is the sequence of amino acid residues. Secondary structure states to particularly stable arrangements of amino acid residues giving rise to repeated structural patterns (Pauling, Corey et al. 1951). Tertiary structure describes all aspects of the three-dimensional folding of a polypeptide. When a protein has two or more polypeptide subunits non-covalently bonded, their arrangement in space is referred to as quaternary structure.





**Figure 1.** Protein structure, from primary to quaternary structure. Figure from ([www.wikipedia.org](http://www.wikipedia.org)).

Predominant functions mediated by proteins include detection of signals induced by the binding of small molecules (Alberts 1989), conformational changes to act as molecular motors, enzymatic reactions that accelerate vital chemical processes, regulation of molecules whose binding to other proteins inhibits or activates their function, signalling via chemically modified peptides or proteins, and the transport of molecules through cellular membranes via selective channels (Voet and Voet 2010).

Given the importance, complexity, and range of functions that proteins mediate, databases have been established which contain protein structures; these databases contain a large number of protein-ligand interactions. Among them the Protein Data Bank (PDB) is probably the more important one, whose significance is continuously being enhanced with the increasing number of structures that register every year.

### **1.2 Proteins targets for drug action**

In general, drugs act by binding particular targets (Imming, Sinning et al. 2006). When drugs are designed to target proteins, often they modify the endogenous function of a protein upon binding (Gohlke and Klebe 2002). For example, when the target is an enzyme, drugs can act as inhibitors to block the binding site for the natural substrate. Drugs can also act in an allosteric manner and bind a protein outside of its active site. In this manner, a drug can disturb the conformation of a binding site to modify the affinity of an enzyme for a particular substrate. In other cases, drugs are agonists which bind in the natural ligand site in an enzyme, thereby inducing its activity (Berg, Tymoczko et al. 2010).

Ligand-binding sites are often only a slight part of a protein's surface. Mutations may change the binding site and affect the affinity or specificity for ligands with little effect on the whole structure change.

Protein-drugs complexes have been defined by the complementarity of shape referred as "lock and key" model (protein called as a lock and the ligand called as key) (Fischer 1894). On top of this complementarity shape, there are some specific interactions, based on the chemical nature of the ligand and receptor. Complementarity has been found to be a key factor in the formation of stable complexes, and extensive experimental and computational

## Introduction

efforts have been undertaken to predict which ligands will bind with high affinity to a particular protein (Tsai, Norel et al. 2001).

With few exceptions, drug targets have included receptors, enzymes, carrier proteins, and ion channels (Imming, Sinning et al. 2006). However, while a single drug or class of drugs often only binds a specific set of targets, their actions are not exclusive to their intended targets. In addition, the dose of a drug can affect its target recognition, thereby providing an opportunity for unintended side effects.

A ligand/drugs bind to the target as inhibitors in different ways including reversible (Segel 1987) and irreversible (Adam, Cravatt et al.) inhibitions.

Reversible inhibition involves the binding and dissociation of an inhibitor. The affinity of the inhibitor affects the on/off rate that is observed. This inhibition is divided to different group involving, competitive, uncompetitive, and non-competitive inhibitors.

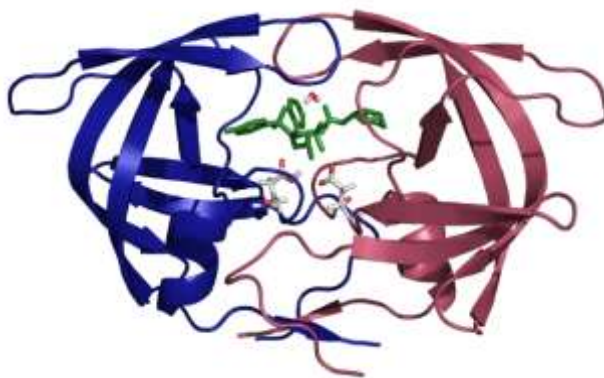
Competitive inhibitors bind active sites of targets, thereby decreasing the binding of other substrates or ligands to the target. Moreover, several inhibitors are designed to mimic an intermediate structure of an enzyme-catalyzed reaction, the binding affinity of the inhibitor may be better than that of the substrate since it provides additional stability for the enzyme to reach a transition state conformation.

Uncompetitive inhibitors bind to an enzyme-ligand complex, in a different area of the active site, in order to prevent the enzyme from reacting with its ligand and forming a product. Alternatively, noncompetitive inhibition involves the binding of a drug to a different region of an enzyme, even if the substrate is already bound. Binding of the drug affects the conformation of the target (enzyme), including the active site in some cases, and this leads to reduced binding efficiency by the substrate.

Some notable classes of reversible inhibitors include protease inhibitors (Hsu, Wang et al. 2006) and protein kinase inhibitors (Bogoyevitch, Barr et al. 2005). The former strongly bind proteases based on the similarity of their structures to that of the natural protease substrates. Moreover, the binding site often involves the active site of the targeted protease. Currently, protease inhibitors represent a very effective class of antiretroviral drugs which are used for the treatment of human immunodeficiency virus (Condra, Schleif et al.). These inhibitors compete with the natural substrate for HIV protease and

## Introduction

non-covalently bind the active site of these enzymes. For example, darunavir (Ghosh, Dawson et al. 2007) was designed on 2006 to form strong interactions with the protease enzyme from several strains of HIV, including strains with multiple resistance mutations to protein inhibitors (see Figure 2).



**Figure 2.** HIV-1 protease dimer with darunavir (green sticks) bound in the active site. A relevant water molecule and two Asp residues are also displayed in ball and stick.

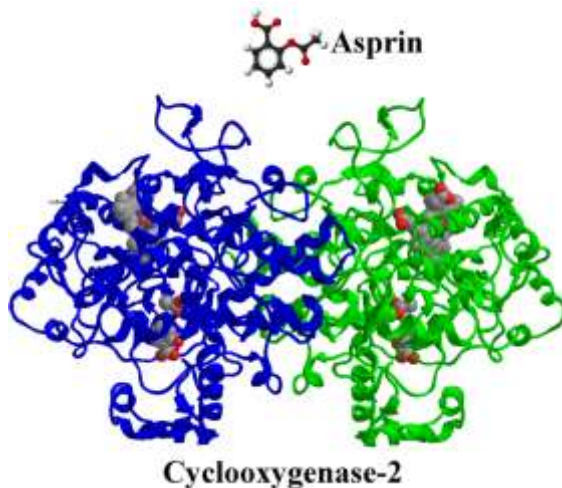
An additional important class of competitive inhibitors target protein kinases in order to compete with substrates at the binding sites where kinase interactions are mediated. However, for this type of target, these inhibitors have to compete with high concentrations of ATP that are present in each cell. The inhibitors are at a disadvantage since the intracellular concentration of a drug is typically much lower than the concentration of ATP (Bogoyevitch, Barr et al. 2005).

Alternatively, irreversible inhibitors bind a target via strong covalent interactions. The binding site may be within the active site of an enzyme, or at a site distant from an active site. As an example, aspirin (Imanishi, Morita et al. 2011) is an irreversible inhibitor that was first prepared in 1853 by the French chemist, Charles Gerhardt. However, its natural form, salicylic acid, is found in plants and has been used for thousands of years. Aspirin has been shown to covalently bond a serine residue in the active site of the cyclooxygenase enzyme.

Another consideration regarding drug targeting strategies is the size of the drug relative to the size of its target. For example, aspirin, with a molecular weight of 180.157 daltons, binds cyclooxygenase-2 (COX-2), an

## Introduction

enzyme with a molecular weight of approximately 70 kDa (see Figure 3). Thus, a 388-fold difference in size exists between this drug and its target.



**Figure 3.** Size comparison between target, COX-2, and its substrate aspirin. We should emphasize that targets can expand largely in size, from ~10 times the drug size to more hundreds of thousands.

### 1.3 Characterization of protein-ligand interactions for drug discovery

Since ancient time the study of drugs and their effects help people to know how to treat or manage disease (Nienhaus 2010). For this reason, pharmacology is considered a very important branch of science.

Pharmacology is the science of drug action on biological system, aiming to provide a deep understanding of the effects of drugs. While remarkable progress has been made in developing new drugs and in understanding how they act, challenges that remain are constant, not only in a better understanding of current diseases but in their constant evolution (new mutations) and appearance of new ones.

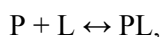
Protein-drug recognition is crucial, not only to improve our understanding of chemistry and biology, but also to advance the treatment of disease. This recognition is complex due to the different type of potential

## Introduction

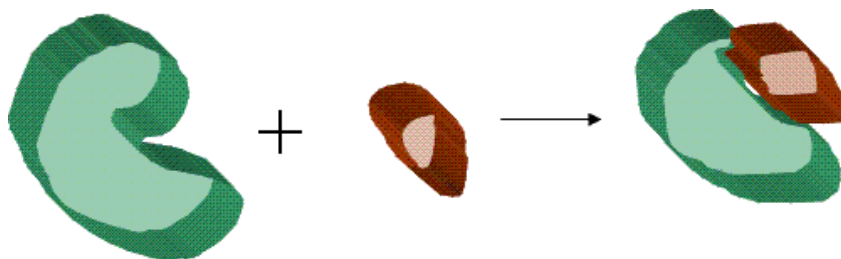
interactions between proteins and other molecules (hydrophobic, electrostatic, etc.). Moreover, it is complex due to the network of interaction between groups of atoms in biomolecules, interaction between molecules and solvent, and the scale of each system. Thus, it is not surprising that several experimental and computational methods have been established for the analysis of the molecular recognition process (Mann, Heywood et al. 1991; Przybylski and Glocker 1996; Lamb and Jorgensen 1997; Mohan, Oldfield et al. 2006). Nevertheless, each method and techniques has its own weaknesses and powers.

### 1.4 Experimental characterization of protein-drug interactions

Let us consider the simple binding reaction,



in which each protein, P, binds one ligand, L, to form a complex, PL (see Figure 4).



**Figure 4.** Diagram view of docking a small molecule to a receptor to produce a complex. Figure from ([www.wikipedia.org](http://www.wikipedia.org)).

In order to understand such process, several experimental and theoretical techniques have been developed. In this section we summarize some of the experimental techniques that are mostly used and that our experimental partners have employed in our collaborative studies. At the structural site, since the solution of angiotensin-converting enzyme and renin in the 1980s (Petrillo and Ondetti 1982), three-dimensional structures of proteins and their interactions with other proteins and substrates have become central for the field of drug discovery. Besides structure, quantifying the equilibrium populations, as well as kinetics of association and dissociation, are also key concepts in protein–ligand interactions (Nienhaus 2010).

## Introduction

The combination of these approaches has facilitated our understanding of the biological roles of proteins, as well as their effects in disease and their implications for drug design.

### **1.4.1 Structural characterization of proteins by experimental techniques**

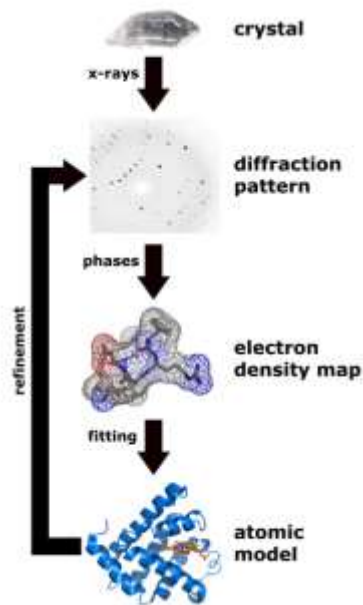
Studying proteins structures is very important because in almost all kind of disease they are targets for different treatments. Therefore methods for studying proteins keep developing both in terms of multiple techniques applied and in terms of improving approaches within specific techniques.

There are several experimental techniques to characterize three dimensional structures of proteins including, X-ray crystallography (Smyth and Martin 2000), NMR (Callaghan 1993), etc. These two methods are briefly explained in the following section.

#### **1.4.1.1 X-ray crystallography**

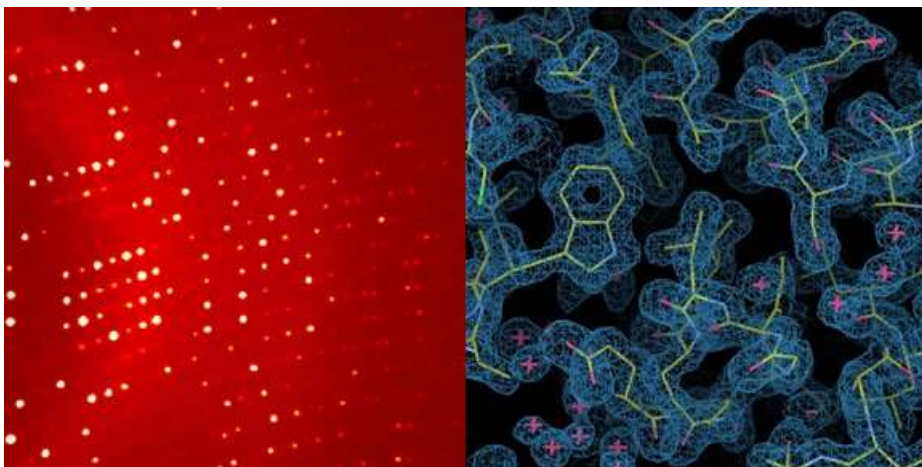
One of the main techniques used to elucidate the three-dimensional structure of large proteins at atomic resolution is x-ray crystallography. The first structure was solved in 1914 and it provided an atomic resolution image of table salt. In 1958, the structure of sperm whale myoglobin was the first x-ray of a protein reported (Kendrew, Bodo et al. 1958). Currently, x-ray crystallography is the most common method used to identify protein binding sites (see Figure 5).

## Introduction



**Figure 5.** Protocol for solving the structure of a protein (or other molecules) by X-ray crystallography. Figure from ([www.wikipedia.org](http://www.wikipedia.org)).

Figure 6 left panel shows a X-ray diffraction image from crystals from the lysozyme protein. The right panel shows the electron density map (blue) calculated from the diffracted X-ray intensities. This map is used to locate atoms and build a chemical model of the structure (yellow lines). Notice, however, that the high level of resolution seen in the right panel is not too frequent.



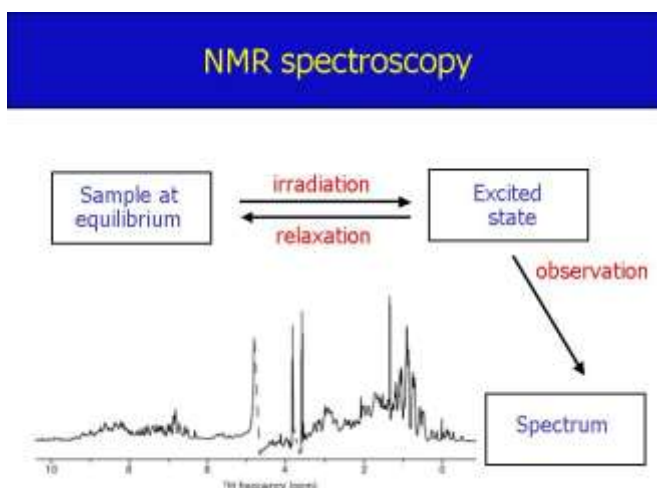
## Introduction

**Figure 6.** Left panel: X-ray diffraction image from lysozyme crystals. Right panel: electron density map and final atomic model. Figure from ([www.sandiego.edu/cas/chemistry](http://www.sandiego.edu/cas/chemistry)).

Studying protein dynamics in solution is not possible with this method (big disadvantage, for example, against NMR techniques—see below). Besides, the X-ray technique is based on the connection of the distribution of waves with electron density, and therefore, hydrogen atoms are typically not catch since they have only one electron, with no enough scattering efficiency. Thus, we need another additional method for the detection of hydrogen (e.g. neutron diffraction, NMR, modeling solutions, etc.).

### 1.4.1.2 Nuclear magnetic resonance (NMR)

Unlike x-ray crystallography, which requires a crystalline sample, NMR uses a small volume of concentrated protein solution to study the physical properties of a protein's conformation in solution. Due to the relative ease of preparing samples for NMR analysis, this approach is widely used in screening applications, as well as in studies of complex proteins (sees Figure 7). However, a disadvantage of this technique is that it is not well-suited for high molecular mass proteins. Additionally, NMR also has difficulties to define the probability of a protein (or parts of it) being in a certain conformation, requiring more complex experimental measurements.



**Figure 7.** A common protocol used to solve the structure of a protein by NMR spectroscopy. Figure from ([www.sciencearchive.org.au](http://www.sciencearchive.org.au)).



## Introduction

NMR spectra are extremely sensitive to interactions between chemical groups and can be relied upon to change when a complex is formed. Consequently, NMR is very widely used in routine screening applications as well as investigation of detailed structural features of complexes.

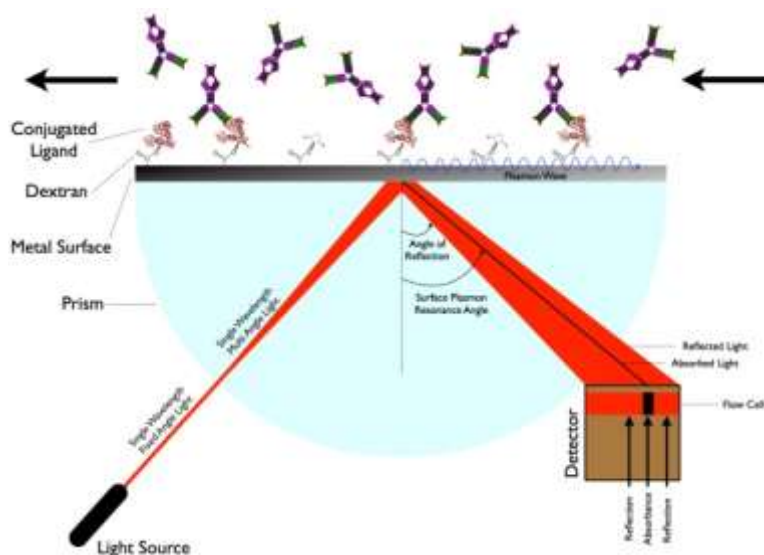
### 1.4.2 Measuring protein-ligand affinities

In addition to the structural characterization that X-ray and NMR techniques provide, kinetic and thermodynamic studies are essential for measuring affinities and for the general understanding of the various types of protein-drug interactions. The following techniques aim to address these points.

#### 1.4.2.1 Surface Plasmon Resonance

SPR (Pattnaik 2005) has been used to describe different types of molecular interactions, including protein-ligand interactions, protein binding of DNA, and protein binding of other proteins. In particular, SPR can characterize the affinity and/or selectivity of these interactions.

For SPR assays, a ligand is immobilized on the surface of a sensor chip and then is incubated with various concentrations of an analyte (see Figure 8). The resulting binding curves are used to calculate association and dissociation stages of the interactions that occur in real-time.



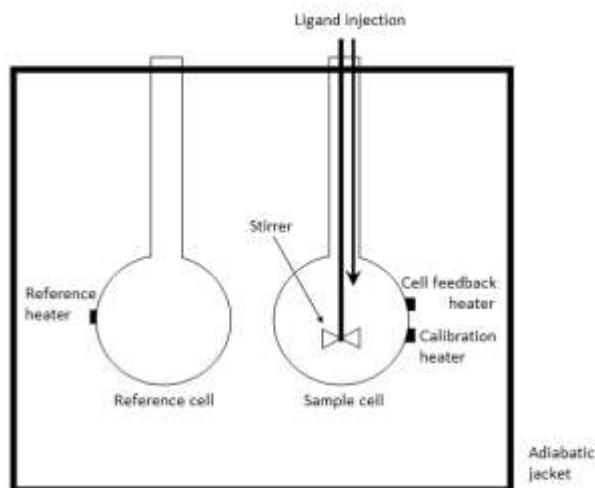
**Figure 8.** Schematic view of the SPR technique. Figure from ([www.wikipedia.org](http://www.wikipedia.org)).

## Introduction

Two important disadvantages of the SPR method are: 1) difficulties to discriminate between specific and non-specific interactions with the sensor surface, and 2) SPR is mass sensitive thus, binding of low molecular weight compounds is more difficult to detect.

### 1.4.2.2 Isothermal titration calorimetry

Physical properties that affect (or result from) the protein-ligand complex formation can be used to determine the features of a reaction including binding, thermodynamic, and kinetic characteristics. ITC (Freire, Mayorga et al. 1990; Lewis and Murphy 2005) measures the heat generated when molecules interact, and this can be used to calculate binding affinity and the thermodynamics of binding. Currently, it is the most commonly used method for studying the binding of small molecules by larger macromolecules, including proteins, DNA, etc. ITC is also used as a secondary screening technique following a first round of high throughput screening. By obtaining a thermodynamic profile for a complex, optimization of small molecule binding can also be achieved (see Figure 9).



**Figure 9.** Schematic view of an ITC instrument. Figure from ([www.wikipedia.org](http://www.wikipedia.org)).

## **1.5 Disadvantage of experimental techniques**

Now the question is whether these experimental techniques are enough to investigate the protein-ligand molecular recognition.

In many protein-ligand complexes, the complementary combination of interactions requires a high degree of dynamics and flexibility (Boehr, Nussinov et al. 2009). As mentioned here, there are several experimental methods by which one can get information about proteins. Different methods are related to different types of information. Even if those methods have had an extraordinary influence in protein-ligands/drugs characterization and drug development, there is still today no experimental technique capable of giving a robust all atom view of the dynamical interactions.

Most experimental techniques rely on data interpretation and fitting. This make difficult in many cases to obtain clear mechanistic information, requiring, in most cases extensive mutational studies. All these effort make it time and money consuming.

To this end, methods that permit the characterization of the protein mechanism (its dynamic, etc.) at atomic resolution are extremely valuable approaches to design molecules targeting protein-drugs interactions (Rosenberg and Goldblum 2006). Theoretical techniques based on all atom force field can be used for this purpose, modelling different molecular conformations that can be tried by a ligand on a protein, together with the dynamics of the processes of transition between them (Weiner, Kollman et al. 1986; Wang, Donini et al. 2001).

## **1.6 Protein dynamics and their effect on molecular interactions**

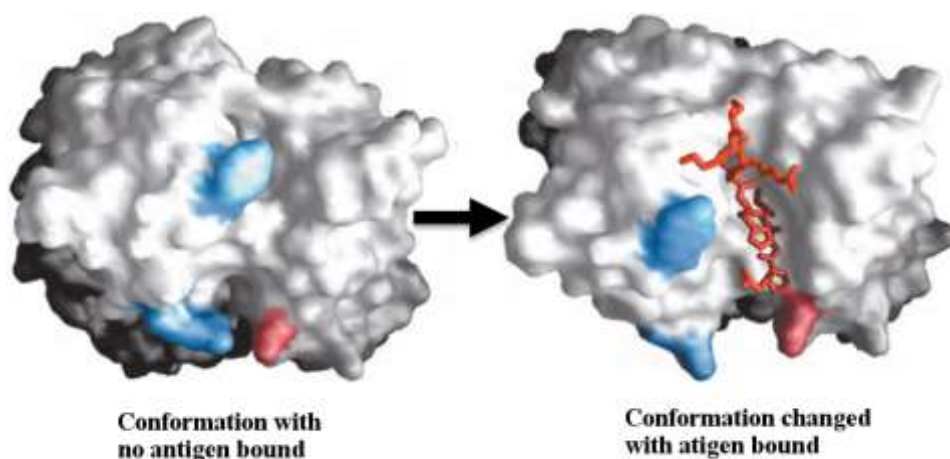
The conformational space of a protein can be described using an energy landscape. In this landscape, different conformations of a protein are populated depending on their energies and the energy barriers that exist between the different states. These conformational states can be further investigated by ligands, which might restrict conformations that a given protein is able to achieve. In some cases, certain conformational changes have been found to be essential for a protein's function (Boehr, Nussinov et al. 2009).

From such a dynamical analysis, a question raises: do the bound conformations (different from the apo one) already exist at some degree in the absence of ligands (Monod-Wyman-Changeux, conformational selection

## Introduction

theory), or do particular ligands induce the receptor to adopt new conformations (Koshland-Nemethy-Filmer induced fit theory) (Changeux and Edelstein 2011) Recent advances in experimental and theoretical techniques have allowed detailed looking into both scenarios.

In general, when two molecules bind each other, their interactions affect the energy landscape of each molecule. Over the last decade, concepts regarding protein-ligand binding mechanisms have expanded to include both a 'rigid docking' model and a 'flexible induced fit' model. Particular ligands induce the receptor to adopt their adequate conformation (induced fit). Clearly, it may be possible to design drugs to stabilize an exact conformation having the desired biological activity (see Figure 10).



**Figure 10.** Schematic views of protein dynamics that were observed following the binding of an antigen to Immunoglobulin G. Figure from (Lehninger Principles of Biochemistry, W. H. Freeman, 2005)

Protein flexibility might play an important role in the binding process during the induced fit procedure (Carlson and McCammon 2000). By using an induced fit method and conformational sampling, specific conformational changes can be identified for protein-ligand interactions. Moreover, understanding the extent of a protein's flexibility can facilitate the design of ligands that induce alternate conformations.

Computational simulation methods can help in addressing the above issues, by obtaining an all atom view of protein dynamics and its possible induced fit. In the following chapter we introduced some of them.

## **1.7 Computational methods to investigate protein structures and protein-ligand interactions**

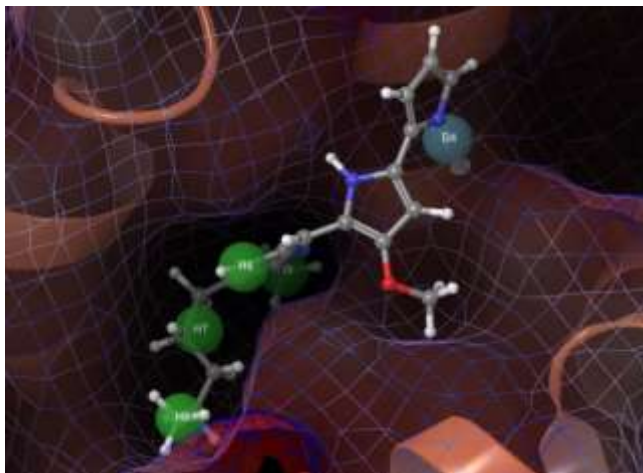
Computational chemistry has increasingly gained a more central and accepted role in the drug discovery process (Rosenberg and Goldblum 2006). Moreover, the field of computational chemistry was recognized with the award of the 2013 Nobel Prize in Chemistry to Michael Levitt, Martin Karplus, and Arieh Warshel “*for the development of multiscale models for complex chemical systems*”, which explore the world of molecules virtually using computers.

In general, computational methods in drug discovery process involve the virtual screening (Huang 2007) (pharmacophore plus docking) of chemical structure libraries containing hundreds of thousands of compounds in order to identify a few drug candidates that exhibit good binding affinity to their target.

In particular, multiple simulation techniques describing, for example, pharmacophores (Mason, Good et al. 2001), homology modelling (Schwede, Kopp et al. 2003), docking simulations (Kitchen, Decornez et al. 2004), molecular mechanics (Bowen and Allinger 2007; Lewars 2010), molecular dynamics (MD) (Berendsen 1988; Rapaport 2004), Monte Carlo simulations (Rathore and de Pablo 2002; Rubinstein and Kroese 2011), normal modes (Case 1994; Alexandrov, Lehnert et al. 2005; Bahar and Rader 2005), etc. have facilitated the use of computational methods in drug discovery. Here, we describe these selected techniques in more detail.

### **1.7.1 Pharmacophores**

A pharmacophore model (Yang 2010) is a method to explain the molecular structures and physical properties which are necessary for interaction of a ligand/drug with a specific target to activate or block its biological response. General pharmacophore types contain aromatic rings, hydrophobic centroids, hydrogen bond (acceptors or donor), cations and anions. These pharmacophoric points may be located in the receptor or found on the ligand itself. This method facilitates the design of new and more potent compounds (see Figure 11).



**Figure 11.** A pharmacophore model of the prodigiosin ligand that binds MCL-1.

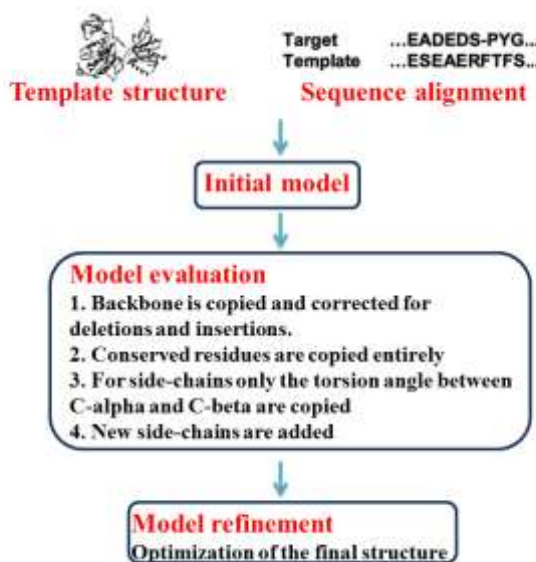
Pharmacophore technique has become a very important tool in drug discovery. A pharmacophore model can be defined either in a structure-based method, by discovering possible interaction between the protein target and ligands or in a ligand-based method, by superposing a set of molecules and searching common chemical features that are necessary for their bioactivity. These methods have successfully been applied in virtual screening, including lead optimization, de novo design and multi-target drug design. A variety of computational tools for pharmacophore modeling and applications have been developed for ligand, receptors or receptor-ligand complex. For example, GLIDE (Halgren, Murphy et al. 2004), MOE (Inc. 2013), etc methods.

### 1.7.2 Homology modeling

Homology modeling or comparative modeling, uses an experimentally defined protein structure as a template for modeling the three-dimensional structure of other proteins that have not been structurally characterized [Browne et al., 1969; Greer, 1981, 1991; Blundell et al., 1987]. However, it is important to note that proteins with similar sequences can still vary in their biological functions. Thus, homology modeling can provide the low-resolution structures that will contain (hopefully sufficient) information about the spatial arrangement of key residues in the protein and guide the design of new experiments. For example, using such model structures can considerably help the design of site-directed mutagenesis experiments.

## Introduction

In general, homology modeling involves the following steps: 1) to find a template(s) for modeling, 2) to align the target sequence with the template(s), and 3) to build the model(s). Several different procedures have been developed to build the model (see Figure 12). The simpler one is to use the backbone from the templates and correct for deletions and insertions. Conserved residues are copied entirely, for other side-chains only the torsion angle between C-alpha and C-beta is copied. Finally new side-chains are added and the structure is optimized. More sophisticated algorithms, however, also exist, using multiple template alignment and experimentally determined distance maps. Most used homology modelling software (and free servers) include: Modeller (Eswar, Webb et al. 2002), SwissPDB (Schwede, Kopp et al. 2003), I-tasser (Mitra, Shultis et al. 2013), Phyre or ROBETTA (Kim, Chivian et al. 2004).



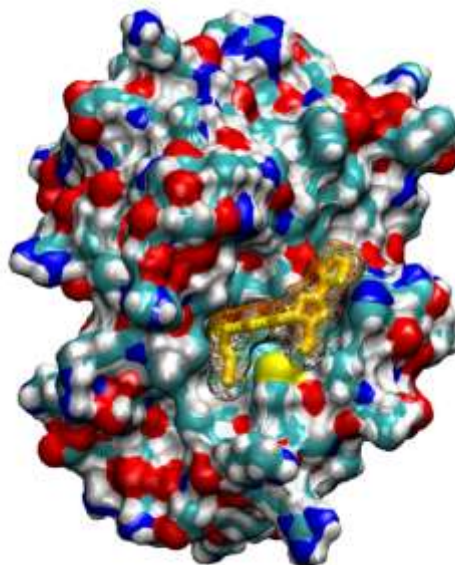
**Figure 12.** Common steps for simple in silico homology modeling.

### 1.7.3 Molecular docking

Molecular docking is a method by which the binding of a small molecule to a protein structure is predicted based on complementarity with respect to structure, charge, and hydrophobicity considerations (Brooijmans and Kuntz 2003; Kitchen, Decornez et al. 2004). To predict the affinity of various compounds, binding energies are estimated to identify the best candidates (see Figure 13). Although this method is far from being perfect,

## Introduction

often missing an accurate estimation of binding energies, is one the most common programs used in drug design and lead optimization.



**Figure 13.** Ligand docking to the active site of the MCL-1 protein.

Docking methods are mainly based on two different steps. First a search algorithm is implemented in order to identify all possible conformations where the ligand candidates interact with the receptor. Multiple search algorithms have been implemented, and almost every month we can find a new docking method or variation of existing ones. Most docking methods perform this search on a grid, previously built on the receptor. This grid describes geometrical characteristics together with biophysical descriptors: polarity, hydrophobicity, etc (Halperin, Ma et al. 2002). Search algorithms on these grids are then performed using stochastic techniques, genetic algorithms, etc. The second step includes the scoring of these configurations in order to discriminate potential binding modes. Scoring is determined using a mathematical function (scoring function) which calculates the energy of the interactions between the protein and the ligand.



## Introduction

There are different types of docking techniques. One type is rigid docking where molecules are defined as rigid items, and their form does not change with docking. Another model uses flexible docking. In this method, (one or more of the) molecules are allowed to be flexible. Most common approaches use rigid receptor but flexible ligands. However, flexibility in the receptor, at the level of side chains, is more and more implemented in docking techniques (Lengauer and Rarey 1996; Brooijmans and Kuntz 2003). However, while flexibility of molecules is critical in drug discovery, it remains a challenge for estimating the binding energies involved, making more difficult to score the different poses.

Docking methods that are commonly used by academic and industry laboratories include Glide (Halgren, Murphy et al. 2004), AutoDock (Morris, Huey et al. 2009), Gold (Jones, Willett et al. 1995), rDock (Ruiz-Carmona, Alvarez-Garcia et al. 2014) and Rosetta (Sircar, Sanni et al. 2011) programs. For example, the program Glide, which is commonly used in our studies, stands for Grid-based Ligand Docking with Energetics, and it allows a ligand to be rapidly docked with a receptor. For this process, Glide has three docking options. The fastest one, high throughput virtual screening (HTVS) uses rigid receptor and ligands and is capable of handling libraries containing millions of compounds. Sampling and scoring of a ligand involves a fraction of a second (milliseconds, etc.). The second option introduces ligand flexibility in a “standard precision” (SP method), taking approximately few seconds per ligand. The third is an extra precision (XP) method which includes extended ligand conformational sampling and improved scoring. This last option is typically applied to smaller libraries, taking up to 5 minutes per ligand. Furthermore, both SP and XP modes can minimize the clustered poses according to OPLS-AA non-bonded interaction energies before scoring them.

Providing fully flexible molecular docking is computationally expensive, especially when the receptor is treated as a fully flexible receptor model. A fully flexible receptor model can vary from thousands to millions of conformations; a complete flexible conformational analysis on both the ligand and the receptor still constitutes a big challenge, which can take a considerable amount of time (in particular if using expensive energy functions). Therefore, due to its computational expense is commonly applied only as a refinement method to few selected compounds.

#### 1.7.4 Introduction to Molecular Mechanics

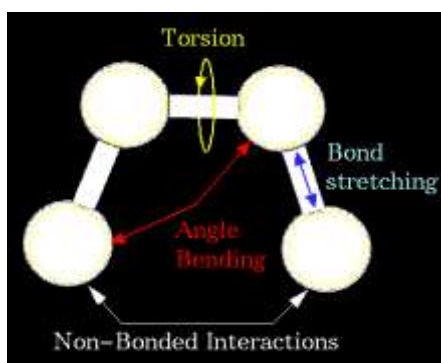
Molecular simulations are mostly done either through a deterministic performance (classical mechanics) (Bowen and Allinger 2007), a stochastic process (Monte Carlo methods) (Rubinstein and Kroese 2011) or a probabilistic mechanics (quantum mechanics) (Lewars 2010). These methods can be used for perturbing geometry from one local minimum to another. Brief descriptions of the ideas in the first two methods are given in the following sections.

Atoms within molecule are not static and present significant motion including angle bending, bond vibration and dihedral angles. As a result of this motion, and as mentioned several times above, large biomolecules have several (thousands) conformations with different energies. In order to distinguish them several different techniques have been developed to estimate molecular structure and energetics. When an atomic level of precision is desired, these have commonly been divided into two main classes, molecular mechanics models and quantum mechanics models. In molecular mechanics techniques molecules are made of spherical partial charged atoms connected together with spring bonds. Atom positions are then described by several geometrical parameters, mainly bonds and angles, which, together with their physical descriptors (charge radii) and a force field equation described uniquely its energy.

Quantum mechanics (QM), in the other hand, solves the Schrödinger equation obtaining in this way a wave function from which to derive the total energy of the system, its electronic distribution and the gradients necessary for its motion (see below). Since the exact solution of the Schrodinger equation (in systems with more than one electron) is impossible, several approximations have been developed. Rigorous approaches can take up to several days to obtain the wave function for a small system (up to dozen of atoms). More approximate methods can give a solution in few minutes for up to few hundred of atoms. Nevertheless its application to a large biological system is still out of the reach today -requiring mixed molecular mechanics and quantum mechanics methods (QM/MM) (Kollman, Massova et al. 2000; Cho, Guallar et al. 2005). In any case, and due to their expense, QM methods are mostly used for defining chemical reactions and other electronic processes, such as electronic excitation or charge transfer, and not to perform exhaustive conformational sampling.

## Introduction

Molecular mechanics describes the energy of a molecule as sum of contributions from distortions of bond distances, bond and torsion angles together with the non-bonded, van der Waals and Coloumbic, interactions (Lewars 2010)(see Figure 14). The set of parameters and equations describing this terms is known as the force field. Its computation cost is subjected by evaluation of non-bonded van der Waals and Coulombic terms, which is set by the square of the number of atoms. In comparison with QM methods, the overall cost for an energy (or its derivative) takes only a fraction of a second, making possible to model the behavior of huge molecular system with millions of atoms, such as proteins, DNA, etc.



**Figure 14.** The atomic view in molecular mechanics. Figure from (<http://c125.chem.ucla.edu/NIH/MolMechanics.htm>)

### 1.7.5 Force field

The interaction between particles (atoms) can be described in terms of either force ( $\mathbf{F}$ ) or a potential (Fahy, Subramaniam et al.). These are equivalent, as the force is the minus derivative of the potential with respect to the degrees of freedom. In a simple word, the purpose of the force field is to explain the energy and forces between the particles of a system.

The common equation of the total energy in force field can be written as:

$$E_{\text{total}} = E_{\text{bonded}} + E_{\text{nonbonded}}$$

where,

## Introduction

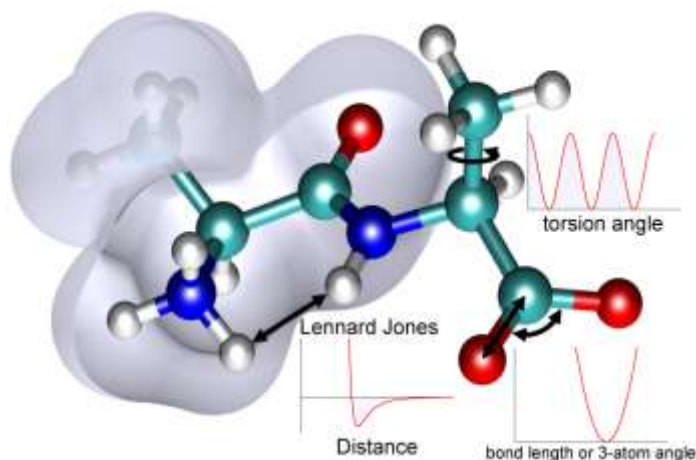
$$E_{bonded} = E_{bond} + E_{angle} + E_{dihedral} \quad \text{and} \quad E_{nonbonded} = E_{electrostatic} + E_{van\ der\ Waal} \quad (1)$$

The functional form of the widely used Assisted Model Building with Energy Refinement (AMBER) force field (Wang, Wolf et al. 2004) has the following expression,

$$E = \sum_{bonds} \frac{1}{2} k_{r_0} (r - r_0)^2 + \sum_{angles} \frac{1}{2} k_{\theta_0} (\theta - \theta_0)^2 + \sum_{dihedrals} \frac{1}{2} V_n [1 + \cos(n\phi - \gamma)] \\ + \sum_{\substack{non-bonded \\ pairs\ i,j}} \epsilon_{i,j} \left\{ \left[ \left( \frac{r_{0ij}}{r_{ij}} \right)^{12} - 2 \left( \frac{r_{0ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right\} \quad (2)$$

Where the first term describes the potential for bonded atom pairs, the second term define for bonded angles, the third the potential for bonded dihedral angles, and the final term the potential for non-bonded atom pairs made up of a Lennard-Jones type diffusive interaction and a coulombic term (see Figure 15 for illustration).

The equilibrium values of the bond distances and bond angles, as well as the equivalent force constants used in the potential energy function, are defined in the force field and are mostly resulting from averaged experimental observations (i.e. crystal structures, infra-red spectroscopy...) and quantum mechanics calculations. Each different molecule and atom type requires its own parameters. Therefore, there are force fields which are parameterized to provide an accurate description of different organic elements or may be parameterized against a particular type of molecules, such as a ligand, DNA or proteins (Cornell, Cieplak et al. 1995; Damm, Frontera et al. 1997).



**Figure 15.** Schematic view of molecular mechanics potential energy.  
Figure from ([www.wikipedia.org](http://www.wikipedia.org))

There are several molecular mechanics force fields that are commonly used, including, the above mentioned AMBER (Wang, Wolf et al. 2004), Optimized Potentials for Liquid Simulation (OPLS/OPLSAA) (Jorgensen, Maxwell et al. 1996), Chemistry at Harvard Macromolecular Mechanics (CHARMM) (Vanommeslaeghe, Hatcher et al. 2010), Gronigen Molecular Simulation (GROMOS) (Oostenbrink, Villa et al. 2004) and Merck Molecular Force Field (MMFF)(Halgren 1996).

### 1.7.6 Molecular dynamics

MD methods create a series of time-dependent points in a trajectory by spreading a set of coordinates and velocities according to the second-law equation of Newton ( $F = ma$ ). Each atom receives an initial velocity and then Newton's laws are applied to spread the dynamics of the system through time. To this aim, to integrate this equation along time, we use the basic relation  $F = -\partial V / \partial R$ , where  $V$  is the potential energy (force field, quantum mechanics, etc.) and  $R$  stands for all the degrees of freedom. Thus, the main computational bottleneck is to compute gradients at each time step for each degree of freedom (several hundred of thousands in a solvated protein system). Moreover, for stability reasons, the time step for integration needs to be sufficiently low (on the order of few femtoseconds), requiring significant computational resources to reach propagation times relevant to biological processes. The method was first published during the 1950s and 1960s by the

## Introduction

original papers of Alder and Wainwright (1957); Gibson, Goland, Milgram, and Vineyard (1960); and Rahman (1964).

Typically, in MD a given temperature will control, through a thermostat algorithm, the different velocities and give (to each degree of freedom) a certain kinetic energy to overcome energy barriers and populate the different minima; more simulation time will produce more exploration. Obviously, for this sampling, a force field is required to define the forces and energies between elements of the system. This technique assumes the ergodicity assumption: any state can be populated if sufficient propagation time is achieved (all accessible microstates have the same probability). However, and due to computer limitations in reaching sufficient long times of propagation, MD simulations often does not provide a complete (exhaustive) sampling due to time limitations (certainly this is not an exclusive MD problem but inherent to most sapling techniques). Moreover, systems often get trapped in local regions of conformational space during a simulation due to the presence of high free-energy barriers. Thus, the key problem is to provide sufficient conformational sampling within a given time. Accordingly, there have been several attempts to improve MD sampling by introducing additional approaches such as: metadynamics (Laio and Parrinello 2002), steered MD (Shen, Shen et al. 2003), replica exchange (Sugita and Okamoto 1999), umbrella sampling (Kumar, Rosenberg et al. 1992), etc. All these techniques aim to speed up the sampling by reducing the conformational space or speeding transition between states.

### **1.7.7 Monte Carlo simulations**

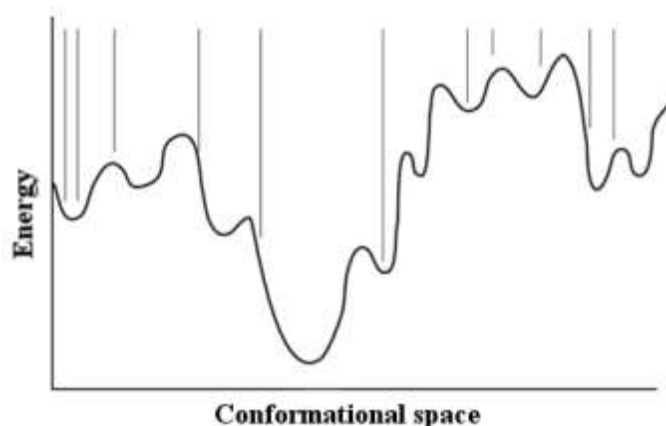
Monte Carlo method (Rathore and de Pablo 2002) is one of the most useful computational algorithms which is based on random sampling to achieve statistical results. Monte Carlo techniques developed in the last half century, and its fundamental idea is to solve the statistic problems repeating random sampling to characterize the features of the particles of a system. Using this method requires to run the simulations several times to obtain the best probability distribution of the object. This method is particularly useful for simulating systems with a large number of degrees of freedom.

As compared to MD simulation, Monte Carlo methods needs less computer time to perform each sampling step for the same system, since typically they do not need to compute forces (as MD does). Nevertheless, the random component of the step involves less probability of sampling important regions (unless an efficient sampling technique is used). Lack of time-

## Introduction

dependent information is also one of the most important limitations in Monte Carlo simulation.

Choosing the step size (and its nature: number of atoms to be modified, simultaneity) in Monte Carlo simulations is one of the main difficulties in its practical implementation, affecting the search towards local or global exploration. Therefore, algorithm may explore some regions more exhaustively than others, as shown in Figure 16, which might require system specific tuning the parameters.



**Figure 16.** Monte Carlo exploration for sampling of conformation space.

### 1.7.7.1 Metropolis Monte Carlo

The Monte Carlo sampling is almost always used in combination with an importance sampling technique. This is true, in particular, for large –lots of degrees of freedom- systems where, otherwise, the system will have serious difficulties to sample low energy areas. Possibly the most importance sampling technique used in biological systems is the metropolis sampling algorithm.

A metropolis Monte Carlo procedure(Binder and Heermann 2010) starts by calculating the energy of the initial structure. The technique then applies a slight change to the molecular system and calculates the new energy. If the movement decreases the energy, then a new structure is automatically taken. If the final energy of the system increases, then the new position of the structure is accepted with probability  $e^{-\Delta E/KT}$ , where  $\Delta E$  is the change in energy, K is the Boltzmann constant in proper units, and T is the temperature

chosen for the calculation. In this way, the outcome of accepted structures will follow the desired Boltzmann distribution.

### 1.7.8 Normal modes

Protein dynamics is in constant evolution, changing from one state to another state. Nevertheless, this motion is often confined within local minima, experiencing small fluctuations. Normal mode analysis (NMA) aims to describe the motion in proteins by approximating a harmonic nature to those minima and obtaining its second derivatives vectors and frequencies. While this, in principle, should apply only to the vicinity of the minima, normal modes from atomic force fields, or from other more approximate methods (see below) have shown to describe more complex conformational transitions in proteins (for example the T to R transition in hemoglobine (Xu, Tobi et al. 2003; Eyal, Yang et al. 2006))

NMA is a method for the analysis of collective motions in biomolecules including proteins. Minimization of the conformational potential energy, calculation of the “Hessian” matrix, which is the second derivatives of the potential energy with respect to the mass-weighted atomic coordinates and the diagonalization of the Hessian matrix are three main steps of NMA. Each of these three steps can be computationally difficult, depending on the size and force field of the molecule.

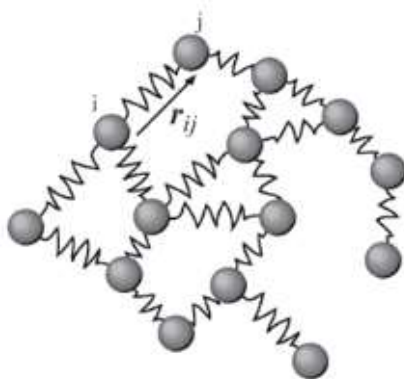
In general, simpler force field representations will lead to faster (but more approximate) normal modes. In these regards, one of the most popular NMA simplifications involves the development of elastic network models (GNM, Gaussian network model). In this algorithm the protein “force field” is significantly simplified: atoms are solely connected by a network of elastic springs. Moreover in most cases this model is applied only to carbon alpha atoms in proteins. There are two main advantages in this method; 1) energy minimization is not needed because the distances of all of the elastic connections are taken to be at their minimum energy length. 2) The diagonalization task is reduced compared with the NMA method because the number of force field terms and atoms is reduced (if using only alpha carbon, for example, from the total number of atoms to the number of residues).

The ANM (anisotropic network model) is possibly the most used version of GNM, accounting for the three coordinates per atom, thus adding directionality. This tool was developed in 2000 (Doruker, Atilgan et al. 2000; Atilgan, Durell et al. 2001) for the normal mode analysis of proteins and



## Introduction

analysis the motions of molecular systems. As mentioned, proteins are represented here as an elastic mass-and-spring network of alpha carbons (see Figure 17).



**Figure 17.** Schematic view of an elastic network. Figure from (Chennubhotla et al., 2005).

From this simplified potential, diagonalization of the second derivatives matrix is readily obtained (few seconds in a commodity machine), being able to use the resulting (approximate) modes in sampling algorithms. While only using the  $C\alpha$ -atoms as nodes might seem an oversimplified approach, Bahar et al. (Eyal, Yang et al. 2006) have shown in their seminal work that motion inferred from this vectors reproduce to a large degree fluctuation observed in crystals (beta factors). Lately, more sophisticated ANM have been developed using internal coordinates (Orellana, Rueda et al. 2010) or all heavy atoms.

### 1.8 Computational methods and drug discovery process

Computer simulation models have provided an additional opportunity to understand atomic detail of receptor-ligand interactions. Within the last twenty years, several *in silico* methods have been developed and applied to the study of biological systems in order to better understand the atomic details of molecular interactions between receptors and drugs. These methods, for example, 1) help in designing new molecules from a rational protein-ligand atomic interaction, 2) shorten the time for screening thousands of compounds, and 3) reduce cost by reducing the amount of reagents used (synthesized) for studies. In particular, structure-based and ligand-based drug design using computational methods and informatics knowledge has made the drug discovery process more efficient.

## Introduction

Correspondingly, remarkable progress has been made in drug design and discovery in many pharmaceutical companies over the past five years, reducing the risk of failure in some cases. This is a key point since companies invest a significant amount of time and money to bring a new drug to market. In silico approaches and in vitro data are used in almost all pharmaceutical companies for drug discovery and development, and they are complemented by pharmacology studies of toxicity and absorption, distribution, metabolism, and excretion (ADME) (Kapetanovic 2008).

In order to rationally design new or improved inhibitors, an understanding of the inhibition mechanisms at an atomic level is very valuable. As we mentioned previously, atomic detailed intermolecular interactions are difficult to obtain from experimental approaches alone. Therefore, computational methods provide an additional source of data.

Rational optimization of a lead compound may be achieved with considerations for chemical modifications that increase the affinity and/or specificity of an inhibitor for a target receptor. As mentioned, expanded computational simulations techniques, based on all atom force field simulations are commonly used to provide mechanistic details on several selected poses.

Moreover, computational techniques describing the protein-ligand interaction mechanism at great detail open the door to study the effects on drugs delivery and binding upon mutations, a very important field in drug discovery. Due to its importance in drug resistance (for example as a consequence of viral evolution or high metabolic rate of cancer cells), or in personalized medicine, where specific patient mutations might have a large influence on which medication to use, it is critical to develop accurate protocols for predicting the effects upon protein mutation.

### **1.9 Successful application of in silico methods**

There are several examples which demonstrate the successful design of inhibitors using in silico methods. These include a kinase inhibitor of the type I transforming growth factor beta (TGH  $\beta$ ) receptor which was developed independently by the companies, Biogen (Singh, Chuaqui et al. 2003) and traditional enzyme and cell-based high-throughput screening by Eli Lilly (Sawyer, Anderson et al. 2003), using virtual screening. Another example is the anti-anxiety, anti-depression 5-HT<sub>1A</sub> agonist (Becker, Dhanoa et al. 2006). Using in silico modeling, this potent and selective compound was designed in

## Introduction

less than two years, and required less than six months for lead optimization and synthesis. Additional examples of new drugs (hits and leads) that have been developed or designed using a combination of screening methods include a dopamine D3 receptor agonist (Varady, Wu et al. 2003), various antibiotics (Olsen, Jost et al. 2006), c-Src/Abl kinase inhibitors (Manetti, Locatelli et al. 2006), a checkpoint-1 kinase inhibitor (Lyne, Kenny et al. 2004), a PPAR $\gamma$  ligand (Lu, Huang et al. 2006), and a MDM2-p53 inhibitor (Lu, Nikolovska-Coleska et al. 2006).

Additional successful examples of target-based virtual screening are as followed; novel, potent and selective CK2 (casein kinase II) inhibitors were obtained by screening a subset of the Novartis database using DOCK (Vangrevelinghe, Zimmermann et al. 2003). The ATP binding site of a human CK2a was inferred via homology modeling (X-ray structure of *Zea mays* CK2a, PDB entry 1DAW). Moreover, Novel BCR-ABL tyrosine kinase inhibitors were identified with a related TBVS workflow using DOCK (Peng, Huang et al. 2003).

There are also several successful examples of ligand-based virtual screening; when using mibefradil, a known T-type calcium channel blocker (IC<sub>50</sub>  $\frac{1}{4}$  1.7 mM) as a query (Schneider, Neidhart et al. 1999), CATS (chemically advanced template search) identified one significant hit with an IC<sub>50</sub> <1 mM, among the top 12 ranked molecules. The same technique, CATS, was applied to identify structurally novel glycogen synthase kinase-3 inhibitors, first by identifying the oxadiazol-pyridyl moiety, a new chemotype, then by synthesizing additional analogs. Compounds with inhibitory activity below 1 mM were identified.

Nevertheless, virtual screening and docking methods have limitation accuracy. For example, calculating protein-ligand binding affinities (scoring functions) has been challenge by the pharmaceutical industry. It has been considered that docking methods presently dock 70 – 80% of ligands to the targets correctly. Recently, one study proposed that current docking and scoring algorithms are not able to identify key interactions and treat them appropriately, false positives, an important topic in structure-based virtual screening, for example (Kapetanovic 2008). Despite these limitations, computational techniques are expected to play a major role in future pharmacology, even more when the development of new drugs involves significant more research studies.

### 1.9.1 Emerging computational techniques in drug discovery

Several computational studies, published in the last recent years, have introduced promising areas for a more accurate description of the protein-ligand interaction.

As mentioned previously QM approaches are more accurate in describing energy functions, capable of better characterizing protein-ligand interaction. As mentioned, this method has size and time limitations, not being possible its application for a large set virtual screening test. Nevertheless, it is suitable for refining and better scoring few poses. In this line, we find several studies in the recent years demonstrating that a better description of the electronic effects of the ligand (and its surrounding active site) helps in discriminating binding poses and in obtaining improved correlation with experimental binding energies (Cho, Guallar et al. 2005).

Improving conformational sampling remains one of the main challenges in biological simulation at an atomic detailed level. Recently, several hardware and software developments, for example from the Shaw research group or from laboratories developing MD simulations on graphical processing units (GPU), have shown that accessing microsecond time MD simulations is now possible (Young, Bank et al. 2009). Resulting information have demonstrated the power and ability of these methods and their successful application to drug discovery and molecular recognition studies. For instance, scientists at Shaw research group performed a completely blind MD simulation of a drug diffusion, active site search and binding, by placing it outside of a kinase protein (Dror, Pan et al. 2011). After running the simulation for a sufficient period of time to sample all of the predicted configurations, the drug was eventually oriented into its binding site in excellent agreement with the crystal structure. Such simulations suggest that computational methods may further be applied in a near future to calculations of protein-drug kinetic binding data, with comparisons made with experimentally determined on and off binding rates.

These computational approaches, however, still represent a significant computational cost (out of the reach of a typical lab) when dealing with complex (partly buried active sites) systems. In the following sections, we will introduce our Monte Carlo approach PELE (Protein Energy Landscape Exploration) (Borrelli, Vitalis et al. 2005; Borrelli, Cossins et al. 2010), which allows for an unbiased search of the protein – ligand dynamics at an accessible computational cost.

## 1.10 The Protein Energy Landscape Exploration (PELE) program

PELE is a hybrid method combining Monte Carlo sampling with protein structure prediction techniques capable of producing rapid and accurate protein and the protein-ligand conformational landscape (Borrelli, Vitalis et al. 2005). PELE was initially developed to provide quick protein-ligand interactions including, biased ligand entrance and exit pathways, induced fit docking, and overall protein dynamics, with less computational cost of molecular dynamics techniques (Borrelli, Cossins et al. 2010).

PELE technique is based on two main steps:

1) **PERTURBATION**. The procedure begins with calculating the energy for the initial structure followed by a perturbation in the system. This step can include a ligand (if present) and the protein backbone perturbation.

Ligand perturbation. If ligands exist, the step begins with the generation of a local perturbation on the ligand. The ligand is perturbed through random rotations and translations. Moreover, ligand internal degrees of freedom are taken into account by building a ligand specific rotamer library. The program can treat several ligands. Several filters are applied to prevent any steric clashes between backbone of the protein and the ligand.

Protein perturbation. The perturbation includes also the backbone of protein (or the backbone surrounding the ligand). To this aim, all atoms are minimized where the alpha carbons are forced to move to a new position by means of a harmonic constraint resulting from a (randomly chosen) low frequency mode from an ANM approach (see above normal modes section). Such a procedure aims to describe the global motion of the protein.

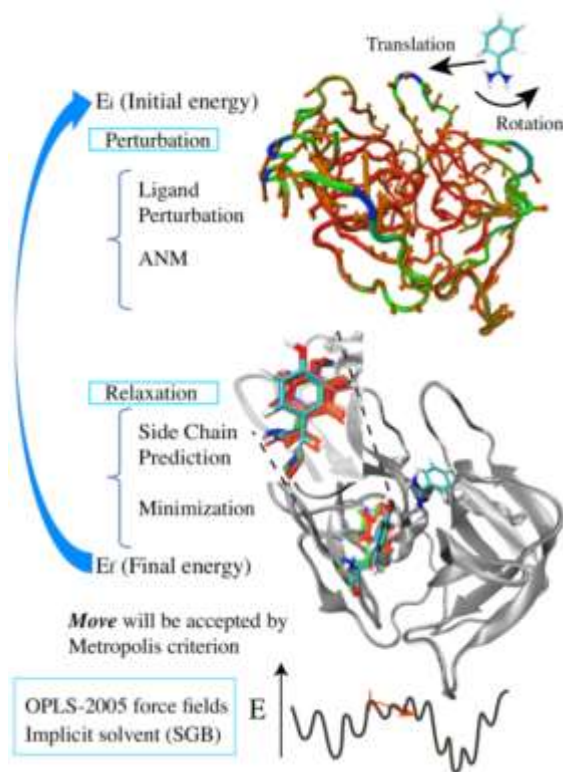
2) **RELAXATION**. Relaxation is based on two protein structure sampling techniques: side chain prediction and minimization.

Side chain sampling. Due to the ligand and protein motions, as a result of the perturbation step, the side-chain step has been developed to reposition residues which underwent a large energy change along the perturbation. For this, we typically select several residues around (within a given radii) the perturbed ligand. In addition, each side chain energy is computed before and after the perturbation. Thus, the user can choose side chains with the largest energy increase (top residues) and predict a better side chain position. The algorithm proceeds by optimally arranging the selected side chains with a rotamer library side-chain optimization at a rotamer resolution of 10° to 30°.

## Introduction

Minimization. The last procedure is the minimization step, which includes (at least) all residues local to the atoms involved in steps 1 and 2. Nevertheless, the user can select any desired region to be included. The truncated Newton minimization algorithm uses a multiscale protocol and it has the option of including a harmonic constraint in those alpha carbons that were modified in the initial ANM perturbation.

Figure 18 shows the heuristic process for the landscape exploration method.



**Figure 18.** Two main steps of PELE program. Figure from ([www.pele.bsc.es/pele.wt](http://www.pele.bsc.es/pele.wt)).

After these steps, the system adopts a new conformation that is accepted or rejected based on a Metropolis criterion. The energy is described with an all-atom OPLS force field with a surface generalized born solvent model (a recent update in PELE allows now to use also the AMBER force field). All the accepted steps will then generate a stochastic trajectory. The combination of protein and ligand perturbations explores efficiently the energy landscape, reproducing large conformational changes along ligand

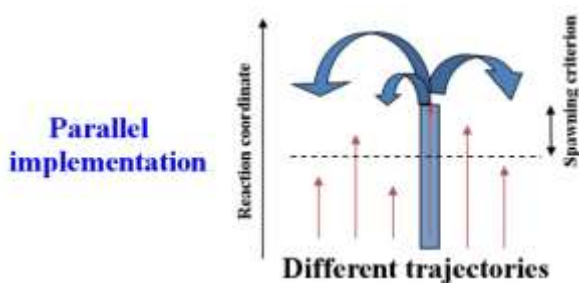
## Introduction

migration. These perturbations, associated to random processes, constitute the major part of the sampling procedure. As mentioned previously, they are maintained along the relaxation (minimization) step, enforcing the stochastic nature of the sampling and the maintenance of (an approximation to) detailed balance.

### 1.10.1 Parallel implementation and spawning

Optimally spawned functions are placed to speed up sampling and establish a collective search between different stochastic trajectories running on different processor. The procedure is capable of interchanging the coordinates between several trajectories by using the MPI communication protocol (see Figure 19). If one of the trajectories is significantly further along to the assumed reaction coordinate, then the trajectory is neglected and picked up from the position of the leading trajectory.

Spawning criteria provide an affecting sampling of the configurational space headed for a particular objective: ligand entrance/exit, protein-ligand binding energy optimization, and so forth.



**Figure 19.** Schematic view of spawning criteria in PELE.

### 1.10.2 PELE's Application

The algorithm and its application to ligand diffusion, protein-ligand interaction and induced fit docking are presented in several papers. Some of these studies described here.

The initial applications revealed the potential of this new technique in mapping microsecond-time-scale processes in a highly efficient way. Ligand exit pathways are successfully modeled for different systems containing ligands of various sizes: carbon monoxide in myoglobin, camphor in cytochrome P450cam and palmitic acid in the intestinal fatty-acid-binding protein (Borrelli et al, 2005); escape pathways are consistent with experimental and theoretical data.

From this early paper it was clear that PELE needed a better backbone and sampling procedure. Therefore, we introduced a new protein perturbation step based on anisotropic network model methodologies, capable of providing significant backbone motion. These PELE developments have been tested on two systems: ubiquitin and T4 lysozyme (T4lyz) which are described in detail in the result section (Cossins, Hosseini et al. 2012).

PELE has recently shown to provide more accurate induced fit results than the state of the art in commercial software. Various protein structure prediction methods applied as part of an induced fit procedure to predict protein-ligand complexes. The best results were obtained with an ANM driven minimized Monte Carlo scheme (PELE) to open the active site followed by a hierarchical rotamer library based refinement to fine tune the details (Borrelli et al, 2010).

In several studies, the method has shown its potential to describe diatomic ligand migration in several globin systems: myoglobin, hemoglobin and the mini-hemoglobin from the sea worm *Cerebratulus lacteus* (Lucas and Guallar 2013). The results clearly show that the simulations are specific to the system providing a different trend in the entrance pathway, as expected from experiments. While Mb presents multiple entrance pathways, populating the well-known xenon cavities, in CerHb the ligand enters the protein only by one apolar channel. In haemoglobin, a clear different trend is observed for the T and R quaternary structures and, furthermore, for the individual different chains. Such detailed information, accessible through the state of the art algorithms in PELE, is computationally inexpensive and available to all non-profit researchers through the BSC web site (<https://pele.bsc.es>)



## Introduction

Gaining absolute binding free energies from unbiased ligand diffusion is an important goal due to its implications in drug discovery. As mentioned, several studies have shown the capability to achieve microsecond molecular dynamics which, combined with a Markov state model analysis, can provide absolute binding free energies. In a recent study, PELE has shown to provide an analogous study through Monte Carlo simulations, instead of MD. The results are in good agreement with experimental data and other molecular dynamics simulations, indicating that PELE can be a useful technique for fast estimates of binding free energies and mechanisms (Takahashi, Gil et al. 2013).

### **1.11 Remaining challenges for computational chemists**

Although computational methods have still not solved most of the problems faced by a medicinal chemist, they have provided significant benefits for the drug discovery process. They also have great potential for further advancing drug discovery and development processes. However, additional work is needed to develop computational methods in order to (more) accurately predict the affinity of a ligand for a protein.

Many aspects are currently open fronts in today's research, for example: 1) development of improved force fields (with better description of dispersion forces, approximate polarization, torsion...); 2) explicit inclusion of water molecules in docking; 3) better techniques to compute entropy changes; 4) improving flexible sampling techniques, etc.

This last point, obtaining better and faster conformational search in protein and protein-ligand dynamics, is the main goal of this thesis. By using and further developing PELE, together with other simulations techniques, we have studied several systems and provided new possibilities when mapping the dynamics of complex protein-ligand systems.

Objectives

## **2 Objectives**

## Objectives

## Objectives

Understanding protein-ligand interactions at a detailed atomic level is an important step towards the rational design of novel inhibitors. One of its main difficulties is the robust (complete) description of the dynamical aspects associated to the protein-ligand interaction, whether conformational selection or induced fit processes (and its combination). Obtaining such a detailed atomic knowledge is very challenging for experimental techniques; computer simulations are ideal tools to accomplish such a task. Against this background, this section reveals the main objectives of this PhD thesis and points out in which particular publications they have been addressed.

The main goal of this thesis is to apply and refine novel computational techniques aiming at a comprehensive description of the protein and protein-ligand energy landscape, advancing into the rational design of novel inhibitors for selected targets.

Specific goals summarize in:

1. Validate our in-house technology PELE (Protein Energy Landscape Exploration) on sampling protein-ligand interactions. To this end, we aim to optimize protocols to map the protein-ligand recognition process for several ligands and targets, many of them suggested by our experimental collaborators, comparing *in silico* result with experimental affinities.
2. Besides protocols and software validation, we aim to develop specific application on biomedical and biotechnological relevant systems. Thus, we aim at adding information for contributing to the mechanistic knowledge of important protein-ligand interactions.
3. Following the previous goal, we aim at the implementation of the atomic detailed knowledge into the rational design of new inhibitors, aiming to enhance specificity and binding strength.
4. An added value of (accurately) describing protein-ligand interactions at a dynamical level, is being able to map possible changes in ligand affinities derived from mutations. We aim to develop protocols in PELE for the description of mutational effects in ligand binding. We tested this part on one of the most well studied systems with important mutational effects: HIV-1 protease.

## Objectives

5. Beside these main objectives based on methods application, we aim to add methodological improvements derived from the application and validation studies.

### **3 Publications**

## Publications

Publications

**3.1 Identification of dual mTORC1 and mTORC2 inhibitors in melanoma cells: prodigiosin vs. obatoclax**

Espona-Fiedler M, Soto-Cerrato V, Hosseini A, Lizcano JM, Guallar V, Quesada R, Gao T, Pérez-Tomás R.

Biochemical Pharmacology, 2011, doi: 10.1016/j.bcp.2011.11.027



## Publications



## Identification of dual mTORC1 and mTORC2 inhibitors in melanoma cells: Prodigiosin vs. obatoclax

M. Espona-Fiedler<sup>a</sup>, V. Soto-Cerrato<sup>a</sup>, A. Hosseini<sup>b</sup>, J.M. Lizcano<sup>c</sup>, V. Guallar<sup>b</sup>, R. Quesada<sup>d</sup>, T. Gao<sup>e</sup>, R. Pérez-Tomás<sup>a,\*</sup>

<sup>a</sup> Cancer Cell Biology Research Group, Department of Pathology and Experimental Therapeutics, Faculty of Medicine, University of Barcelona, E-08007 Barcelona, Spain

<sup>b</sup> ACREA Joint MSC-IMB Research Program in Computational Biology, Barcelona Supercomputing Center, E-08034 Barcelona, Spain

<sup>c</sup> Institute of Neuroscience and Department of Biochemistry and Molecular Biology – Faculty of Medicine, Autonomous University of Barcelona, E-08193 Barcelona, Spain

<sup>d</sup> Department of Chemistry, Faculty of Science, University of Burgos, E-09001 Burgos, Spain

<sup>e</sup> Markey Cancer Center, Department of Molecular and Cellular Biochemistry, University of Kentucky, Lexington, KY 40536-0509, United States

### ARTICLE INFO

Article history:  
Received 28 October 2011  
Accepted 25 November 2011  
Available online 6 December 2011

**Keywords:**  
Prodigiosin  
Obatoclax  
Melanoma  
PI3K/AKT  
mTOR inhibitors  
mTOR complexes

### ABSTRACT

The PI3K/AKT/mTOR signaling pathway regulates cell proliferation, survival and angiogenesis. The mammalian target of rapamycin (mTOR) is a protein kinase ubiquitously expressed within cells that regulates cell growth and survival by integrating nutrient and hormonal signals. mTOR exists in two complexes, mTORC1 and mTORC2. Hyperactivation of the mTOR protein has been linked to development of cancer, raising mTOR as an attractive target for cancer therapy. Prodigiosin (PG) and obatoclax (OBX), two members of the prodiginines family, are small molecules with anticancer properties which are currently under clinical trials. In the present paper, we demonstrate that mTOR is a molecular target of both prodiginines in melanoma, a highly drug-resistant cancer model. The inhibition of mTORC1 and mTORC2 complexes by PG or OBX resulted in a loss of AKT phosphorylation at S473, preventing its full activation, with no significant effect on T308. The strongest activity inhibition (89%) was induced by PG on mTORC2. Binding assays using Surface Plasmon Resonance (SPR) provide kinetic and affinity data of the interaction of these small molecules with mTOR. In addition, *in silico* modeling produced a detailed atomic description of the binding modes. These results provide new data to understand the mechanism of action of these molecules, and provide new structural data that will allow the development of more specific mTOR inhibitors for cancer treatment.

© 2011 Elsevier Inc. All rights reserved.

### 1. Introduction

Prodigiosin (PG) and obatoclax (OBX) are two prodiginine family members which have emerged as promising anticancer drugs and are currently in clinical trials. Prodiginines are bacterial metabolites with a pyrrolylpyrromethene skeleton which have shown immunosuppressive and anticancer properties. PG has shown apoptotic activity against several cancer cell types with low cytotoxicity in non-malignant cells. The National Cancer Institute ([www.dtp.nci.nih.gov](http://www.dtp.nci.nih.gov)) tested prodigiosin (and some of its derivatives) against a collection of ~60 cell lines with an average IC<sub>50</sub> (for PG) of 2.1 μM [1]. It has been described that the apoptotic

process triggered by PG is mediated through the mitochondrial pathway and involves the induction of the proapoptotic gene BAX-1 [2]. Nevertheless, the molecular target of this agent is still unknown. Further studies are also necessary to understand the mechanism of action of OBX. OBX is a synthetic indolylprodiginosin derivative, which was developed by GeminX Pharmaceuticals (recently acquired by Cephalon) and was described as a BH3 mimetic drug [3].

In the present report we identify the mammalian target of rapamycin (mTOR) as a new molecular target of the prodiginines. mTOR is an evolutionarily conserved serine/threonine protein kinase which is constituted by two signaling complexes: mTOR complex 1 (mTORC1) and mTOR complex 2 (mTORC2). Both complexes have specific effects on distinct cellular functions, such as controlling mRNA translation, ribosome biogenesis, autophagy and metabolism [4–6]. mTORC2 phosphorylates AGC kinases such as AKT, serum- and glucocorticoid-induced protein kinase-1 (SGK1) and protein kinase C-α (PKCα) [7–9]. AKT is one of the best-known downstream effectors of phosphatidylinositol-3 kinase (PI3K). Complete AKT activation depends on

Abbreviations: PG, prodigiosin; OBX, obatoclax; Rap, rapamycin; mTOR, mammalian target of rapamycin; PI3K, phosphoinositide 3-kinase; AKT/PKB, protein kinase B.

\* Corresponding author at: Cancer Cell Biology Research Group, Department of Pathology and Experimental Therapeutics, University of Barcelona, Pavelló Central, 5a planta, LR 5101C/Feixa Llarga s/n, E 08907 L'Hospitalet de Llobregat, Barcelona, Spain. Tel.: +34 934004288; fax: +34 934024288.

E-mail address: [rperez@ub.edu](mailto:rperez@ub.edu) (R. Pérez-Tomás).

phosphoinositide-dependent kinase-1 (PDK1) and mTORC2, which phosphorylate AKT at two key sites: the activation loop (T308) and the C-terminal hydrophobic motif (S473), respectively [10,11]. mTOR signaling is regulated through a network of feedback loops, protein partners, substrates, and regulators [12,13]. Among them, PRAS40 (proline-rich AKT substrate 40 kDa) is a key regulator of mTORC1. Moreover, in contrast to mTORC2, which contains rictor (rapamycin-insensitive companion of mTOR), mTORC1 contains raptor (regulatory associated protein of mTOR), which positively regulates mTOR activity and functions as a scaffold for recruiting mTORC1 substrates [14,15]. PRAS40 interacts with raptor in insulin-deprived cells and inhibits the activation of mTORC1 pathway [16]. mTORC1 regulates protein synthesis through S6-kinase and the translation repressor protein 4E-BP1. mTORC1 phosphorylates the hydrophobic motif of p70S6K on T389 [17]. On the other hand, phosphorylation of 4E-BP1 at S65 by mTORC1 prevents the binding of 4E-BP1 to the eIF4E translation initiation factor activating cap-dependent translation [13].

Interest in identifying and developing new mTOR inhibitors has increased since the second generation of mTOR inhibitors showed encouraging results in the treatment of different types of cancer, including melanoma [18]. Melanoma is an extremely aggressive disease with high metastatic potential and notoriously strong resistance to cytotoxic agents. Development of resistance has been related to the presence of different feedback loops that link both PI3K/AKT/mTOR and mitogen activated protein kinase (MAPK) pathways. These pathways are critical to melanoma progression and both are deregulated in melanoma, but not in normal cells [19,20]. Thus, compounds that counteract these feedback loops are considered in cancer therapy.

Here, we report that prodiginines inhibit both mTORC1 and mTORC2 complexes and thus counteract the SGK-1/IRS-1 negative feedback loop in melanoma. Moreover, binding assays provide data on the stability and affinity of the interaction between these small molecules and mTOR. In addition, we describe several of the recognition motifs involved in these interactions by *in silico* model.

## 2. Material and methods

### 2.1. Reagents

Prodiginosin (2-methyl-3-pentyl-6-methoxyprodiginosene) was provided by Dr. R.J. Schultz of the National Cancer Drug Synthesis and Chemistry Branch Chemotherapeutic Agents Repository (Bethesda, MD). Obatoclox, a synthetic indol-containing prodiginine, was provided by Dr. Roberto Quesada of the University of Burgos (Supplementary Fig. 1). Rapamycin (FRAP1/mTOR inhibitor) was purchased from Invitrogen (Carlsbad, CA). All stock solutions were diluted in DMSO and stored at  $-20^{\circ}\text{C}$ .

### 2.2. Cell lines and culture conditions

Human melanoma cancer cell lines SK-MEL-28 and SK-MEL-5 were purchased from American Type Culture Collection (Manassas, VA).

Stable control, mTOR knockdown and raptor knockdown human colon cancer cells SW480 were generated using lentivirus-based shRNA targeting mTOR or raptor as described in [21].

All cell types were cultured in Dulbecco's Modified Eagle's Medium (DMEM, Biological Industries, Beit Haemek, Israel) supplemented with 10% heat-inactivated fetal bovine serum (FBS; Life Technologies, Carlsbad, CA), 100 U/ml penicillin, 100  $\mu\text{g}/\text{ml}$  streptomycin, and 2 mM L-glutamine all from Biological Industries. Cells were grown at  $37^{\circ}\text{C}$  in a 5%  $\text{CO}_2$  atmosphere.

### 2.3. Cell viability assay

Cell viability was determined by MTT assay using 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (Sigma-Aldrich Chemical Co., St. Louis, MO) [see Supplementary material and methods].

### 2.4. Hoescht staining

Cell morphology was evaluated by fluorescence microscopy following Hoescht 33342 DNA staining (Sigma-Aldrich Chemical Co., St. Louis, MO). Cells ( $2 \times 10^5$  cells/ml) were treated or not with PG for 24 h. They were washed in PBS and resuspended in 2  $\mu\text{g}/\text{ml}$  Hoescht 33342 and incubated for 30 min at  $37^{\circ}\text{C}$  in the dark. Then, cells were washed in PBS and examined under a Carl Zeiss Jena microscope.

### 2.5. Kinase profiling

Kinase profiling was performed by The National Centre for Protein Kinase Profiling (MRC Protein Phosphorylation Unit, Dundee, UK). All kinase assays were carried out using a radioactive ( $^{32}\text{P}$ -ATP) filter-binding assay in duplicate. Screening was carried out at 10  $\mu\text{M}$  and industry standard QC procedures were used to validate each assay.

### 2.6. Immunoblot analysis

Cells were treated with prodiginines before insulin (Sigma-Aldrich Chemical Co., St. Louis, MO) stimulation. Adherent and floating cells were lysed in buffer (50 mM Tris pH 7.5, 60 mM glycerophosphate, 20 mM sodium pyrophosphate, 2 mM EGTA, 5 mM EDTA, 30 mM NaF, 1 mM orthovanadate, 1 mM DTT, 1% Triton X-100, 1 mM PMSF, 5  $\mu\text{M}$  pepstatin A, 10  $\mu\text{M}$  leupeptin). Protein concentration was determined with the BCA protein assay (Pierce, Rockford, IL) using bovine serum albumin as standard. 40  $\mu\text{g}$  of protein extracts was separated by SDS-PAGE and transferred to Immobilon-P membranes (Millipore, Bedford, MA). Immunoblots were developed with primary antibodies according to the manufacturer's instructions (see Supplementary material and methods). DMSO was used as a control.

### 2.7. Immunoprecipitation and non-radioactive kinase activity assay

mTORC1 and mTORC2 complexes were immunoprecipitated from SK-MEL5 cells (see Supplementary material and methods). Immunoprecipitates were assayed against recombinant protein AKT1 and p70-S6K1 (Invitrogen, Carlsbad, CA), respectively, in a final volume of 30  $\mu\text{l}$  containing 50 mM Tris pH 7.5, 10 mM magnesium chloride and 10  $\mu\text{M}$  ATP which was incubated for 30 min at  $30^{\circ}\text{C}$  with gentle shaking. Assays were stopped by addition of 5  $\mu\text{l}$  Laemmli buffer and samples were then heated to  $100^{\circ}\text{C}$  for 5 min. Samples were loaded on SDS-PAGE gel and analyzed by immunoblotting.

### 2.8. Surface Plasmon Resonance (SPR) assays

SPR assays were performed using Biacore T-100 (see Supplementary material and methods), which is a sensitive, high-performance, flow-cell-based SPR biosensor used for the analysis of protein-protein or protein-small molecule interactions [22]. This system incorporates software wizards which assist with the analysis of every interaction parameter, including kinetic and affinity evaluation and determination of binding specificity.

## 2.9. Theoretical methods

Using two separate computational approaches, (1) Blast + Modeller and (2) I-TASSER, we developed a homology modeling procedure for mTOR. Both approaches identified templates from the PI3K protein kinase family with 24–28% identity, and both built similar models. Additionally, two control *in silico* models from the PI3K family, PDK1 (pdbcode:2PE1) and PKC-alpha (pdbcode:3IW4), were chosen for the protein–ligand simulations (see Supplementary material and methods).

After preparing the models, we performed a cavity search with SiteMap, which confirmed the ATP binding site as the top ranked binding cavity, followed by initial rigid ligand docking with Glide [23]. For PDK1 and PKC-alpha we docked the crystallographic ligands together with PG. For mTOR we docked PG, OBX and PP242, a ligand that inhibits mTOR with an  $IC_{50}$  of 8 nM [24]. The induced fit was modeled by 600 iterations with PELE (Protein Energy Landscape Exploration), a stochastic method of mapping large conformational rearrangements and induced fit events in protein–ligand interactions [25]. To map the change in affinity after the protein–ligand induced fit, the PELE results were clustered and representative structures were redocked with Glide.

## 2.10. Statistical analysis

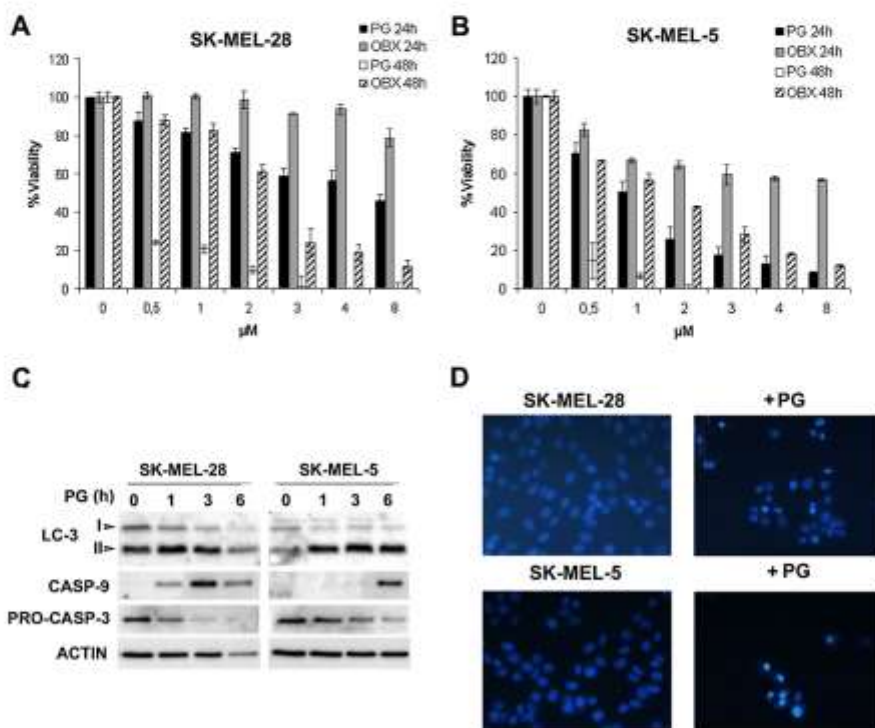
For analysis of activity kinase, results are expressed as the mean  $\pm$  S.D. of three independent experiments. Statistical analysis (ANOVA) was carried out with the Statgraphics plus 5.1. statistical package.  $P < 0.05$  and  $P < 0.01$  are represented by \* and \*\*, respectively.

## 3. Results

## 3.1. Prodigonines induce cell death in melanoma cells

To determine whether prodiginines could be potential melanoma chemotherapeutic agents, we first examined their effect on two melanoma cell lines obtained from different stages of melanoma progression according to the broadly accepted Clark model [26]. We examined SK-MEL-28 (radial growth phase) cells derived from an *in situ* melanoma. We then examined the next progression stage SK-MEL-5 (vertical growth phase) cells derived from a metastatic site (axillary node) of a melanoma-bearing patient.

To compare the cytotoxic-inducing potential of PG and OBX on melanoma cells, we treated SK-MEL-28 and SK-MEL-5 cells with



**Fig. 1.** Cell death in PG- or OBX-treated SK-MEL-28 and SK-MEL-5 cells. (A) Cytotoxic effect of PG is higher than that of OBX. Cells were treated with a range of PG or OBX concentrations (0–8 µM) for 24 and (B) 48 h. Cell viability was determined by MTT assay. The percentage of viable cells was calculated as the ratio of  $A_{550}$  between treated and control cells. Values are shown as mean  $\pm$  S.E.M. of three independent experiments performed in triplicate. (C) PG induces activation of both autophagic and apoptotic mechanisms. Cells were treated with PG  $IC_{50}$  and cell extracts were assayed for LC-3, caspase-9 and procaspase-3 expression by immunoblotting. Actin was used as loading control. (D) PG induces apoptotic body formation and nuclear condensation. This was observed through Hoechst staining in cells treated with PG  $IC_{50}$  for 24 h.

both compounds at concentrations ranging from 1  $\mu\text{M}$  to 8  $\mu\text{M}$  for 24 and 48 h. Cell viability was reduced by PG in a dose-dependent manner. PG showed a half inhibitory concentration ( $\text{IC}_{50}$ ) value of  $4.51 \pm 0.47 \mu\text{M}$  and  $1.02 \pm 0.15 \mu\text{M}$  in SK-MEL-28 and SK-MEL-5 cells, respectively (Fig. 1A and B). In contrast to PG, OBX had little effect on cell viability at 24 h. At 48 h of treatment, OBX showed an  $\text{IC}_{50}$  value of  $2.2 \mu\text{M} \pm 0.43$  and  $1.8 \mu\text{M} \pm 0.21$  in SK-MEL-28 and SK-MEL-5 cells, respectively.

Previous studies showed that OBX mediates cell death through the induction of autophagy and subsequent activation of apoptosis [27]. Therefore, we analyzed which cell death mechanism was triggered by PG, using caspase-3, -9 and LC-3 proteins as apoptotic and autophagic markers, respectively. After PG treatment, the cytosolic form of LC-3 (LC-3 I) disappeared and the form conjugated to phosphatidylethanolamine (LC-3 II) appeared. In a dose-dependent manner (Fig. 1C). It has been reported that LC-3 II is recruited to autophagosomal membranes, thus participating in autophagy [28]. LC-3 II was induced at shorter time exposures than the activation of caspase-9, indicating that both processes are triggered, but that autophagy precedes apoptosis. Moreover, caspase-3 proteolysis and the formation of apoptotic bodies corroborate the apoptotic process (Fig. 1D).

In addition, these results demonstrate that the cytotoxic effect mediated by PG is higher than by OBX in melanoma cells, especially in SK-MEL-5.

### 3.2. Regulation of PI3K/AKT/mTOR and MAPK pathways is mediated by prodiginosin

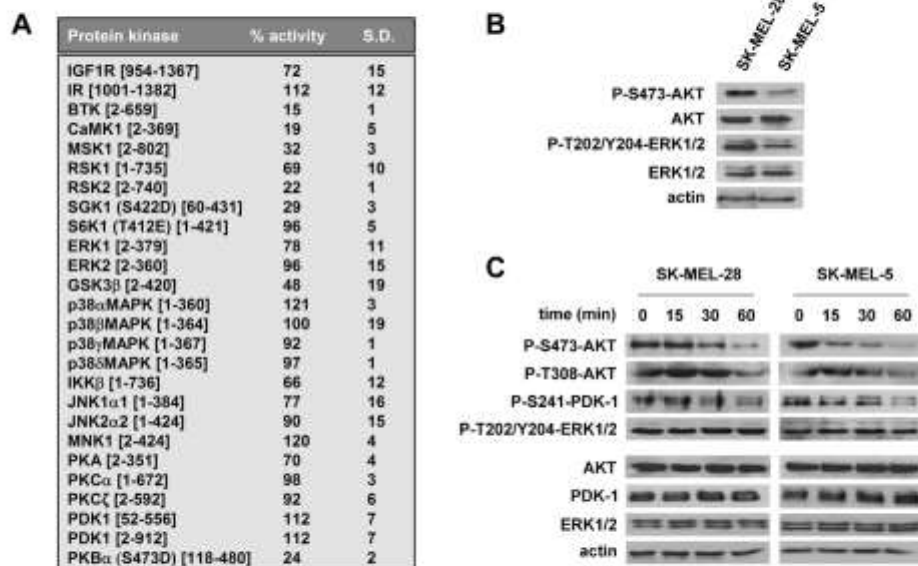
To analyze the potential of prodiginosin as protein kinase regulators a kinase profiling procedure was performed. The results

showed that PG (10  $\mu\text{M}$ ) suppressed the activity of few protein kinases from a panel of 65 protein kinases encoded by the human genome. The proteins that were significantly inhibited (>70%) were AKT, ribosomal S6 kinase (RSK)-2, mitogen and stress activated protein kinase (MSK)-1, serum glucocorticoid-inducible kinase (SGK)-1, calmodulin-dependent kinase (CaMK)-1 and Bruton's tyrosine kinase (Btk) [29,30]. All these proteins participate in PI3K/AKT/mTOR and MAPK pathways. Interestingly, other proteins closely related to these pathways such as PDK1, insulin growth factor (IGF)-1 receptor, MAPK protein kinases (extracellular related kinases (ERK)-1/2, p38, c-Jun N-terminal kinase (JNK)-1/2) and inhibitor of nuclear factor kappa-B kinase (IKK) were not inhibited. Unfortunately, other proteins of interest such as mTOR were not included in this kinase profiling (Fig. 2A). These results characterize PG as a multi-kinase inhibitor, and they provide a range of new potential targets of prodiginosin, all of which play a critical role in the control of the cell cycle and tumor progression.

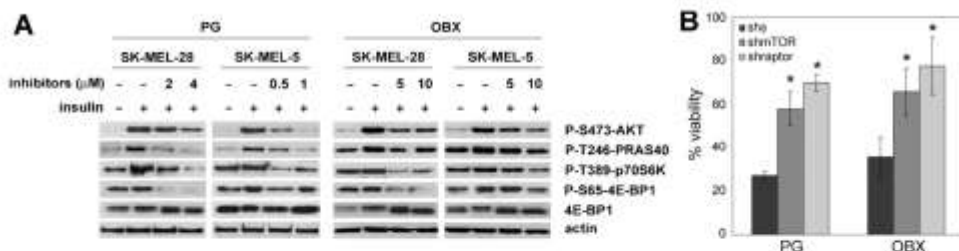
To elucidate the specific molecular mechanisms that induce the cytotoxic effect of prodiginosin, and based on the kinase profiling results, we studied the effect of PG and OBX on PI3K/AKT/mTOR and MAPK pathways in melanoma cells. Both pathways regulate melanoma cell death and proliferation [26].

We first examined the activation of the two key effector kinases of these pathways, AKT and ERK1/2. Phosphorylation levels of these kinases were higher in SK-MEL-28 than in SK-MEL-5 cells (Fig. 2B). These differences in the up-regulation of pro-survival protein kinases might explain why resistance to prodiginosin treatment depends on the cell line.

We then analyzed the effect of PG on both pathways. Cells were treated with PG at the  $\text{IC}_{50}$  concentrations established as described above. PG reduced AKT phosphorylation on S473 in a time-



**Fig. 2.** Effect of PG on PI3K/AKT/mTOR and MAPK survival pathways. (A) PG induces inhibition of some PDK and MAPK-related protein kinases in vitro. Kinase assays were performed by the MRC Protein Phosphorylation Unit. PG 10  $\mu\text{M}$  was screened in duplicate against a panel of 65 protein kinases. All assays were carried out using a radioactive ( $^{32}\text{P}$ -ATP) filter-binding assay. (B) PI3K/AKT and MAPK pathways are activated in SK-MEL-28 more strongly than in SK-MEL-5 melanoma cells. Phosphorylation and total protein levels of AKT and ERK1/2 were visualized by Immunoblotting. Actin was used as a loading control. (C) PG-induced AKT but not ERK1/2 dephosphorylation. SK-MEL-28 and SK-MEL-5 cells were treated with PG at a concentration of  $\text{IC}_{50}$  for 15 min to 1 h. Phosphorylation levels were visualized by Immunoblotting. Actin was used as a loading control.



**Fig. 3.** Prodiginines target mTOR pathway. (A) Effect of prodiginines on critical effectors of PI3K/AKT/mTOR pathway in melanoma cells. SK-MEL-28 and SK-MEL-5 cells were deprived for 24 h and treated with PG or OBX at the indicated concentrations for 1 h before insulin-stimulation at 1  $\mu$ M for 30 min. Cell extracts were analyzed by Immunoblotting. Actin was used as loading control. (B) Inhibition of mTOR activity prevents from prodiginines-mediated cell death. Stable control and knockdown cells were treated with PG or OBX at 4  $\mu$ M for 48 h. The percentage of viable cells was calculated as the ratio of  $A_{570}$  between treated and control cells. Values are shown as mean  $\pm$  S.E.M. of three independent experiments performed in triplicate. Statistical significance is shown as \* $P < 0.01$ .

dependent manner. Nevertheless, only a slight effect was observed on T308, and PG had no effect on PDK-1 or ERK1/2 (Fig. 2C). These results, together with the kinase profiling, indicate that mTOR signaling might be down-regulated in the presence of prodiginines.

### 3.3. Inhibition of mTOR signaling by prodiginines in melanoma cells

After observing significant inhibitory effects of PG on AKT, we sought to confirm the inhibition of the AKT/mTOR/p70S6K signaling pathway by prodiginines in melanoma cells. Thus, we examined the effect of prodiginines on the main regulators and substrates of this pathway. As shown in Fig. 3A, both PG and OBX inhibited mTORC2 activity, leading to an inhibition of the insulin-stimulated phosphorylation of AKT and PRAS40. Nevertheless, it was necessary to use higher doses of OBX (10  $\mu$ M) to obtain same effects as PG at 1 h of treatment. Insulin-stimulated phosphorylation of PRAS40 at T246 by AKT suppresses its mTORC1 inhibitory activity. Therefore, insulin stimulation activates mTORC1 and increases p70S6K phosphorylation [16]. After treatment with prodiginines, p70S6K and 4E-BP1, which are directly regulated by mTORC1 were also dephosphorylated. Dephosphorylation of mTOR effectors suggests that prodiginines inhibit mTOR pathway.

To further confirm that prodiginines target mTOR pathway, we first determined whether knockdown of mTOR (sh-mTOR) or raptor (sh-raptor) in SW-480 cells prevented the cytotoxic effect induced by prodiginines. Knockdown of mTOR or raptor resulted in 47% or 74% reduction in endogenous protein expression, respectively, compared with stable control cells (sh- $\phi$ ) (Supplementary Fig. 2A and B). The functional depletion of mTOR pathway was confirmed by decreased phosphorylation levels of p70S6K and AKT in mTOR and raptor knockdown cells (Supplementary Fig. 2C). According to protein depletion levels in both cell lines, results showed higher p70S6K inhibition in raptor knockdown cells. Complete AKT inhibition was observed after treatment with PG or OBX at 4  $\mu$ M for 1 h. After 6 h of treatment with both prodiginines, protein expression down-regulation was induced in both mTOR effectors.

We next assessed whether depletion of mTOR activity prevented from prodiginines-mediated cell death. Stable control cells and knockdown cells were treated with a range of concentrations (0–4  $\mu$ M) of PG or OBX for 24 or 48 h (Supplementary Fig. 3A and B). Results showed greater cytotoxic effects in sh- $\phi$  cells after treatment with PG than OBX, as observed in melanoma cells. Nevertheless, the most significant results were obtained after 48 h of treatment. At 4  $\mu$ M of PG or OBX, sh-mTOR and sh-raptor cell death was reduced in a ~35–40% compared with sh- $\phi$  cells (Fig. 3B), demonstrating that mTOR complexes are critical for prodiginines cytotoxic effect.

### 3.4. Prodiginines counteract the activation of the S6K-1/IRS-1 negative feedback loop through mTORC1 and mTORC2 inhibition

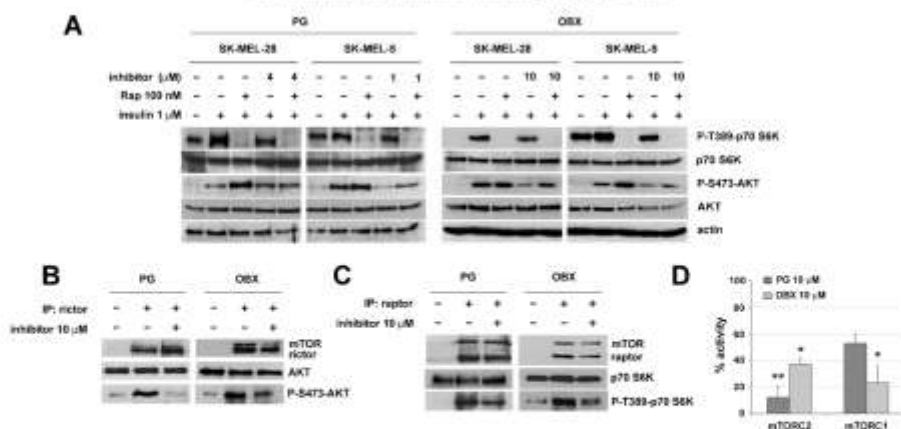
We further examined the ability of PG and OBX to inhibit mTORC1 and mTORC2 complexes in melanoma cells. Inhibition of mTORC1 induces activation of S6K-1/IRS-1 negative feedback [19]. Dual inhibition of mTORC1 and mTORC2 might counteract this mechanism. With this in mind, we first compared the effect of both compounds with that of rapamycin, which inhibits mTORC1, leading to an increase in AKT phosphorylation through an IGF-1R-dependent mechanism. As expected, rapamycin alone decreased phospho-p70S6K levels, while increasing phospho-AKT levels. The greatest effect of rapamycin was seen at 100 nM (Supplementary Fig. 4). As a further comparison, we co-treated both cell lines with PG (IC<sub>50</sub>) or OBX (10  $\mu$ M) and 100 nM rapamycin for 1 h, before insulin stimulation. In contrast to both prodiginines alone, treatment with rapamycin induced a stronger decrease in the phospho-p70S6K levels, suggesting that mTORC1 is not completely inhibited by PG or OBX (Fig. 4A). Nevertheless, in both cell lines, PG rather than OBX counteracted the activation of this feedback loop, leaving AKT quite dephosphorylated.

To evaluate the inhibition on mTORC2 and mTORC1 complexes, further kinase activity assays were performed. We isolated active mTORC2 and mTORC1 complexes from melanoma cells after insulin stimulation. Active mTORC2 and mTORC1 were immunoprecipitated from the lysates using anti-rictor and anti-raptor antibodies, respectively. mTORC2 activity was analyzed using recombinant AKT as substrate. The inhibition of mTORC2 by PG or OBX resulted in a loss of AKT phosphorylation (Fig. 4B). We next measured the effect of both molecules on mTORC1 activity using p70S6 as a substrate (Fig. 4C). Both prodiginines markedly inhibited both mTOR complexes. The strongest inhibition (89%) was induced by PG on mTORC2 (Fig. 4D).

### 3.5. Kinetic characterization

In order to further characterize the prodiginines as mTOR inhibitors, we first monitored the interaction between mTOR and prodiginines by real-time interaction analysis. We used Surface Plasmon Resonance (SPR) assays, which allow kinetic and affinity evaluation and determination of binding specificity between proteins and small molecules [22].

For binding experiments, we first immobilized the recombinant protein mTOR (aa 1360–2549) on a sensor surface. The analytes (PG and OBX) were then injected in solution over the surface. Changes in SPR response were detected even at nanomolar concentrations. The interactions of small molecule inhibitors with

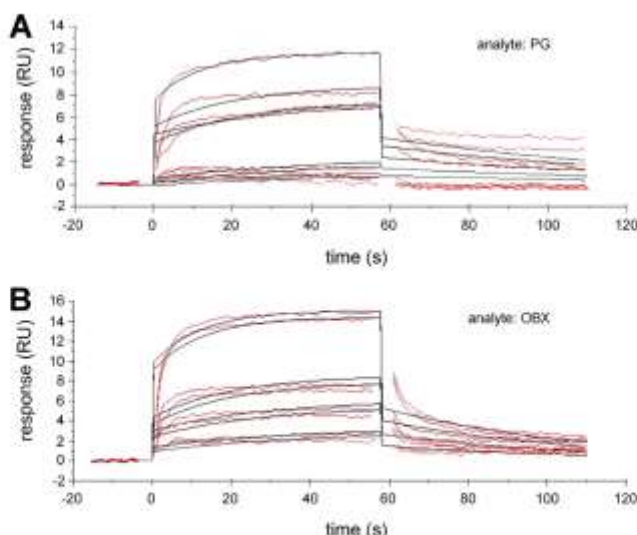


**Fig. 4.** Prodinigines counteract SGK/IGF-1R negative feedback loop through mTORC1 and mTORC2 inhibition. (A) Effect of PG on SK-MEL-28 and SK-MEL-5 cells. Before insulin stimulation, deprived cells were pre-incubated with rapamycin 100 nM for 15 min and prodiginines were then added at the indicated concentrations for 1 h. (B) Prodinigines inhibit mTORC2 and mTORC1 complexes. Active mTORC2 and mTORC1 complexes were immunoprecipitated from insulin-stimulated SK-MEL-5 cells using anti-rictor or (C) anti-raptor antibodies, respectively. Immunoprecipitated were incubated with recombinant AKT or p70S6K (20 ng), 50 mM Tris pH 7.5, 10 mM magnesium chloride, 10 μM ATP and PG or OBX, both at 10 μM for 30 min. Kinase reaction was stopped with Laemmli buffer (3×) and phosphorylation levels of p70 S6K or AKT were visualized by immunoblotting. (D) Data is expressed as the percentage of phospho-AKT or phospho-p70S6K levels (normalized by AKT or p70S6K levels), respectively, and shown as the mean ± S.D. Statistical significance is shown as \*P < 0.05; \*\*P < 0.01.

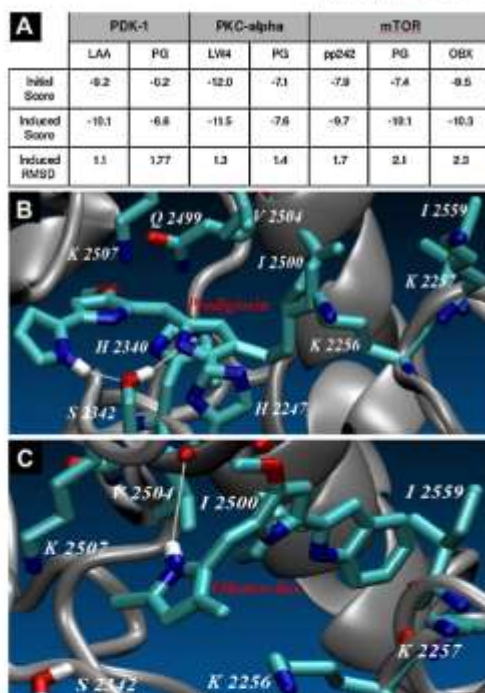
mTOR were analyzed, providing kinetic and affinity data in the nanomolar range (Fig. 5A and B). The same order of magnitude in the affinity data ( $K_D$ ) and the low dissociation rates ( $k_{off}$ ) suggest that both molecules form similar specific and stable binders.

Small molecule specificity was tested by performing interaction analysis between these compounds and another protein kinase:

AKT. In this case, the data did not fit this interaction model using PG as analyte. Moreover, using OBX no high-affinity saturable component was obtained (Supplementary Fig. 5). These marked differences in the characteristics of the interactions compared to mTOR suggest that stable binding does not occur between prodiginines and AKT.



**Fig. 5.** Kinetic characterization of PG and OBX binding to mTOR by Surface Plasmon Resonance (SPR) assays. (A) PG or (B) OBX were injected at a range of concentrations between 0.06 μM and 1 μM over immobilized mTOR (aa 1360–2549). Sensograms and curve fit (black lines) are shown. Kinetics and affinity evaluation of these sensograms showed affinity ( $K_D = k_{off}/k_{on}$ ) constants of 416 ± 49 nM for PG and 355 ± 121 nM for OBX, respectively. RU, resonance units.



**Fig. 6.** In silico model. (A) Initial and after PELE (induced fit) scores obtained with Glide for the different systems and ligands. Also shown is the RMSD along the induced fit process. RMSD was computed as all heavy atom within 10 Å of the initial docking position (units in angstroms). (B) Detailed view of the PG-mTOR and (C) OBX-mTOR induced fit complex.

### 3.6. In silico docking prediction

To further characterize the interaction of prodiginines with mTOR, we developed a homology model of the active site sequence (residues 2131–2516). PDK1 and PKC-alpha were used as control systems. Fig. 6A shows the Glide docking scores before and after the PELE run modeling the induced fit procedure, together with the active site RMSD along the fit process, for all systems and ligands. As expected, we obtain good initial docking scores for the two crystallographic ligands in PDK1 and PKC-alpha,  $-9.2$  and  $-12.0$ , respectively. Furthermore, when comparing the docked structure for each ligand with its crystal structure we obtained an RMSD of  $\sim 0.3$  Å. Such agreement, however, is expected when a ligand is docked into its crystal structure. The induced fit procedure does not significantly change the scores, and only introduces slight changes to the protein-ligand structures in the active site:  $1.1$  Å and  $1.3$  Å for LAA and LW4, respectively. Thus, the control test with both crystallographic ligands indicates good affinities and small induced fit reorganization. Next, we applied the same protocol to PG and found significantly lower initial binding scores:  $-6$  to  $-7$ . The initial models, however, were taken from the crystal structures with bound ligands. Thus, they are biased towards the crystal ligands, requiring some induced fit in order to adapt to PG. For this purpose PELE was used [25]. The induced fit procedure increased the RMSD but did not substantially improve the affinity for PG; the

control experiments support the absence of inhibition in PDK1 and PKC observed in the kinase profiling.

Regarding the mTOR ligand docking, the initial docking score of pp242, PG and OBX is about  $-7$  to  $-8$ , similar to the value measured for PG in our *in silico* control assays. For mTOR, however, the induced fit procedure (the same used for PDK1 and PKC-alpha) introduced significant changes. We observe a clear increase in binding affinity along with a significant active site adjustment, the RMSD increases to  $2.1$  and  $2.3$  for PG and OBX, respectively. For all three ligands we obtained scores which are similar to those obtained for the crystal ligands in PDK1 and PKC-alpha. Thus, our simulation studies corroborate recent observations on pp242 [24] and also support our experimental findings with PG and OBX.

In Fig. 6B and C we display the detailed atomic view of the induced fit of the PG- and OBX-mTOR complex, where we can observe several recognition motifs. For example, in the fit PG-mTOR complex, the hydrophobic environment around the pentyl side chain includes Ile2500, Ile2559 and Val2504. We also observe a stacking interaction between His2340 and the PG pyrrole rings. The most interesting feature, however, is the “H-bond ring” created by the alcohol side chain in Ser2342 and the two extreme pyrrole nitrogens. *In silico* mutation of this serine to a glycine reduces the glide score by 3 units, pointing to the importance of this interaction [31].

### 4. Discussion

Prodiginosin (PG) and obatoclax (OBX), both belong to the prodiginine family, which are pyrrole alkaloids of bacterial origin. Both molecules are promising candidates as anticancer drugs, since they have pro-apoptotic activity in a broad range of human cancer cell lines [1]. Although they are currently in pre-clinical and clinical trials, respectively, further studies are necessary to identify the molecular target involved in their anticancer activity. It was hypothesized that OBX has more complex effects on melanoma cells than merely binding to and inhibiting antiapoptotic Bcl-2 family proteins [3]. In the present report, we considered other survival signals to further understand the mechanism of action of prodiginines. We studied the prodiginine-mediated cytotoxic effect in different stages of melanoma progression, PI3K/mTOR and MAPK signaling pathways are involved in growth and progression in melanoma. Deregulation of these pathways is associated with resistance to apoptosis, increased cell growth, cell proliferation and cell energy metabolism. It also confers to melanoma resistance to many chemotherapeutic agents [26]. Prodiginines overcome this resistance, and are thus cytotoxic in both cell lines. Moreover, deregulation of these pathways might explain the higher  $IC_{50}$  value of SK-MEL-28 cells compared to SK-MEL-5 cells.

Melanoma cells were more sensitive to PG than to OBX. Our results are consistent with previous reports where melanoma cells rendered more sensitive to OBX-induced apoptosis only in combination with molecules that induce reticulum stress, but not as a single agent [32]. It was also described that OBX induces autophagy before apoptosis [27]. Our results show that PG also activates both mechanisms. Moreover, cytotoxic effect of prodiginines was partially prevented by mTOR and rapTOR proteins depletion, concluding that prodiginines-induced mTOR pathway inhibition leads to cell death. We have characterized the effect of both prodiginines on critical elements of mTOR signaling such as AKT, PRAS40, p70 S6K and 4E-BP1, as well as on different compensatory mechanisms which link PI3K/mTOR and MAPK signaling [18,20]. Our results demonstrate that prodiginines target mTOR pathway. In contrast to rapamycin, prodiginines counteract the feedback loop triggered by mTORC1 inhibition, which leads to AKT activation through IGF-1/IRS-1 signaling [19]. The effective-



ness of prodiginines relies on their ability to inhibit mTOR activity. Therefore, the suppression of mTORC1 as well as mTORC2 activity implies a reduction of phosphorylation of AKT at S473.

The kinetics and affinity evaluation revealed high-affinity binding between mTOR and prodiginines. Nevertheless, although mTOR is a validated target for the treatment of cancer, it might be necessary to develop more effective prodiginine-derivates with improved chemical properties that increase their bioavailability.

The second generation of mTOR inhibitors targets the ATP site. These compounds suppress AKT phosphorylation, leading to a stronger antiproliferative effect than that of rapamycin [18,33]. Therefore, prodiginines effect might be closer to the second generation of mTOR inhibitors. Thus, it will be interesting to compare prodiginines with other small molecules such as AZD-8055 (AstraZeneca), INK-128 (Intellilike), OSI-027 (OSI pharmaceuticals) or pp242 which also block both mTORC1 and mTORC2 complexes. pp242 is structurally similar to prodiginines [24]. *In silico* models suggest that prodiginines, like pp242, could interact in the active-site of mTOR and provided several of the recognition motifs involved in their interaction. Moreover, these models also illustrate the importance of the alcohol side chain and the two extreme pyrrole nitrogens of the prodiginines for the interaction on Ser2342 of mTOR.

These findings described here by PG and OBX as mTORC1 and mTORC2 inhibitors contribute to our understanding of the molecular mechanisms of action of both molecules and provide data about their structural properties that will allow the development of more-effective mTOR inhibitors in the future.

#### Acknowledgements

The authors thank Marta Taulés for technical assistance from Centres Científics i Tecnològics (Universitat de Barcelona) and Robin Rycroft for language assistance. This work was supported by a research grant from Spanish government and the European Union (FIS-PI10/00338).

#### Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bcp.2011.11.027.

#### References

- Perez-Tomas R, Vinas M. New insights on the antitumoral properties of prodiginines. *Curr Med Chem* 2010;17:2222–31.
- Soto-Cerrato V, Vinals F, Lambert JR, Kelly JA, Perez-Tomas R. Prodiginin induces the proapoptotic gene NAG-1 via glycogen synthase kinase-3beta activity in human breast cancer cells. *Mol Cancer Ther* 2007;6:362–9.
- Nguyen M, Marcellus RC, Roulston A, Watson M, Serfass L, Murrby Madiraju SR, et al. Small molecule obatoclax (GX15-070) antagonizes MCL-1 and overcomes MCL-1-mediated resistance to apoptosis. *Proc Natl Acad Sci U S A* 2007;104:19512–7.
- Sarbasov DD, Ali SM, Kim DH, Guertin DA, Latek RR, Erdjument-Bromage H, et al. Rictor, a novel binding partner of mTOR, defines a rapamycin-insensitive and raptor-independent pathway that regulates the cytoskeleton. *Curr Biol* 2004;14:1296–302.
- Wullschlaeger S, Loewith R, Hall MN. TOR signaling in growth and metabolism. *Cell* 2006;124:471–84.
- Guertin DA, Sabatini DM. The pharmacology of mTOR inhibition. *Sci Signal* 2009;2:pe24.
- Ali SM, Sabatini DM. Structure of S6 kinase 1 determines whether raptor-mTOR or rictor-mTOR phosphorylates its hydrophobic motif site. *J Biol Chem* 2005;280:19445–8.
- Frodin M, Anzal TL, Dummer BA, Jensen CJ, Deak M, Gammeltoft S, et al. A phosphoserine/threonine-binding pocket in AGC kinases and PKI1 mediates activation by hydrophobic motif phosphorylation. *EMBO J* 2002;21:5396–407.
- García-Martínez JM, Alessi DR. mTOR complex 2 (mTORC2) controls hydrophobic motif phosphorylation and activation of serum- and glucocorticoid-induced protein kinase 1 (SGK1). *Biochem J* 2008;416:375–85.
- Alessi DR, Andjelkovic M, Caudwell B, Cron P, Morrice N, Cohen P, et al. Mechanism of activation of protein kinase B by insulin and IGF-1. *EMBO J* 1996;15:6541–51.
- Manning BD, Cantley LC. AKT/PKB signaling: navigating downstream. *Cell* 2007;129:1261–74.
- Schmelzle T, Hall MN. TOR, a central controller of cell growth. *Cell* 2000;103:253–62.
- Fingar DC, Salama S, Tsou C, Harlow E, Blenis J. Mammalian cell size is controlled by mTOR and its downstream targets S6K1 and 4EBP1/eIF4E. *Genes Dev* 2002;16:1472–87.
- Hara K, Maruki Y, Long X, Yoshino K, Oshiro N, Hidayat S, et al. Raptor, a binding partner of target of rapamycin (TOR), mediates TOR action. *Cell* 2002;110:177–80.
- Kim DH, Sarbasov DD, Ali SM, Kim JE, Latek RR, Erdjument-Bromage H, et al. mTOR interacts with raptor to form a nutrient-sensitive complex that signals to the cell growth machinery. *Cell* 2002;110:163–75.
- Sancak Y, Thoreen CC, Peterson TR, Lindquist RA, Kang SA, Spooner E, et al. PRAS40 is an insulin-regulated inhibitor of the mTORC1 protein kinase. *Mol Cell* 2007;25:903–13.
- Jeffries HR, Farnagali S, Dennis PB, Rainbird C, Pearson RB, Thomas G, et al. Rapamycin suppresses 5' TOP mRNA translation through inhibition of p70S6k. *EMBO J* 1997;16:3693–704.
- Huang S, Houghton PJ. Targeting mTOR signaling for cancer therapy. *Curr Opin Pharmacol* 2009;3:371–7.
- Wan X, Harkavy B, Shen N, Grubbar P, Helman LJ. Rapamycin induces feedback activation of Akt signaling through an IGF-1R-dependent mechanism. *Oncogene* 2007;26:1932–40.
- Carracedo A, Ma L, Teruya-Feldstein J, Rogo F, Salmena L, Alimonti A, et al. Inhibition of mTORC1 leads to MAPK pathway activation through a PI3K-dependent feedback loop in human cancer. *J Clin Invest* 2008;118:3065–74.
- Liu J, Stevens PD, Gao Y. mTOR-dependent regulation of PHUPP expression controls the rapamycin sensitivity in cancer cells. *J Biol Chem* 2011;286(8):6510–20.
- Nordin H, Jungnelius M, Karlsson R, Karlsson OP. Kinetic studies of small molecule interactions with protein kinases using biosensor technology. *Anal Biochem* 2005;340:359–68.
- Glide. version 5.6. Schrödinger, New York, NY: LLC; 2010.
- Feldman ME, Apse B, Uotila A, Loewith R, Knight ZA, Ruggiero D, et al. Active-site inhibitors of mTOR target rapamycin-resistant outputs of mTORC1 and mTORC2. *PLoS Biol* 2009;7:371–83.
- Borrelli K, Cossins B, Gaillar V. Exploring hierarchical refinement techniques for induced fit docking with protein and ligand flexibility. *J Comput Chem* 2010;31:1224–35.
- Miller AJ, Mihm Jr MC, Melanoma. *N Engl J Med* 2006;355:51–65.
- Pan J, Cheng C, Verstovsek S, Chen Q, Jin Y, Cao Q. The BHD-mimetic GX15-070 induces autophagy, potentiates the cytotoxicity of carboplatin and 5-fluorouracil in esophageal carcinoma cells. *Cancer Lett* 2010;293:167–74.
- Kabaya Y, Mizushima N, Yamamoto A, Ohsumi T, Yoshimori T, LC3, GABARAP and GATE16 localize to autophagosomal membrane depending on form-II formation. *J Cell Sci* 2004;117:2805–12.
- Dalby KN, Morrice N, Caudwell FB, Avruch J, Cohen P. Identification of regulatory phosphorylation sites in mitogen-activated protein kinase (MAPK)-activated protein kinase-1a/p90rsk that are inducible by MAPK. *J Biol Chem* 1998;273:1496–505.
- Hanada M, Feng J, Hemmings BA. Structure, regulation and function of PKB/AKT – a major therapeutic target. *Biochim Biophys Acta* 2004;1697:3–16.
- Tiron MM. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys Rev Lett* 1996;77:1905–8.
- Jiang CC, Wroblewski D, Yang F, Hershey P, Zhang XD. Human melanoma cells under endoplasmic reticulum stress are more susceptible to apoptosis induced by the BHD mimetic obatoclax. *Neoplasia* 2009;11(9):945–55.
- Liu Q, Thoreen C, Wang J, Sabatini D, Gray NS. mTOR mediated anti-cancer drug discovery. *Drug Discov Today Ther Strateg* 2009;6:47–55.

## Supplementary Material

### Identification of dual mTORC1 and mTORC2 inhibitors in melanoma cells: prodigiosin vs. obatoclax

España-Fiedler M, Soto-Cerrato V, Hosseini A, Lizcano JM, Guallar V, Quesada R, Gao T, Pérez-Tomás R.

#### Material and Methods

##### 1. Cell Viability Assay

SK-MEL-28 and SK-MEL-5 were plated in triplicate wells ( $1.5 \times 10^4$  cells/ml) in 100  $\mu$ l of growth medium in 96-well plates and allowed to grow for 24 h. After 24 h or 48 h of treatment with PG or OBX, 10  $\mu$ M of MTT was added to each well for an additional 4 h. DMSO was added as a control. The blue MTT formazan precipitate was dissolved in 100  $\mu$ l of isopropanol: 1N HCl (24:1). The absorbance at 570 nm was measured on a multiwell plate reader. Cell viability was expressed as a percentage of the control, and data are shown as the mean value  $\pm$  S.D. of three independent experiments.

##### 2. Immunoblot Analysis

The following antibodies were obtained from Cell Signaling Technology (Beverly, MA): anti-LC-3, anti-caspase-9, anti-procaspase-3, anti-actin, anti-phospho PDK-1 (Ser241), anti-phospho AKT (Ser473), anti-phospho Akt (Thr308), anti-phospho PRAS40 (Thr246), anti-phospho ERK1/2 (Thr202/Tyr204), anti-phospho p70S6K (Thr389), anti-phospho 4E-BP1 (Ser65), anti-rictor, anti-raptor, anti-AKT, anti-mTOR, anti-p44/42 MAP kinase, anti-PDK-1, p70S6K and 4E-BP1. We used secondary antibodies conjugated to horseradish peroxidase (Santa Cruz Biotechnology, Santa Cruz, CA), and signal was detected using the enhanced chemiluminescence detection kit (GE Healthcare Bio-Sciences AB, Uppsala, Sweden).

### **3. Immunoprecipitation Assay**

Cells were lysed in CHAPS buffer (50 mM Tris pH 7.5, 120 mM NaCl, 1 mM EDTA, 10 mM sodium pyrophosphate, 50 mM NaF, 0.3% CHAPS, 1 mM PMSF, 5  $\mu$ M pepstatin A, 10  $\mu$ M leupeptin). 200  $\mu$ l of cell lysate was incubated with primary antibody (dilution 1:100) with gentle shaking overnight at 4°C. 20  $\mu$ l of Protein A Agarose beads (Cell Signaling Technology, Beverly, MA) was added and incubated with gentle rocking for 3 h at 4°C. As a negative control, cell lysates were also incubated with Protein A Agarose beads alone. Cell lysates were centrifuged for 30 s at 4°C. Pellet were washed twice in 500  $\mu$ l of CHAPS buffer and 3 times in a buffer containing 50 mM Tris pH 7.5, 40 mM NaCl, 2 mM EDTA. Immunoprecipitate activity was analyzed following the kinase assay procedure.

### **4. Surface Plasmon Resonance (Rodger, Lodwick et al.) Assays**

GST-tagged human recombinant mTOR (aa 1360-2549) (Invitrogen, Carlsbad, CA) was covalently attached to a CM5 sensor chip (GE Healthcare Bio-Sciences AB, Uppsala, Sweden) according to amine-coupling protocol, exploiting primary amine groups on the ligand after activation of the surface with 1-ethyl-3 (3-dimethylaminopropyl) carbodiimide (EDC) and N-hydroxysuccinimide (NHS). The excess of reactive groups were deactivated with ethanolamine. mTOR was diluted to 4.5 ng/  $\mu$ l in acetate buffer pH 3.9, and immobilized to a level of 11.897 RU into flow-cell 4. Flow cell 1 was activated and blocked with GST and assigned as reference. RU corresponds to a measure of a 0.0001 degree-shift in the refractive index.

Compounds were stored as stock solutions in 100% dimethyl sulfoxide (DMSO) at -20°C. The compounds were diluted with 1.05-fold concentrated assay buffer without DMSO to prepare the highest experimental concentration of each compound in a final buffer composition of 10 mM HEPES pH 7.4, 150 mM NaCl, 100 mM MgCl<sub>2</sub>.

Assay buffer was also used as the instrument running buffer and for further sample dilution. The run was started with three start-up cycles, in which assay buffer was injected instead of sample, followed by sample injection cycles. Zero concentration samples were used as blanks. A typical sample injection cycle consisted of a 60 s sample injection (30  $\mu$ l/min), 120-

## Publications

200 s of buffer flow (dissociation phase), and a 30 s buffer injection to check for sample carryover. The complexes were regenerated using 25 mM NaOH in every cycle prior to sample injection. Moreover, between sample series, a solvent correction cycle was run according to the instrument manual (26) to adjust for referencing errors due to refractive index mismatches between running buffer and samples.

Experiments were performed with the instrument temperature (flow cell, sensor chip, and sample compartment temperature) set to 25 °C and the flow cell temperature set to 20 °C. For kinetic and affinity evaluations, Biacore T100 evaluation software was used for subtraction of reference and blank data along with solvent correction as well as for curve fitting.

Small molecule binding specificity to AKT was analyzed using human recombinant histidine-tagged AKT (Invitrogen, Carlsbad, CA) attached to another CM5 sensor chip according to the same protocol. Flow cell 1 was activated and blocked without GST and assigned as reference.

## 5. Theoretical Methods

**5.1. mTOR homology modelling.** Since there is no crystal structure available for any mTOR complex, we produced a homology model of the active site sequence, residues 2131-2516. For this purpose we used two approaches. The first involved a BLAST search followed by multiple alignment model generation with Modeller. For the second approach we used an automatic multiple-threading alignment server from the University of Michigan, I-TASSER, which was ranked as the top server for protein structure prediction in the recent CASPS experiments. Both approaches identified templates from the PI3K protein kinase family with 24-28% identity. The alpha carbon RMSD between the top poses of each method (excluding the common gap regions) is  $<2.5\text{\AA}$ . The ATP binding region is even more conserved and the RMSD, when aligning the top two models within  $5\text{\AA}$  of the ATP binding region, is  $<1.0\text{\AA}$ . For all templates there is a gap of approximately 30 residues. In all models, however, the predicted gap is not near the ATP binding region. Based on visual inspection of this area, however, we chose the top I-TASSER model. I-TASSER also uses a secondary structure prediction algorithm, and it produced a less protruding gap area. Furthermore, we performed a 10 ns molecular dynamics trajectory using the OPLS2005 force field with implicit SGB solvent, where only the

## Publications

gap area was allowed to move for further refinement of this segment. In any case, the gap area was always far from the ATP binding region.

**5.2. Control *in silico* models.** Two other protein kinases from the PI3K family, PDK1 (pdbcode:2PE1) and PKC-alpha (pdbcode:3IW4), were selected for *in silico* control experiments. These two kinases have crystallographic structures with bound inhibitors, LAA and LW4, respectively. While docking scores gave a measure of the binding strength, results of the docking experiments for these bound crystallographic ligands indicated what scores could be expected for inhibitors in the PI3K family.

All the system were prepared with the protein wizard from Schrödinger, which adds hydrogens and optimizes the hydrogen bond network by dihedral rotation of Asn, Gln, Tyr, Cys, Thr, Ser, and histidine protonation/rotation.

**5.3. Ligand docking and induced fit procedure.** After preparation we performed a cavity search with SiteMap, which confirmed the ATP binding site as the top ranked binding cavity in all three systems. Initial rigid ligand docking was performed with Glide using the XP score (23). For PDK1 and PKC-alpha we docked the crystallographic ligands together with PG. For mTOR we docked PG, OBX and PP242. Following the rigid docking we performed 600 iterations of induced fitted adjustment with PELE (Protein Energy Landscape Exploration), a stochastic method capable of mapping large conformational rearrangements and induced fit events in protein-ligand interactions (25). PELE's algorithm is based on three main steps:

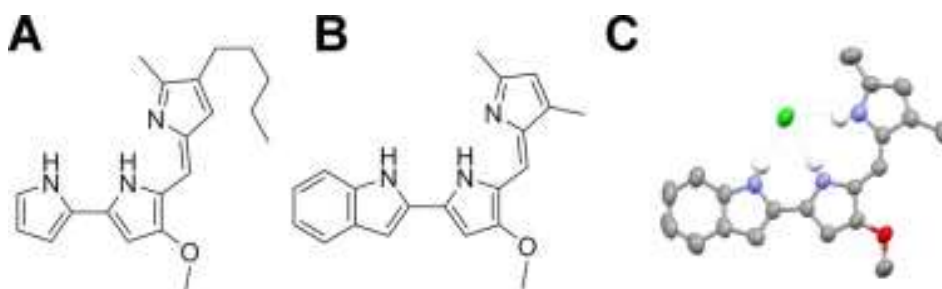
1) Localized perturbation. After an energy calculation for the initial structure, the procedure begins with the generation of a perturbation in the system. The perturbation might include a ligand translation and rotation and a quick minimization where the alpha carbons are forced by a harmonic constraint to a new position. This new position is derived from a small displacement in a low frequency mode (or a combination of modes) resulting from an anisotropic network model approach, a simple model for normal mode analysis (31).

2) Side-chain sampling. The algorithm proceeds by placing all side chains local (within 3Å) to the ligand with a rotamer library side chain optimization at a rotamer resolution of 10°.

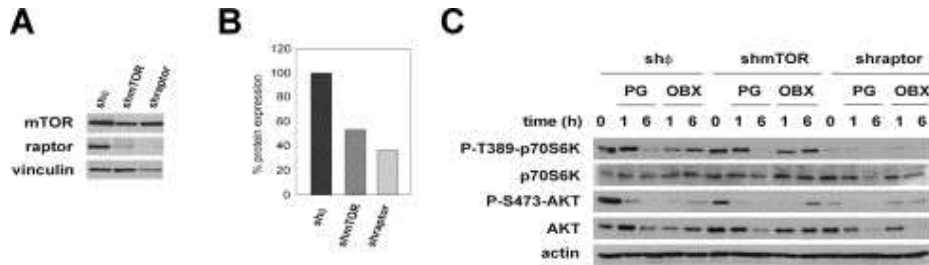
3) Minimization. The last step involves the minimization of a region including, at least, all residues local to the atoms involved in the first two steps.

These three steps comprise a move which is accepted (defining a new minimum) or rejected on the basis of a Metropolis criterion. The collection of accepted steps forms a stochastic trajectory. We clustered the trajectory (based on carbon alpha RMSD) in 5 groups and selected the median in each group. For each cluster representative we redocked all ligands. Thus for each system and ligand we have two docking scores, before (initial docking) and after the induced fit.

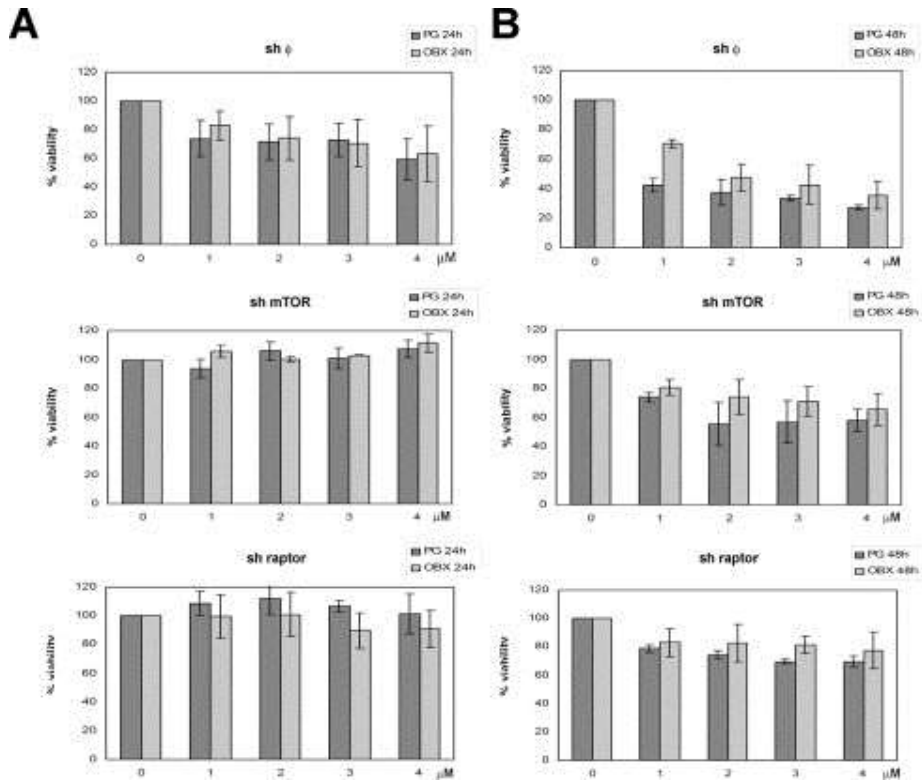
### Supplementary Figures



**Supplementary Fig. 1. Chemical structure of prodiginines.** (A) Side-on view of 2-methyl-3-pentyl-6-methoxyprodigosene (prodigiosin) and (B) the synthetic indol-containing prodiginine (obatoclax) showing the planar arrangement of the pyrrole rings. (C) Representation of the solid state structure of obatoclax.HCl. The structure of obatoclax as the hydrochloride salt was determined by single crystal X-Ray diffraction. The *tris*-heterocyclic skeleton of OBX is essentially flat, with the three NH groups oriented in the same directions and forming hydrogen bonds with the chloride anion (N-H $\cdots$ Cl distances 3.17-3.18 Å).

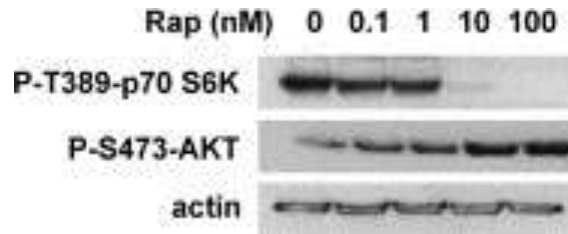


**Supplementary Fig. 2. (A) Knockdown of endogenous mTOR and raptor.** Stable control (sh-φ), mTOR (sh-mTOR) or raptor (sh-raptor) knockdown cells were analyzed for mTOR and raptor expression. Vinculin was used as the loading control. Immunoblots were quantified and normalized to the control cells. **(B)** Data is expressed as the percentage of protein expression. **(C) Inhibition of mTOR effectors in knockdown cells.** Cells were treated with PG or OBX at 4 μM for 0-6 h. Phosphorylation and total protein expression of AKT and p70S6K were detected by Immunoblotting. Actin was used as loading control.

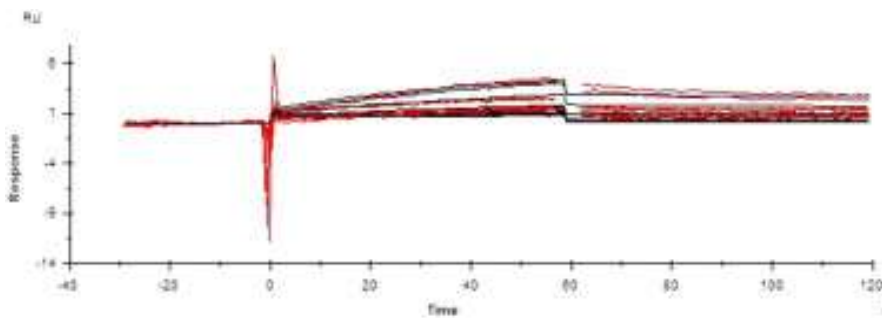


**Supplementary Fig. 3. (A) Cell viability in PG- or OBX-treated SW-480 knockdown cells.** Stable control (sh- $\phi$ ), mTOR (sh-mTOR) or raptor (sh-raptor) knockdown cells were treated with PG or OBX (0-4  $\mu\text{M}$ ) for 24 or **(B)** 48 h. The percentage of viable cells was calculated as the ratio of  $A_{570}$  between treated and control cells. Values are shown as mean  $\pm$  S.E.M. of three independent experiments performed in triplicate.





**Supplementary Fig. 4. Rapamycin dose-response curve.** SK-MEL-5 cells were treated with rapamycin at 100 nM for 30 min. Cell extracts were analyzed by Immunoblotting using actin as loading control.



**Supplementary Fig. 5. Kinetic characterization of OBX binding to AKT by Surface Plasmon Resonance (Rodger, Lodwick et al.) assays.** OBX were injected at a range of concentrations between 0.06  $\mu$ M to 1  $\mu$ M over immobilized AKT. Sensograms and curve fit (black lines) are shown.

**3.2 Molecular Interactions of Prodiginines with the BH3 Domain of Anti-Apoptotic Bcl-2 Family Members**

Ali Hosseini, Margarita Espona-Fiedler, Vanessa Soto-Cerrato, Roberto Quesada, Ricardo Pérez-Tomás, Victor Guallar

PLOS ONE, 2013, DOI: 10.1371/journal.pone.0057562

## Publications

# Molecular Interactions of Prodiginines with the BH3 Domain of Anti-Apoptotic Bcl-2 Family Members

Ali Hosseini<sup>1,2</sup>, Margarita Esposa-Fiedler<sup>2,3</sup>, Vanessa Soto-Cerrato<sup>2</sup>, Roberto Quesada<sup>3</sup>, Ricardo Pérez-Tomás<sup>2</sup>, Víctor Guallar<sup>1,4\*</sup>

**1** Joint BSC-IRB Research Program in Computational Biology, Barcelona, Spain, **2** Cancer Cell Biology Research Group, Department of Pathology and Experimental Therapeutics, University of Barcelona, Barcelona, Spain, **3** Department of Chemistry, University of Burgos, Burgos, Spain, **4** Institutació Catalana de Recerca i Estudis Avançats, Barcelona, Spain

## Abstract

Prodigiosin and obatoclax, members of the prodiginines family, are small molecules with anti-cancer properties that are currently under preclinical and clinical trials. The molecular target(s) of these agents, however, is an open question. Combining experimental and computational techniques we find that prodigiosin binds to the BH3 domain in some BCL-2 protein families, which play an important role in the apoptotic programmed cell death. In particular, our results indicate a large affinity of prodigiosin for MCL-1, an anti-apoptotic member of the BCL-2 family. In melanoma cells, we demonstrate that prodigiosin activates the mitochondrial apoptotic pathway by disrupting MCL-1/BAK complexes. Computer simulations with the PELE software allow the description of the induced fit process, obtaining a detailed atomic view of the molecular interactions. These results provide new data to understand the mechanism of action of these molecules, and assist in the development of more specific inhibitors of anti-apoptotic BCL-2 proteins.

**Citation:** Hosseini A, Esposa-Fiedler M, Soto-Cerrato V, Quesada R, Pérez-Tomás R, et al. (2013) Molecular Interactions of Prodiginines with the BH3 Domain of Anti-Apoptotic Bcl-2 Family Members. PLoS ONE 8(2): e57562. doi:10.1371/journal.pone.0057562

**Editor:** Aleksey Porollo, University of Cincinnati College of Medicine, United States of America

**Received:** July 17, 2012; **Accepted:** January 26, 2013; **Published:** February 27, 2013

**Copyright:** © 2013 Hosseini et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported by a research grant from the Spanish government and the European Union (FP7-01000338) and from the ERC-2009-AdG 25027-PELE European project. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: victor.guallar@bcac.es

† These authors contributed equally to this work.

## Introduction

In order to advance in molecular target therapies, it is important to elucidate the target and the atomic detailed mechanisms of protein-drug interactions. A wide set of experimental techniques, such as crystallography, NMR, calorimetry, etc. together with theoretical docking efforts aim to address this issue. Recently, we have turned our attention in solving the molecular target and the binding mechanism for prodigiosin (PG). PG, a bacterial metabolite from the prodiginine family (see Figure 1), has shown apoptotic activity against several cancer cell types with low cytotoxicity in non-malignant cells. The National Cancer Institute (dtp.nci.nih.gov) tested prodigiosin against a collection of ~60 cancerous cell lines with an average half maximal inhibitory concentration (IC<sub>50</sub>) of 2.1 μM [1]. Furthermore, recent studies elucidated that PG triggers apoptosis by the intrinsic pathway [2], provoking the increase in the pro-apoptotic NAG-1 protein and the negative cell cycle regulator p21 [3], and inducing down-regulation of the inhibitor of apoptosis survivin [4], SKP2 [5] and RAD51 [6] proteins. Nevertheless, the direct molecular target(s) of this agent is still an open question.

Apoptosis, the programmed cell death that controls removal of damaged cells, is extremely well regulated by the extrinsic and intrinsic pathways. The key regulator proteins of the intrinsic pathway are known as the BCL-2 family, being BAX and BAK the pro-apoptotic members responsible for the mitochondrial outer membrane permeabilisation. Cytochrome c is then released to the

cytosol allowing the activation of procaspase-9, leading to cell death [7]. In survival conditions, BAX and BAK are sequestered by the anti-apoptotic family members: BCL-2, BCL-X<sub>L</sub>, MCL-1, BCL-W, BCL-B and BCL2A1. These anti-apoptotic proteins share four regions of high sequence similarity known as the BCL-2 Homology (BH) domain, (BH1, BH2, BH3 and BH4). When a stress stimulus occurs, BH3-only proteins bind to the BH3 domain of anti-apoptotic BCL-2 proteins, displacing and releasing pro-apoptotic BAK or BAX, committing the cell to death. [7,8].

BH3-mimetic molecules have emerged as promising anti-cancer drugs since they are able to directly reverse the evasion of apoptosis [9]. This is the case for Obatoclax (OBX), which binds to a broad spectrum of BCL-2 family members [10–13]. To elucidate whether PG also behaves as a BH3-mimetic drug and how these interactions occur at the molecular level, we combined again immunoprecipitation assays with *in silico* modeling [14]. We focused on MCL-1, BCL-xL and BCL-2, three anti-apoptotic proteins representative of selectivity patterns among BH3 domains within the BCL-2 family that have centered most of recent studies [15–17]. We find that PG antagonizes MCL-1 by binding to the BH3 domain triggering BAK release. Furthermore, we obtain an atomic detailed description of PG's interaction in the BH3 domain. Altogether, these results show, for the first time, the BH3 mimetic nature of PG and provide a detailed atomic view of the molecular interactions of prodiginines (PGs) with the BH3 domain of several antiapoptotic BCL-2 proteins.

## Materials and Methods

### Reagents

Prodigiosin (2-methyl-3-pentyl-6-methoxyprodigiosene) was provided by Dr. R. J. Schultz of the National Cancer Drug Synthesis and Chemistry Branch Chemotherapeutic Agents Repository (Bethesda, MD). Obatoclax, a synthetic indol-containing prodiginine, was prepared by acid catalyzed condensation of 2-(4-methoxy-5-vinyl-1H-pyrrol-2-yl)-1H-indole and 2,4-dimethyl-1H-pyrrole [18]. All stock solutions were diluted in DMSO and stored at  $-20^{\circ}\text{C}$ .

### Cell Lines and Culture Conditions

Human melanoma cancer cell line SK-MEL-5 was purchased from the American Type Culture Collection (Manassas, VA). Cells were cultured in Dulbecco's Modified Eagles Medium (DMEM, Biological Industries, Beit Haemek, Israel) supplemented with 10% heat-inactivated foetal bovine serum (FBS; Life Technologies, Carlsbad, CA), 100 U/ml penicillin, 100  $\mu\text{g}/\text{ml}$  streptomycin, and 2 mM L-glutamine, all from Biological Industries. Cells were grown at  $37^{\circ}\text{C}$  in a 5%  $\text{CO}_2$  atmosphere.

### Immunoblot Analysis

After their respective treatments, adherent and floating cells were lysed in immunoprecipitation (IP) buffer (50 mM Tris (pH 8.0), 60 mM KCl, 1 mM EDTA, 1 mM DTT, 0.5% Nonidet P-40 (NIDEPAL), 10 mM sodium vanadate, 50 mM NaF, 1  $\mu\text{g}/\text{ml}$

aprotinin, 1  $\mu\text{g}/\text{ml}$  leupeptin, 1  $\mu\text{g}/\text{ml}$  pepstatin and 0.1 mM PMSF) or lysis buffer (0.1% SDS, 1% NP-40, 0.5% sodium deoxycholate, 50 mM NaF, 40 mM  $\beta$ -glycerophosphate, 200  $\mu\text{M}$  sodium orthovanadate, 1 mM phenylmethylsulfonyl fluoride and serine and cysteine protease inhibitor cocktail (Roche 11836170001)) for MCL-1 overexpression analysis. Total cell extracts were centrifuged at  $12000\times g$  for 10 min at  $4^{\circ}\text{C}$ . Protein concentration was determined with the BCA protein assay (Pierce, Rockford, IL) using bovine serum albumin as standard. 40  $\mu\text{g}$  of protein extracts were separated by SDS-PAGE and transferred to Immobilon-P membranes (Millipore, Bedford, MA). They were then incubated with primary antibodies anti-MCL-1, BCL-2, BAK, BAX, and actin (Santa Cruz Biotechnology, Inc., Santa Cruz, CA), anti-Caspase-9 and PARP (Cell Signalling, Beverly, MA) or anti-Vinculin (Sigma-Aldrich Chemical Co., St. Louis, MO) according to the manufacturer's instructions. Antibody binding was detected with secondary antibodies conjugated to peroxidase and the ECL detection kit (Amersham, Buckinghamshire, UK).

### Co-immunoprecipitation from Cells

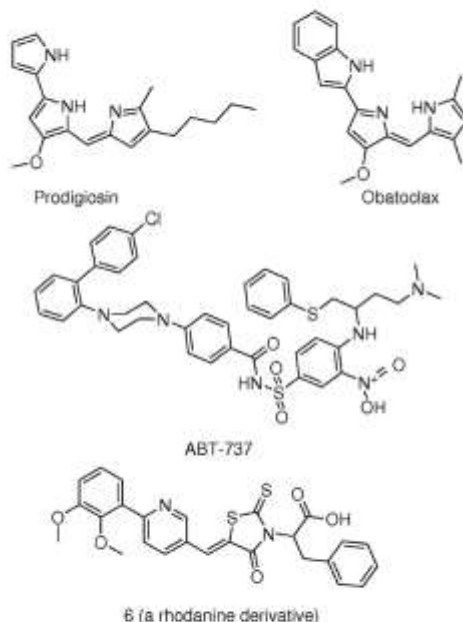
Cell extracts (800  $\mu\text{g}$  of protein) were brought to a volume of 1 ml with IP buffer and incubated with 2  $\mu\text{g}$  of anti-MCL-1 antibody (Santa Cruz Biotechnology), anti-BAK (Cell Signalling Technology) or anti-BAX antibody (Invitrogen, Carlsbad, CA) overnight at  $4^{\circ}\text{C}$ . Immune complexes were precipitated by incubation with protein A-coated agarose beads (Sigma-Aldrich Chemical Co.) previously equilibrated with IP buffer, and washed three times with 0.5 ml of IP buffer. Immunoprecipitated proteins were loaded on a 12% SDS-PAGE gel and analyzed by Immunoblot using anti-MCL-1, anti-BAK and anti-BAX antibodies (Santa Cruz Biotechnology).

### MCL-1 Overexpression and Cell Viability Evaluation

SK-MEL-5 cells were seeded in 6-well plates and allowed to grow up to 70% confluence. Before transfection, growth media was replaced by Optimum media without FBS (Invitrogen) and 1  $\mu\text{g}$  of plasmidic DNA was transfected to cells using 20  $\mu\text{l}$  of lipofectin reagent (Invitrogen) per condition. Empty pcDNA3-GFP vector or pTOPO(MCL1) plasmid (Addgene plasmid 21605 [19]), were used. After 20 h of transfection, the media was replaced by complete media with or without 2 or 20  $\mu\text{M}$  PG or OBX, respectively. After 24 h of treatment, cells were resuspended and 100  $\mu\text{l}$  of each condition were passed in triplicate to a 96-well plate. Cell viability was determined using the methyl-thiazole-tetrazolium (MTT) assay [20]. Briefly, 10  $\mu\text{M}$  of MTT (Sigma Chemical Co.) was added to each well for an additional 2 h. The blue MTT formazan precipitate was dissolved in 100  $\mu\text{l}$  of isopropanol: 1N HCl (24:1). The absorbance at 570 nm was measured on a multiwell plate reader. Cell viability was expressed as a percentage of non-treated cells and data are shown as the mean value  $\pm$  S.D. of two independent experiments. Statistical analysis (ANOVA and LSD tests) was carried out with the STATGRAPHICS Centurion XVII, statistical package.  $P < 0.05$  and  $P < 0.01$  were represented with \* and \*\*, respectively. To determine transfection efficiency, immunoblot analyses were performed to assess MCL1 protein expression levels in each condition.

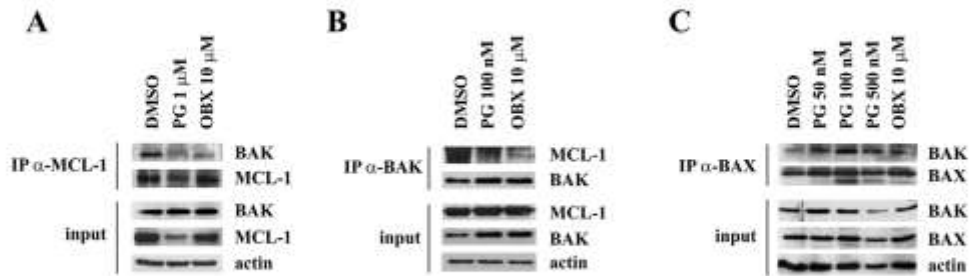
### Analyses with Isolated Mitochondria

SK-MEL-5 cells were treated with PG (100 nM), OBX (10  $\mu\text{M}$ ) or DMSO (vehicle) at  $37^{\circ}\text{C}$  for 24 h. Cells were lysed in ice-cold 25 mM Tris (pH 6.0), 250 mM sucrose, 1 mM EDTA, 0.05% digitonin, 1 mM DTT, 1  $\mu\text{g}/\text{ml}$  aprotinin, 1  $\mu\text{g}/\text{ml}$  leupeptin,



**Figure 1. The ligands used.** The ligands used in this study are shown: prodigiosin, obatoclax, ligand number 6 (a rhodanine derivative) and ABT-737.

doi:10.1371/journal.pone.0057562.g001



**Figure 2. PG is a BH3-mimetic molecule.** (A and B) PG disrupts constitutive MCL-1/BAK interaction. Cell lysates were subjected to immunoprecipitation with anti-MCL-1 or anti-BAK antibody after PG treatment at 1  $\mu$ M (8 h) and 100 nM (24 h), respectively. (C) PG permeabilizes the outer mitochondrial membrane. BAX was immunoprecipitated from PG (50, 100 and 500 nM), OBX (10  $\mu$ M) or DMSO-treated cells for 24 h and then BAK/BAX complex formation was analyzed by immunoblot with anti-BAK and anti-BAX antibodies. doi:10.1371/journal.pone.0057562.g002

1  $\mu$ g/ml pepstatin, 0.1 mM PMSE. Samples were then centrifuged at 13000 $\times$ g for 5 min at 4°C. Mitochondrial fraction (pellet) was isolated, washed once and resuspended with lysis buffer. Total cell lysate, mitochondrial and cytosolic fractions were analyzed by Immunoblot using cytochrome c, porin and actin antibodies (Santa Cruz Biotechnology).

**Computational Methods**

Computational docking was modeled by combining PELE (Protein Energy Landscape Exploration) [21] with Glide [22]. To map protein-ligand conformational changes and induced fit we used our in house program PELE, a Monte Carlo algorithm where new trial configurations are produced with sequential ligand (and protein) perturbation, side chain prediction and minimization steps. Ligand perturbation includes a ligand specific rotamer library. Trial configurations are then filtered with a Metropolis acceptance test, where the energy is described with an all-atom OPLS force field with a surface generalized Born solvent model. PELE has recently shown to provide more accurate induced fit

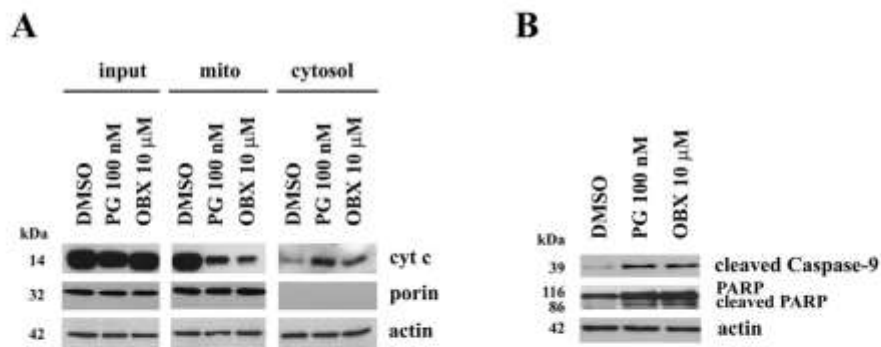
results than the state of the art commercial software [23], and to reproduce the conformational sampling obtained in microsecond molecular dynamics trajectories with two orders of magnitude reduction in computational cost [24].

We have modeled PG and OBX binding to three different anti-apoptotic BCL-2 members for which crystal structures and inhibitors (as controls) are known: MCL-1, BCL-2 and BCL-xL.

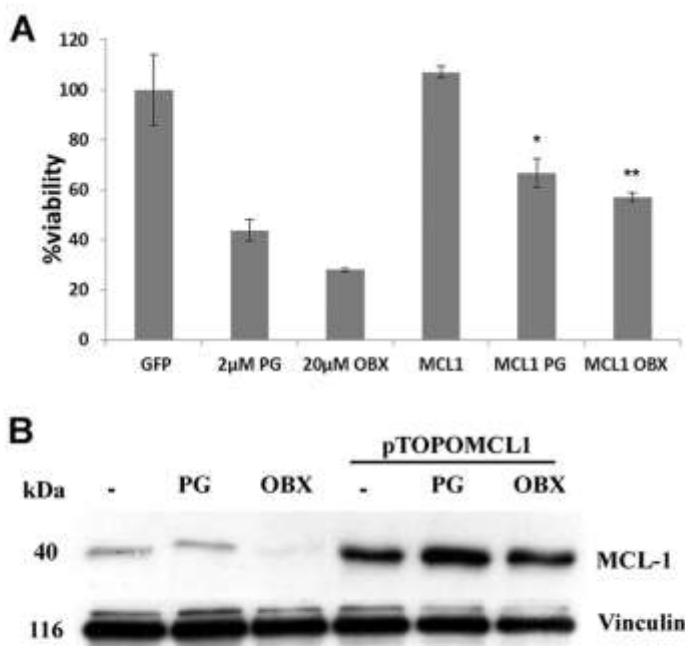
**MCL-1.** To model MCL-1 we used the crystal structure bound to a BH3-peptide, pdb code 2NLA [25]. For the control simulation we used the ligand named 6, a derivative of rhodamine that has an IC50 of 0.25  $\mu$ M [26]. Additionally, OBX has been shown to bind to MCL-1 with an IC50 of 2.9  $\mu$ M [10-13].

**BCL-2.** To model BCL-2 the crystal structure bound to the 43B peptide (BH3 mimic), pdb code 1YSW was used [27]. For the control simulation we used the ligand ABT-737, with an IC50 of 0.12  $\mu$ M [10,28].

**BCL-xL.** To model BCL-xL we used a crystal structure bound to a known inhibitor, ABT-737, with pdb code 2YXJ [28,29]. For



**Figure 3. PG induces activation of the intrinsic apoptotic pathway.** (A) PG induces cytochrome c release to the cytosol. Cytochrome c release from SK-MEL-5 isolated mitochondria after PG and OBX treatment. Mitochondria were isolated from PG (100 nM), OBX (10  $\mu$ M) or DMSO-treated cells for 24 h. Cytochrome c release from mitochondria to cytosol was analyzed by immunoblot using the mitochondrial marker porin as a quality control of the isolation process. (B) Activation of caspases. Cells were treated with PG (100 nM) or OBX (10  $\mu$ M) for 24 h and total cell lysates were analyzed by immunoblot. Actin was used as loading control. doi:10.1371/journal.pone.0057562.g003



**Figure 4. Effect of MCL-1 overexpression on cell viability.** (A) MCL-1 overexpression partially blocks PGs cytotoxicity. 5X-MEL-5 cells were transfected with 1  $\mu$ g of pTOPOMCL1 plasmid and, after 20 h, cells were treated with PGs (2 and 20  $\mu$ M, respectively) for an additional 24 h. Then cell viability was assessed by the MTT assay. (B) Analysis of MCL-1 protein levels. After MCL-1 overexpression and PGs treatment, MCL-1 protein levels were analyzed by immunoblot. Vinculin was used as a loading control. doi:10.1371/journal.pone.0057562.g004

the control simulation we used the ligand present in the initial crystal, ABT-737, with an IC50 of 0.06  $\mu$ M [10,29].

All ligands used in this study are shown in Figure 1.

For all systems we removed the crystallographic ligands and prepared the protein with Schrodinger's Protein Wizard [30]. This algorithm builds hydrogen-bonded clusters and performs 100000 Monte Carlo moves by reorienting hydroxyl and thiol groups, water molecules, amide groups of Asn and Gln, and the imidazole ring in His. The algorithm also predicts protonation states of His, Asp, Glu, Lys and Arg. Each possibility is scored based on the total number of hydrogen bonds and their quality (relative to an idealized hydrogen bond). In particular, all Asp, Glu, Lys and Arg kept their anionic state. Histidines 224, 252 and 277 in MCL-1,

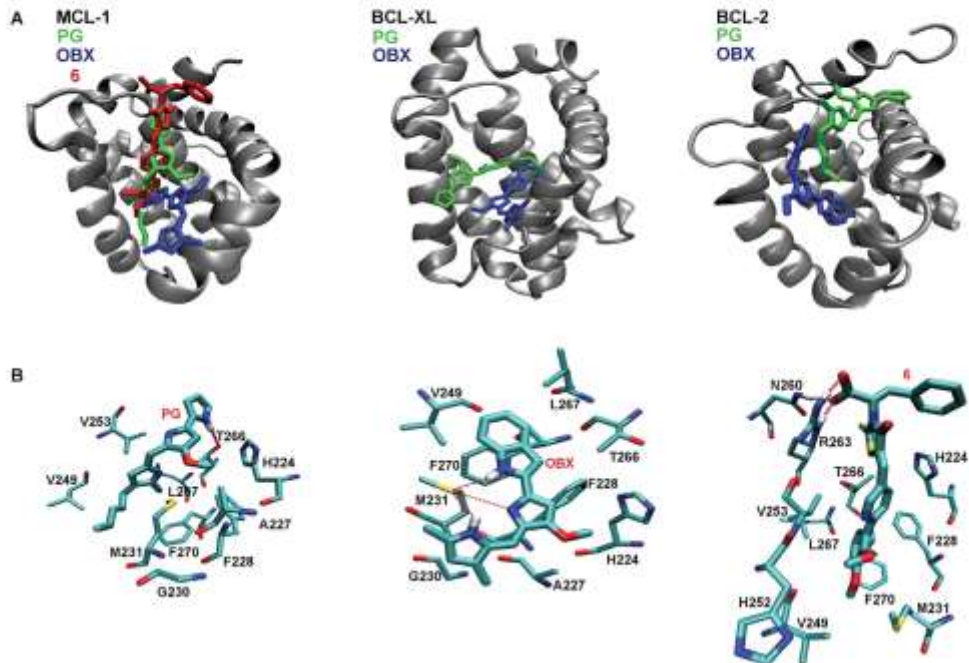
and 117 and 183 in BCL-2 were epsilon protonated; all other histidines kept the default delta protonation. The ligand's atomic charges were derived from the electrostatic potential fitting of a single point DFT/B3LYP calculation with the 6-31G\*\* basis set.

**Ligand docking and induced fit procedure.** After the protein and ligand's preparations, we performed a cavity search with SiteMap [31], which confirmed the BH3 domain as the top ranked binding cavity in the three systems. Initial rigid docking was performed with Glide [22] using the extra precision (XP) scoring function [32], currently viewed as the state of the art docking software/procedure. Correlation of Glide scores with binding affinities, however, can only be done at a qualitative level. While score values below  $-8/-9$  indicate a good binder, a more

**Table 1.** Before and after induced fit docking scores.

Ligands	MCL-1			BCL-xL			BCL-2		
	PG	OBX	6	PG	OBX	ABT	PG	OBX	ABT
Initial Score	-4.3	-2.9	-6.0	-7.0	-3.4	-7.9	-6.3	-6.0	-8.5
Final Score	-8.6	-8.8	-8.7	-7.4	-8.3	-13.9	-7.9	-8.8	-9.6
RMSD	3.0	4.0	7.4	3.0	5.5	1.4	4.8	4.5	4.0

Also shown is the ligand RMSD along the induced fit process. doi:10.1371/journal.pone.0057562.t001



**Figure 5. Induced fit docked structures.** Top panels (A): final induced fit structures in MCL-1 (left) BCL-XL (center) and BCL-2 (right). Bottom panels (B): detailed view of the molecular interactions of PG, OBX and 6 with MCL-1. doi:10.1371/journal.pone.0057562.g005

quantitative assessment requires a system specific control. Thus, for each protein we docked an inhibitor with known (good) binding affinities. Comparing the values predicted for these control ligands with those obtained for PG, we could estimate more accurately their binding strength.

Following the rigid docking we performed 11 independent PELE trajectories, each including 600 iterations (24 hours), of induced fitted adjustment. We then clustered the trajectory in 5 groups, based on the ligand's heavy atom root mean square deviation (RMSD), and selected the median in each group. For each representative cluster we re-docked all ligands with Glide. Thus, for each system and ligand we have two XP score values: an initial one biased to the crystal structure and a final score after the induced fit (the largest score from the 5 clusters), which aims to adapt the protein to each specific ligand.

## Results

### PG Disrupts MCL-1/BAK Complex in Melanoma Cells

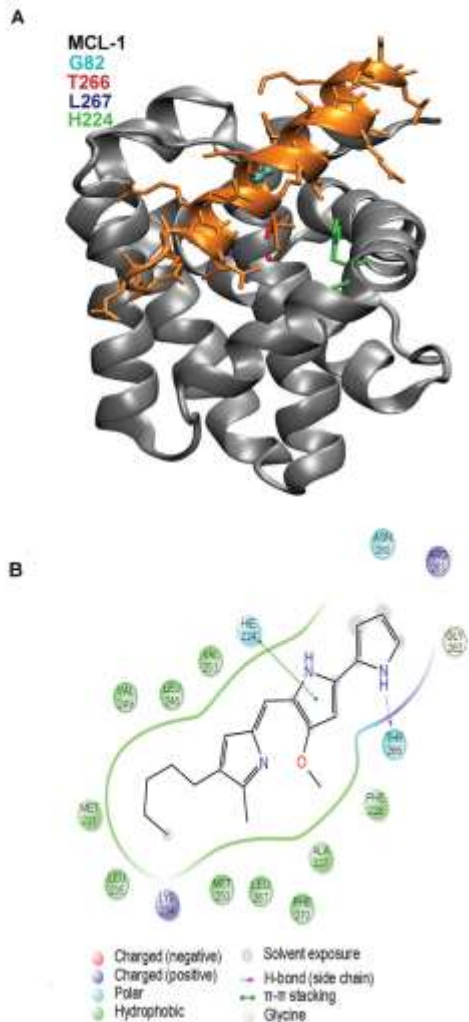
BCL-2 family member MCL-1 binds to and regulates BAK within the mitochondrial outer membrane [33] and the BH3 mimetic molecule OBX has been reported to inhibit this constitutive interaction [34]. To examine whether PG could also alter this binding, MCL-1/BAK complex was co-immunoprecipitated from PG, OBX or DMSO-treated SK-MEL-5 cells. In DMSO-treated cells, MCL-1 was co-immunoprecipitated with BAK using MCL-1 antibody, indicating that these proteins

heterodimerize in basal conditions. Alternatively, treatment with PG or OBX resulted in a complete release of BAK from MCL-1, showing the BH3 mimetic nature of PGs (Figure 2A). As MCL-1 levels were downregulated by 1  $\mu$ M PG, lower doses were used to corroborate these results. Similarly, BAK immunoprecipitation resulted in the appearance of MCL-1/BAK complex in non-treated cells, whilst this binding disappeared in treated cells and no protein level modifications were observed (Figure 2B). Finally, since BAK and BAX heterodimerize when they are released from their anti-apoptotic proteins MCL-1 and BCL-2 (in order to form pores in the mitochondrial membrane), the formation of this complex was analyzed in treated cells. As observed in Figure 2C, after 24 h of treatment, PG as well as OBX induced BAK/BAX complex formation, indicating mitochondrial membrane permeabilization.

### Mitochondrial Apoptosis is Triggered after PGs Treatment

At the same time that MCL-1/BAK complex was disrupted by PGs treatment and the mitochondrial membrane pore was formed, we observed cytochrome c release from the mitochondria to the cytosol (Figure 3A). This protein binds to Apaf-1 and caspase-9 to form the apoptosome and this complex facilitates caspase-9 activation by proteolysis. In Figure 3B we can observe cleaved caspase-9 appearance after PGs treatment, indicating its activation. Moreover, we also observe cleavage of the caspase





**Figure 6. MCL-1 contact map with PG.** (A) Crystal structure of the mNoxaB BH3 peptide bound to MCL-1. In cyan we show Gly82, the 11 helix residue located in the h3 position, Leu267 in blue, Thr266 in red and His224 in green, from MCL-1 are also shown. (B) Pharmacophore analysis of the binding interactions. Residues within 3 Å from the ligand have been included in the analysis.  
doi:10.1371/journal.pone.0057562.g006

substrate called PARP, corroborating caspases activation. Altogether these results demonstrate the activation of the intrinsic apoptotic pathway after PGs treatment.

**MCL-1 Overexpression Partially Abrogates PGs Induced Cell Death**

To elucidate whether MCL-1 was involved in the cytotoxic effect triggered by PGs, MCL-1 was overexpressed in SK-MEL-5 cells. After 20 h from transfection, PGs treatment was added during 24 h and cell viability was analyzed by MTT assay (Figure 4). PGs-induced cytotoxic effect was significantly blocked by overexpressing MCL-1. These results might suggest that PGs are not able to disrupt all the MCL-1/BAK complexes when MCL-1 is overexpressed, preventing some BAK protein to form mitochondrial membrane pores, though being lower the apoptotic effect. Figure 4 B shows MCL-1 protein levels at basal or overexpressing conditions after PGs treatment.

**Molecular Docking**

Table 1 summarizes the docking results for MCL-1, BCL-xL, and BCL-2. Besides from PG and OBX, for each anti-apoptotic target we have docked a control ligand known to be a good binder. For each ligand and target we show the initial docking score, biased towards the initial (crystal) structure, and the final score as a results of the induced fit (after PELE simulation). For MCL-1 the initial docking was performed after removing the BH3 peptide from the crystal structure. Thus, we expect a significant RMSD change and an improvement in the docking score along the induced fit process. Clearly for all ligands we observe a large RMSD, ranging from 3 to 7, and a significant increase in the score. Interestingly, similar scores (—9) are obtained for the control, ligand 6, and for the two prodiginines under study, OBX and PG.

For BCL-xL, the initial crystal structure used to model the target already has the control ligand ABT-737 bound to it. Thus, as expected, we observe the lowest induced fit RMSD for this ligand, 1.4 Å. Additionally, we find good initial and (very good) final scores for ABT-737 a potent inhibitor of BCL-xL from Abbott Laboratories with an IC50 of 0.06 μM [10]. For OBX we observe 5.5 Å RMSD change along the PELE simulation, together with a large improvement of its binding score. For PG, we observe a lower final score and a medium induced fit, pointing to micromolar rather than nanomolar activity.

In BCL-2 our model was derived from a peptide (43B) bound crystallographic structure. Accordingly, we observe again significant induced fit RMSD changes and improvements in the scores. For ABT-737 we observe good initial scores and the lower RMSD, possibly as a result of its large size and excellent BH3 helix mimetic properties. As expected from its IC50 of 0.12 μM [10], the final score is -9.6. OBX is again the second best scorer followed by PG.

Figure 5 shows the induced fit structures obtained after modeling with PELE. The left panel in Figure 5A compares the final structures for OBX, PG and 6 in the BH3 binding domain of MCL-1. The center and right panels show the final structures obtained for OBX and PG in BCL-xL and BCL-2. For simplicity, we superimpose the final position for each ligand into a consensus ribbon representation. In all three proteins, PG and OBX use slightly different regions of the BH3 domain and present different protein-ligand interactions. Figure 5B shows the protein-ligand interactions for PG, OBX and 6 with MCL-1. All three ligands bind to the hydrophobic core defined by Val253, Val249, Ala227 and Leu267. PG forms a hydrogen bond with Thr266. OBX makes a hydrogen bond with Met231 and ligand 6 makes hydrogen bonds with Asn260 and Arg263. Despite sharing a similar molecular skeleton, PG and OBX present important differences in the hydrogen bond network. As seen in Figure 1, a hydrogen bond acceptor ring in PG turns into a hydrogen bond

donor in OBX, explaining the differences in the binding modes. In supporting information we provide a pharmacophore analysis (Supplementary Figure S1) and the atomic coordinates (in pdb format) for the docked structures shown in Figure 5A, the best scoring structures for each ligand and protein.

## Discussion

The proapoptotic agent prodigiosin has shown an average  $IC_{50}$  value of 2.1  $\mu$ M when tested against a collection of ~60 cancerous cell lines [1]. Thus, it seems a good candidate as a base drug to carry on further development. For this, it is necessary to elucidate its target(s) and obtain an atomic detailed description of its binding mechanism. In a previous study we analyzed the interactions of PG and OBX with several kinases, demonstrating that the mammalian target of rapamycin (mTOR) is a molecular target of both PGs in melanoma [14]. Moreover, these results also showed that inhibition of mTOR was accompanied by the activation of both cell death mechanisms, autophagy and apoptosis. Here we show how PG, as indicated previously for OBX [10], binds to the BII3 domain of the anti-apoptotic BCL-2 family, synergizing the proapoptotic effects induced by mTOR inhibition. In particular, in melanoma cells, we demonstrate that PG disrupts the interaction between MCL-1 and BAK, allowing the formation of BAK/BAX complex and the subsequent cytochrome *c* release to the cytosol, which mediates the mitochondrial apoptosis activation. Moreover, MCL-1 has been identified as a molecular target directly involved in PGs induced cell death, since its overexpression is able to decrease PGs cytotoxicity.

In order to characterize the protein-drug complex, we used recent advances in protein-ligand recognition software. It is clear from the results that the induced fit process is essential in order to get good binding affinities. Rigid docking into the crystals, for example, would fail to recognize ligand 6 as a nanomolar binder in MCL-1. Furthermore, the induced fit simulations allow comparing the ABT-737 scores with the experimental  $IC_{50}$  in BCL-xL and BCL-2. Overall, the results with the two prodiginines and the three BCL-2 family members indicate good protein-ligand interactions when compared to control ligands. In MCL-1, in particular, PG scores are similar to the control ligand, suggesting an  $IC_{50}$  in the hundreds of nanomolar range.

PG, OBX and ligand 6 bind in a specific region of the MCL-1 BH3 domain, defined by a hydrophobic core including Val253, Val249, Ala227 and Leu267. This region corresponds closely to the h1–h3 position of the BH3 peptides. Figure 6A shows the interaction of a BH3 helix peptide with MCL-1. Peptide residue 11, located at the h3 position of the helix (G82 in the crystal structure, shown in cyan), is a conserved small residue in contact with the hydrophobic core, which has been shown to be important for ligand binding [35,36]. Additional studies by Chen et al. [15] and Day et al. [37] have also underlined the importance of positions h1 and h3 for binding to MCL-1 and BCL-xL. Computational studies also pointed to this hydrophobic core in the MCL-1 ligand recognition [26,38]. Interestingly, in BCL-xL, the two prodiginines have more distant binding modes, centered in the h1 h2 region, in agreement with the larger degree of flexibility observed in this end of the helix for BCL-xL [35,36].

Figure 6B shows a pharmacophore analysis for the PG/MCL-1 binding interactions (the other ligand and proteins are shown in

supporting information). Clearly, the combination of the hydrophobic interactions (in green) together with the hydrogen bond to Thr266 play a key role in PG's binding. Stacking to His224 is also underlined in the pharmacophore analysis [15]. To further analyze the role of Thr266, we have modeled a single mutation T266A in the active site of MCL-1. The final score, associated with a 4.2Å RMSD, is reduced to -6.3 (instead of a -8.6 for the wild type), indicating the importance of this residue for binding. Obviously, confirmation of these binding modes simulations and residue analysis will require future directed mutagenesis.

The identification of PG as a new BH3 mimetic molecule, together with previous results demonstrating the potential of PG as an inhibitor for both mTOR complexes [14], evidences the potential of PG as a chemotherapeutic agent. In fact, emergent molecular therapies are focused on molecules that are able to target multiple proteins involved in cell survival. Molecules such as PP242 (ATP mimetic inhibitor) [39] or ABT-263 (analog of ABT-737) [40] have shown similar successful results as OBX in clinical trials [11,34]. Nevertheless, combinational strategies are still necessary to improve the effect of these molecules. Based on previous results which markedly enhanced OBX-mediated cell death [12,13], we might consider combining PG with the ER stress inducers such as tunicamycin [41], cisplatin [42] or sorafenib [43,44] which markedly enhanced apoptotic cell death.

Altogether, our results demonstrate, for the first time, that MCL-1 is a molecular target of PG involved in its cytotoxic effect and that this is due to the capacity of PG to displace activating BH3 proteins from the pocket of MCL-1 triggering BAK oligomerization and the subsequent cytochrome *c* release-mediated apoptosis.

## Supporting Information

The coordinates (in pdb format) and pharmacophore analyses are provided for the best scoring structures for each ligand and protein.

## Supporting Information

**Figure S1 Pharmacophore analysis.** Pharmacophore analysis for the binding interactions of all residues with all proteins. Residues within 3 Å from the ligand have been included in the analysis.

(PDF)

**File S1 The coordinates (in pdb format) are provided for the best scoring structures for each ligand and protein.**

(PDB)

## Acknowledgments

The authors thank José Luis Rosa and M. Fátima Lucas for his expert technical advice.

## Author Contributions

Conceived and designed the experiments: VS RP VG. Performed the experiments: AH ME. Analyzed the data: VSR RQ RP VG. Contributed reagents/materials/analysis tools: AH RQ ME. Wrote the paper: AH VSR RP VG.

## References

1. Perez-Tomas R, Vinas M (2010) New Insights on the Antitumoral Properties of Prodiginines. *Current Medicinal Chemistry* 17: 2222–2231.
2. Soto-Cerrato V, Llagostera E, Montaner B, Scheffler GL, Perez-Tomas R (2001) Mitochondria-mediated apoptosis operating irrespective of multidrug resistance

- in breast cancer cells by the anticancer agent prodigiosin. *Biochem Pharmacol* 68: 1345–1352.
- Soto-Cerrato V, Vinals F, Lambert JR, Kelly JA, Perez-Tomas R (2007) Prodigiosin induces the proapoptotic gene NAG-1 via glycogen synthase kinase-3beta activity in human breast cancer cells. *Mol Cancer Ther* 6: 362–369.
  - Ho TF, Pong YT, Chuang SM, Lin SC, Fong BI, et al. (2009) Prodigiosin down-regulates survivin to facilitate paclitaxel sensitization in human breast carcinoma cell lines. *Toxicol Appl Pharmacol* 235: 253–260.
  - Hsieh HY, Shieh JJ, Chen CJ, Pan MY, Yang SY, et al. (2012) Prodigiosin down-regulates SKP2 to induce p27/KIP1 stabilization and antiproliferation in human lung adenocarcinoma cells. *Br J Pharmacol* 166: 2095–2108.
  - Liu GH, Lin SC, Yang SY, Pan MY, Liu YW, et al. (2012) Prodigiosin-induced cytotoxicity involves RAD51 down-regulation through the JNK and p38 MAPK pathways in human breast carcinoma cell lines. *Toxicol Lett* 212: 83–89.
  - Willis SN, Fletcher JJ, Kaufmann T, van Delft MF, Chen L, et al. (2007) Apoptosis Initiated When Bcl-2 Family Members Engage Multiple Bcl-2 Homologs, Not Bax or Bak. *Science* 315: 856–859.
  - Kazi A, Sun J, Doi K, Sung S-S, Takahashi Y, et al. (2011) The BH3  $\alpha$ -Helical Mimic BH3-M6 Disrupts Bcl-XL, Bcl-2, and MCL-1 Protein-Protein Interactions with Bax, Bak, Bcl-1, or Bim and Induces Apoptosis in a Bax- and Bim-dependent Manner. *Journal of Biological Chemistry* 286: 9382–9392.
  - Chongladda TN, Letai A (2008) Mimicking the Bcl-2 domain to kill cancer cells. *Oncogene* 27: S149–S157.
  - Zhai D, Jin C, Satterthwaite AC, Reed JC (2006) Comparison of chemical inhibitors of antiapoptotic Bcl-2-family proteins. *Cell Death Differ* 13: 1419–1421.
  - Parikh SA, Kantarjian H, Schimmer A, Walsh W, Asatiani E, et al. (2010) Phase II Study of Obatoxax Mesylate (GX15 070), a Small-Molecule BCL-2 Family Antagonist, for Patients With Myelofibrosis. *Clinical Lymphoma Myeloma and Leukemia* 10: 285–289.
  - Konopleva M, Watt J, Contractor R, Tsao T, Harris D, et al. (2008) Mechanisms of Antileukemic Activity of the Novel Bcl-2 Homology Domain-3 Mimetic GX15 070 (Obatoxax). *Cancer Research* 68: 3413–3420.
  - Nguyen M, Marcellus RC, Rounston A, Watson M, Serfass L, et al. (2007) Small molecule obatoxax (GX15 070) antagonizes MCL-1 and overcomes MCL-1-mediated resistance to apoptosis. *Proceedings of the National Academy of Sciences* 104: 19512–19517.
  - Espuna-Fiedler M, Soto-Cerrato V, Hosseini A, Lizcano JM, Guallar V, et al. (2012) Identification of dual mTORC1 and mTORC2 inhibitors in melanoma cells: Prodigiosin vs. obatoxax. *Biochemical Pharmacology* 83: 489–496.
  - Chen L, Willis SN, Wei A, Smith BJ, Fletcher JJ, et al. (2005) Differential Targeting of Prosurvival Bcl-2 Proteins by Their BH3-Only Ligands Allows Complementary Apoptotic Function. *Molecular cell* 17: 393–403.
  - Certo M, Moore VDG, Nishino M, Wei G, Kosmeyer S, et al. (2006) Mitochondria primed by death signals determine cellular addiction to antiapoptotic BCL-2 family members. *Cancer cell* 9: 351–363.
  - Ku B, Liang C, Jung JU, Oh B-H (2011) Evidence that inhibition of BAX activation by BCL-2 involves its tight and preferential interaction with the BH3 domain of BAX. *Cell Res* 21: 627–641.
  - Diaz de Grenu B, Hernández PI, Espuna M, Quiñero D, Light ME, et al. (2011) Synthetic Prodigiosin Obatoxax (GX15 070) and Related Analogues: Anion Binding, Transmembrane Transport, and Cytotoxicity Properties. *Chemistry – A European Journal* 17: 14074–14083.
  - Maurer U, Chavet C, Wagnon AS, Dejardin E, Green DR (2006) Glycogen synthase kinase-3 regulates mitochondrial outer membrane permeabilization and apoptosis by destabilization of MCL-1. *Mol Cell* 21: 749–760.
  - Mosmann T (1983) Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays. *J Immunol Methods* 65: 55–63.
  - Borrelli K, Vitalis A, Alcantara R, Guallar V (2005) PELE: Protein Energy Landscape Exploration. A Novel Monte Carlo Based Technique. *Journal of Chemical Theory and Computation* 1: 1304–1311.
  - (2010) Glide, version 5.6 ed. New York, NY: LLC: Schrödinger.
  - Borrelli KW, Cossins B, Guallar V (2010) Exploring hierarchical refinement techniques for induced fit docking with protein and ligand flexibility. *Journal of Computational Chemistry* 31: 1224–1235.
  - Cossins BP, Hosseini A, Guallar V (2012) Exploration of Protein Conformational Change with PELE and Meta-Dynamics. *Journal of Chemical Theory and Computation* 8: 959–965.
  - Czabotar PE, Lee EF, van Delft MF, Day CL, Smith BJ, et al. (2007) Structural insights into the degradation of Mcl-1 induced by BH3 domains. *Proceedings of the National Academy of Sciences* 104: 6217–6222.
  - Bernardo PH, Sivaraman T, Wan K-F, Xu J, Krishnamoorthy J, et al. (2010) Structural Insights into the Design of Small Molecule Inhibitors That Selectively Antagonize Mcl-1. *Journal of Medicinal Chemistry* 53: 2314–2318.
  - Oltersdorf T, Elmore SW, Shoemaker AR, Armstrong RC, Augeri DJ, et al. (2005) An inhibitor of Bcl-2 family proteins induces regression of solid tumours. *Nature* 435: 677–681.
  - van Delft MF, Wei AH, Mason KD, Vandenberg CJ, Chen L, et al. (2006) The BH3 mimetic ABT-737 targets selective Bcl-2 proteins and efficiently induces apoptosis via Bak/Bax if Mcl-1 is neutralized. *Cancer cell* 10: 389–399.
  - Lee EF, Czabotar PE, Smith BJ, Deshayes K, Zobel K, et al. (2007) Crystal structure of ABT-737 complexed with Bcl-2: implications for selectivity of antagonists of the Bcl-2 family. *Cell Death Differ* 14: 1711–1713.
  - (2011) Protein Preparation Wizard. Epic version 2.2, Impact version 5.7, Prime version 3.0 ed. New York, NY: Schrödinger, LLC.
  - (2011) SiteMap, version 2.5 ed. New York, NY: Schrödinger, LLC.
  - Friesner RA, Murphy RB, Repasky MP, Frye LL, Greenwood JR, et al. (2006) Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein–Ligand Complexes. *Journal of Medicinal Chemistry* 49: 6177–6196.
  - Willis SN, Chen L, Dawson G, Wei A, Naik E, et al. (2005) Proapoptotic Bak is sequestered by Mcl-1 and Bcl-xL, but not Bcl-2, until displaced by BH3-only proteins. *Genes & Development* 19: 1291–1305.
  - Trudel S, Li ZH, Raus J, Tiedemann RF, Wei XY, et al. (2007) Preclinical studies of the pan-Bcl inhibitor obatoxax (GX151 070) in multiple myeloma. *Blood* 109: 5430–5438.
  - Boersma MD, Haase HS, Peterson-Kaufman KJ, Lee EF, Clarke OB, et al. (2011) Evaluation of Diverse  $\alpha/\beta$ -Backbone Patterns for Functional  $\alpha$ -Helix Mimicry: Analogues of the Bim BH3 Domain. *Journal of the American Chemical Society* 134: 315–323.
  - Smits C, Czabotar PE, Hinds MG, Day CL (2008) Structural Plasticity Underpins Promiscuous Binding of the Prosurvival Protein Bcl-2. *Structure (London, England : 1993)* 16: 818–829.
  - Day CL, Chen L, Richardson SJ, Harrison PJ, Huang DCS, et al. (2005) Solution Structure of Prosurvival Mcl-1 and Characterization of Its Binding by Proapoptotic BH3-Only Ligands. *Journal of Biological Chemistry* 280: 4738–4744.
  - Acoca S, Cui Q, Shore GC, Purisima EO (2011) Molecular dynamics study of small molecule inhibitors of the Bcl-2 family. *Proteins: Structure, Function, and Bioinformatics* 79: 2624–2636.
  - Feldman ME, Apsel B, Uotila A, Loewith R, Knight ZA, et al. (2009) Active-Site Inhibitors of mTOR Target Rapamycin-Resistant Outputs of mTORC1 and mTORC2. *PLoS Biol* 7: e1000038.
  - Gandhi L, Canidigbe DR, Ribeiro de Oliveira M, Bonomi P, Gandara D, et al. (2011) Phase I Study of Navitoxax (ABI-263), a Novel Bcl-2 Family Inhibitor, in Patients With Small-Cell Lung Cancer and Other Solid Tumors. *Journal of Clinical Oncology* 29: 909–916.
  - Dong L, Jiang CC, Thorne RF, Croft A, Yang F, et al. (2011) Ets-1 mediates upregulation of Mcl-1 downstream of NBP-1 in human melanoma cells upon ER stress. *Oncogene* 30: 3716–3726.
  - Li J, Viallet J, Haura E (2008) A small molecule pan-Bcl-2 family inhibitor, GX15 070, induces apoptosis and enhances cisplatin-induced apoptosis in non-small cell lung cancer cells. *Cancer Chemotherapy and Pharmacology* 61: 525–534.
  - Martin AP, Park MA, Mitchell C, Walker T, Rahmani M, et al. (2009) BCL-2 Family Inhibitors Enhance Histone Deacetylase Inhibitor and Sorafenib Lethality via Autophagy and Overcome Blockade of the Extrinsic Pathway to Facilitate Killing. *Molecular Pharmacology* 76: 327–341.
  - Rahmani M, Aust MM, Atkisson F, Williams DC, Ferreira-Gonzalez A, et al. (2012) Inhibition of Bcl-2 anti-apoptotic members by obatoxax potently enhances sorafenib-induced apoptosis in human myeloid leukemia cells through a Bim-dependent process. *Blo* 119: 6089–6098.

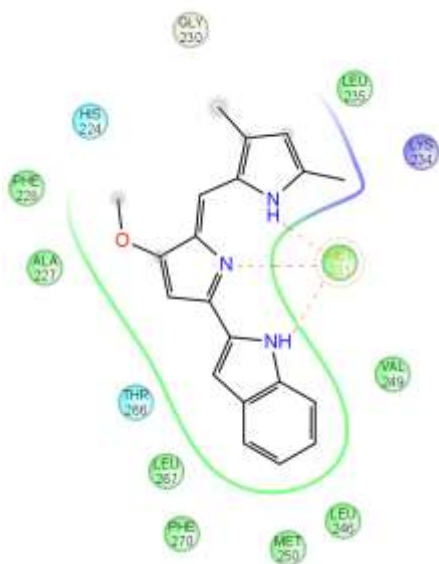
## Supplementary Material

### Molecular Interactions of Prodiginines with the BH3 Domain of Anti-Apoptotic Bcl-2 Family Members

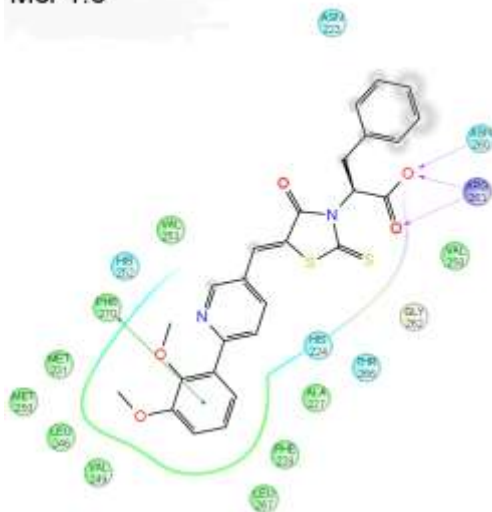
Ali Hosseini, Margarita Espona-Fiedler, Vanessa Soto-Cerrato, Roberto Quesada, Ricardo Pérez-Tomás, Victor Guallar

#### Pharmacophore analysis of the binding interactions

Mcl-1.OBA

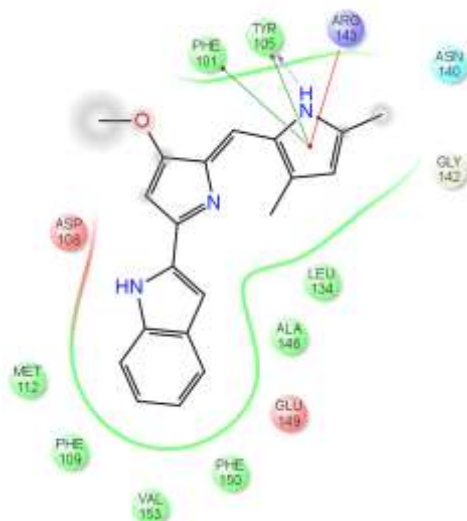


Mcl-1.6

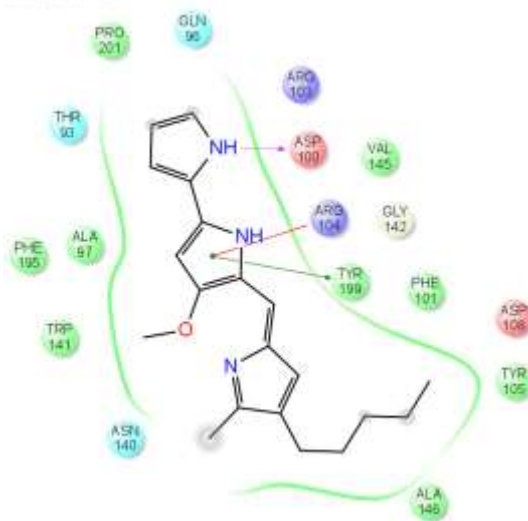


## Publications

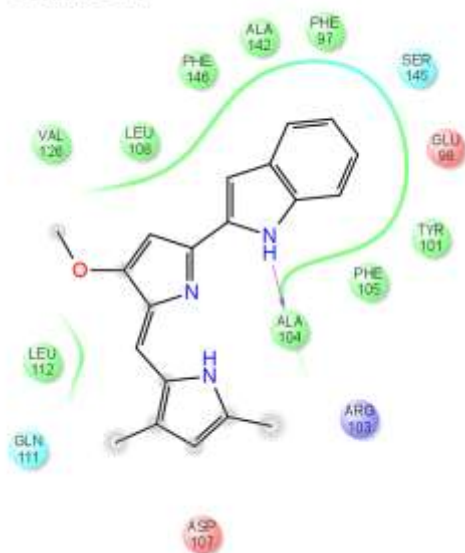
### Bcl-2.OBA



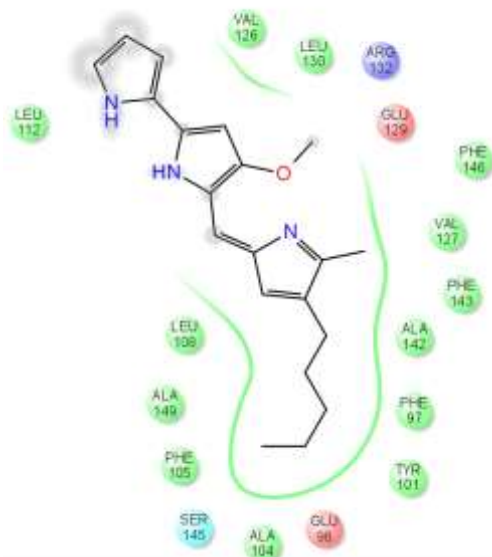
### Bcl-2.PG



### Bcl-xl.OBA



### Bcl-xl.PG



- Charged (negative)
- Charged (positive)
- Polar
- Hydrophobic
- Glycine

- Metal
- H<sub>2</sub>O Water
- Hydration site
- ✗ Displaced hydration site
- ↔  $\pi$ - $\pi$  stacking

- $\pi$ -cation
- H-bond (backbone)
- H-bond (side chain)
- Metal coordination
- Solvent exposure

**Figure S1.**

Pharmacophore analysis. Pharmacophore analysis for the binding interactions of all residues with all proteins. Residues within 3 Å from the ligand have been included in the analysis.

## Publications

Publications

### **3.3 Atomic Picture of Ligand Migration in Toluene 4-Monooxygenase**

Ali Hosseini, Moran Brouk, Maria Fatima Lucas, Fabian Glaser, Ayelet, Fishman, and Victor Guallar

Physical Chemistry B, 2014, DOI: 10.1021/jp502509a



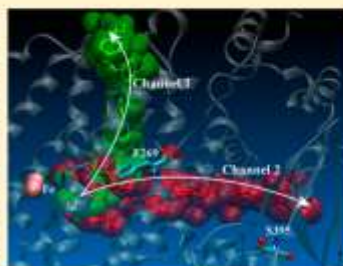
## Publications

## Atomic Picture of Ligand Migration in Toluene 4-Monooxygenase

Ali Hosseini,<sup>†</sup> Moran Brouk,<sup>‡</sup> Maria Fatima Lucas,<sup>†</sup> Fabian Glaser,<sup>§</sup> Ayelet Fishman,<sup>\*;‡</sup> and Victor Guallar<sup>\*;†;||</sup><sup>†</sup>Department of Life Sciences, Barcelona Supercomputing Center, Nexus II Building, 08034 Barcelona, Spain<sup>‡</sup>Department of Biotechnology and Food Engineering, Technion-Israel Institute of Technology, Haifa 32000, Israel<sup>§</sup>Bioinformatics Knowledge Unit, Lorry I. Lokey Interdisciplinary Center for Life Sciences and Engineering, Technion-Israel Institute of Technology, Haifa 32000, Israel<sup>||</sup>Institució Catalana de Recerca i Estudis Avançats, Passeig Lluís Companys 23, E-08010 Barcelona, Spain

## Supporting Information

**ABSTRACT:** Computational modeling combined with mutational and activity assays was used to underline the substrate migration pathways in toluene 4-monooxygenase, a member of the important family of bacterial multicomponent monooxygenases (BMMs). In all structurally defined BMM hydroxylases, several hydrophobic cavities in the  $\alpha$ -subunit map a preserved path from the protein surface to the diiron active site. Our results confirm the presence of two pathways by which different aromatic molecules can enter/escape the active site. While the substrate is observed to enter from both channels, the more hydrophilic product is withdrawn mainly from the shorter channel ending at residues D285 and E214. The long channel ends in the vicinity of S395, whose variants have been seen to affect activity and specificity. These mutational effects are clearly reproduced and rationalized by the *in silico* studies. Furthermore, the combined computational and experimental results highlight the importance of residue F269, which is located at the intersection of the two channels.



## INTRODUCTION

Bacterial multicomponent monooxygenases (BMMs)<sup>1</sup> are a family of proteins containing a nonheme carboxylate-bridged diiron center capable of activating molecular oxygen ( $O_2$ ) for the oxidation of several hydrocarbon substrates. These proteins present a common architecture with three or four components including a multisubunit dimeric hydroxylase component, where the diiron center and catalytic active site are contained.<sup>1,2</sup> Toluene 4-monooxygenase (T4MO) is a soluble four-component BMM that oxidizes toluene with ~95% regioselectivity at the *para* position.<sup>3–5</sup> The reaction occurs in the toluene 4-monooxygenase hydroxylase which contains alpha, beta, and gamma subunits.<sup>6</sup> T4MO is of particular interest in industry given the high number of substrates that can be oxidized along with the elevated specificity.<sup>7</sup>

Different studies of these systems, and related diiron active site enzymes, show that mutagenesis in the active site can change the regioselectivity for reactions with toluene and other nonphysiological substrates.<sup>7–9</sup> Moreover, using directed evolution, the T4MO variant S395C showed different activity and specificity.<sup>4</sup> The production of hydroxytyrosol, a phenol with high antioxidant and anticarcinogenic activities obtained via double hydroxylation of 2-phenylethanol (PEA), showed a 15-fold improvement in the mutated protein in comparison to the wild-type specie. In contrast to other cases, this residue is located 30 Å away from the active site, near the interface of

subunits  $\alpha$  and  $\gamma$ , suggesting that its catalytic influence should be the result of a change in the active site dynamics or in the ligand delivery. Several crystallographic studies of the BMM superfamily proposed a common channel through the  $\alpha$  subunit connecting the diiron center to the surface; this pathway involves surface residues D285 and E214 in T4MO.<sup>10–12</sup> Additional reports described two other hydrophobic cavities, one near the active site pocket and another which is located near the interface of the  $\alpha$  and  $\gamma$  subunits.<sup>1,6,13,14</sup> Measurements of oxygen migration by the formation rates of a peroxodiiron(III) intermediate suggested that the two cavities may connect via movement of the protein components thus regulating oxidation rate.<sup>13</sup> This was further supported by the findings describing changes in the volume of the pockets upon binding of the effector protein to the hydroxylase, thus restricting free access from the solvent to the active site.<sup>16</sup>

To understand the atomic detail of ligand migration in T4MO as well as the involvement of S395 in the catalytic rate enhancement, we have modeled the ligand migration pathways with all-atom computational techniques. Recently, due to its

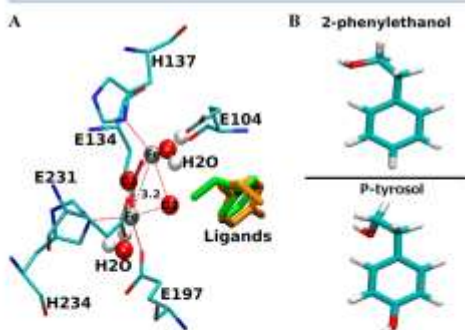
Special Issue: William L. Jorgensen Festschrift

Received: March 12, 2014

Revised: May 5, 2014

implications in drug design and enzyme engineering there has been an increasing interest in applying molecular dynamics techniques for mapping ligand diffusion, entrance and exit.<sup>25,16</sup> To this aim, many biased approaches, such as steered molecular dynamics<sup>17</sup> or metadynamics,<sup>18</sup> have been applied. Using special purpose machines or graphical processor units, a nonbiased search accessing microsecond time scale simulations is also possible.<sup>19,20</sup> These computational approaches, however, still represent a significant computational cost (out of the reach of a typical lab) when dealing with complex systems such as T4MO. Our Monte Carlo approach PELE (Protein Energy Landscape Exploration) allows for an unbiased search of the protein–ligand dynamics, which includes mapping the exit and entrance pathways at an accessible computational cost.<sup>21–23</sup>

Along the simulations, two different substrates were evaluated: the reactant PEA and the first hydroxylation product, *p*-tyrosol (*p*-tyr) (see Figure 1B). Our results confirm the



**Figure 1.** Panel A: T4MO active site view. Panel B: the two substrates (reactant and product) used in this work.

presence of two pathways, previously observed,<sup>1,6,13,14</sup> by which aromatic molecules can enter/escape the active site. While the substrate enters from both channels, the more hydrophilic products are withdrawn mainly from the shorter channel ending at residues D285 and E214. Furthermore, computational studies were complemented with mutational and activity assays. The combined results highlight the importance of residue F269, which is located at the intersection of the two channels.

## METHODS AND MATERIALS

**Computational Work.** The starting coordinates were taken from the Protein Data Bank (PDB) with entry 3DHG.<sup>1</sup> The protein preparation wizard algorithm<sup>24</sup> implemented in the Schrödinger software package was employed to prepare the initial structure. The algorithm predicts protonation states of histidines, aspartic acid, and glutamic acid and optimizes the hydrogen bond network. In this work, all charged residues (at pH 7) kept their initial state. Histidines 4, 36, 37, 137, 163, 174, 247, 329, 332, and 484 were epsilon protonated, and 328, 421, and 447 were double protonated; all other histidine residues kept their default delta protonation state. Two water molecules, generally found in the active site of T4MO, one in the first coordination sphere of an iron atom and the other in the active site pocket, were kept in the system.

Since our main interest is the ligand migration pathways, and due to the lack of accurate details in the T4MO reaction mechanism, we modeled the diiron center (directly from the crystal) as Fe<sub>2</sub>O<sub>2</sub> (Figure 1). Charges were obtained from the OPLS2005 force field where, in this case, we partly reduce the polarization in the active site by modifying the Fe and O charges to 2+ and 1−, respectively. All distances in the iron coordination sphere were constrained to their initial crystal value. Likewise, the Fe–Fe distance was maintained at 3.2 Å. As seen in Figure 1A, reactant and product molecules were placed in a similar position in the active site. This assumption is based on the high degree of complementarity of the T4MO active site that shows that toluene and other aromatic ring substrates bind in a similar fashion.<sup>14,25</sup>

Ligand diffusion was studied with our in-house program PELE.<sup>22,26</sup> This technique is based on a Monte Carlo algorithm, where new trial configurations are produced with sequential ligand and protein perturbation, side chain prediction, and minimization steps. Trial configurations are then filtered with a Metropolis acceptance test, where the energy is described with an all-atom OPLS force field<sup>27</sup> with a surface-generalized Born solvent model.<sup>28</sup> PELE has recently been shown to provide more accurate induced fit results than state of the art commercial software and to reproduce the conformational sampling obtained in microsecond molecular dynamics trajectories with a two-order reduction in computational cost.<sup>23</sup>

Two different sets of simulations, describing the ligand exit and entrance, were performed for the wild-type and the S395C, F269V, and F269W mutants.

**Ligand Exit.** A total of 10 independent runs were produced for each simulation, where the ligand initially placed in the active site, after Glide<sup>29</sup> docking, explores randomly possible exit paths. Each run is based on 12 trajectories (running on different cores) that search collectively for an exit route. The collective search uses the distance from the ligand center of mass to the distance of the ligand to its active site position to couple the different trajectory search. In particular, if a processor is 3 Å behind the leading processor (the one with the furthest distance to the point), it will abandon its current coordinates and receive the leading ones. Thus, there is no bias in the direction that the ligand should take. All 10 runs were interrupted after 24 h of CPU or when (if) the ligand reached a solvent-accessible surface area larger than 0.4; i.e., the ligand enters the solvent.

**Ligand Entrance.** Studies were also done for the migration of the substrate from the surface of the protein to the active site for both channels identified in the exit exploration. In this case the ligands were placed at the solvent and in the vicinity of the exit point as described in the exit paths. Again, we performed 10 independent runs with 12 trajectories in each run, where the trajectories search collectively to move toward the active site (with a 3 Å distance to the distance of the ligand to its active site position).

Fpocket<sup>30,31</sup> was employed to measure the cavity size and their hydrophobic nature. This method uses a Monte Carlo approach to find a path where a sphere of variable radius can pass. Moreover, Fpocket also allows pocket detection on a large set of structures forming a trajectory.

## EXPERIMENTAL WORK

**Chemicals.** 2-Phenylethanol (PEA) and *o*-, *m*-, and *p*-tyrosol were purchased from Sigma-Aldrich Chemical Co.

(Sigma-Aldrich, Rehovot, Israel). Hydroxytyrosol was obtained from Cayman Chemical Co. (MI, USA). All standards were prepared as stock solutions in ethanol. All materials used were of the highest purity available and were used without further purification.

**Bacterial Strains and Growth Conditions.** *Escherichia coli* TG1 (*supE hsdΔ5 thi Δ(lac-proAB) F' [traD36 proAB<sup>+</sup> lacI<sup>q</sup> lacZΔM15]*) with the plasmid constructs was routinely cultivated at 37 °C in Luria–Bertani (LB) medium<sup>52</sup> supplemented with kanamycin at 100 μg/mL to maintain the plasmids. To stably and constitutively express the toluene monooxygenase genes from the same promoter, the expression vector pBS(Kan)T4MO (henceforth T4MO) was constructed as described earlier.<sup>33,34</sup> All experiments were conducted as described previously.<sup>4,33,35</sup> Shortly, overnight cells were diluted to an optical density (OD) at 600 nm of 0.1 and grown to an OD of 1.3. The exponentially grown cells were centrifuged (8000g for 10 min at 25 °C) and resuspended in potassium phosphate buffer (100 mM, pH 7.0). Cells were subsequently used in biotransformation protocols as described later on.

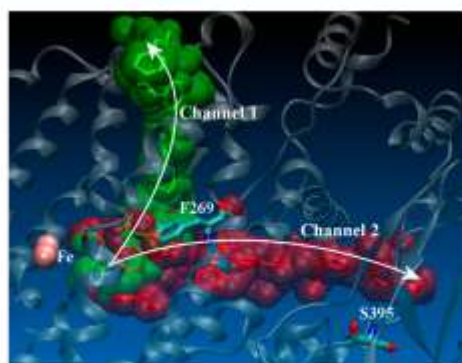
**Construction of T4MO Mutants.** Variant S395C was obtained through random mutagenesis as described by Brouk and Fishman.<sup>4</sup> Variants F269V and F269W were constructed using site-directed mutagenesis at the T4MO *tnoA* gene via overlap extension PCR.<sup>35</sup> Briefly, three oligonucleotide primer pairs were designed (Supporting Information, Table 1) to generate the desired mutations. For generating the F269V and F269W mutations, the first mutated PCR fragment was amplified using primers T4MO-befEcoRI-front and T4MO\_269\_V\_rear and T4MO-befEcoRI-front and T4MO\_269W\_rear, respectively. The second mutated PCR fragment was amplified using T4MO\_269\_V\_front and T4MO-ABRear and T4MO\_269W\_front and T4MO-ABRear, respectively. The PCR program consisted of an initial denaturation at 94 °C for 2 min, followed by 25 cycles of 94 °C for 45 s, 55 °C for 45 s, and 72 °C for 2.2 min, with a final extension at 72 °C for 8 min. The two fragments were combined during the final reassembly PCR in a 1:1 molar ratio using the outer primers T4MObefEcoRI Front and T4MOABRear. The assembling PCR was programmed similarly to the above PCR program, with extension at 72 °C for 3.15 min instead of 2.2 min. The assembled PCR fragment was ligated into WT T4MO, after the double digestion of both vector and insert with EcoRI and AatII, replacing the corresponding fragment in the original plasmid. The resulting plasmid library was electroporated into *E. coli* TG1 cells. Plasmid DNA was isolated using a Mini Kit (Qiagen, CA, USA), and verification of the mutations was done by sequencing with primer T4MO seq 1.

**Whole-Cell Enzymatic Biotransformations.** Whole-cell activity assays were performed as described previously.<sup>4,35</sup> The biotransformation was carried out in screw-capped 16 mL glass vials containing 2 mL of cells and 0.25 mM substrate (added from a 100 mM stock solution in ethanol). All the vials were shaken at 600 rpm (Vibramax 100, Heidolph, Nuremberg, Germany) at 30 °C. The reaction was stopped periodically (a vial was sacrificed) by filtration of the cells and analysis by HPLC. The negative control used in these experiments was TG1/pBS(Kan) (a plasmid without the monooxygenase). The initial transformation rates were determined by sampling at 3–20 min intervals during the first 2–5 h. The specific activity (nmol/min/mg protein) was calculated as the ratio of the initial transformation rate and the total protein content, 0.24 [mg

protein/mL/OD<sub>600 nm</sub>]<sup>9,36</sup>. Activity data reported in this paper are based on at least three independent results. Analytical methods were described previously.<sup>4,35</sup>

## RESULTS

**Ligand Migration in the Wild Type.** We started our computational study by searching possible migration pathways connecting the active site and the protein surface. For this purpose, we placed both the reactant (PEA) and the first hydroxylation product (p-tyr) at the active site cavity and modeled their exit using PELE. As seen in Figure 2, we found

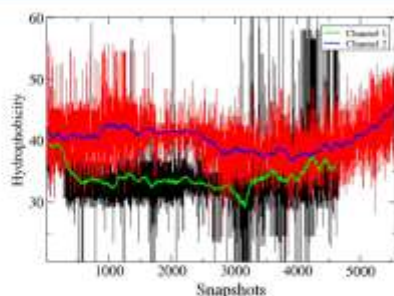


**Figure 2.** Two observed migration paths obtained with the PELE simulations. Channel 1 is shown in green, while channel 2 is shown in red.

two different channels crossing T4MO's  $\alpha$ -subunit and adopting a nearly 90° disposition. One channel (hereafter known as channel 1) is a traverse of about 23 Å to the protein's surface providing the shortest path between the active site and solvent. This channel is formed mainly by H96, F269, Q204, D211, E214, and D285. The second route, channel 2, is significantly longer, 35 Å, and presents a larger hydrophobic nature: F269, P390, V335, W167, W338, and S395. As seen in Figure 2, both channels share the active site pathway section and bifurcate around F269.

From the 10 simulations performed for the reactants, we observe a similar number of migrations in either pathway (40% by channel 1 and 60% by channel 2). In the case of the products, however, we see that 90% of the runs leave by channel 1, while only 10% follow channel 2 (see Table 2). As mentioned above, visual inspection of both channels indicates the presence of more polar residues in channel 1. This might be the main reason why the products prefer this exit pathway; the product presents an extra hydroxyl group in the para position. To quantify better the polar nature of each pathway we used Fpocket,<sup>30,31</sup> which allows us to compute the hydrophobicity of both channels. As seen in Figure 3, channel 1 is significantly less hydrophobic than channel 2.

**Role of S395.** Figure 2 indicates the location of S395, in close vicinity to the solvent exit point in channel 2. As mentioned previously, variant S395C, obtained by directed evolution, shows a 15-fold improvement in PEA oxidation in comparison to the wild-type specie.<sup>4</sup> Performing saturation mutagenesis at position 395 revealed three additional variants



**Figure 3.** Comparison of channel 1 and channel 2 hydrophobicities for the different snapshots produced by PELE along the exit pathways.

with improved activity on PEA. The product distribution obtained from PEA oxidation displayed a decrease in the regioselectivity by the variants and ability to form *o*-tyrosol which wild-type cannot (results not shown). Thus, residue S395, which is distant from the active site, has a strong influence on both activity and selectivity of T4MO.

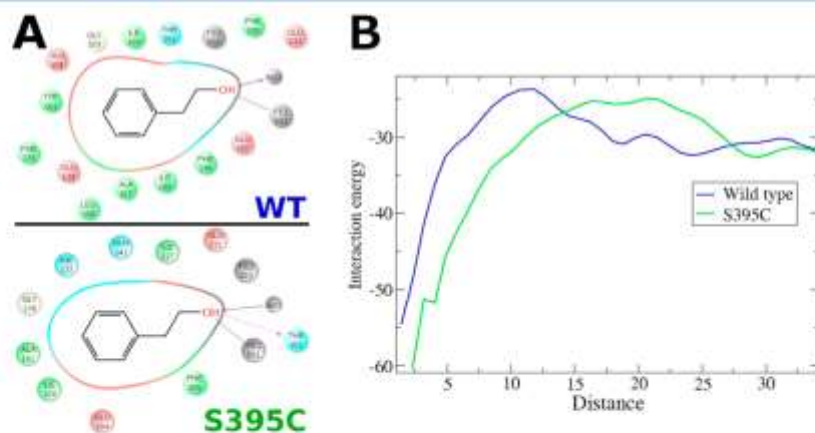
To understand the possible influence of the S395C mutation on ligand delivery, we computed the average protein–substrate interaction energy along channel 2 for the wild-type and the mutant. Figure 4 shows both interaction energy plots, presenting clear binding profiles for both systems, with lower energies at the active site distance ( $\sim 5$  Å). Distances to the active site were measured using the initial (Glide) docked ligand center of mass as a reference. The main differences between the wild-type and the mutant, however, are observed precisely at the active site and at the 15–25 Å distance segment from the active site, where residues W167 and W338 are located. These two residues play an important role as gatekeepers in ligand egression through channel 2. Replacement of the serine by a cysteine reduces the interaction of the

ligand with the protein (less favorable interaction) at this 15–25 Å segment, facilitating in this way its passage through the channel. At the active site, however, we observe the opposite effect, an increase in interaction energy for the mutated species. Interestingly, the minimum at the active site is shifted in the mutant, pointing to an increase in volume. Panel A in Figure 4 compares the main contacts of the protein with the substrate at the active site, where we observe a slight increase in the number of interactions in the mutant.

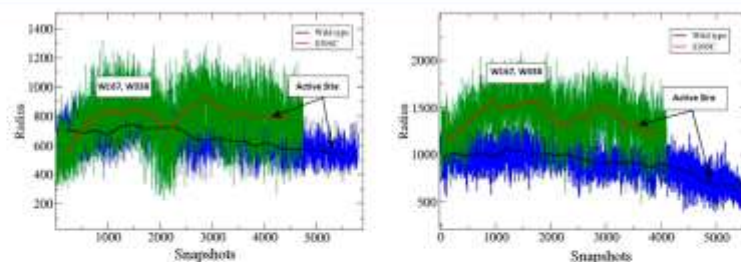
PELE uses a random walk approach, producing hundreds of new conformations that are accepted or rejected based on their energy change (following a Metropolis algorithm). Under the same conditions, the number of steps required to accomplish a task indicates qualitatively its “level of difficulty”. Thus, for example, by comparing the number of steps required by different ligands, or by introducing a mutation in the protein, we can obtain an approximation to the system dynamics. In the case of the wild type the substrate reached the active site in 700 steps, while for S395C only 500 steps are needed (average numbers for the 10 runs). This translates into a significantly faster migration for S395C when compared to the wild-type systems, in agreement with the lower protein–substrate interaction energy observed in Figure 4.

To complete the analyses we have inspected the differences in the cavity radius along the entrance in channel 2. Figure 5 shows the radius for two different runs, where we combined all 12 trajectories for each run. As mentioned above, the number of steps to reach the active site is clearly lower in the mutant. Moreover, and in agreement with the interaction energy and the quicker substrate migration, the pathway for the mutant presents a larger cavity radius. The main variation occurs close to positions W167/W338 and at the entrance to the active site, the regions where we observed the larger interaction energy differences. Thus, the Fpocket analysis confirms an increase in volume for the mutant at the active site.

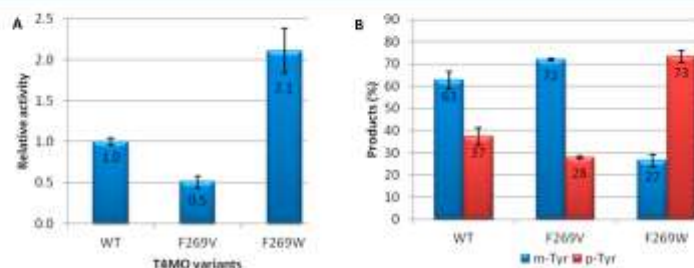
**Role of F269.** Due to the strategic position of F269 at the bifurcation point observed in the modeled pathways, we



**Figure 4.** (A) Different contacts observed for the reactant in the active site of wild-type (WT) and variant S395C. (B) Interaction energies versus distance for the substrate entrance along channel 2. The distances, in Å, correspond to the distance of the ligand to its active site position. Energies are in kcal/mol.



**Figure 5.** Cavity radius in two different runs for all snapshots along the channel 2 entrance as viewed for wild-type and variant S395C. The presence of the ligand (along the snapshots) in the active site and next to W167/W338 is underlined with boxes.



**Figure 6.** Relative activity on PEA (A) of WT and T4MO F269 variants and the product distribution after 24 h (B). The activity was determined via HPLC analysis over a 5 h time period, and regioselectivity was determined via GCMS analysis after 24 h, with an initial PEA concentration of 0.25 mM.

**Table 1.** Initial PEA Oxidation Rate and Product Formation Rates of WT T4MO and the F269 Variants

rate ( $\mu\text{mol}/\text{min}/\text{mg}$ proteins)	T4MO variant		
	WT	F269V	F269W
PEA oxidation	$0.062 \pm 0.002$	$0.131 \pm 0.017$	$0.032 \pm 0.004$
m-tyr formation	$0.141 \pm 0.008$	$0.060 \pm 0.0007$	$0.010 \pm 0.001$
p-tyr formation	$0.120 \pm 0.002$	$0.223 \pm 0.003$	ND <sup>a</sup>

<sup>a</sup>ND: The p-tyr formation rate by the F269V could not be calculated due to its low activity.

performed the following *in vitro* and *in silico* mutations: F269V and F269W. We substituted position F269 to valine and tryptophan with the expectation that the smaller residue would allow better and faster discharge of the product, whereas the larger residue would slow down catalysis. In fact, the experimental results show quite the opposite (Figure 6A). The F269V mutant was very slow and quite inactive. The product distribution was similar to wild-type (Figure 6B). However, the F269W variant was more active than WT (2.1-fold), and the regioselectivity changed. The enzyme favors *p*-tyrosol rather than *m*-tyrosol. This latter fact is unique since all the mutants we have generated until now (probably tens of different variants)<sup>4,35,37</sup> were always *pro-meta* except for position 395 which also influenced the regioselectivity. The change in selectivity is the result of different formation rates of the two products (Table 1).

To further understand the influence of residue 269, we performed *in silico* analysis of PEA and *p*-tyr movement after placing them in the active site. As observed for the wild-type protein, for both F269 mutants we find again channel 1 and channel 2 as the only possible migration pathways. Moreover,

as observed previously, the products have a strong preference for channel 1. Interestingly, however, the preferred channel for PEA's migration changes with the mutation at residue 269. In particular, our calculations show that upon mutation of phenylalanine 269 to tryptophan the main exit pathway shifts from an almost even ratio to exclusively channel 1. A summary of all results is depicted in Table 2.

Besides shifting the ratio of migration pathways, we find that the mutation at position 269 strongly affects the dynamics of both ligands in channel 1. Table 3 lists the number of

**Table 2.** Distribution of Ligand Exit by the Two Channels for the Reactant and Product Species<sup>a</sup>

	PEA		<i>p</i> -tyr	
	channel 1	channel 2	channel 1	channel 2
wild-type	40	60	90	10
F269V	50	50	70	30
F269W	100	0	90	10

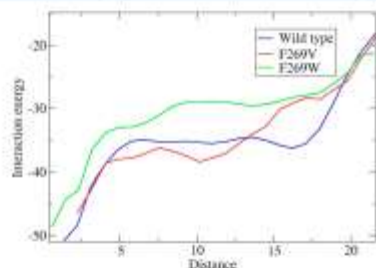
<sup>a</sup>Values are in %.

**Table 3.** Average Steps Required for the Migration of the Ligands through Channel 1

	PEA	<i>p</i> -tyr
wild-type	175 ± 21.21	130 ± 22.52
F269V	210 ± 24.08	150 ± 15.81
F269W	100 ± 30.33	75 ± 11.18

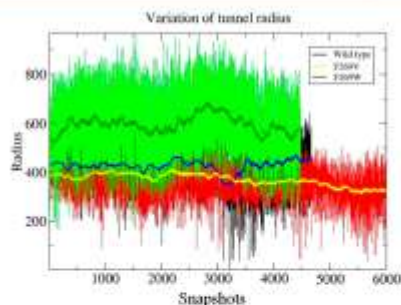
simulation steps needed to displace the ligand from the active site to the protein's surface for all species and ligands along channel 1. Clearly the product requires significantly less steps, which correlates with its larger polarity, capable of creating new contacts along the polar cavity. The most interesting result, however, is how mutation at F269 influences the ligand's diffusion. In the case of the wild-type the reactant molecule finds the solvent in an average of 175 steps, while for F269V 210 steps are needed. Migration was particularly fast (~100 steps) for the F269W when compared to the other two systems.

The difficulty in ligand migration shown by PELE agrees nicely with the experimental relative activities. To gain more insight into the substrate diffusion toward the active site, we modeled PEA's entrance along channel 1 for the wild-type and the two mutants. Figure 7 shows the interaction energy along

**Figure 7.** Interaction energies versus the distance for the substrate entrance through channel 1 for wild-type and F269 variants. The distances, in Å, correspond to the distance of the ligand to its active site position. Energies are in kcal/mol.

the entrance process, where we observe considerable differences among the different systems. Placing a tryptophan residue results in substantial lower interaction energy, facilitating the migration of the reactants from the solvent into the catalytic active site. For the valine mutant, however, we see a slight increase in the interaction energy, which explains the increase in simulation time required to complete the migration.

Finally, Figure 8 shows the changes in the radius for the entrance along channel 1 for one run (other runs give analogous results). While the volume of the tunnel is largely increased for the tryptophan mutant, changing the phenylalanine to a valine slightly decreases the radius along the migration cavity. Thus, contrary to our initial intuition, inserting a tryptophan at the "bifurcation" position 269 resulted in a substantial increase of the volume along channel 1, easier migration of the substrate to the active site, and overall increase of the activity.

**Figure 8.** Cavity radius for the wild-type and the F269 variants along the entrance simulations in channel 1.

## DISCUSSION

Our dynamic simulation experiments confirm the presence of two important passages for ligand entry/exit to the active site of T4MO. One begins in the active site pocket and extends to the protein surface close to E214 and E285 (channel 1), which was described previously by Sazinsky et al.<sup>11</sup> for the analogous ToMO. The second (channel 2) is a longer passage connecting the active site to the protein surface by passing W167 and exiting close to S395 and is in good agreement with recent experimental data.<sup>1,14</sup> While Song et al. suggested that this latter channel is the major pathway by which dioxygen enters the active site,<sup>11</sup> our experimental and computational work suggest that it is also a means for substrate entrance/exit. The differences in hydrophobicity of these two channels explain the preference of the more hydrophilic product to exit by channel 1, the less hydrophobic among the two. PELE simulations also revealed the involvement of W167 and W338 in the restriction of aromatic molecules within the enzyme. The presence of cysteine rather than serine at the entrance of channel 2 in the variant caused an increase in the radius of the channel precisely around W167 and W337 (Figure 5), and this was corroborated with less favorable interactions with the substrate (Figure 4). Faster rates of reaction with PEA may be possible, as well as different orientation and positioning of the substrate in the active site, leading to the differences observed in regioselectivity. In particular, an increase in the volume at the active site of the mutant could enable several options for orientation of the substrate thus changing the regioselectivity. These two residues are conserved in ToMO, and indeed a site-specific mutation in this enzyme, W167E, enabled a 3-fold increase in oxygen transfer rate.<sup>14</sup> The crystal structure of the wild-type and the variant revealed that the indole ring of W167 obstructs the channel, while the glutamic acid residue formed hydrogen bonds with adjacent residues causing an opening of the channel.<sup>14</sup>

While in the previous case Monte Carlo simulations were used to explain the experimental results observed for S395C, the opposite scenario occurred with residue F269. This position, conserved in both T4MO and ToMO, has not been reported in the literature. However, visual inspection of the T4MO's active site revealed that residue F269 is positioned in the intersection of the two channels. This finding motivated further experimental studies, and the mutants F269V and F269W were produced. Interestingly, modification to the smaller valine and the larger tryptophan revealed apparent

contradictory results. While we expected a correlation between the size of this residue and the difficulties in migration through the channel, the opposite was seen. To rationalize these findings additional PELE simulations were performed for the mutated proteins. The results revealed local and propagated conformational changes in the protein's structure which explain the experimental results. The tryptophan at position 269 increased the radius of channel 1 rather than reduced it, leading to better access to and from the active site by this pathway and possible different orientation within the pocket which may explain the changes in regioselectivity of variant F269W. Determination of the crystal structure of this variant could substantiate this assumption.

It is concluded that mutations in distant locations in the protein may alter the activity and selectivity of nonheme monoxygenases. The combination of experimental and computational approaches facilitates understanding of structure–function correlations and assists in designing improved biocatalysts.

## ■ ASSOCIATED CONTENT

### Supporting Information

Further information on the preparation of variants F269V and F269W. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## ■ AUTHOR INFORMATION

### Corresponding Authors

\*E-mail: [victor.guallar@bsc.es](mailto:victor.guallar@bsc.es)

\*E-mail: [afishman@tx.technion.ac.il](mailto:afishman@tx.technion.ac.il)

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This work was supported by grants from the European Research Council—2009-Adg25027-PELE European project, the Spanish Ministry “Juan de la Cierva” for M.F.L. The Minna Kroll Memorial Research Fund is gratefully acknowledged for supporting the work performed at the Fishman lab (grant 2018004).

## ■ REFERENCES

- (1) Bailey, L. J.; McCoy, J. G.; Phillips, G. N.; Fox, B. G. Structural Consequences of Effector Protein Complex Formation in a Diron Hydroxylase. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105* (49), 19194–19198.
- (2) Lindqvist, Y.; Huang, W.; Schneider, G.; Shanklin, J. Crystal Structure of Delta9 Stearoyl-acyl Carrier Protein Desaturase from Castor Seed and Its Relationship to Other Di-iron Proteins. *EMBO J.* **1996**, *15* (16), 4081–4092.
- (3) Loustos, G. T.; Mitchell, K. H.; Studts, J. M.; Fox, B. G.; Oeville, A. M. Crystal Structures and Functional Studies of T4moD, the Toluene 4-Monooxygenase Catalytic Effector Protein. *Biochemistry* **2005**, *44* (19), 7131–7142.
- (4) Brook, M.; Fishman, A. Protein Engineering of Toluene Monoxygenases for Synthesis of Hydroxytyrosol. *Food Chem.* **2009**, *116* (1), 114–121.
- (5) Fishman, A.; Tao, Y.; Rui, L. Y.; Wood, T. K. Controlling the Regioselective Oxidation of Aromatics via Active Site Engineering of Toluene Para-Monooxygenase of *Ralstonia Pickettii* PRO1. *J. Biol. Chem.* **2005**, *280* (1), 506–514.
- (6) McCormick, M. S.; Sazinsky, M. H.; Condon, K. L.; Lippard, S. J. X-ray Crystal Structures of Manganese(II)-Reconstituted and Native Toluene/o-Xylene Monoxygenase Hydroxylase Reveal Rotamer Shifts in Conserved Residues and an Enhanced View of the Protein Interior. *J. Am. Chem. Soc.* **2006**, *128* (47), 15108–15110.
- (7) Notoomista, E.; Scognamiglio, R.; Troncone, L.; Donadio, G.; Pezzella, A.; Di Donato, A.; Izzo, V. Tuning the Specificity of the Recombinant Multicomponent Toluene o-Xylene Monoxygenase from *Pseudomonas* sp. Strain OX1 for the Biosynthesis of Tyrosol from 2-Phenylethanol. *Appl. Environ. Microbiol.* **2011**, *77* (15), 5428–5437.
- (8) Pikus, J. D.; Studts, J. M.; McClay, K.; Steffan, R. J.; Fox, B. G. Changes in the Regioselectivity of Aromatic Hydroxylation Produced by Active Site Engineering in the Diron Enzyme Toluene 4-Monooxygenase. *Biochemistry* **1997**, *36* (31), 9283–9289.
- (9) Vardar, G.; Wood, T. K. Protein Engineering of Toluene-o-Xylene Monoxygenase from *Pseudomonas stutzeri* OX1 for Synthesizing 4-Methylresorcinol, Methylhydroquinone, and Pyrogallol. *Appl. Environ. Microbiol.* **2004**, *70* (6), 3253–3262.
- (10) Vardar, G.; Wood, T. K. Alpha-subunit Positions Methionine 180 and Glutamate 214 of *Pseudomonas stutzeri* OX1 Toluene-o-Xylene Monoxygenase Influence Catalysis. *J. Bacteriol.* **2005**, *187* (4), 1511–1514.
- (11) Sazinsky, M. H.; Bard, J.; Di Donato, A.; Lippard, S. J. Crystal Structure of the Toluene/o-Xylene Monoxygenase Hydroxylase from *Pseudomonas stutzeri* OX1: Insight Into the Substrate Specificity, Substrate Channeling, and Active Site Tuning of Multicomponent Monoxygenases. *J. Biol. Chem.* **2004**, *279* (29), 30600–30610.
- (12) Vardar, G.; Wood, T. K. Protein Engineering of Toluene-o-Xylene Monoxygenase from *Pseudomonas stutzeri* OX1 for Enhanced Chlorinated Ethene Degradation and o-Xylene Oxidation. *Appl. Microbiol. Biotechnol.* **2005**, *68* (4), 510–517.
- (13) Song, W. J.; Gucinski, G.; Sazinsky, M. H.; Lippard, S. J. Tracking a Defined Route for O<sub>2</sub> Migration in a Dioxygen-activating Diron Enzyme. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108* (36), 14795–14800.
- (14) Bailey, L. J.; Acheson, J. F.; McCoy, J. G.; Elsen, N. L.; Phillips, G. N.; Fox, B. G. Crystallographic Analysis of Active Site Contributions to Regioselectivity in the Diron Enzyme Toluene 4-Monooxygenase. *Biochemistry* **2012**, *51* (6), 1101–1113.
- (15) Ruscio, J. Z.; Kumar, D.; Shukla, M.; Prisant, M. G.; Murali, T. M.; Onufriev, A. V. Atomic Level Computational Identification of Ligand Migration Pathways Between Solvent and Binding Site in Myoglobin. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105* (27), 9204–9209.
- (16) Lin, T.-L.; Song, G. Efficient Mapping of Ligand Migration Channel Networks in Dynamic Proteins. *Proteins: Struct., Funct., Bioinf.* **2011**, *79* (8), 2475–2490.
- (17) Grubmüller, H.; Heymann, B.; Tavan, P. Ligand Binding: Molecular Mechanics Calculation of the Streptavidin Biotin Rupture Force. *Science* **1996**, *271* (5251), 997–999.
- (18) Gervasio, F. L.; Laio, A.; Parrinello, M. Flexible Docking in Solution Using Metadynamics. *J. Am. Chem. Soc.* **2008**, *127* (8), 2600–2607.
- (19) Shan, Y.; Kim, E. T.; Eastwood, M. P.; Dror, R. O.; Seoliger, M. A.; Shaw, D. E. How Does a Drug Molecule Find Its Target Binding Site? *J. Am. Chem. Soc.* **2011**, *133* (24), 9181–9183.
- (20) Buch, I.; Giorgino, T.; De Fabritius, G. Complete Reconstruction of An Enzyme-inhibitor Binding Process by Molecular Dynamics Simulations. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108* (25), 10184–10189.
- (21) Takahashi, R.; Gál, V. A.; Guallar, V. Monte Carlo Free Ligand Diffusion with Markov State Model Analysis and Absolute Binding Free Energy Calculations. *J. Chem. Theory Comput.* **2014**, *10* (1), 282–288.
- (22) Borrelli, K.; Vitalis, A.; Alcantara, R.; Guallar, V. PELE: Protein Energy Landscape Exploration. A Novel Monte Carlo Based Technique. *J. Chem. Theory Comput.* **2005**, *1* (6), 1304–1311.
- (23) Cossins, B. P.; Housseini, A.; Guallar, V. Exploration of Protein Conformational Change with PELE and Meta-Dynamics. *J. Chem. Theory Comput.* **2012**, *8*, 959–965.
- (24) Sastry, G. M.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and Ligand Preparation: Parameters, Protocols,



and Influence on Virtual Screening Enrichments. *J. Comput. Aided Mol. Des.* **2013**, *27* (3), 221–234.

(25) Moe, L. A.; Hu, Z.; Deng, D.; Austin, R. N.; Groves, J. T.; Fox, B. G. Remarkable Aliphatic Hydroxylation by the Diiron Enzyme Toluene 4-Monooxygenase in Reactions with Radical or Cation Diagnostic Probes Norcaradiene, 1,1-Dimethylcyclopropane, and 1,1-Diethylcyclopropane. *Biochemistry* **2004**, *43* (50), 15688–15701.

(26) Borrelli, K. W.; Cossins, B.; Guallar, V. Exploring Hierarchical Refinement Techniques for Induced Fit Docking with Protein and Ligand Flexibility. *J. Comput. Chem.* **2010**, *31* (6), 1224–1235.

(27) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118* (45), 11225–11236.

(28) Yu, Z. Y.; Jacobson, M. P.; Josovitz, J.; Rapp, C. S.; Friesner, R. A. First-shell Solvation of Ion Pairs: Correction of Systematic Errors in Implicit Solvent Models. *J. Phys. Chem. B* **2004**, *108* (21), 6643–6654.

(29) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem.* **2004**, *47* (7), 1750–1759.

(30) Schmidtko, P.; Bidon-Chanal, A.; Luque, F. J.; Barril, X. MDpocket: Open Source Cavity Detection and Characterization on Molecular Dynamics Trajectories. *Bioinformatics* **2011**, *27*, 3276–3285.

(31) Le Guilloux, V.; Schmidtko, P.; Tuffery, P. Fpocket: An Open Source Platform for Ligand Pocket Detection. *BMC Bioinf.* **2009**, *10* (1), 168.

(32) Sambrook, J. R. D. *Molecular Cloning: A Laboratory Manual*, 3rd ed.; Cold Spring Harbor Laboratory Press, N. Y., 2011.

(33) Tao, Y.; Fishman, A.; Bentley, W. E.; Wood, T. K. Oxidation of Benzene to Phenol, Catechol, and 1,2,3-trihydroxybenzene by Toluene 4-monooxygenase of *Pseudomonas mendocina* KR1 and Toluene 3-monooxygenase of *Ralstonia pickettii* PKO1. *Appl. Environ. Microbiol.* **2004**, *70* (7), 3814–3820.

(34) Canada, K. A.; Iwashita, S.; Shim, H.; Wood, T. K. Directed Evolution of Toluene Ortho-monooxygenase for Enhanced 1-naphthol Synthesis and Chlorinated Ethene Degradation. *J. Bacteriol.* **2002**, *184* (2), 344–349.

(35) Brook, M.; Derry, N.-L.; Shainsky, J.; Zelas, Z. B.-B.; Boyko, Y.; Dabush, K.; Fishman, A. The Influence of Key Residues in the Tunnel Entrance and the Active Site on Activity and Selectivity of Toluene-4-monooxygenase. *J. Mol. Catal. B: Enzym.* **2010**, *66* (1–2), 72–80.

(36) Tao, Y.; Fishman, A.; Bentley, W. E.; Wood, T. K. Altering Toluene 4-monooxygenase by Active-site Engineering for the Synthesis of 3-methoxycatechol, Methoxyhydroquinone, and Methylhydroquinone. *J. Bacteriol.* **2004**, *186* (14), 4705–4713.

(37) Brouk, M.; Nov, Y.; Fishman, A. Improving Biocatalyst Performance by Integrating Statistical Methods into Protein Engineering. *Appl. Environ. Microbiol.* **2010**, *76* (19), 6397–6403.

## Supplementary Material

## Atomic Picture of Ligand Migration in Toluene 4-Monooxygenase

Ali Hosseini, Moran Brouk, Maria Fatima Lucas, Fabian Glaser, Ayelet Fishman, and Victor Guallar

**Table 1.** Primers used for site directed mutagenesis and sequencing of the F269 residue in the *tmoA* gene in TG1/pBS(Imanishi, Morita et al.)T4MO

Primer	Nucleotide sequence <sup>a</sup>
mutagenesis	
T4MObefEcoRI Front	5' - CCATGATTACGCCAAGCGCG-3'
T4MOABRear	5' -TCCATGCTCTTCACTGTTGAC-3'
T4MO_269V_Front	5' - CAATTGGCGTGCCTGGCGTCTAG <u>T</u> IGCGG-3'
T4MO_269V_Rear	5' -GGCCCGGTTAGTACCGCA <u>ACT</u> AGACGCCAG-3'
T4MO_269W_Front	5' -CAATTGGCGTGCCTGGCGTCTA <u>TGGG</u> CGGTAC-3'
T4MO_269W_Rear	5' -CGGCCCGGTTAGTACCGCC <u>AT</u> AGAGGCCAG-3'
Sequencing	
T4MO seq 1	5' -CCCGCATGAATACTGTAAGAAGGATCGC-3'

<sup>a</sup> Positions subjected to mutagenesis are underlined and bases modified to obtain the desired point mutation are indicated in bold.

## Publications

Publications

### **3.4 Exploration of Protein Conformational Change with PELE and Meta-Dynamics**

Benjamin P. Cossins, Ali Hosseini, and Victor Guallar

Journal of Chemical Theory and Computation, 2012,


DOI: 10.1021/ct200675g

## Publications

## Exploration of Protein Conformational Change with PELE and Meta-Dynamics

 Benjamin P. Cossins,<sup>†</sup> Ali Hosseini,<sup>†</sup> and Victor Guallar<sup>\*,†,‡</sup>
<sup>†</sup>Joint BSC-IRB Research Program in Computational Biology, Barcelona Supercomputing Center, c/Jordi Girona 29,08034 Barcelona, Spain

<sup>‡</sup>Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig Lluís Companys 23, 08010 Barcelona, Spain

 Supporting Information

**ABSTRACT:** Atomistic molecular simulation methods are now able to explore complex protein or protein–ligand dynamical space in a tractable way with methods such as meta-dynamics or adaptive biasing force. However, many of these methods either require a careful selection of reaction coordinates or the knowledge of an initial pathway of some kind. Thus, it is important that effective methods are developed to produce this pathway data in an efficient fashion. PELE, a proven protein–ligand sampling code, has been developed to provide rapid protein sampling in highly flexible cases, using a reduced network model eigen problem approach. The resulting method is able to rapidly sample configuration space with very general driving information. When applied to ubiquitin, PELE was able to reproduce RMSD and average force data found in molecular dynamics simulations. PELE was also applied to explore the opening/closing transition of T4 lysozyme. A meta-dynamics exploration using a low energy pathway validated that the configurations explored by PELE represent the most populated regions of phase space. PELE and meta-dynamics explorations also discovered a low free energy region where a large cross-domain helix of T4 lysozyme is broken in two. There is previous NMR evidence for the validity of this unfolded helix region.

### 1. INTRODUCTION

Recent years have brought the realization of important milestones in atomistic protein simulation. Simulations of relatively large time scale events such as protein folding, protein–ligand association, and large scale conformational change are becoming tractable and predictive.<sup>1–5</sup> These advances rely on accelerated methods which use simplified pathway coordinates to explore complex many-dimensional processes and/or molecular simulation techniques able to efficiently use large numbers of computer processors.<sup>6,7</sup> Whatever the combination of methods used, there is a need to perform hundreds of nanoseconds worth of conformational sampling. The scale of computer power needed for systems of interest is not available to all, and so for the majority, the problem of rapidly obtaining realistic dynamic information on proteins remains.

Methods able to quickly probe large scale protein conformational changes based on molecular dynamics such as steered MD<sup>8,9</sup> (SMD) and essential dynamics sampling (EDS)<sup>10–12</sup> have been used to direct MD in a direction of interest through clever constraints or restraints. There are many examples of steered or biased MD simulations being used to find a pathway for further free energy analysis with umbrella sampling or other such methods.<sup>9,13,14</sup> A recently developed method, temperature-enhanced essential dynamics replica exchange, seems able to steer large biomolecular MD simulations through temperature control of specific essential space modes while maintaining Boltzmann weighting.<sup>15</sup> Other advanced methods such as the finite temperature string method<sup>16</sup> and transition path sampling<sup>17,18</sup> attempt to sample defined pathways using molecular dynamics.

Alternative pathway building methods have been developed based on minimization rather than MD. A family of methods based on the nudged elastic band method (NEB)<sup>19–22</sup> have been used to find pathways between two experimental structures of the same protein. NEB methods in general work on the basis of the minimization of a series of intermediate configurations between two end point protein structures. Every intermediate configuration is connected to the previous and next by springs which keep the structure of the path while allowing minima to be found. NEB methods have been used to find probable low energy pathways of protein conformational change, which can then be used in conjunction with free energy methods to give predictive information.<sup>23</sup>

We present here a novel methodology capable of producing accurate and quick conformational sampling, and of providing reliable initial pathways for free energy methods. The methodology is a new development of the Protein Energy Landscape Exploration (PELE) program. PELE, a Monte Carlo (MC) based method, has thus far been used to characterize the exit pathways of bound molecules from proteins and for protein–ligand docking.<sup>24–26</sup> We introduced a new protein perturbation step based on anisotropic network model methodologies, capable of providing significant backbone motion.

These PELE developments have been tested on two systems: ubiquitin (Ubi) and T4 lysozyme (T4lyz). Both systems were chosen due to their small size and the amount of experimental and computational studies on their dynamics. For Ubi, a 76 residue regulatory protein, we have compared the PELE

Received: September 25, 2011

Published: January 27, 2012

conformational search with three 1  $\mu$ s MD trajectories. MD simulations of this length have been shown to map the conformational space of native Ubi, through comparison with NMR data.<sup>37</sup> For T4lyz, an enzyme found in *Escherichia coli* (*E. coli*), which has been infected with the bacteriophage T4, we perform meta-dynamics calculations<sup>28,29</sup> using a pathway defined after a PELE exploration of the opening and closing pathways of the two domains presents in T4lyz.

## 2. METHODS

### 2.1. Anisotropic Network Model (ANM) driven PELE.

The methodology has been built on the foundations of the PELE program which has thus far been used to characterize ligand exit pathways from proteins.<sup>24–26</sup> PELE uses a Monte Carlo (MC) scheme where new trial configurations are produced with sequential system perturbation, side chain prediction, and minimization steps. Trial configurations are then filtered with a Metropolis acceptance test. Using the MPI protocol, multiple trajectories can produce simultaneous trial configurations that share information and interchange coordinates toward one common goal.

**2.1.1. Perturbation Step.** Previous PELE studies used ligand translation and rotation for the perturbation step. Now, a protein perturbation step has been added to the PELE MC scheme that uses a reduced harmonic model of the system to calculate the probable major motions. The reduced harmonic model and the application of the calculated motions to perturb the system can take a number of forms.

**$\alpha$ -Carbon ANM Perturbation.** The initial reduced harmonic model used was the  $\alpha$ -carbon anisotropic network model (PELE-CA) similar to that described by Tirion.<sup>30</sup> The PELE-CA model uses the  $\alpha$ -carbons (CAs) to create a network model with identical springs connecting all CAs within a predefined cutoff (15 Å for PELE-CA in this study). A decay parameter (2.5 for PELE-CA in this study) is also included such that the force constant of the PELE-CA springs is decreased with the interaction distance. A similar decay parameter has been shown to improve the accuracy of ANM models in a recent study.<sup>31</sup> Extra user defined atoms and/or automatically chosen atoms from any ligand can also be included in the network. The Hessian matrix diagonalization of this reduced harmonic potential results in a series of eigenvectors that have been shown to describe the vibrational dynamics of proteins and their complexes.<sup>32,33</sup> Of course, for this model, movement vectors are acquired only for CAs. Thus, PELE-CA calculated motions are applied to the all atom (AA) system through the addition of harmonic restraints to CAs in the directions of the chosen PELE-CA modes. The system is then minimized for this biased potential, which has the effect of moving the system in the desired direction. A prototype version of this PELE-CA methodology has been used with success in a recent study of ligand migration in a truncated hemoglobin.<sup>25</sup>

**Heavy-Atom ANM Model.** A more elaborate alternative reduced harmonic model has been developed which uses all heavy atoms (HAs) of the AA system to create an elastic network model (PELE-HA). Pairs of HAs which have covalent or disulfide bonds between them are given a very high force constant spring (300 kcal mol<sup>-1</sup>) while nonbonded HA pairs are given a spring with a far weaker force constant (0.5 kcal mol<sup>-1</sup>; HA pairs more distant than 50 Å are not included). This model has similarities to recently published ANM and GNM models.<sup>34,35</sup> Again, a force constant decay parameter is applied such that longer springs are softer (3 for PELE-HA in this

study). This more complex harmonic model gives more detailed protein motions; however, these motions are calculated more slowly (2–10 s for PELE-CA and 2 min for PELE-HA on a 2000 atom system).

For the PELE-HA model, movement vectors are acquired for all heavy atoms. Hydrogen movement vectors are found by copying the vectors of atoms to which the hydrogens in question are covalently bonded. For this AA perturbation scheme, the system is not minimized with a biased potential following a chosen mode, as performed in the PELE-CA. Perturbation steps along given vectors are directly applied on all atoms, followed by a minimization where we constrain all CAs to the final perturbed position. Additionally, PELE-HA uses a nested-move consisting of a large number (50) of small randomly directed moves with move sizes scaled by the eigenvalue of the mode. Nested moves with an energy above a simple predefined threshold are rejected; however, in practice almost all moves are accepted. Of course the final trial configuration may be rejected by the acceptance test at the end of the PELE step.

**Choosing ANM Modes.** The problem of which ANM modes to apply is solved in a number of different ways. Single ANM modes or combinations can be chosen at random from a range of 1 to  $N$  modes where only these  $N$  modes are calculated to save time. This random sampling of modes can also be biased by the frequency of the modes such that the lowest frequency modes are chosen more often. In this study, modes were chosen at random from the top 10 ( $N = 10$ ) for PELE-CA and from the top 50 ( $N = 50$ ) using a biased distribution for PELE-HA.

The magnitude of the displacement down the calculated movement vectors can also be chosen at random from a predefined range of distances which can be biased by the frequency of the modes in question such that lower frequency modes use larger displacements. Of course in the case of PELE-CA, displacements discussed here are in fact the length of applied harmonic restraints.

If a target structure is provided, a PELE simulation can move toward this target. This is achieved by picking ANM modes which display a predefined level of directional overlap with the vectors of each atom of the initial structure to the equivalent atoms of the target. Directional overlap is simply the dot product of the mode vector with the vectors to the target structure. A random number can then be used to discard some overlapping modes such that sequential trials have differing overlapping modes. When a target structure is not available the spawning methodology of PELE is applicable, as described below in the Minimization Step section.

**2.1.2. Side Chain Step.** Due to larger protein motion, as a result of the new perturbation, the side-chain optimization step has been developed to pick residues which underwent a large energy change along the perturbation. Each side chain residue energy is computed before and after the perturbation. Thus, the user can choose side chains with the largest energy increase (top residues) and predict a better side chain position (using the algorithms developed in the PLOP program<sup>36</sup>). There is also the option to choose fewer top residues and select a spherical cluster around them. The user can control the number of high-energy residues optimized and the radius of the clusters around these residues. In this study, all PELE searches used two high energy residues and clusters with radii of 3 Å. In our experience, the number of side chains to be optimized should not exceed 25. Due to the approximate (but quicker) side chain

prediction algorithms, larger groups of side chains will result in high energy states and a low Metropolis acceptance.

**2.1.3. Minimization Step.** The last step in the PELE sampling algorithm uses a truncated Newton minimization of all of the system (or of those parts of the system included in the first two steps). While the minimization allows us to attain energetically accessible regions, it might also revert the structure to the initial conditions. In order to improve the conformational sampling, we have added the possibility to constrain the CAs to their final perturbed position within the final minimization. After this minimization, the trial configuration is accepted or rejected as stated above.

The multiprocessor parallel implementation of PELE shares predefined metrics across identical replica calculations. For example, several processors running identical PELE calculations can perform a random search with ANM protein sampling, those which manage to move away from an initial structure will export their coordinates to the processors lagging behind (within a given threshold).<sup>24</sup> This multiprocessor structure is able to drive the sampling to find new low energy configurations away from the initial. This "spawning" algorithm can be used to drive many different metrics. In this study, one of our spawning criteria for T4Lyz is the distance between the  $\alpha$ -carbons of GLU22 and ARG137. Hence, a multiprocessor search would allow all processors to search the conformational space of T4Lyz randomly until one processor has a GLU22-ARG137 distance which is 4 Å smaller than the largest GLU22-ARG137 distance found. At this point, the processor with the smallest GLU22-ARG137 distance takes on the configuration with the largest GLU22-ARG137 distance and continues to search.

**2.2. Free Energy Methods.** Meta-dynamics is a free energy methodology developed by the Parrinello group and is able to explore the free energy differences of atomistic systems using a series of driving collective variables. Here, we use meta-dynamics to explore the pathways found with PELE as a form of validation. A very brief description of meta-dynamics follows; more information can be found in various reviews.<sup>29,37</sup>

The concept of meta-dynamics is based around finding a collective variable of a system which, in a coarse-grained fashion, describes a property of interest within a simulation. Gaussians of width  $\sigma_i$  are then added to the collective variable in a history dependent manner such that the system is less likely to return to a region of the collective variable which has already been explored. The free energy surface of the region explored can then be found through the negative of the total biasing potential which was added through the simulation.

More recent developments of meta-dynamics (well tempered meta-dynamics) have given the possibility of faster and more complete convergence<sup>38,39</sup> through the adaptive adjustment of the Gaussian deposition rate  $\omega$ , achieved by rescaling  $W$ , the height of Gaussians.

Another development by the same group provides two functions of a possible reaction coordinate  $R$  which allow for efficient exploration of otherwise complex collective variables:<sup>28</sup>

$$s(R) = \lim_{\lambda \rightarrow \infty} \frac{\int_0^1 t e^{-\lambda(R-R(t))^2} dt}{\int_0^1 e^{-\lambda(R-R(t))^2} dt} \quad (1)$$

and

$$z(R) = \lim_{\lambda \rightarrow \infty} -\frac{1}{\lambda} \ln \int_0^1 e^{-\lambda(R-R(t))^2} dt \quad (2)$$

In our case,  $R$  is a series of protein  $\alpha$ -carbon coordinates for conformations of a protein which describe the motion of interest and  $(R - R(t))^2$  is the mean square displacement between the present configuration and the defined path. For sufficiently large  $\lambda$ , these functions allow a meta-dynamics search of the path defined by  $R$  such that the  $s$  path defines the distance along  $R$  and the  $z$  path defines the distance from  $R$ . In this way, a representation of the free energy landscape of  $R$  and surrounding regions can be calculated through

$$F(s, z) = -\frac{1}{\beta} \ln \left\langle \delta(s - s(R)) \delta(z - z(R)) \right\rangle \quad (3)$$

**2.3. Simulation Setup.** T4Lyz and Ubi structures used were PDB codes 3DMV and 1AAR, respectively. For T4Lyz, all heteroatoms including ligands, ions, and water were removed, hydrogens were added, and HIS 31 was protonated at both the  $\Delta$  and  $\epsilon$  positions. For Ubi, again all heteroatoms were removed, hydrogens were added, and no residues received extra protons. The structures were then minimized with constrained  $\alpha$ -carbons using a truncated-Newton algorithm and SGB solvation of PLOP.<sup>16</sup>

The MD simulation of Ubi was performed with GROMACS 4.0.5<sup>6</sup> and three different force fields: OPLSAA, Amber, and Charm. The system was solvated with a periodic cubic box, containing 8813 TIP3P<sup>40</sup> molecules and particle mesh Ewald (PME) long-range electrostatics.<sup>41</sup> Lennard-Jones interactions were cut off at 9 Å with a switching function, and Coulomb interactions were cut off at 10 Å. Temperature and pressure equilibrations preceded the 1  $\mu$ s of production at 300 K and 1 atm where system configurations were saved every 10 ps.

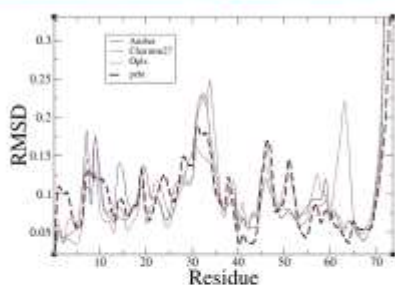
Meta-dynamics simulations were performed with GROMACS 4.0.5 patched with Plumed.<sup>42</sup> The OPLSAA force-field was used along with PME and similar cutoff and equilibration arrangements to the GROMACS simulations described above.

### 3. RESULTS

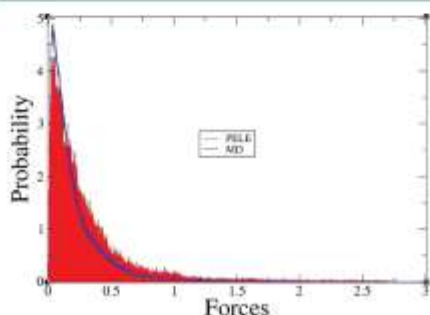
In order to study the capabilities of PELE being able to reproduce local fluctuations, PELE-CA and PELE-HA explorations of Ubi have been compared to molecular dynamics. Eight processors were allowed to freely explore (no search guidance) conformations around the initial structure for 24 h; an additional 24 h did not produce any significant change. Modes were chosen randomly from the top 50 in terms of frequency, and move sizes had a maximum of 1.5 Å and were biased not to be too small.

Figure 1 shows the RMSD from the initial structure for individual residues from PELE and three 1  $\mu$ s (each with a different force field) explicitly solvated MD simulations. While there are a few differences for most residues, in particular between the different force fields, there is an overall quantitative agreement. Remarkably, PELE captures most of the maximum and minimum fluctuations at a fraction of the MD cost. This agreement can also be seen when comparing average atomistic simulation forces. Bond length forces are of course very different owing to the minimization of PELE. Nonbonded, 1-4 interaction, torsional, and angle forces are very similar to those seen in Figure 2 and the Supporting





**Figure 1.** Plot of residue RMSD to average structures for MD using the OPLSAA, AMBER, and CHARMM27 force fields and OPLSAA PELE simulations of ubiquitin.



**Figure 2.** Histograms of atomistic nonbonded forces from PELE (red) and MD (blue).

Information. These comparisons suggest that the conformational space explored by PELE is very similar to that of MD.

Table 1 shows the results of a cluster analysis of PELE and MD ubiquitin trajectories. As seen when comparing PELE with MD, but also when comparing the different MD force fields, all simulations produce a similar number of clusters and agree on the percentage population of the top 4 and top 1 cluster. It is remarkable that PELE, a quick exploration technique with approximately 2 orders of magnitude less computational cost, can reproduce this semiquantitative analysis. Additionally, PELE has the most clusters with a 2 Å cutoff, suggesting that PELE finds more significantly different conformations.

Both PELE-CA and PELE-HA protocols as defined in the Methods section were applied to open T4lyz. After the initial minimization, PELE starts with a local exploration (100

iterations of free search) of the initial closed structure in order to relax the system. Then, searches of the opening process using driving metrics were performed. PELE-CA and HA searches used spawning on energy and atom–atom distance (GLU22 and ARG137  $\alpha$ -carbons). Once fully opened, with an  $\alpha$  carbon RMSD of 5 Å from the initial structure, another short local search of the opened structure is carried out. Then, the driving metrics are reversed, and searches return toward the initial structure.

PELE-CA and -HA searches using 32 processors and drawing modes from the top 50 biased by frequency were able to open and close T4lyz 1–2 times in 48 h. The potential energy data (Figure 3) seem to suggest low energy regions at the closed and open states. The energy profile clearly indicates that PELE-CA and -HA calculations seem very similar and essentially provide the same trajectories. Figure 4 shows the projection of the conformers of the PELE-HA trajectory on the PCA vectors calculated from the same trajectories. The opening–closing motion is very clear in vector 1, and there seems to be a small correlation with this motion for vectors 2 and 5. For the other PCA vectors, there seems to be more of a relaxation through the simulation. For these PELE trajectories, the eigenvalue of the first PCA vector is 20 times larger than any other, and this is borne out in the projection analysis (Figure 4). A visual comparison of the PCA vectors calculated for both PELE-CA and -HA and those found through a similar procedure using MD<sup>43</sup> suggested a good level of similarity.

The red circles in the HA plots of Figure 3 show the configurations chosen by binning all configurations by RMSD to the initial and choosing the lowest energy configuration in each bin. These configurations were visualized and assessed to ensure maximal equality in RMSD between all. If a chosen configuration was deemed to be too far from the required RMSD equality, it was removed, and the gap was filled using the Morph Server of Krebs and Gerstein.<sup>44</sup> The four open T4lyz structures with the same structure as that used here (150 L)<sup>45</sup> were within 1.3 Å of one of the configurations of the final pathway used in the meta-dynamics analysis. PELE projected these conformations without any previous knowledge of the open structure.

Figure 5 shows the free-energy landscape around the pathway taken from the PELE search. The  $x$  axis shows the S-path (see eq 1), which starts at 4 and ends at 7.5 units. The  $y$  axis shows the Z-path (see eq 2), which describes the distance from the S-path in Å<sup>2</sup>. Figure 6 shows graphic representations of the important regions found in Figure 5. The lowest free-energy regions (regions 1–3) of phase space were found close to the path found by PELE and correspond directly to the low energy regions in Figure 3. The two lowest energy regions suggest that T4lyz is able to open and close with energy barriers of around 10 and 30 kJ mol<sup>-1</sup>, respectively.

**Table 1.** Table of Results for a Series of Cluster Analyses of Ubiquitin Trajectories Using PELE and MD with Different Force Fields<sup>a</sup>

	PELE			MD-OPLS			MD-AMBER			MD-CHARMM		
	1	1.5	2	1	1.5	2	1	1.5	2	1	1.5	2
no. of clusters	98	19	8	238	20	6	157	14	4	110	12	4
% in top 4 clusters	39	86	95	38	88	99	44	90	100	60	96	100
% in top cluster	17	56	80	14	60	97	14	56	86	34	65	96
no. clusters for 90% of frames	33	5	2	54	4	1	28	4	2	16	2	1

<sup>a</sup>This analysis was carried out with the cluster GROMACS<sup>6</sup> tool, using the linkage algorithm and  $\alpha$ -carbon RMSD.

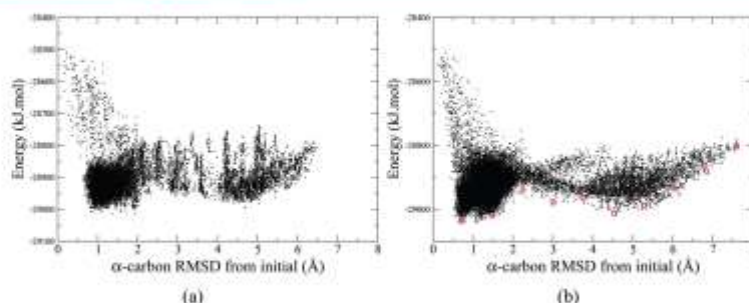


Figure 3. Potential energy against  $\alpha$ -carbon RMSD from the initial structure for (a) PELE-CA and (b) PELE-HA searches targeted with spawning on  $\alpha$ -carbon RMSD and atom. Red circles represent configurations chosen by the low energy binning procedure and used to create the meta-dynamics path after visual analysis.

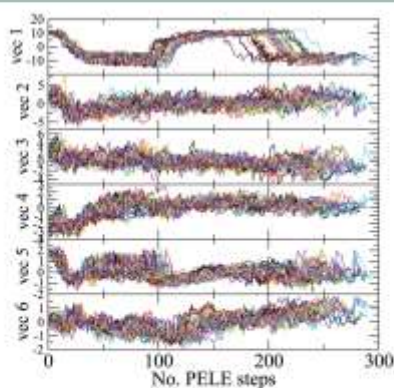


Figure 4. Projections (in nm) of PELE-HA trajectories on their PCA eigenvectors. Each PELE MPI process is colored differently. The PELE search started in the closed state; hence, the vec1 plot 10 corresponds to the closed state and  $-10$  to the open state.

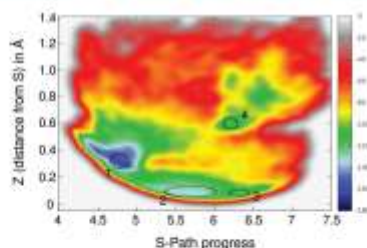


Figure 5. Free energy map of the PELE pathway and surrounding region explored using meta-dynamics. The free energy is depicted by color in  $\text{kJ mol}^{-1}$ .

Additionally, the large cross domain helix was found to break in the most open conformations (region 3), and this helix was found to become even more broken in an isolated region, which may not be significantly populated (region 4).

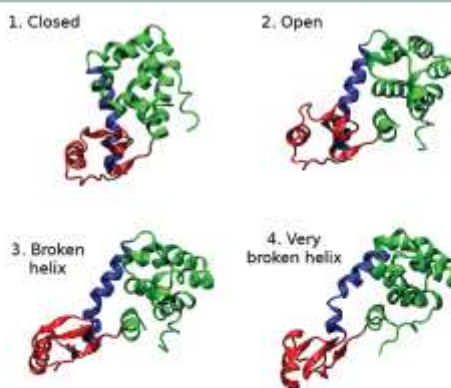


Figure 6. T4Lyz structures representing the different states found in the meta-dynamics analysis and referenced in Figure 5.

#### 4. DISCUSSION

Combining protein structure prediction algorithms with a Monte Carlo sampling technique, we introduced PELE, a program capable of describing the protein energy landscape associated with ligand migration in proteins. By adding a backbone perturbation following normal modes, we have expanded these capabilities to explore local dynamics and large conformational changes. A comparison of residue based RMSD and average force decomposition for Ubi suggests that PELE can rapidly give an idea of the protein conformational space. Employing the spawning methodology, PELE was able to rapidly explore the large conformational space of T4Lyz.

Both PELE-CA and PELE-HA seem suitable for the purposes of rapid exploration of protein large conformational change and fully explored the open/close transitions in T4Lyz in just 1–2 days usage of a small cluster. The minimum RMSD between the pathway denoted by red circles in Figure 3 and the open and closed states in Figure 6 was 1.55 Å and 1.80 Å, respectively. We should emphasize that PELE projected the open conformation, and the pathway toward it, without any knowledge of the open structure. As stated, the conformational exploration used the GLU22 and ARG137 atom distance as

spawning criteria. Equivalent results, however, were obtained when using as spawning criteria the RMSD to the initial structure, indicating the robustness of the landscape and of the search procedure. The overall method could be very useful as in the absence of initial pathway information, MD would need significant computational resources. The low free-energy regions of T4Lyz phase-space found by meta-dynamics were very similar to the low potential energy pathway found by PELE. However, if accurate free energies are necessary, then PELE should be combined with another more rigorous exploration technique and possible experimental methods.

Our calculations suggest that the main cross-domain helix is broken as T4Lyz becomes more open. This helix becomes slightly broken at the extremity of the probably well populated open region (region 3 of Figure 3) and very broken in the little populated region 4. Any functional significance is not clear, but there is experimental NMR evidence based on a loss of stability and hydrogen bonds in the helix.<sup>66,67</sup> Another simulation analysis of T4Lyz<sup>33</sup> has been carried out, and although this study only uses 1 ns trajectories, the PCA analysis provides similar modes and eigenvalues to those found in the present study and analyzed in the PCA projection analysis. While there is some controversy, experimental studies<sup>68,69</sup> seem to suggest that, in solution, T4Lyz is generally more open than closed. Our energy profiles indicate a larger basin (and more minima) for the open state. Using the OPLS-AA force field, however, the closed state is 3 kcal mol<sup>-1</sup> more stable. Thus, it is difficult to predict the preferred state of the system without a kinetic analysis (which is not the aim of this study).

The protein sampling capabilities of PELE are an addition to the molecular modeling tool-kit, which allows rapid sampling of interesting protein motions on the atomistic scale. These capabilities are complemented well by those of meta-dynamics, but PELE is also useful for any application which requires protein sampling such as protein–ligand and protein–protein docking. A public Web server for PELE protocols will be made available in January 2012 at <http://pele.bsc.es>.

## ■ ASSOCIATED CONTENT

### Supporting Information

A series of histograms comparing atomistic forces from PELE and MD simulations of ubiquitin. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## ■ AUTHOR INFORMATION

### Corresponding Author

\*E-mail: [victor.guallar@bsc.es](mailto:victor.guallar@bsc.es).

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

The authors would like to thank Santi Esteban-Martin and Xavier Salvatella for helpful discussions and for supplying some MD trajectories of ubiquitin. Also the European Research Council for grant ERC-2009-Adg 25027-PELE.

## ■ REFERENCES

- Piana, S.; Laio, A. *J. Phys. Chem. B* **2007**, *111*, 4553–4559.
- Kubitzki, M. B.; de Groot, B. L. *Structure* **2008**, *16*, 1175–1182.
- Pietrucci, F.; Marinelli, F.; Carloni, P.; Laio, A. *J. Am. Chem. Soc.* **2009**, *131*, 11811–11818.
- Bertozzi, A.; Cavalli, A.; Branduardi, D.; Gervasio, F. L.; Recanatini, M.; Parrinello, M. *J. Am. Chem. Soc.* **2009**, *131*, 244–250.

- Lemkul, J. A.; Brann, D. R. *J. Phys. Chem. B* **2010**, *114*, 1652–1660.
- Spoel, D. D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, *26*, 1701–1718.
- Bowers, K. J.; Chow, E.; Xu, H.; Dror, R. O.; Eastwood, M. P.; Gregersen, B. A.; Klepeis, J. L.; Kolossvary, I.; Moraes, M. A.; D., F.; Salmon, J. K.; Shan, Y.; Shaw, D. E. *Proc. ACM/IEEE Conf. Supercomput.* **2006**.
- Shen, L.; Shen, J.; Luo, X.; Cheng, F.; Xu, Y.; Chen, K.; Arnold, E.; Ding, J.; Jiang, H. *Biophys. J.* **2003**, *84*, 3575–3590.
- Cuenet, M. A.; Michielin, O. *Biophys. J.* **2008**, *95*, 3575–3590.
- Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. *Proteins: Struct., Funct., and Bioinf.* **1993**, *17*, 412–425.
- Amadei, A.; Linssen, A. B.; de Groot, B. L.; van Aalten, D. M.; Berendsen, H. J. *J. Biomed. Struct. Dyn.* **1996**, *13*, 615–625.
- Daidone, I.; Amadei, A.; Roccatano, D.; Nola, A. D. *Biophys. J.* **2003**, *85*, 2865–2871.
- Liu, Z.; Xu, Y.; Tang, P. *J. Phys. Chem. B* **2006**, *110*, 12789–12795.
- Huang, H.; Ozkirimli, E.; Post, C. B. *J. Chem. Theory Comput.* **2009**, *9*, 1304–1314.
- Kubitzki, M. B.; de Groot, B. L. *Biophys. J.* **2007**, *92*, 4262–4270.
- Vanden-Eijnden, E.; Venturoli, M. *J. Chem. Phys.* **2009**, *130*, 194103.
- Dellago, C.; Bolhuis, P. G.; Chandler, D. *J. Chem. Phys.* **1999**, *110*, 6617.
- Bolhuis, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L. *Annu. Rev. Phys. Chem.* **2002**, *53*, 291–318.
- Elber, R.; Karplus, M. *Chem. Phys. Lett.* **1987**, *139* (380), 375.
- Fischer, S.; Karplus, M. *Chem. Phys. Lett.* **1992**, *194*, 252–261.
- Henkelman, G.; Uberuaga, B. P.; Jonsson, H. *J. Chem. Phys.* **2000**, *113*, 9901–9904.
- Ignacio Fdez. Galvín, F.; Field, M. J. *J. Comput. Chem.* **2008**, *29*, 139–143.
- Arora, K.; Brooks, C. L. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 18496–18501.
- Borrelli, K. W.; Vitalis, A.; Alcantara, R.; Guallar, V. *J. Chem. Theory Comput.* **2005**, *5*, 1304–1311.
- Guallar, V.; Lu, C.; Borrelli, K.; Egawa, T.; Yeh, S. *J. Biol. Chem.* **2009**, *284*, 3106–3116.
- Borrelli, K. W.; Cossins, B.; Guallar, V. *J. Comput. Chem.* **2010**, *31*, 1224–1235.
- Lange, O. F.; van der Spoel, D.; de Groot, B. L. *Biophys. J.* **2010**, *99*, 647–655.
- Branduardi, D.; Gervasio, F. L.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, 054103.
- Leone, V.; Marinelli, F.; Carloni, P.; Parrinello, M. *Curr. Opin. Struct. Biol.* **2010**, *20*, 148–154.
- Tirion, M. M. *Phys. Rev. Lett.* **1996**, *77*, 1905–1908.
- Eyal, E.; Yang, L.; Bahar, I. *Bioinformatics (Oxford, England)* **2006**, *22*, 2619–2627.
- Bahar, I.; Atilgan, A. R.; Erman, B. *Folding Des.* **1997**, *2*, 173–81.
- Atilgan, A. R.; Durell, S. R.; Jernigan, R. L.; Demirel, M. C.; Keskin, O.; Bahar, I. *Biophys. J.* **2001**, *80*, 505–15.
- Lu, M.; Ma, J. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 15358–15363.
- Rueda, M.; Bottegoni, G.; Abagyan, R. *J. Chem. Inf. Model.* **2009**, *9*, 358–363.
- Jacobson, M. P.; Friesner, R. A.; Xiang, Z.; Honig, B. *J. Mol. Biol.* **2002**, *320*, 597–608.
- Laio, A.; Gervasio, F. L. *Rep. Prog. Phys.* **2008**, *71*, 126601.
- Barducci, A.; Bussi, G.; Parrinello, M. *Phys. Rev. Lett.* **2008**, *100*, 020603.
- Bonomi, M.; Barducci, A.; Parrinello, M. *J. Comput. Chem.* **2009**, *30*, 1615–1621.
- Price, D. J. III; C., L. B. *J. Chem. Phys.* **2004**, *121*, 10096–10103.

- (41) Darden, T.; Perera, L.; Li, L.; Pedersen, L. *Structure* **1999**, *7*, R55–R60.
- (42) Bonomi, M.; Branduardi, D.; Bossi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietracchi, F.; Brogna, R. A.; Parrinello, M. *Comput. Phys. Commun.* **2009**, *180*, 1961–1972.
- (43) de Groot, B. L.; Hayward, S.; van Aalten, D. M.; Amadè, A.; Berendsen, H. J. *Proteins: Struct., Funct., Bioinf.* **1998**, *31*, 116–127.
- (44) Krebs, W. G.; Gerstein, M. *Nucleic Acids Res.* **2000**, *28*, 1665–1675.
- (45) Zhang, X.; Matthews, B. *Protein Sci.* **1994**, *3*, 1031–1039.
- (46) McIntosh, L. P.; Wand, A. J.; Lowry, D. F.; Redfield, A. G.; Dahlquist, F. W. *Biochemistry* **1990**, *29*, 6341–6362.
- (47) Llinas, M.; Gillespie, B.; Dahlquist, F. W.; Margussee, S. *Nat. Struct. Mol. Biol.* **1999**, *6*, 1072–1078.
- (48) Mchaourab, H. S.; Oh, K. J.; Fang, C. J.; Hubbell, W. L. *Biochemistry* **1997**, *36*, 307–316.
- (49) Goto, N. K.; Skrynnikov, N. R.; Dahlquist, F. W.; Kay, L. E. *J. Mol. Biol.* **2001**, *308*, 745–764.

## Publications

## Supplementary Material

### Exploration of protein conformational change with PELE and Meta-dynamics

Benjamin P. Cossins, Ali Hosseini, and Victor Guallar

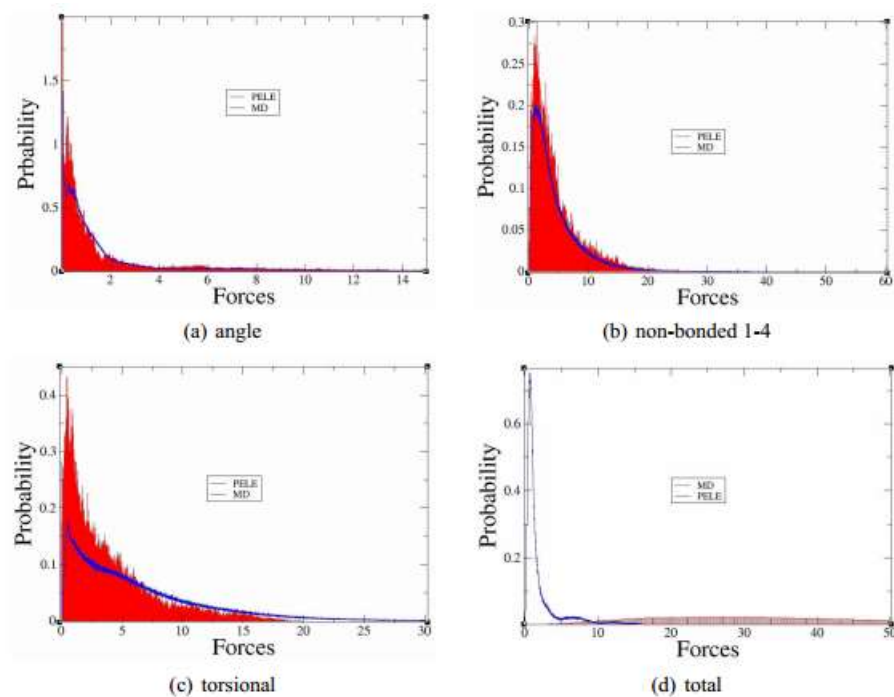


Figure 1: Histograms of atomistic forces from PELE (red) and MD (blue) simulations of ubiquitin.

## Publications

Publications

### **3.5 Computational Prediction of HIV-1 Resistance to Protease Inhibitors**

Ali Hoseini, Andreu Alibés, Marc Noguera-Julian, Roger Paredes, Robert Soliva, Modesto Orozco and Victor Guallar

In preparation (2014)



## Publications

## Computational Prediction of HIV-1 Resistance to Protease Inhibitors

Ali Hoseini<sup>1</sup>, Andreu Alibés<sup>1,2</sup>, Marc Noguera-Julian<sup>3,4,5</sup>, Roger Paredes<sup>3,4,5</sup>, Robert Soliva<sup>1</sup>, Modesto Orozco<sup>1,2,6</sup> and Victor Guallar<sup>1,7\*</sup>

\* Correspondence to V.Guallar. victor.guallar@bsc.es

<sup>1</sup> Joint BSC-CRG-IRB Research Program in Computational Biology. Barcelona Supercomputing Center. c/ Jordi Girona 29, 08034 Barcelona, Spain.

<sup>2</sup> Institute for Research in Biomedicine (IRB Barcelona), Baldiri i Reixac 8, Barcelona 08028, Spain

<sup>3</sup> IrsiCaixa AIDS Research Institute, Hospital Universitari Germans Trias i Pujol, Badalona, Catalonia, Spain

<sup>4</sup> Universitat Autònoma de Barcelona, Catalonia, Spain

<sup>5</sup> Universitat de Vic – Universitat Central de Catalunya, Vic, Catalonia, Spain

<sup>6</sup> Departament de Bioquímica. Universitat de Barcelona. Avgda Diagona 647. 08029, Spain

<sup>7</sup> Institució Catalana de Recerca i Estudis Avançats, Passeig Lluís Companys 23, 08010 Barcelona, Spain

### Abstract

Antiretroviral drug resistance is a major obstacle to end the HIV/AIDS pandemic. Protease inhibitors (PIs), one of the mainstays of HIV therapeutics, block HIV-1 protease (PR), which cleaves the Gag and Gag–Pol HIV-1 polyproteins to yield mature infectious virions. Development of mutations in HIV-1 PR hinders the activity of these drugs, making anti-AIDS therapy less efficient, and forcing changes in drug prescription. Most resistance assessments used in the clinic and epidemiologic surveillance to date rely on expert-based rules to interpret predefined sets of stereotypical mutations, and are applied by the clinical community to design alternative therapies. Such approach, exclusively information-driven, is powerful, but cannot capture the effects of new polymorphisms impacting virus susceptibility and fitness, and cannot be applied for new drugs. Computational modeling of PI-PR interactions could provide an unbiased, wider, and more general assessment of PI resistance, and could be made available to clinicians and caregivers through the Internet. In the present proof-of-concept study, we create a protocol involving sequence comparison and all-atom protein-ligand induced

## Publications

fit simulations to predict PR resistance at the molecular level. We first compared our predictions with experimentally determined 50% inhibitory concentrations ( $IC_{50}$ ) of darunavir, amprenavir, ritonavir and indinavir from reference PR structures of the wild type molecular clone HIV-1 NL4.3 and various mutant PRs displaying different PI resistance levels. We then performed analyses on a large set of variants harboring more than 10 PR mutations. Finally, several mutant sequences of HIV-1 PR isolated from real patients were analyzed for amprenavir and darunavir. Our computational approach detected all genotype changes triggering high-level PI resistance, even those involving a large number of mutations.

## Introduction

Antiretroviral therapy (ART) is one of the most effective interventions in medicine and, in particular, in HIV treatment. In ideal conditions of treatment, ART transforms a deadly disease into a chronic pathology, allowing patients to achieve a life expectancy (Rodger, Lodwick et al. 2013) and quality similar to that of non-HIV-infected individuals.(Hogg, Heath et al. 1998; Palella, Delaney et al. 1998) Such efficacy, however, can be offset by HIV's ability to develop mutations conferring antiretroviral resistance in the presence of drug selective pressure(Larder and Kemp 1989; Cohen 1992; Condra, Schleif et al. 1995; Rhee, Fessel et al. 2005) and by patient-to-patient transmission of resistant viruses. In resource-limited settings, where the HIV/AIDS toll is higher, ART is often provided without virological monitoring,(Gilks, Crowley et al. 2006; Keiser, Orrell et al. 2008) which is being associated with an alarming increase in drug resistance.(Hamers, Wallis et al. 2011; Hamers, Sigaloff et al. 2012; Organization 2012)

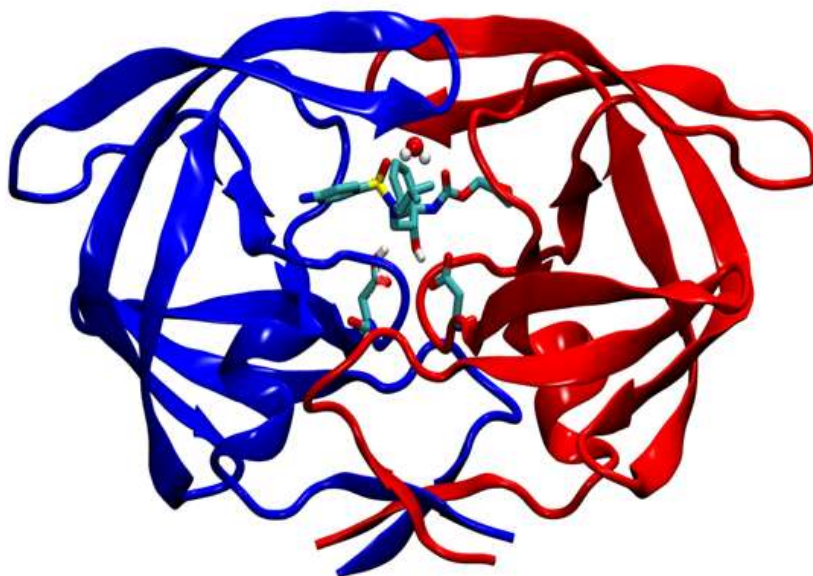
Antiretroviral drug resistance testing is key for clinical management(Johnson, Calvez et al. 2013) and epidemiologic surveillance,(Gupta, Hill et al. 2009; Sigaloff, Calis et al. 2011; Organization 2012; Sigaloff, Hamers et al. 2012; Paredes, Marconi et al. 2013) but it is not trivial to assess. Caregivers often rely on expert-based rules to interpret predefined sets of stereotypical mutations. Such procedure assigns susceptibility scores to different drugs for each mutation encountered after sequencing, from a previously defined list; individual mutation scores are then added into a global score for each drug and combination of drugs. Such scores are usually translated into a susceptible-intermediate-resistant (Sircar, Sanni et al.) interpretation. Interpretation rules based on a predefined list of mutations

## Publications

rely on previous knowledge and are limited in the mutation scope, so they cannot take into account the effect of non-predefined polymorphisms, mutational interactions, or mutation effects on different genetic backgrounds in the virus. Moreover, interpretation rules cannot provide resistance assessments to new molecules inhibiting the same target/s until new knowledge is generated, which can require years of “blind” treatment.

In an attempt to build a universal computational tool for resistance prediction that would not rely on prior knowledge on drug resistance, we have developed an automatic protocol combining sequence alignment with protein-ligand induced fit sampling techniques. In this first proof-of-concept study we focused on the HIV-1 PR, but the process could be later extended to other enzymes, HIV-1 subtypes or viruses.

HIV-1 protease (PR) is a 22 kDa dimeric aspartic protease that contributes to the maturation progress of the virus, cleaving the polyprotein precursor into functional viral proteins. Being one of the main targets in the treatment against HIV, the development of nine FDA-approved protease inhibitors, along with numerous crystal structures were established in the last two decades.(Kim, Baker et al. 1995; Turner, Strohbach et al. 1998). The active form of HIV-1 protease is a homodimer, with ~99 residues per chain. The catalytic residues Asp25 and Asp125 lie at the bottom of the binding cavity shielded by a pair of flaps (residues 44-55 and 144-155).(Tie, Boross et al. 2004) Main features common to all inhibitors are the existence of a hydroxyl group in contact with Asp25/125 residues, and a water molecule that mediates contact between a conserved carbonyl of the inhibitors and the protease amide groups of Ile50/Ile150 located in the flaps (Figure 1). Several resistant mutations have been described along the whole protease chain. Many mutations might occur simultaneously within a single variant highlighting the importance of complex cooperative effects. Typically these changes confer resistance not only by direct contact with the inhibitor, but also by subtle changes in the structure/dynamics incidentally affecting the active site. Obviously while these mutations block or reduce drug inhibition, they also need to have minor effects in the normal function of the enzyme which in the absence of compensatory mutations can generate changes in viral fitness.(Tantillo, Ding et al. 1994; Condra, Schleif et al. 1995; Gubareva, Bethell et al. 1996)



**Figure 1.** The HIV-1 protease dimer. Cartoon diagram of NL4-3 protease sequence showing monomers in blue and red. Residues Asp25 and Asp125 and APV are shown as sticks colored by atom type. A relevant water molecule is displayed as ball and stick.

In this study, we used all existing HIV-1 PR crystal structures available in the protein data bank and PELE (Protein Energy Landscape Exploration, a technique capable of reproducing complex protein-ligand induced fit (Borrelli, Vitalis et al. 2005; Borrelli, Cossins et al. 2010)), to explore the impact of mutations in inhibitor binding. Our protocol was first tested by comparing calculated and experimental binding energies of several FDA-approved inhibitors with engineered PR variants containing different amino acid mutations. (Koh, Nakata et al. 2003; Shenderovich, Kagan et al. 2003) Then, we used our approach to predict *ab initio*, without any informational-driven bias, the resistance profile of 44 clinical HIV-1 PR variants to amprenavir (APV) and darunavir (DRV). Our technique is fast, universal, and contrary to previous structure-based computational methods (Shenderovich, Kagan et al. 2003) which were limited to variants with a reduced number of mutations, it can predict PI resistance to variants containing more than 10 mutations.

## Methods

### *Systems used*

#### *References structures*

To build the homology models of the wild type NL4-3 and HXB2 reference sequences (which were considered as non-resistant genotypes unaltered by the selective pressure of any PI), we used the crystal structure of HIV-1 protease bound to BE6 inhibitor (pdb code 1W5Y(Lindberg, Pyring et al. 2004)), with only 2 and 1 mutations per chain, respectively.

#### *HIV-1 protease-inhibitor complexes*

Several crystal structures with measured catalytic activities were used to tune our computational protocol. For the APV ligand, the PDB entries used were: 3NU3(Shen, Wang et al. 2010) a WT reference with analogous relative affinity as NL4-3, 3S43(Tie, Wang et al. 2012) a triple mutant with 15 fold increase in inhibition constant ( $K_i$ , involving a decrease in affinity), 3NU5(Shen, Wang et al. 2010) a single mutant with 30 increase in inhibition constant. For the DRV ligand, the PDB entries used were, 2IEN(Tie, Boross et al. 2004) a WT reference with analogous relative affinity as NL4-3, and 3EM6(Mittal, Bandaranayake et al. 2013) a double mutant with 4 fold higher dissociation constant.

#### *Modeling HIV-1 protease mutants*

Due to the strong correlation between sequence similarity and structure, our strategy to model mutants with no solved structure was to use the closest existing crystal structure in terms of sequence. We found ~450 crystal structures of HIV-1 protease which can be used as templates for building a given clinical sequence. Thus, modeling mutants was performed in two steps: (1) a search for the crystal structure with the highest similarity to the mutant sequence using BLAST(Altschul, Gish et al. 1990), and (2) building our model by replacing each mutant using the Maestro(Sastry, Adzhigirey et al. 2013) software. Mutations were introduced simultaneously in the two protease chains. Using this procedure several sets of mutants were built: *The first set* was selected from the work by Koh et al.,(Koh, Nakata et al. 2003) were three NL4-3 variants: (L10I, G48V, I54V, L90M), (L10F, V32I, M46I, I54M, A71V, I84V) and (L10F, D30N, K45I, A71V, T74S) were tested against four inhibitors DRV, APV, IDV and RTV. *The second set* was obtained from the

## Publications

study by Shenderovich et al.(Shenderovich, Kagan et al. 2003) For this case, we selected the more difficult set of sequences, with more than 10 mutations (and a maximum number of 17 mutations per monomer, giving a total of 15 sequences) against IDV inhibitor. *The third set* included 21 and 23 PR FASTA sequences with different resistance profiles to APV and DRV, respectively, obtained from routine genotypic resistance testing (TRUGENE® HIV-1 Genotyping Assay, Siemens Healthcare, Barcelona, Spain) in the HIV Unit and irsiCaixa AIDS Research institute, Hospital Universitari Germans Trias i Pujol, Badalona, Spain. This last set of studies was performed as a blind test, where the modeling team had only access to the sequence but not to the resistance score.

### ***Systems preparations***

All systems were prepared with Schrodinger's Protein Wizard.(Sastry, Adzhigirey et al. 2013) This algorithm builds hydrogen-bonded clusters and performs 100000 Monte Carlo moves by reorienting hydroxyl and thiol groups, water molecules, amide groups of Asn and Gln, and the imidazole ring of His, to correct for typical crystal structure refinement errors. The algorithm also predicts protonation states of His, Asp, Glu, Lys and Arg. Each possibility is scored based on the total number of hydrogen bonds and their quality (relative to an idealized hydrogen bond). In this work, Asp25 was protonated in all structures, whereas Asp125 was considered ionized. Histidines 69 and 169 were either epsilon or double protonated depending on the structural environment; all other histidines kept delta protonation. A water molecule, generally found in all protease-inhibitor crystal structures was kept. The water preserved mediates a contact between the P2/P1' carbonyl oxygen atoms from the inhibitors and the amide groups of Ile50/Ile50'.

Once the all atom model was built, the ligand was initially docked in the active site using Glide.(Halgren, Murphy et al. 2004) The top score model (XP scoring) obtained by glide was selected followed by a molecular mechanics minimization using Schrodinger's Protein Wizard and the OPLS2005 force field to remove any possible geometric clashes. In order to keep the system close to its initial geometry, a restrain was applied to all heavy atoms (allowing a maximum displacement of 0.3Å). All ligands were quantum mechanically minimized separately in the gas phase using the DFT/B3LYP and 6-31G\*\* basis set level of theory. The atomic charges defining ligand electrostatics were then derived from the electrostatic potential fitting at the same level of theory.

### ***Assessing the induced fit and binding energy***

To map the protein-ligand conformational sampling we used PELE (Borrelli, Vitalis et al. 2005), which implements a Monte Carlo algorithm where new trial configurations are produced with sequential ligand (and protein) perturbation, side chain prediction and minimization steps. Ligand perturbation includes a ligand specific rotamer library (Borrelli, Cossins et al. 2010). Trial configurations are then filtered with a Metropolis acceptance test, where the energy is described with an all-atom OPLS force field with a surface generalized born solvent model (Yu, Jacobson et al. 2004). PELE has recently shown to provide some competitive advantages with respect to *state of the art* induced fit commercial software and to reproduce the conformational sampling obtained in microsecond molecular dynamics trajectories (Espona-Fiedler, Soto-Cerrato et al. 2012; Hosseini, Espona-Fiedler et al. 2013).

A total of 12 independent MC trajectories were produced for each inhibitor and mutant sequence. Trajectories were interrupted after 12h of CPU, providing approximately a total of 6000 Monte Carlo steps and ~2000 accepted minima. PELE's binding energy was then obtained by averaging the interaction energies of all accepted minima (approximately 2000 snapshots).

## **Results**

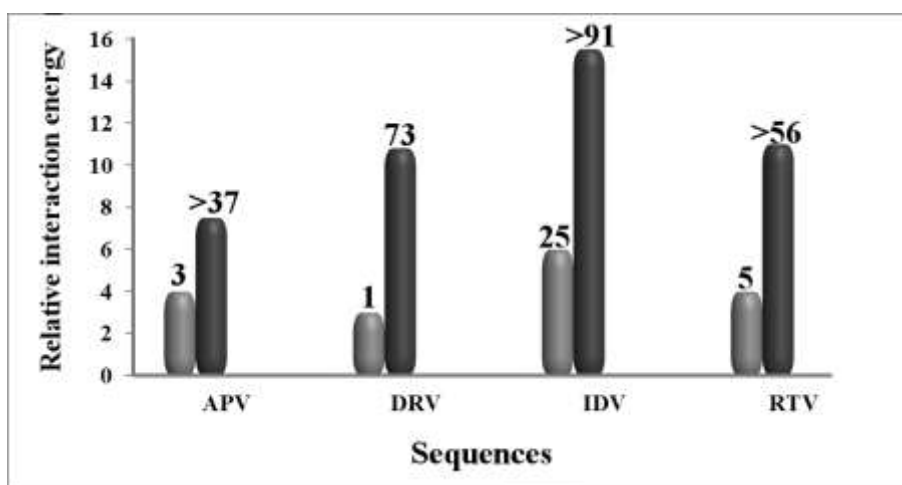
In the first part of the study, we validate the method by comparing *in vitro* IC<sub>50</sub> results for specific mutations with our theoretical predictions. In the second part, attention was centered on predicting how mutations in HIV-1 protease taken from real patients affect drug resistance against APV and DRV.

### ***Validation for the set of data with known experimental binding affinity (IC<sub>50</sub>)***

We first calculated the interaction energies of APV, DRV, IDV and RTV for three different sequences (each ligand) among the set described in Koh et al. (Koh, Nakata et al. 2003). Besides NL4-3, for each drug we selected the mutation with the lowest and highest IC<sub>50</sub> value relative to the value for NL4-3 (see Table 3 in Koh et al. (Koh, Nakata et al. 2003)). Figure 2, summarizes PELE's relative increase in binding energy for each ligand and



target, compared to the reference NL4-3 (the higher the increase in binding energy, the greater the resistance conferred by the mutation). The experimental relative (with respect to NL4-3) increase in  $IC_{50}$  is also shown with numbers in Figure 2. Clearly, PELE's simulations can distinguish between the high (dark gray) and low (light gray) resistance mutant in each specific inhibitor. Moreover, the relative increase in PELE's binding energy correlates nicely with the experimental increase in resistance in this set of data.

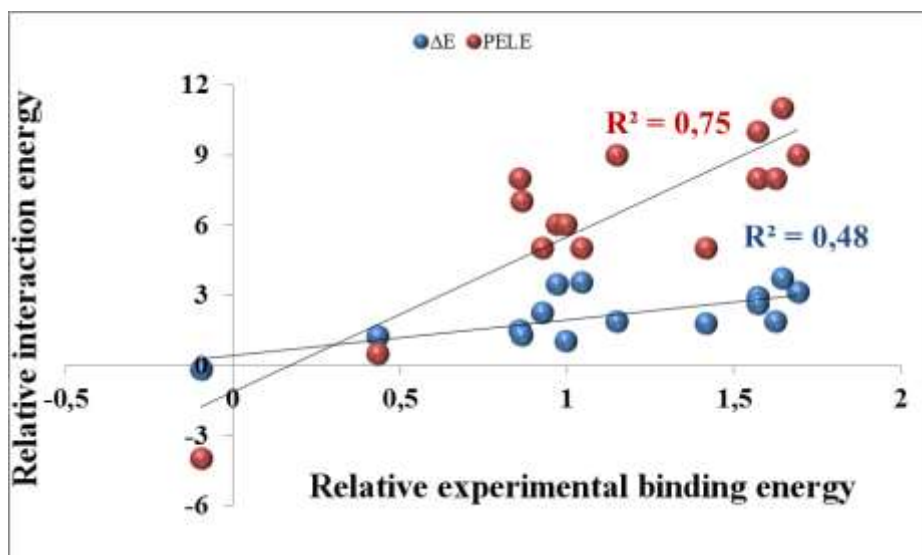


**Figure 2.** PELE's relative binding energies (in kcal/mol) predicted for the high (dark gray) and low (light gray) resistance HIV-1 PR mutants for APV, DRV, IDV and RTV observed by Koh et al. (Koh, Nakata et al. 2003). Experimental relative (with respect to NL4-3) increase in  $IC_{50}$  is shown with numbers above each bar.

For a second test, we used a subset of sequences derived from the work of Shenderovich et al. (Shenderovich, Kagan et al. 2003). Using known and in-house prepared mutations, these authors developed possibly the most comprehensive computational predictor to date. However, as noticed by the same authors, the quality of predictions correlates negatively with the increase of number of mutations. From their test set, we selected all the sequences with more than 10 mutations for the IDV inhibitor (a total of 15 sequences, which represent a specially difficult set for prediction (Shenderovich, Kagan et al. 2003)). Figure 3 shows the correlation between our PELE relative interaction energy estimates, using again NL4-3 as the reference zero value, and the experimental relative binding energy. Additionally we include the estimates

derived from Shenderovich et al. (Shenderovich, Kagan et al. 2003), also referred to NL4-3. We should notice that in Shenderovich et al., all mutant models were derived from one crystal and that they used simpler side chain (and backbone) sampling algorithms.

Despite the difficulty of this set of 15 compounds, the method outlined here behaves quite well, with a coefficient of determination equal to 0.75 (PELE's  $p$ -value =  $2.5e^{-05}$  and  $\Delta E$ 's  $p$ -value = 0.004), improving previous predictions significantly.



**Figure 3.** Correlation between modeled changes in relative binding energies (in kcal/mol) obtained in Shenderovich et al. (Shenderovich, Kagan et al. 2003) ( $\Delta E$ , blue beads) and PELE (red beads) with changes from experimental relative binding energy.

### ***Prediction results for APV and DRV (clinically isolated data)***

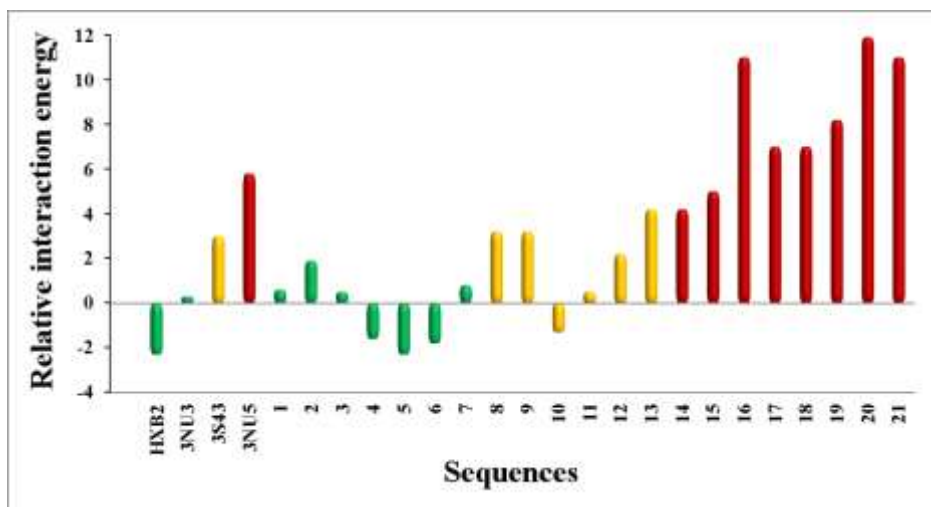
After testing our computational protocol with publicly available data, we proceeded to perform “blind predictions” of the resistance in clinical samples. Importantly, each of these variants, taken from HIV infected patients, contained a large number of mutations in each monomer (15-25) when compared to the reference NL4-3 (therefore, in some of the simulated systems the protease bore as many as 50 mutations, representing an incredible difficult test). The data for the clinical isolates was divided (and ordered) based on the resistance scores calculated from expert assessments in HIVdb (Liu and Shafer 2006). Samples were categorized as sensitive (S) when

## Publications

HIVdb resistance scores were below 20, intermediate (I) resistance when scores were between 20 and 50, and resistant (R) when scores were 50 or higher.

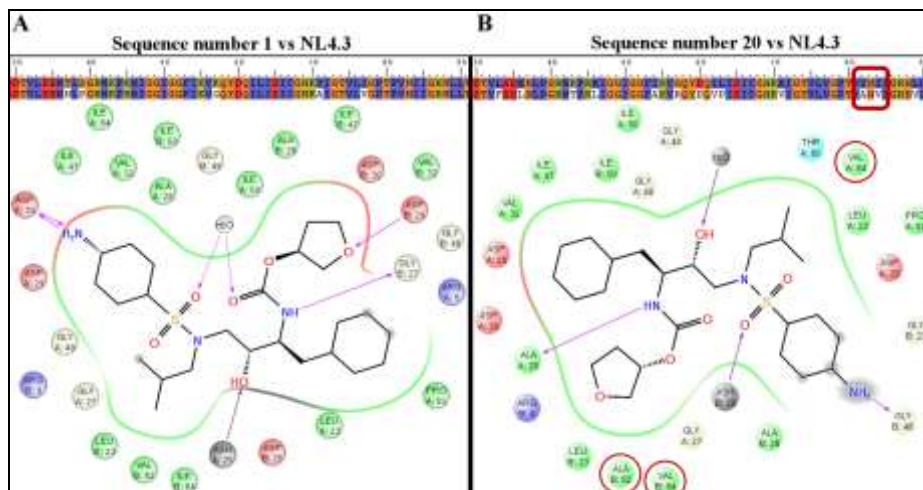
Figure 4 shows the results for APV where again we computed PELE's relative binding energies to the reference compound NL4-3. For this compound we had initially run a benchmark with four variants containing known resistance-related mutations: two consensus reference proteins (the sequence from pdb structure 3NU3 and the reference sequence HXB2, which were both considered non-resistant), the single mutation I50V with 30 fold increase in  $K_i$  (pdb id 3NU5) and the triple mutant V32I/I47V/V82I with 15 fold increase in  $K_i$  (pdb id 3S43). The data from this benchmark has also been included in Figure 4 (left-hand side). As can be seen, the two additional non-resistant reference sequences (the one from 3NU3 and HXB2) show equal or (slightly) better interaction to that found for NL4-3. This result shows how our computational protocol predicts reference sensitive sequences other than NL4-3 as sensitive, and similarly, the two APV-resistant reference sequences (3S43 and 3NU5), with a 15 and 30 fold increase in  $K_i$ , respectively, were correctly predicted as interfering drug-binding (note that the impact of the mutation in predicted binding, matches the  $K_i$  increase found experimentally).

For the sequences extracted from the patients (1-21), it can be seen that all sensitive sequences (green; columns 1-7) are predicted to be sensitive by our method, and all highly resistant sequences are predicted to be highly resistant (red, columns 14-21). Sequences classified as intermediate resistance (in yellow) offer a less clear differential profile, but there is only one pair (13/14) where PELE will have some difficulty to predict the resistance level produced by the mutation.



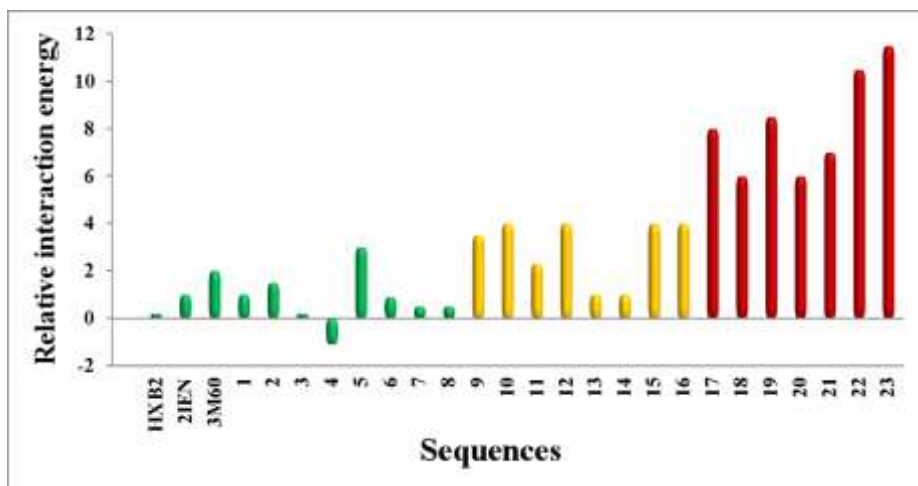
**Figure 4.** PELE's relative change in APV binding energies (kcal/mol). Sensitive, intermediate and resistance HIVdb values for each sequence are shown in green, yellow and red colors, respectively.

Due to the large number of mutations in each sequence, assigning individual effects is not a trivial task. Figure 5 compares the protein-APV interactions for sequences number 1 and 20, with a ~12 kcal difference in relative affinity (HIVdb scores 0 and 150, respectively). Sequence 1 presents 6 mutations (per chain) compared to NL4-3 while sequence 20 has 16 mutations. Clearly, sequence 20 shows a reduction in the number of interactions. Mutations on residues Val82 and Val84 affect the binding mode and orientation of the inhibitor. In particular, these mutations influence the proper interaction of APV with polar residues, Asp25, Asp29 and Asp30, significantly decreasing the affinity of the ligand. Asp25, for example, plays an important role in ligand recognition by making a strong interaction with the hydroxyl group of almost all inhibitors.



**Figure 5.** Protein-APV interaction diagram for sequences number 1, panel A, and 20, panel B. Important residues, discussed in the main text, are underlined in red circles. Partial sequence alignment to NL4-3 is also shown for both systems.

Figure 6 shows the result of equivalent calculations for DRV. For this compound we also modeled three additional variants containing known resistance-related mutations: two consensus reference proteins (the one taken from pdb structure 2IEN and the reference sequence HXB2) and the double mutant I50L/A71V (3M60), which has 4 fold higher dissociation constant. As in the APV case, reference and sensitive compounds show little or no effect when compared to NL4-3. Again, the most remarkable result is the correlation between PELE relative binding energy and the estimation of resistance, where all highly resistant variants are clearly identified.



**Figure 6.** PELE's relative change in DRV binding energies (kcal/mol). Sensitive, intermediate and resistance HIVdb values for each sequence is shown in green, yellow and red colors, respectively.

## Discussion

The primary aim of this study was to develop an automatic computational protocol for rapid discrimination between resistant and sensitive HIV-1 protease variants. To this aim we have used PELE, an atomic-resolution sampling algorithm combining a stochastic Monte Carlo procedure with protein structure prediction techniques, which is specially suited for induced-fit docking problems. (Espona-Fiedler, Soto-Cerrato et al. 2012; Hosseini, Espona-Fiedler et al. 2013)

Since most of the mutant structures have not been crystallized, one of the main questions is how to generate an all atom model for each sequence. Since we were comparing the binding energies to the NL4-3 reference sequence, one simple strategy would have been to use its structure as a template (in Modeller, (Fiser, Do et al. 2000; Eswar, Webb et al. 2002) I-TASSER, (Mitra, Shultis et al. 2013) etc.). However, many sequences have up to 20 mutations (when compared to NL4-3) in each chain, which could introduce large errors in building the models. The large number of HIV-1 X-ray crystal structures enabled better strategies. One obvious is the use of multiple templates in homology modeling, but our initial analysis indicated

## Publications

that it is better to start from a well-defined single template than to combine different crystal structures. Thus, we decided to create a subset of several crystal structures to which our sequences were compared and to use the crystal structure with the best sequence alignment to our target (to minimize the number of mutations modelled) as template. By doing so, we reduced the largest number of mutations (in terms of model building) to a maximum of 12 per chain.

Our initial two tests compared the predicted relative binding energy with systems having experimental affinities. As seen in Figures 2 and 3, PELE is capable of distinguishing between resistant and sensitive sequences. Moreover, it shows a good correlation to experimental affinities even for sequences with a large number of mutations, improving existing theoretical prediction tools. Next, we tested our protocol prediction capabilities in a blind test with patient sequences for two well-known HIV-1 inhibitors, APV and DRV. The results showed again noteworthy correlations between PELE's calculated changes in binding energy and the index obtained by HIVdb scores. In all cases we could distinguish the sequences with strong resistance, having predicted relative binding energies  $>5\text{kcal/mol}$ . This value is also consistently being observed in the initial two tests. The method showed problems to discriminate between mutants with moderate resistance, but this is the segment where rule-based methods (used here as reference) are expected to show the poorest performance.

The majority of resistance-related mutations are conservative substitutions among residues Leu10, Val32, Ile54, Val82, Ile84 and Leu90. Our simulations indicated that these mutations do not introduce large changes in the structure of the complex, but modify subtle van der Waals and hydrogen bond interactions between the ligand and active site amino acids (see Figure 5). Such resistance mechanisms are observed for other systems, (Skálová, Dohnálek et al. 2006; Ali, Bandaranayake et al. 2010) and agree with the fact that none of these mutations significantly disrupts enzyme activity. However, even small, the subtle conformational changes induced by mutations are important in defining accurate binding, making it necessary the use of: i) specific homology models, ii) induce fit relaxation. The later point is clearly shown by the fact that PELE's scores after homology modeling (or after only a short exploration of sampling) did not correlate with susceptibility data (Table 1). The quality (and necessity) of the induced fit conformational sampling obtained by PELE is also evident when predicting the interaction

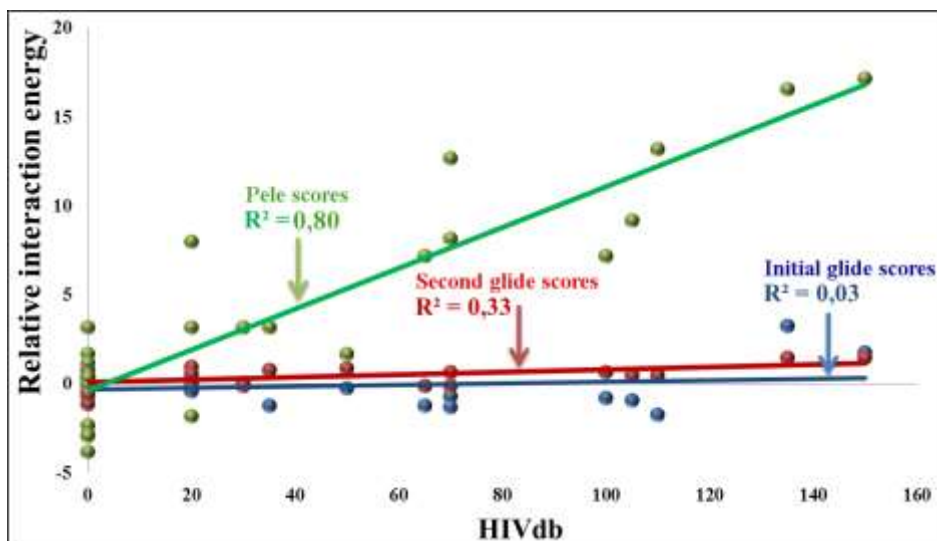
## Publications

with a general scoring function. For this, we used Glide XP from Schrodinger (Halgren, Murphy et al. 2004). Figure 7 shows relative Glide scores before and after the induced fit for the 21 clinical sequences in APV (plus NL4-3 and HXB2). Each score is the average of 50 structures obtained after clustering PELE's trajectories. Clearly the correlation with HIVdb scores significantly improves after conformational sampling with PELE. While the final Glide scores still have low correlation, 0.33, they can identify most of the high resistance sequences. Remarkably, the correlation increases to 0.80 when using PELE's all-atom binding energy (as score), revealing the importance of an explicit treatment of all-atom interactions (Borrelli, Cossins et al. 2010). Notice also that the clustering procedure only reduces the overall PELE's correlation from 0.82 to 0.80.

Table 1. Evolution of the coefficient of determination between PELE's interaction energies and HIVdb values along the conformational sampling for the 21 APV patient sequences.

	First-score	Step-50	Step-100	Step-200	Step-300	Total
$R^2$	0.43	0.64	0.72	0.75	0.81	0.82





**Figure 7.** Comparison between PELE and Glide (initial and final) correlation scores, against HIVdb values, for the 21 APV patient sequences (using 50 representative structures after clustering).

## Conclusions

Overall, this study shows how computational techniques are capable of quantitatively discriminating resistance variants of HIV-1 protease. Our protocol, combining sequence alignment to current pdb structures and state of the art protein-ligand induced fit sampling algorithms, shows great promise as an automatic tool for a quick prediction. The entire protocol can be run in less than 24 hours in a small commodity workstation, and is based on biophysical first principles. Moreover, it is able to trace the effect of novel mutations and the binding of new drugs. In summary, modeling drug-target interactions holds the potential to provide less biased and more accurate assessments of antiretroviral drug resistance, which could improve clinical management of HIV-infected subjects.

## **Acknowledgment**

The authors thank Victor Gil Sepulveda for technical assistance and Mark D. Shenderovich for providing data and helpful discussion. Computational resources were provided by the Barcelona Supercomputing Center. This work was supported by the European Research Council (Grant ERC-2009-Adg 25027-PELE), the Spanish MINECO (BIO2012-32868) and the Severo Ochoa Program (R.S.). MO is an ICREA Academia Researcher.

**Keywords:** HIV-1 PR, PELE, drug resistance, antiretroviral therapy, molecular modeling

## References;

1. Rodger, A. J.; Lodwick, R.; Schechter, M.; Deeks, S.; Amin, J.; Gilson, R.; Paredes, R.; Bakowska, E.; Engsig, F. N.; Phillips, A.; Grp, E. S. *Aids* 2013, 27(6), 973-979.
2. Hogg, R. S.; Heath, K. V.; Yip, B.; Craib, K. J. P.; O'Shaughnessy, M. V.; Schechter, M. T.; Montaner, J. S. G. *Jama-Journal of the American Medical Association* 1998, 279(6), 450-454.
3. Palella, F. J.; Delaney, K. M.; Moorman, A. C.; Loveless, M. O.; Fuhrer, J.; Satten, G. A.; Aschman, D. J.; Holmberg, S. D. *New England Journal of Medicine* 1998, 338(13), 853-860.
4. Condra, J. H.; Schleif, W. A.; Blahy, O. M.; Gabryelski, L. J.; Graham, D. J.; Quintero, J. C.; Rhodes, A.; Robbins, H. L.; Roth, E.; Shivaprakash, M.; Titus, D.; Yang, T.; Teppler, H.; Squires, K. E.; Deutsch, P. J.; Emini, E. A. *Nature* 1995, 374(6522), 569-571.
5. Cohen, M. L. *Science* 1992, 257(5073), 1050-1055.
6. Larder, B. A.; Kemp, S. D. *Science* 1989, 246(4934), 1155-1158.
7. Rhee, S. Y.; Fessel, W. J.; Zolopa, A. R.; Hurley, L.; Liu, T.; Taylor, J.; Nguyen, D. P.; Slome, S.; Klein, D.; Horberg, M.; Flamm, J.; Follansbee, S.; Schapiro, J. M.; Shafer, R. W. *Journal of Infectious Diseases* 2005, 192(3), 456-465.
8. Gilks, C. F.; Crowley, S.; Ekpini, R.; Gove, S.; Perriens, J.; Souteyrand, Y.; Sutherland, D.; Vitoria, M.; Guerma, T.; De Cock, K. *Lancet* 2006, 368(9534), 505-510.
9. Keiser, O.; Orrell, C.; Egger, M.; Wood, R.; Brinkhof, M. W. G.; Furrer, H.; van Cutsem, G.; Ledergerber, B.; Boulle, A.; Shcs; Int Epidemiol Databases, E. *Plos Medicine* 2008, 5(7), 1102-1111.
10. Hamers, R. L.; Wallis, C. L.; Kityo, C.; Siwale, M.; Mandaliya, K.; Conradie, F.; Botes, M. E.; Wellington, M.; Osibogun, A.; Sigaloff, K. C. E.; Nankya, I.; Schuurman, R.; Wit, F. W.; Stevens, W. S.; van Vugt, M.; de Wit, T. F. R.; Paser. *Lancet Infectious Diseases* 2011, 11(10), 750-759.
11. Hamers, R. L.; Sigaloff, K. C. E.; Wensing, A. M.; Wallis, C. L.; Kityo, C.; Siwale, M.; Mandaliya, K.; Ive, P.; Botes, M. E.; Wellington, M.; Osibogun, A.; Stevens, W. S.; de Wit, T. F. R.; Schuurman, R.; PharmAccess African Studies, E. *Clinical Infectious Diseases* 2012, 54(11), 1660-1669.
12. Organization, W. H. WHO HIV Drug Resistance Report 2012; World Health Organization, 2012.

## Publications

13. Johnson, V. A.; Calvez, V.; Gunthard, H. F.; Paredes, R.; Pillay, D.; Shafer, R. W.; Wensing, A. M.; Richman, D. D. *Topics in antiviral medicine* 2013, 21(1), 6-14.
14. Gupta, R. K.; Hill, A.; Sawyer, A. W.; Cozzi-Lepri, A.; von Wyl, V.; Yerly, S.; Lima, V. D.; Guenthard, H. F.; Gilks, C.; Pillay, D. *Lancet Infectious Diseases* 2009, 9(7), 409-417.
15. Sigaloff, K. C. E.; Calis, J. C. J.; Geelen, S. P.; van Vugt, M.; de Wit, T. F. R. *Lancet Infectious Diseases* 2011, 11(10), 769-779.
16. Sigaloff, K. C. E.; Hamers, R. L.; Wallis, C. L.; Kityo, C.; Siwale, M.; Ive, P.; Botes, M. E.; Mandaliya, K.; Wellington, M.; Osibogun, A.; Stevens, W. S.; van Vugt, M.; de Wit, T. F. R.; PharmAccess African Studies, E. *Journal of Infectious Diseases* 2012, 205(11), 1739-1744.
17. Paredes, R.; Marconi, V. C.; Lockman, S.; Abrams, E. J.; Kuhn, L. *Journal of Infectious Diseases* 2013, 207, S93-S100.
18. Kim, E. E.; Baker, C. T.; Dwyer, M. D.; Murcko, M. A.; Rao, B. G.; Tung, R. D.; Navia, M. A. *Journal of the American Chemical Society* 1995, 117(3), 1181-1182.
19. Turner, S. R.; Strohbach, J. W.; Tommasi, R. A.; Aristoff, P. A.; Johnson, P. D.; Skulnick, H. I.; Dolak, L. A.; Seest, E. P.; Tomich, P. K.; Bohanan, M. J.; Horng, M. M.; Lynn, J. C.; Chong, K. T.; Hinshaw, R. R.; Watenpaugh, K. D.; Janakiraman, M. N.; Thaisrivongs, S. *Journal of Medicinal Chemistry* 1998, 41(18), 3467-3476.
20. Tie, Y. F.; Boross, P. I.; Wang, Y. F.; Gaddis, L.; Hussain, A. K.; Leshchenko, S.; Ghoshl, A. K.; Louis, J. M.; Harrison, R. W.; Weber, I. T. *Journal of Molecular Biology* 2004, 338(2), 341-352.
21. Tantillo, C.; Ding, J. P.; Jacobomolina, A.; Nanni, R. G.; Boyer, P. L.; Hughes, S. H.; Pauwels, R.; Andries, K.; Janssen, P. A. J.; Arnold, E. *Journal of Molecular Biology* 1994, 243(3), 369-387.
22. Gubareva, L. V.; Bethell, R.; Hart, G. J.; Murti, K. G.; Penn, C. R.; Webster, R. G. *Journal of Virology* 1996, 70(3), 1818-1827.
23. Borrelli, K.; Vitalis, A.; Alcantara, R.; Guallar, V. *Journal of Chemical Theory and Computation* 2005, 1(6), 1304-1311.
24. Borrelli, K. W.; Cossins, B.; Guallar, V. *Journal of Computational Chemistry* 2010, 31(6), 1224-1235.
25. Shenderovich, M. D.; Kagan, R. M.; Heseltine, P. N. R.; Ramnarayan, K. *Protein Science* 2003, 12(8), 1706-1718.
26. Koh, Y.; Nakata, H.; Maeda, K.; Ogata, H.; Bilcer, G.; Devasamudram, T.; Kincaid, J. F.; Boross, P.; Wang, Y. F.; Ties, Y. F.; Volarath, P.; Gaddis, L.; Harrison, R. W.; Weber, I. T.; Ghosh, A. K.; Mitsuya, H. *Antimicrobial Agents and Chemotherapy* 2003, 47(10), 3123-3129.

## Publications

27. Lindberg, J.; Pyring, D.; Lowgren, S.; Rosenquist, A.; Zuccarello, G.; Kvarnstrom, I.; Zhang, H.; Vrang, L.; Classon, B.; Hallberg, A.; Samuelsson, B.; Unge, T. *European Journal of Biochemistry* 2004, 271(22), 4594-4602.
28. Shen, C.-H.; Wang, Y.-F.; Kovalevsky, A. Y.; Harrison, R. W.; Weber, I. T. *FEBS Journal* 2010, 277(18), 3699-3714.
29. Tie, Y.; Wang, Y.-F.; Boross, P. I.; Chiu, T.-Y.; Ghosh, A. K.; Tozser, J.; Louis, J. M.; Harrison, R. W.; Weber, I. T. *Protein Science* 2012, 21(3), 339-350.
30. Mittal, S.; Bandaranayake, R. M.; King, N. M.; Prabu-Jeyabalan, M.; Nalam, M. N. L.; Nalivaika, E. A.; Yilmaz, N. K.; Schiffer, C. A. *Journal of Virology* 2013, 87(8), 4176-4184.
31. Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. *Journal of Molecular Biology* 1990, 215(3), 403-410.
32. Sastry, G. M.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. *J Comput Aided Mol Des* 2013, 27(3), 221-234.
33. Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. *J Med Chem* 2004, 47(7), 1750-1759.
34. Yu, Z.; Jacobson, M. P.; Josovitz, J.; Rapp, C. S.; Friesner, R. A. *The Journal of Physical Chemistry B* 2004, 108(21), 6643-6654.
35. Espona-Fiedler, M.; Soto-Cerrato, V.; Hosseini, A.; Lizcano, J. M.; Guallar, V.; Quesada, R.; Gao, T.; Perez-Tomas, R. *Biochemical Pharmacology* 2012, 83(4), 489-496.
36. Hosseini, A.; Espona-Fiedler, M.; Soto-Cerrato, V.; Quesada, R.; Perez-Tomas, R.; Guallar, V. *Plos One* 2013, 8(2), 1-8.
37. Liu, T. F.; Shafer, R. W. *Clinical Infectious Diseases* 2006, 42(11), 1608-1618.
38. Eswar, N.; Webb, B.; Marti-Renom, M. A.; Madhusudhan, M. S.; Eramian, D.; Shen, M.-y.; Pieper, U.; Sali, A. In *Current Protocols in Bioinformatics*; John Wiley & Sons, Inc., 2002.
39. Fiser, A.; Do, R. K. G.; Šali, A. *Protein Science* 2000, 9(9), 1753-1773.
40. Mitra, P.; Shultis, D.; Brender, J. R.; Czajka, J.; Marsh, D.; Gray, F.; Cierpicki, T.; Zhang, Y. *PLoS Comput Biol* 2013, 9(10), e1003298.
41. Ali, A.; Bandaranayake, R. M.; Cai, Y.; King, N. M.; Kolli, M.; Mittal, S.; Murzycki, J. F.; Nalam, M. N. L.; Nalivaika, E. A.; Özen, A.; Prabu-Jeyabalan, M. M.; Thayer, K.; Schiffer, C. A. *Viruses* 2010, 2(11), 2509-2535.

## Publications

42. Skálová, T.; Dohnálek, J.; Dušková, J.; Petroková, H.; Hradílek, M.; Souček, M.; Konvalinka, J.; Hašek, J. *Journal of Medicinal Chemistry* 2006, 49(19), 5777-5784.

## Publications

### **3.6 Ligand recognition in steroid hormone receptors: from conserved plasticity to binding mechanism**

Karl Edman, Ali Hosseini, Anders Hogner, Magnus K Bjursell, Anna Aagaard, Stefan Bäckström, Cristian Bodin, Lisa Wissler, Tina Jellesmark-Jensen, Anders Cavallin, Ulla Karlsson, Ewa Nilsson, Daniel Lecina, Ryoji Takahashi, Christoph Grebner, Matti Lepistö & Victor Guallar

In preparation (2014)



## Publications

## **Ligand recognition in steroid hormone receptors: from conserved plasticity to binding mechanism**

Karl Edman, Ali Hosseini, Anders Hogner, Magnus K Bjursell, Anna Aagaard, Stefan Bäckström, Cristian Bodin, Lisa Wissler, Tina Jellesmark-Jensen, Anders Cavallin, Ulla Karlsson, Ewa Nilsson, Daniel Lecina, Ryoji Takahashi, Christoph Grebner, Matti Lepistö & Victor Guallar

### **Abstract**

Steroid hormone receptor drugs have been available for more than half a century, but the ligand binding mechanism has remained elusive. We solved X-ray structures of both the glucocorticoid and mineralocorticoid receptors to identify a conserved plasticity at the interface of helices 3,7 and 11 extending the ligand binding pocket towards the receptor surface. Since none of the endogenous ligands exploit this region, we hypothesized that it constitutes an integral part of the binding event. Extensive all atom unbiased ligand exit and entrance simulations together with structural principal component analysis, corroborate a ligand entry trajectory that gives the observed structural plasticity a key functional role. Our findings reveal why evolution has conserved the capacity to open up this region and provide a new aspect to the selection pressure that has formed this receptor family.

### **Introduction**

Biological functions originate from, and are maintained by, a combination of genomic drift and selection. The traditional method to derive evolutionary relationships is to compare primary sequences, tertiary structures, and protein function. However, while changes in the amino acid sequence and placement of key residues provide useful insights into lineage, this only provides the basic framework for mechanistic detail. A more complete functional understanding requires protein plasticity to be considered. Moreover, comparing protein flexibility of related systems adds an important dimension when exploring evolutionary trajectories (13. NSMB 2013).

## Publications

The steroid receptor family consist of five closely related receptors: the mineralocorticoid receptor (MR), the glucocorticoid receptor (GR), the androgen receptor (AR), the progesterone receptor (PR), and the estrogen receptors (ER $\alpha$  and ER $\beta$ ) (Fig. 1a). They all bind cholesterol derivatives and play a critical role in fundamental biological processes, ranging from pregnancy, early development, to the stress response, and electrolyte homeostasis (1. Evans 1988, 2. Mange 1995). Continual pharmaceutical efforts have resulted in several efficacious drugs, such as prednisolone (GR)(3. Cole 2006), eplerenone (MR) (4. Gravez 2013), bicalutamide (AR) (5. Shelley 2008), drospirenone (PR) (6. Sitruk 2010), and tamoxifen (ER)(53. Br. J. Ph. 2013). However, target class-related side-effects limit the prescription of these drugs in many indications and the scope for further improvement is considered to be high (14. Burris 2013, 15. Bertocchio 2011).

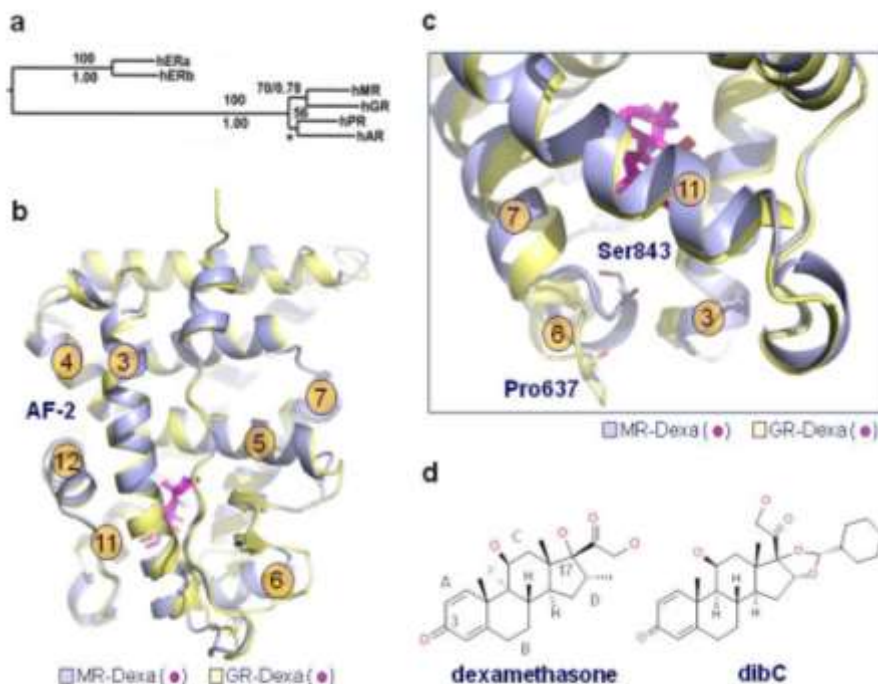
The receptors share a common architecture with three separate domains: the N-terminal domain (NTD), the DNA binding domain (DBD) and the ligand binding domain (LBD). Besides recognizing the ligand pharmacophore, the LBD also contains the activation function-2 (AF-2), which is important for transmitting ligand binding information and partially driving the co-regulator interaction fingerprint (28. Gronemeyer. 2004). In the resting state, the receptors are associated with chaperone proteins in the cytoplasm. Ligand activation leads to a partial release of chaperone proteins, followed almost always by nuclear translocation. In the nucleus, the receptors will dimerize and form ligand and context specific protein complexes, resulting in activation and/or repression of gene transcription.

The increasing number of receptor-ligand X-ray structures has provided valuable understanding of molecular drivers for the different pharmacological responses. All steroid receptor LBD structures exhibit the typical 3-layered alpha helical fold that fully encloses the various compounds in the ligand binding pocket (LBP), Figure 1b (7. Bledsoe 2002, 8. Williams 1989, 9. Fagart 2005, 10. Matias 2000). Within the receptor, specific polar interactions are important determinants of hormone specificity. The combination of polar interactions with shape match between the compound and LBP provides the basis for ligand potency and allosteric modulation of the auxiliary surfaces, primarily AF-2 and the dimerization surface, directing the functional response.

## Publications

When overlaying the steroid receptors, the largest structural difference in proximity to the ligand is located in the region where helices 3, 7 and 11 meet (24. Li 2005). Figure 1c shows a detailed comparison of GR to its paralog MR. An outward tilt of the helix 6–7 (H6-H7) interface in GR results in an expanded ligand binding pocket, which was thought to reflect that the most highly potent GR ligands, such as budesonide and fluticasone furoate, contain a large  $17\alpha$  substituent (11. Li 2005). Despite the smaller pocket in MR, several ligands with large  $17\alpha$  substituents on the steroidal D-ring, such as desisobutyrylciclesonide (dibC, the active metabolite of the pro-drug ciclesonide), are more potent in our MR binding assay than the endogenous agonist aldosterone.

To build a detailed understanding around the plasticity of this region, we determined the high-resolution X-ray structures of MR and GR in complex with both dexamethasone (Dexa) and dibC (Fig. 1d). The structures revealed that with a large  $17\alpha$  substituent, MR is fully capable of adopting an open structural conformation. Why has nature preserved the capacity to open up this region, even though it is not exploited by the endogenous ligands? We propose that the observed plasticity is an integral part of the ligand entry mechanism.



**Figure 1 Evolutionary relationship of the steroid receptors with structural comparison of GR and MR LBD.** (a) Evolutionary relationship of the steroid hormone receptors (ER, MR, GR, PR and AR). (b) GR (yellow) in complex with dexamethasone (magenta) overlaid on MR (lightblue) in complex with dexamethasone (magenta). The AF-2 surface is located where helices 3,4 and 12 meet. (c) Details near the region where helices 3, 7, and 11 meet. (d) The chemical structures of dexamethasone and dibC. The steroidal A, B, C and D rings and positions 3 and 17 are marked on the dexamethasone structure.

To investigate this hypothesis we performed comprehensive all atom simulations. Computer modeling has advanced considerably; specialized hardware and software today can perform microsecond time-scale simulations, overcoming previous sampling deficiencies (17. Dror 2013, 18. Jensen 2012, 36. Shan 2014). Spontaneous ligand binding events have been investigated in exposed active sites for kinases (19. Shan 2011), GPCRs (17. Dror 2013) and proteases (20. Fabritiis 2011). Notably, PELE (Protein Energy Landscape Exploration) (12. Borrelli 2005), a technological development combining Monte Carlo algorithms with protein structure prediction techniques, is capable of performing such studies rapidly for fully occluded binding pockets (21. Madadkar-Sobhani 2013, 22. Takahashi 2014).

This offers a unique opportunity to explore unbiased ligand/protein dynamics of complex systems that require larger structural rearrangements.

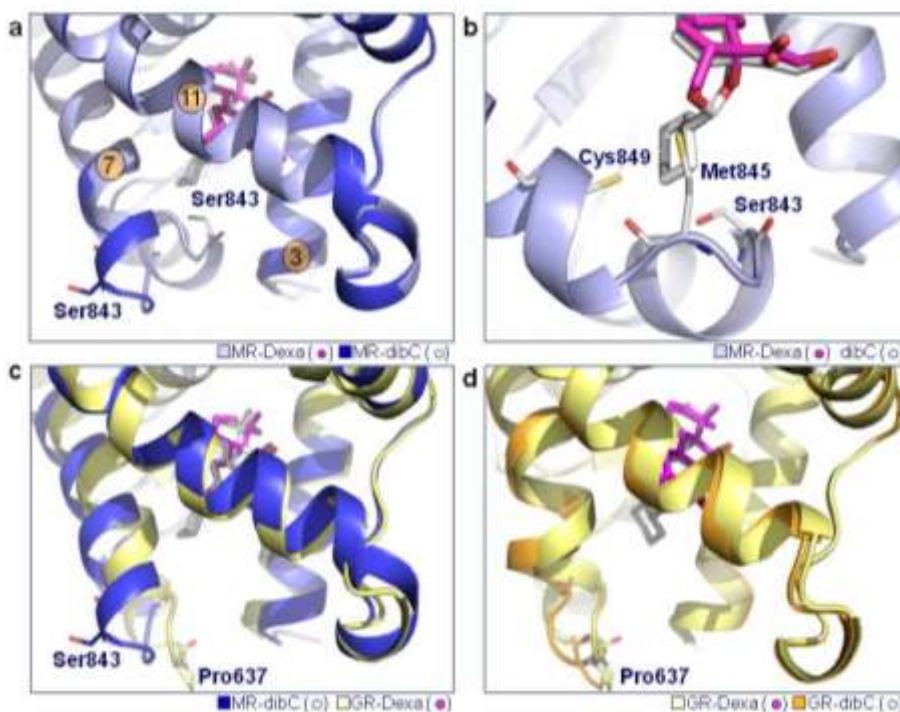
## Results

### A conserved plasticity

Dexamethasone was originally developed as a GR specific agonist (53. Br.J Pharmacol 2013), but was later shown to be a functional MR antagonist with low nanomolar affinity (54. FEBS Lett 1999). The X-ray structure of MR in complex with dexamethasone (MR:Dexa) is similar to the corresponding GR:Dexa structure (normalised RMSD of 0.37 Å for 100 C $\alpha$  atoms). However, examining in detail the region where helices 3, 7 and 11 meet confirms that the 17 $\alpha$  sub-pocket is considerably smaller in the MR structure compared with the GR structure (Fig. 1c). This is reflected in the total volume of the MR:Dexa LBP, which is approximately 393Å<sup>3</sup> compared with 450Å<sup>3</sup> in the GR:Dexa structure. Supplementary Figure 1a shows the 2mFo-DFc electron density from the MR:Dexa LBP,

It has been proposed that structural differences in the loop between Helices 6 and 7 are primarily due to replacement of Ser843<sup>MR</sup> by Pro637<sup>GR</sup>, which alters the geometrical constraints of this region and allows the receptor to adopt a more open conformation (24. Li 2005). However, despite the limited size of the MR sub-pocket, dibC has higher affinity than aldosterone in our scintillation proximity assay (Vangrevelinghe, Zimmermann et al.) using tritiated aldosterone and MR LBD fusion protein (K<sub>i</sub> for dibC is 0.18 nM compared to 1.0 nM for aldosterone). To study the structural flexibility associated with large 17 $\alpha$  substituents, we determined the complex structures of MR:dibC and GR:dibC.

The structure of MR:dibC superimposes well on the MR:Dexa structure (normalised RMSD of 0.28 Å for 100 C $\alpha$  atoms). Moreover, in the LBP, dibC is placed in a nearly identical position to dexamethasone with all polar interactions conserved (Supplementary Fig. 1b). However, while these two receptor conformations are closely related, dibC induces a large rearrangement of the H6-H7 loop region, essentially extending the LBP towards the receptor surface (Fig. 2a).



**Figure 2. Comparison of the complex structures of the MR:Dexa, MR:dibC, GR:Dexa, and GR:dibC.** (a) MR (light blue) in complex with dexamethasone (magenta) overlaid on MR (dark blue) in complex with dibC (white). (b) The cyclohexyl motif of dibC come into direct conflict with residues from H7, enforcing a new structural state. (c) MR (dark blue) in complex with dibC (white) superimposed on GR (yellow) in complex with dexamethasone (magenta). (d) GR (yellow) in complex with dexamethasone (magenta) overlaid on GR (orange) in complex with dibC (white).

Specifically, side chains of Ser843<sup>MR</sup>, Met845<sup>MR</sup> and Cys849<sup>MR</sup> in the MR:Dexa complex occupy the same volume as the cyclohexyl motif of dibC, forcing the receptor to adopt a new conformation (Fig. 2b). This leads essentially to a repositioning of helix 6 and an extension of helix 7. While Ser843<sup>MR</sup> was previously buried within the protein and engaged in a hydrogen bond to the backbone nitrogen of Met845<sup>MR</sup>, it is now exposed to the solvent, forming the new start of helix7 (Fig. 2a).

The size of the 17 $\alpha$  pocket in the MR:dibC complex increases significantly (total LBP volume 539 Å<sup>3</sup>) and the superposition on the GR:Dexa structure shows that this region now adopts a more closely related structural state (Fig. 2c). However, it is interesting to note that GR in complex

## Publications

with dibC (Fig. 2d) expands the 17 $\alpha$  pocket even further (total LBP volume 551  $\text{\AA}^3$ ). While the plasticity in the H6-H7 loop region seems to be conserved across these two receptors, the details of the ligand driven rearrangements are clearly different.

To gain insights of how the flexibility in the H6-H7 region is conserved across the steroid receptor family, we performed principle component analysis (PCA) for all X-ray structures from the public domain (PDB) for each receptor. This technique allows visualization of the variance between structures as a set of normal modes. While the description of this variance will be highly dependent on what regions of the LBP are exploited by the various ligands, the mode describing H6-H7 motion is one of the dominant features (Supplementary Fig. 2). However, for MR the H6-H7 motion is only prominent if we include the MR:dibC structure from this work, emphasizing that the MR:dibC structure describes a novel structural conformation.

### Modeling non biased entry and exit pathways

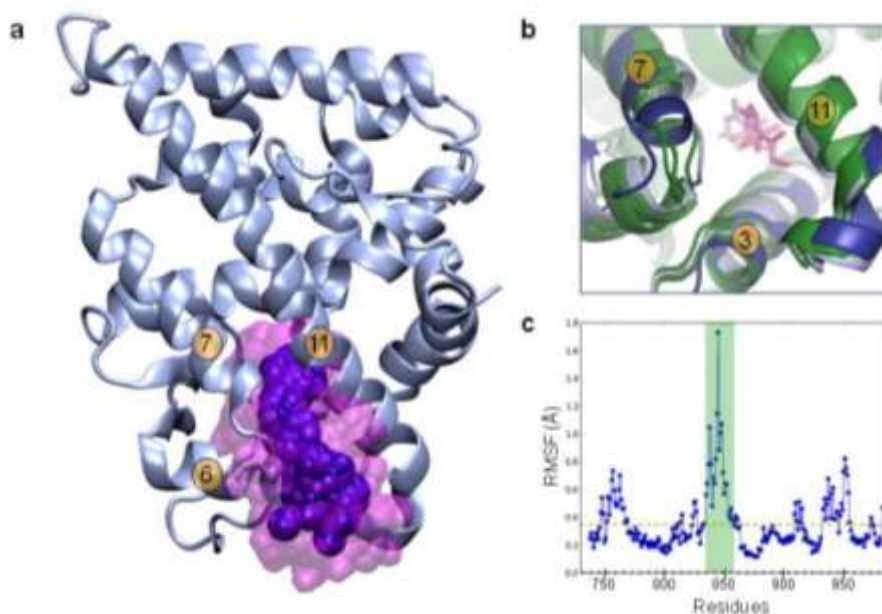
To study the ligand exit and entry pathways, we performed two different types of unbiased simulations. The first protocol explored ligand escape routes using the MR and GR X-ray complex structures as the starting position. In the second protocol, the ligand was randomly placed in the bulk solvent and allowed to freely migrate along the protein surface. All simulations were completed in the presence and absence of a co-factor peptide at the AF-2 site (NCOA1 residues 1430-1441 for MR and NCOA2 residues 741-753 for GR). In addition, both the wild-type protein sequences and the specific mutants present in the X-ray structures were used.

### Ligand dissociation

For all permutations of both MR and GR, we performed three separate exit simulations. In all instances we observed only one exit trajectory perforating the surface where helices 3, 7 and 11 meet. Figure 3a illustrates the MR:Dexa exit pathway simulation with the array of dexamethasone positions superimposed on the initial MR structure. Notably, ligand motion is coupled with significant rearrangement of the protein backbone along the migration pathway. In particular, the loop connecting helices 6-7 is clearly shifted outwards to accommodate ligand release (Fig. 3b). Interestingly, the simulated protein movements are reminiscent of the observed differences between the MR:Dexa and MR:dibC structures shown in light and dark blue, respectively. To better quantify the plasticity in the H6-H7 region, we



calculated the root mean square fluctuations (RMSF) along the exit trajectory. As seen in Figure 3c, the motion in this region is considerably larger than the rest of the protein.



**Figure 3. Ligand exit pathway for the MR:Dexa complex.** (a). The ligand center of mass is highlighted in blue beads, all other atoms of the ligand being shown in transparent spacefill. (b) Detail of the backbone rearrangement along the exit pathway. The MR:Dexa and MR:dibC X-ray structures are shown in light and dark blue, respectively, with dexamethasone in the LBP in magenta. Three cartoon snapshots from the exit simulations are shown in green. (c) Residues RMSF fluctuations against the average structure along the MR:Dexa exit pathway plotted for each residue. The dotted line denotes the average RMSF across the LBD. Helices 6 and 7 are marked with green shade.

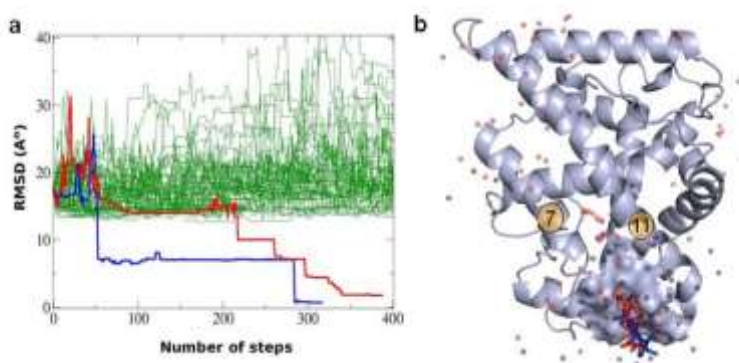
Supplementary Figures 3, 4 and 5 show the equivalent exit pathways for the GR:Dexa, MR:dibC, and GR:dibC. Based on the complete set of ligand dissociation simulations it is clear that both MR and GR have the same ligand exit pathway. In addition, while ligand exit is associated with similar protein motions, the fluctuations in the H6-H7 region are significantly larger for MR than for GR. This is in agreement with the idea that GR would require smaller rearrangements as the receptor is more open to begin with.

### Ligand association

To investigate ligand entry into MR and GR, we randomly placed dexamethasone in the bulk solvent and released it to freely probe the protein surface. For each receptor we performed five runs with 64 independent trajectories over 48 hours. Each run yielded 1-2 trajectories where the ligand entered the LBP. In all runs the ligand is free to move without any predefined search direction.

Figure 4a shows the evolution of the distance between the ligand's center of mass and the binding site (obtained from the ligand's center of mass at its crystal bound structure) for one of the MR:Dexa runs. When studying the distance progression in detail, it is clear that most of the trajectories explore the receptor surface with some excursions into the bulk solvent. However, the blue and red trajectories enter the LBP at steps ~50 and ~210, respectively. Figure 4b shows the ligand center of mass along these trajectories superimposed on the initial protein structure with the entry to the LBP denoted by a surface representation. Supplementary Movie shows the simulation of ligand entry into MR where dexamethasone was initially placed in the bulk solvent.

While the entry along the blue trajectory is relatively fast, the red trajectory demonstrates the non-biased nature of the simulation, probing a large portion of the receptor surface before finding the entrance pathway. In keeping with the ligand escape simulations for all runs in both systems, trajectories entering the LBP pierce the protein surface at the H3-H7-H11 junction (the corresponding figure for a GR:Dexa run is shown in Supplementary Fig. 6).



**Figure 4. Unbiased simulation of dexamethasone entering MR.** (a) Each line represents the ligand's RMSD (heavy atom) to the bound crystal ligand for a different trajectory. Two of the trajectories represented by blue and red lines enter the LBP at step 52 and 214, respectively. (b) The ligand center of mass for the two trajectories that enter the LBP are shown as red and blue spheres. The region where the ligands enter the LBP is emphasized as a surface with two ligands shown in stick representation.

While the mutants used in the X-ray structures did not influence the simulations significantly, removal of co-factor peptide at the AF-2 resulted in larger fluctuations in both the helix 12 and the H3-H7-H11 junction along the exit and entrance trajectories; the ligand migration pathway remained unchanged. Supplementary Figure 7 shows in detail the larger amplitude motion of helix 12 in absence of the co-regulator peptide.

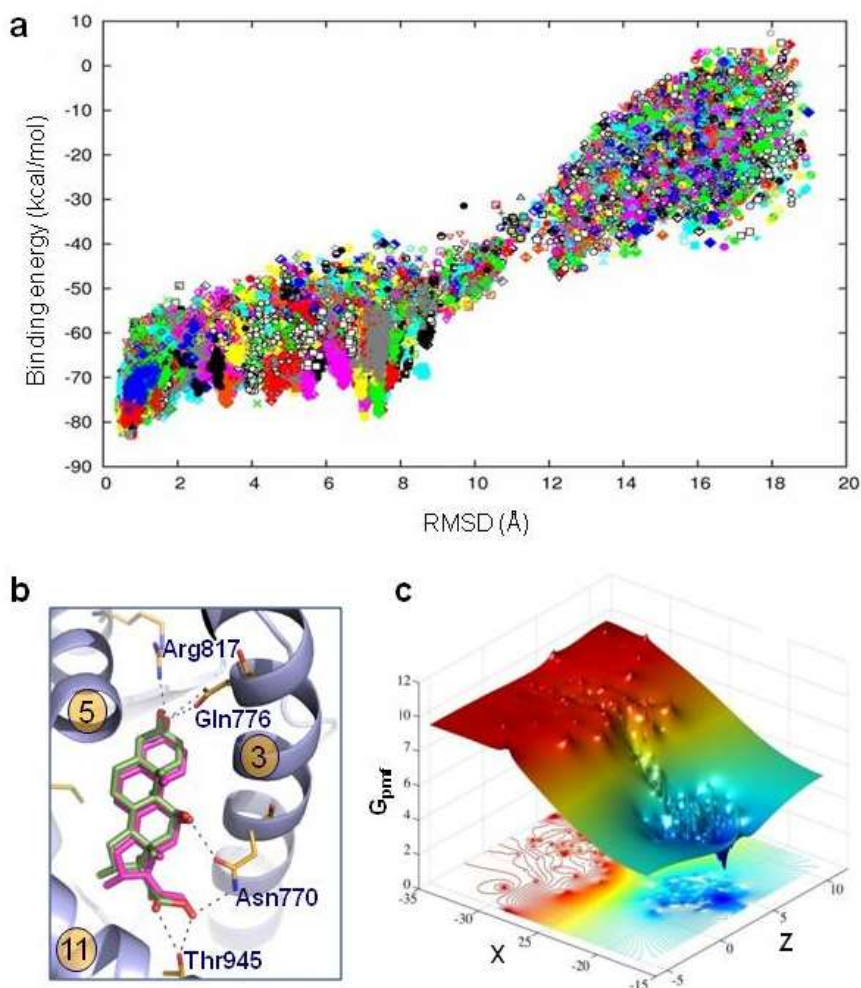
#### Active site ligand refinement and binding free energy.

Once the entrance path to the MR LBP had been located, we refined the free search with local enhanced sampling to obtain a precise pose for the best binder. This procedure does not add any bias in the ligand search direction, but it limits the sampling to the region around the entrance point (typically 10-15 Å). Figure 5a shows the interaction energy profile plotted against the ligand (heavy atom) RMSD to the bound crystal structure for the MR:Dexa refining process (400 trajectories). The lowest binding energies are derived from poses located within 0.75 Å RMSD of the X-ray ligand conformation. The sampling places dexamethasone in the accurate orientation with the A-ring 3-keto moiety pointing toward the Arg817<sup>MR</sup>-Gln776<sup>MR</sup> pair from helices 5 and 3, and the D-ring hydroxyacetyl approaching the Asn770<sup>MR</sup> on the N-terminal half of helix3 (Fig. 5b). Studying the protein-ligand interaction energy plot in more detail (Fig. 5a), it is interesting to note that the surface

## Publications

exploration exhibit a local minima at about 12Å from the LBP. In the crystal structure of GR:Dexa and GR:dibC, this site is occupied by a steroid-like CHAPS molecule that is part of the crystallization condition (Supplementary Fig. 8). In addition, for MR a non-steroidal antagonist has been observed at this position (27. Hasui .2011). It is tempting to speculate that it may correspond to a peripheral binding site at the H3-H7-H11 junction and that the energy barrier located at the 11-12Å segment reflects the energy cost associated with the surface crossing event.

Since sampling along the local refinement process is fast (each trajectory running in 12-24 hours in a single core), it facilitates running hundreds of trajectories. Based upon Markov State Model (MSM) analysis, we used this data to calculate the absolute binding free energies for MR:Dexa and MR:dibC (22. Takahashi . 2014). While absolute values might be slightly shifted due to the absence of an exhaustive surface/bulk exploration, relative values should be in reasonable agreement, because both ligands share entry point and binding site. Figure 5c shows a 2D projection of the potential mean field (PMF) obtained for MR:Dexa along the 400 refinement trajectories. The red area corresponds to the bulk exploration whereas the global minimum, shown in blue, corresponds to ligand positions near the crystallographic structure, at 0.75 Å heavy atom ligand RMSD as seen in Figure 5a. Integration of the PMF volume at the active site gives a binding free energy of -7.5 kcal/mol for dexamethasone and -9.3 kcal/mol for dibC. The difference in binding free energy of 1.8 kcal/mol is in quantitative agreement with the experimental difference of 2.09 kcal/mol (derived from the  $K_i$  values of 6.3 nM for dexamethasone and 0.18 nM for dibC).



**Figure 5. Refined ligand binding simulations and estimated binding free energy** (a) The protein-ligand interaction energy (kcal/mol) plotted against the ligands (heavy atoms) RMSD to the bound crystal along the 400 refinement trajectories in MR:Dexa. (b) MR (blue) in complex with dexamethasone (magenta) overlaid on the lowest interaction energy structure after the refined exploration (Kempf, Marsh et al.). (c) X-Z 2D projection of the PMF obtained in the MSM analysis for the same process.

### The evolutionary view

The ligand entry and exit mechanism establishes a functional role for helices 6 and 7 as a gatekeeper. As GR is constitutively open due to the Ser843<sup>MR</sup> to Pro637<sup>GR</sup> substitution, one would expect evolution to leave a

## Publications

differential signature on the protein sequences. Provided that the residues are not involved in any direct downstream protein-protein contacts and that the influence of the potential peripheral binding site is minimal, we hypothesized that the H6-H7 amino-acid sequence in GR should be less constrained, because ligand entry does not require a significant rearrangement of this region. In contrast, the other steroid receptors must be able to efficiently switch between a more open and closed conformation. To explore this, sequence clusters for each receptor were downloaded from the OrthoDB database (64. Waterhouse 2013) by searching for the human gene and selecting the vertebrate subset. The sequences for each receptor were then aligned using ClustalX version 2.0 (65. Larkin 2007). Each position was then assigned a variability score based upon the number of different amino acids at that position across the various species (i.e. if one position is perfectly conserved in all species it would get a score of 1). For each receptor the average variability score of the H6-H7 region was then compared to the average score of the LBD (Table 1).

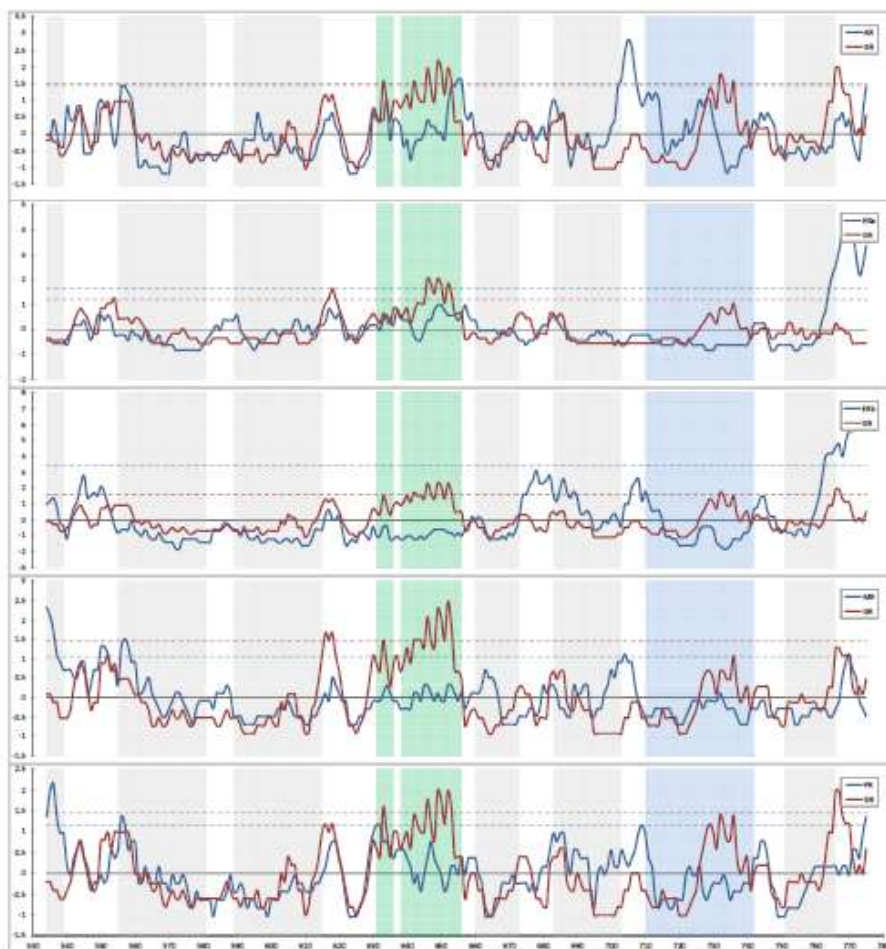
	GR	MR	PR	AR	ER $\alpha$	ER $\beta$
LBD	2.2	2.1	2.1	2.5	1.9	3.6
H6-H7	4.0	1.8	2.8	3.2	2.4	3.3

**Table 1.** Variability score across the steroid receptor family. The residues included in calculating the average for the LBD are selected based on a structural overlay of the receptor family. The H6-H7 region was defined as the respective sequence corresponding to Glu837<sup>MR</sup>-Leu848<sup>MR</sup>, selected based upon a structural overlay.

In this broad selection of receptor sequences from different species, it is encouraging to see that GR has the highest variability score in the H6-H7 region relative the average across the receptor LBD. However, it is important to note that the initial selection from the OrthoDB yields different number of species sequences for the various receptors.

To be able to compare the variability across the receptors in better detail, we then looked at the species overlap of GR with each of the other receptors in turn. The species that were present for both receptors were selected. However, sequences that exhibited less than 20% homology to the human H6-H7 region were filtered away for each receptor, to remove all dubious sequences from the comparison. Using this set a new variability score was calculated. All receptor sequences were overlaid on the GR sequence

using X-ray structures to define the equivalent positions. Finally we plotted the variability score against the amino acid sequence for all receptor pairs (Fig. 6). To reduce the noise in the graphs, we averaged the variability scores for each position using a 5 amino acid sliding window.



**Figure 6. Evolutionary conservation of the LBD for the steroid receptors.** The graphs show average normalized and smoothed (sliding window of 5 AA) amino acid variability score for pairwise comparisons of AR, ER $\alpha$ , ER $\beta$ , MR, PR (blue) vs GR (red) plotted against the GR amino acid sequence. High variability scores indicate less conservation. Helices 1-12 are demarked using vertical bars (green: H6-7; blue: H10-11; gray: all others;).

Figure 6 confirms that important structural elements of the receptors are relatively conserved. For example, the AF-2 surface (H12, H4 and the c-terminal end of H3), which is directly involved in the protein-protein

## Publications

interaction transmitting the ligand activation signal, shows a consistently low variability score for all receptors. However, H6-7 shows greater variability in GR relative to all other receptors. This supports our hypothesis that this region has a differential selection pressure across the family. Interestingly, GR also has a segment of higher variability near the C-terminal end of H11. This region sits directly across from the N-terminal end of H7 (Fig. 1c) and it is conceivable that amino acid sequences of these regions may well co vary with each other. Finally, GR appears to be more conserved than the other receptors near the loop between H9 and H10. The functional rationale behind this observation remains to be determined.

## Discussion

The fundamental role and mechanism of action of steroid receptors have been studied extensively over several decades, yet the details of the ligand recognition and binding mechanisms have remained unclear.

### A conserved structural plasticity

By comparing the structures of MR and GR in complex with dexamethasone and dibC, we identified an intrinsic capacity to open up the H6-H7 region. While the GR:Dexa structure adopts an open conformation compared with the MR:Dexa complex, the MR:dibC structure is able to extend the ligand binding pocket significantly and adopt a structural state akin to the GR:Dexa arrangement. The GR:dibC complex shows that the local flexibility can be pushed even further. The structural differences between GR and MR have been attributed to the change from Ser843<sup>MR</sup> to Pro637<sup>GR</sup> (24. Li 2005). Comparing the MR:Dexa structure with MR:dibC, Ser843<sup>MR</sup> is shifted from an internal position in the ligand binding pocket to a position at the protein surface at the tip of helix 7. Recent data suggests that phosphorylation of this residue affects both ligand binding and receptor translocation into the nucleus (29. Shibata 2013). The structural changes observed here explain how the receptor may use the local plasticity to make Ser843<sup>MR</sup> available for modification.

Studies of the ancestral corticoid receptor, AncCR, revealed that the Ser106<sup>AncCR</sup> (corresponding to Ser843<sup>MR</sup>) to Pro637<sup>GR</sup> switch was a permissive mutation that facilitated a subsequent Leu111<sup>AncCR</sup> (corresponding



## Publications

to Leu848<sup>MR</sup>) to Gln642<sup>GR</sup> mutation (30. Bridgham Science 2006). This is an example of conformational epistasis and has played an important role for the evolution of the GR hormone selectivity (30. Bridgham Science 2006, 31. Ortlund Science 2007). As both GR and MR demonstrate a similar capacity to form the open conformation, it is likely that the AncCR also exhibited the same flexibility. Hence, as GR evolved from AncCR, the Ser<sup>AncCR</sup> to Pro637<sup>GR</sup> mutation would primarily serve to select a subset of pre-existing structural states, rather than creating a completely new arrangement. The importance of conformational selection over induced fit has provided mechanistic insights for several biological systems (32. Changeux Nature 2013); it is conceivable that evolution through mutation operates in an analogous way.

The structure-activity relationship (SAR) of 17- $\alpha$  substituted PR, AR and ER ligands (35. Andrieu 2015, 56. Nettles EMBO 2007) to the different steroid receptors suggests that plasticity near the H3-H7-H11 interface is conserved in the steroid receptor family. Indeed, the PCA using all structures in the public domain, confirms that the ability to rearrange the H6-H7 region is one of the most prominent features across the family. However, in the bound state none of the endogenous ligands exploit this region. We propose that the flexibility near the interface of H3-H7-H11 has been conserved throughout evolution due to its essential role in the ligand binding event.

### The ligand entry and exit pathways

We performed two sets of unbiased simulations exploring both ligand binding and dissociation using the protein–ligand sampling code PELE (21. Madadkar-Sobhani 2013). In all simulations, entry and exit trajectories pass through the H3-H7-H11 junction. Protein conformational changes along these trajectories reflect the rearrangements observed when comparing the closed *versus* the open crystal structures of MR and GR. Specifically, the outward bending motion of the H6-H7 region is qualitatively similar to the observed perturbations caused by the large 17- $\alpha$  cyclohexyl substituent in the dibC complex structures. These results indicate that large-amplitude protein motions of helix 12, as suggested by apo and holo crystallographic nuclear hormone receptors (37. Moras 1998, 38. Yen 2001, 39. Brzozowski 1997), are not required for ligand entry. Instead, the conformation of the LBD is likely to resemble the ligand bound agonistic conformations of the receptors during the ligand entry step (25. Capelli 2013, 40. Batista 2013). We confirm that small scale vibrations combined with a structural rearrangement of H6-H7 region are enough to identify an energetically favorable pathway to allow the ligands

## Publications

to diffuse into the LBP. In contrast to other modeling studies using biased protocols, we do not observe multiple ligand entry or exit pathways (25. Capelli 2013, 41. Sonoda 2008, 42. Aci-Sèche 2011). It is conceivable that even though there is likely to be cross-talk between the different regions of the receptors, physically separating the ligand entry plasticity from the structural modulation of the AF-2 surface, would provide greater freedom to create a more versatile genomic response.

With the modeling results in hand, we could ascribe a mechanistic role to the H6-H7 plasticity. Further investigations of the ancestral receptors revealed that the AncCR ligand selectivity profile could not be recovered with direct reversal of key amino acids of GR to their ancestral states, because of additional epistatic mutations of residues near H7. (33. Bridgham Nature 2009). This is in agreement with our hypothesis that since GR has a constitutively open structure, it is probably more forgiving for mutations in the H6-H7 area. Attempting to reverse evolution from GR to AncCR would enforce more specific constraints on H6-H7 dynamics for functional ligand entry, because the new protein must be able to efficiently switch in between the open and the closed state. This is corroborated by our findings that the mutational frequency of the GR H6-H7 region is significantly higher than for the corresponding region in the other steroid receptors.

It is firmly established that steroid receptors depend on a number of chaperone and co-chaperone proteins for correct folding capable of high-affinity hormone binding (44. Grad 2007). Although the ligand entry function is likely to have evolved before the synergies with chaperone proteins, these proteins will nevertheless limit the access to the receptors and thereby form boundary conditions for any ligand entry hypothesis. Mutation and peptide competition studies suggest that Hsp90 is interacting at the AF-2 surface (45 Ricketson 2007, 46. Fang 2006). In addition, co-chaperones have been mapped to interact with regions surrounding the C-terminal end of H1 and the N-terminal end of H3 (48. Caamaño 1998), and with the loop that connects them (49. Cluning 2013). Taken together, these observations contradict the idea that ligands would enter the LBP through a structural rearrangement of H12 (37. Moras 1998, 38. Yen 2001, 39. Brzozowski 1997). In our studies, the ligand entry trajectory was not affected by removal of co-regulator peptide, which allowed H12 to move more freely. While none of the suggested chaperone and co-chaperone interaction surfaces overlap with our

## Publications

binding trajectory, the proximity provides a good rationale for why chaperone binding could directly facilitate ligand association.

### The energy landscape

After the free ligand exploration of the MR and GR surfaces identified a common passage from the solvent into the LBP, a local enhanced sampling of the MR binding event allowed a detailed view of the binding energy landscape associated with ligand entry. The simulations result in productive binding with the lowest protein-ligand interaction energy having an RMSD of  $\sim 0.75$  Å from the X-ray complex (Fig. 5b). Careful analysis of the binding energies along the entry trajectory revealed a local minima  $\sim 12$  Å away from the LBP. This site overlaps with small molecule binding observations in both MR and GR X-ray structures. While the peripheral binding site could be a crystallization artifact, our simulations indicate a local minimum functioning as a pre-docking site, increasing the local concentration of the ligands near the access channel to the LBP. The exhaustive local sampling also facilitated estimation of the relative binding free energies of dexamethasone and dibC to MR. While we have only investigated a limited set of compounds, these values show remarkable agreement with experimental values; accurate prediction of relative binding free energies is a topic of great interest to the pharmaceutical industry.

The tremendous growth in the number of available X-ray structures from increasingly more advanced protein classes and complexes provides a plethora of snapshots of molecular mechanism in action. However, to bridge the gap to detailed mechanistic insights, and to establish evolutionary relationships between different functions, orthogonal data from biochemical experiments and *in silico* modeling are required. Based on information from several X-ray structures, unbiased simulations and bioinformatic analysis, we have uncovered the ligand binding mechanism into the occluded LBP of steroid hormone receptors. While it is difficult to derive any mechanistic details from primary protein sequences, the ligand binding mechanism enabled us to focus in on the H6-H7 region and to discover that the evolutionary pressures to maintain the ligand entry function had left a differential fingerprint on the amino acid sequences for different species across the steroid receptor family. Ligand binding to the steroid receptors marks the first step in a chain of events that in the end triggers both broad genomic and non-genomic mechanisms. Understanding the details of ligand association and dissociation may facilitate the rational design of molecules

## Publications

that exploit the plasticity of the entry and exit processes to a greater extent. This could yield ligands with different modes of action, such as antagonists that block nuclear translocation or agonists with extended receptor occupancy and a prolonged pharmacological response.

## Methods

### **Protein expression and purification**

#### **GR:Dexa**

The cDNA sequence encoding the human glucocorticoid receptor ligand binding domain (NR3C1; aa500-777, GR-LBD) with three introduced mutations N517D, F602S, C638D and an N-terminal 6-histidine tag followed by a thrombin cleavage site was cloned into a pFastBac-HTb vector (Life Technologies). Recombinant baculovirus was generated using the Bac-to-Bac expression system (Life Technologies) and High Five cells (Life Technologies) were infected followed by suspension culture in Express Five medium (Gibco) for 48h at 27°C, the last 24h in the presence of 10 µM dexamethasone, after which cells were collected by centrifugation.

All protein purification steps were performed at 4°C. Cells were lysed in buffer A (50 mM Tris pH 8.0, 2.5 mM DTT, 1% CHAPS, 50 µM dexamethasone, 10% glycerol) supplemented with Complete EDTA-free protease inhibitor cocktail (Roche) followed by affinity purification using Ni-NTA beads (Qiagen). Protein was eluted in buffer A supplemented with 150 mM NaCl and 300 mM imidazole, and subjected to size exclusion chromatography using a HiLoad 26/60 Superdex 200 gel filtration column equilibrated in buffer A. Five-fold molar excess of a TIF2 peptide, KENALLRYLLDK (Innovagen) was added, the N-terminal 6-histidine tag was removed using thrombin-agarose (Sigma) and subsequently the free 6-histidine tag was removed. The protein was thereafter passed over a Q Sepharose fast-flow ion-exchange column (GE Healthcare) equilibrated in buffer A and stored at -80 °C. Approximately 5.4 mg protein was obtained from 10 L High Five cells.

## **GR:dibC**

A pFastBac (Invitrogen) construct encoding human GR-LBD (amino acids 500-777) with the mutations N517D, V571M, F602S, C638D and an N-terminal, thrombin cleavable 6-His tag was used to generate baculoviruses in Sf9 cells (Invitrogen). GR-LBD encoding viruses were used to infect High Five cells (Invitrogen) at a density of 2-3x10E6 cells/ml and a MOI of 3 in a Wave Bioreactor at 27°C. 24 hours post-infection, dexamethasone was added to a final concentration of 10 µM. The cells were harvested by centrifugation 48 hours post-infection, washed in PBS and stored at -80°C until lysis.

Cells were resuspended in lysis buffer (50 mM Tris-Cl pH 8.0, 10% glycerol, 1% CHAPS, 2.5 mM DTT, Complete EDTA-free protease inhibitor cocktail (Roche) and 50 µM dexamethasone) and lysed by 5x1 min passes in a polytron homogeniser. The cell-lysate was clarified by centrifugation at 18500 g for 90 minutes and batch-bound to Ni-NTA Superflow (Qiagen) for 1.5 hours at 4°C. The IMAC resin was packed in a column, washed with wash buffer (50 mM Tris pH8.0, 60 mM NaCl, 30 mM imidazole, 10% glycerol, 1% CHAPS, 2.5 mM DTT and 50 µM dexamethasone) and GR-LBD was step eluted with elution buffer (50 mM Tris pH 8.0, 30 mM NaCl, 300 mM imidazole, 10% glycerol, 1% CHAPS, 2.5 mM DTT and 50 µM dexamethasone).

The eluate was loaded on a HiLoad 26/60 Superdex 200 size exclusion column equilibrated in gel filtration buffer (50 mM Tris-Cl pH 8.0, 10% glycerol, 1% CHAPS, 2.5 mM DTT and 50 µM dexamethasone). GR containing fractions were pooled and a 3-fold excess of co-activator NR-box peptide (KENALLRYLLDK, human NCoA2, residues 740-751) was added. The His-tag was cleaved over night at 4° C with Thrombin-agarose (Sigma) and removed by negative IMAC using Ni-NTA. The protein was finally polished through Q Sepharose FF (GE Healthcare) equilibrated in gel filtration buffer, flash-frozen in liquid nitrogen and stored at -80°C.

## **MR**

Human MR-LBD (amino acids 735-984) with the mutations C808S, C910S (and S810L in the case of dibC), an N-terminal, TEV cleavable 6-HN tag and a C-terminal thrombin cleavable co-activator peptide PQAQQKSLQQLLTE (residues 1427-1441 of the Nuclear Receptor

## Publications

Coactivator 1 (NCoA1) was codon optimized and cloned into pET24a(+). Recombinant human MR-LBD was expressed in *Escherichia coli* BL21 Star™ (DE3) (Invitrogen) cells. Cells were grown in terrific broth (TB) at 37°C until OD<sub>600</sub>=0.5-1.0, chilled on ice for 30 minutes and 100 µM of dexamethasone (Alfa Aesar) or dibC was added. Cells were shaken at 16°C for 30 minutes before protein production was induced using 0.1 mM isopropyl β-D-thiogalactopyranoside (IPTG) for an additional 24-48 hours.

After harvest, cells were resuspended in lysis buffer (30 mM Na-Hepes pH 7.5, 150 mM NaCl, 20 mM imidazole, 100 mM arginine-Cl, 10% glycerol, 1% CHAPS, 1 mM TCEP) containing 20 µM of respective ligand, EDTA-free Complete protease inhibitor cocktail (Roche) and 0.05 g/ml of CelLytic™ Express (C1990, SIGMA). Cells were lysed by rotation at room-temperature for 15 minutes. The lysate was cleared by centrifugation at 48000 g for 20 minutes and loaded onto Ni-Sepharose FF (GE Healthcare) equilibrated in lysis buffer. After washing, protein was step eluted by the addition of one column volume (CV) of lysis buffer containing 0.5 M Arginine-Cl followed by 5 CV of elution buffer (30 mM Na-Hepes, pH 7.5, 150 mM NaCl, 500 mM imidazole, 500 mM arginine-Cl, 10% glycerol, 1% CHAPS, 1 mM TCEP, 20 µM of respective ligand). Size exclusion chromatography was performed on a HiLoad Superdex 200 column (GE Healthcare) equilibrated in 20 mM Na-Hepes pH 6.7, 150 mM NaCl, 0.5 M arginine-Cl, 10% glycerol, 0.1% CHAPS, 1 mM TCEP and 2 µM dexamethasone or dibC.

Finally, MR-LBD co-expressed with dexamethasone was diluted 10x in 20 mM Tris-HCl pH 8.0, 10 mM CaCl<sub>2</sub>, 20 µM dexamethasone, cleaved with TEV protease and Thrombin CleanCleave Kit (SIGMA), purified by reverse IMAC on Ni-Sepharose FF and concentrated to 15 mg/ml. MR-LBD co-expressed with dibC was diluted 15x in 10 mM Tris-HCl pH 8.5, 20 µM dibC, 1mM TCEP and concentrated to 7 mg/ml.

## Crystallization

### **GR:Dexa**

A tube with 1.0 mg of GR(500-777)N517D, F602S, C638D was thawed and washed three times in the concentrator tube with 3.5 ml of 10 mM Tris pH 8.5, 2.5 mM DTT, 45µM dexamethasone. A fivefold molar excess of co-activator NR-box peptide (KENALLRYLLDKDD, human NCoA2, residues 740-753) was added and the complex was concentrated to 9 mg/ml.

## Publications

Crystals were grown at 4°C in hanging drops using 1 µl of protein and 1 µl of well solution (10% PEG8000, 10% ethylene glycol and 0.1 M HEPES pH 7.5). Crystals were frozen in liquid nitrogen with 20% ethylene glycol as cryo protectant prior to data collection.

### **GR:dibC**

A tube with 5.0 mg's of GR(500-777)N517D, V571M, F602S, C638D was thawed and concentrated to about 1.5 ml. The protein was washed three times in the concentrator tube with 10 ml of 10 mM Tris pH 8.5, 2.5 mM DTT (buffer B) to remove excess of dexamethasone and thereafter diluted to a final volume of 6 ml. dibC was added to a final concentration of 0.25 mM to boost ligand exchange prior to dialysis. Dialysis was performed using two Slide-A-Lyzer dialysis cassettes in a beaker containing buffer B and 60 µM of dibC. Dialysis solution was exchanged after 20, 28 and 46 hours before harvesting the sample. The protein was concentrated to 1 ml and buffer was exchanged to fresh buffer B using a NAP10 column. A twofold molar excess of co-activator NR-box peptide (KENALLRYLLDKDD, human NCoA2, residues 740-753) was added and the complex was concentrated to 9 mg/ml.

Crystals were grown at 4°C in hanging drops using 2 µl of protein and 1 µl of well solution (10% PEG8000, 20% ethylene glycol and 0.1 M HEPES pH 7.5). Crystals appeared as rod like crystals after 1-2 days but continued to grow for one to two weeks. Crystals were frozen in liquid nitrogen without any cryo protectant prior to data collection.

### **MR:Dexa**

Crystals of MR-LBD co-expressed and purified with dexamethasone were grown by sitting drop vapor diffusion in 30% PEG4000, 0.1 M NaCl, 0.2 M Pipes pH 7.4. The crystals appeared after one day and grew to its full size after three days.

Crystals were cryo-protected by transfer to a cryo solution of well solution supplemented with 20% glycerol and flash frozen in liquid nitrogen prior to data collection.

## **MR:dibC**

Crystals of MR-LBD co-expressed and purified with dibC were grown by sitting drop vapor diffusion in 18% PEG4000, 0.14 M LiSO<sub>4</sub>, 85 mM Tris pH 8.5, 15% glycerol. The crystals appeared after ten days and continued to grow for several weeks. Crystals were flash frozen in liquid nitrogen prior to data collection.

### **Data collection and Structure determination**

The MR:Dexa data were collected using an Rigaku FRE rotating anode (wavelength 1.54 Å). The GR:Dexa data were collected at the ID14:4 beam line at the ESRF (wavelength 0.94 Å). The MR:dibC and GR:dibC data were collected at the ID29 beam line at the ESRF (wavelength 0.98 Å). All data sets were collected from a single crystal at 100K. The MR data sets were integrated with XDS (57. Kabsch 2010) and the GR data sets were integrated with Mosflm (58. Leslie 2007). All data sets were merged with SCALA (59. Evans, 2006) from the CCP4 suite (60. Collaborative Computational Project, Number 4). The MR and GR structures were solved with PHASER (61. McCoy 2007) using PDB entry 2AA2 and 1M2Z as starting models, respectively. The structures were refined using the BUSTER (62. BUSTER) and manual rebuilding using Coot (63. Emsley 2004). The GR:Dexa structure had 1 (0.39%) Ramachandran outlier while the other structures did not have any outliers. All figures were prepared using PyMOL ([www.pymol.org](http://www.pymol.org)). Crystallographic coordinates and structure factor amplitudes have been deposited into the protein data bank (MR:Dexa 4uda , MR:dibC 4udb , GR:Dexa 4udc , GR:dibC 4udd).

### **Mineralocorticoid receptor ligand competition binding assay**

The human mineralocorticoid receptor ligand binding domain (NR3C2; aa729-984) with an N-terminal maltose binding protein (MBP) tag was expressed using the Bac-to-Bac expression system (Life Technologies). High Five cells were co-infected with recombinant P23 co-chaperone baculovirus followed by suspension culture in Express Five medium (Gibco) for 48h at 27°C. Cells were lysed in lysis buffer (10 mM Tris-HCl pH 7.4, 0.5 mM EDTA, 2.5 mM DTT, 10% glycerol, 20 mM Na<sub>2</sub>MoO<sub>4</sub>, Complete protease inhibitor (Roche)) followed by centrifugation and the supernatant was stored at -80°C. Compound binding was assessed using a ligand competition binding scintillation proximity assay (Vangrevelinghe,



## Publications

Zimmermann et al.). Compounds were incubated with MR-High Five cell lysate (7 $\mu$ g/ml) and 5 nM  $^3$ H-aldosterone (Perkin Elmer (NET419250UC) in assay buffer (10 mM Tris-HCl, 0.5 mM EDTA, 20 mM Sodium molybdate dehydrate, 10 % Glycerol, 0.1 mM DTT) for one hour before addition of 2.5 mg/ml anti-rabbit SPA PS beads (Perkin Elmer RPNQ0299) and 2  $\mu$ g/ml rabbit anti-MBP antibodies (Abcam ab9084) followed by incubation at room temperature for 8 hours before detection of signal using a LeadSeeker imaging system (GE Healthcare).

### **Sequence homology analysis**

Sequence clusters for each receptor were downloaded from the OrthoDB database (64. Waterhouse 2013) by searching for the human ENS gene ID and selecting the vertebrate subset. For each receptor, the sequences were filtered to remove sequences with a length two standard deviations below the average or that contained more than 100 'X' (unknown amino acids). The sequences for each receptor were then aligned using ClustalX version 2.0 (65. Larkin 2007), then further filtered to only keep sequences with an intact H6-7 region (max 1 indel or 'X' and  $\geq$  20% homology to the human H6-7 sequence). In order to remove bias stemming from the inclusion of sequences from different species, subsets were generated where the same species were included for pairs of GR with either of (AR, ER $\alpha$ , ER $\beta$ , MR, PR). The paired subsets were realigned and the resulting alignments were analyzed and scored using custom perl scripts as follows; for each position in the alignment, a conservation score was calculated by counting the number of different types of amino acids (i.e. if an alignment position contained 5F, 3Y and 9L, then the score is 3). Averages were calculated for the entire protein and for specific subsets thereof, including DBD, LBD and the H6-7 loop region using the coordinates from Table 1.

### **PELE simulations**

#### **Systems setup**

Initial coordinates for GR and MR were obtained from the crystals presented here. Three different models were prepared for studying the ligand exit, entry and protein rearrangement: 1) the crystallographic GR (F602S and C638D) and MR (C808S and C910S) mutants, 2) the wild type by reverting the mutations with the maestro software, and 3) the wild type in absence of

## Publications

the peptide cofactor. All structures were preprocessed with the protein preparation wizard available in the Schrödinger package (50. Madhavi 2013) adding hydrogen atoms and optimizing the hydrogen bond network. Final visual inspection ensured that the predicted states were coherent with the system under study and maximized hydrogen bond formation.

## PELE sampling

Long-timescale dynamics associated with the free ligand diffusion are computationally feasible with PELE (12. Borrelli 2005. 51. Benjamin 2012). PELE uses a combination of a Monte Carlo approach with protein structure prediction methods. Three main steps define the algorithm: 1) protein backbone and ligand perturbation, 2) specific side-chain sampling, and 3) global minimization. Ligand perturbations involves a random rotation and translation, while protein perturbations is based on the displacement of  $\alpha$ -carbon according to an anisotropic network model (ANM) (52. Bahar 1998). The side-chain sampling step involves arranging all side chains adjacent to the ligand within a predefined distance of 6 Å of the ligand's center of mass. The last stage involves the minimization of a region including, at least, all residues local to the atoms involved in the perturbation and side-chain steps. Finally, the new structure is accepted or rejected based on a Metropolis test. The program uses an OPLS (Optimized Potentials for Liquid Simulations) all-atom force field with an implicit surface-generalized Born (SGB) continuum solvent model.

## Simulation protocols

Ligand exit. From the crystallographic prepared models, the exit protocol included random ligand's translations of 0.8 Å and rotation of 0.2 radians. The backbone perturbation included the lowest 6 ANM modes with maximum displacements of each alpha carbon up to 1Å. The perturbed mode, randomly selected, was updated every 6 Monte Carlo steps and included a 40% mixing of the remaining 5 modes. The ligands displacement was randomly selected at each step. A spawning criteria of 4Å was used: any ligand whose center of mass is 4Å behind the structure with the center of mass farthest coordinates (with respect to the initial position), in any direction, will abandon its position and continue the execution with the coordinates from the leading (farthest) one. Thus, all processors search collectively, with no bias in direction, for an effective escape path. Simulations were finished after the

## Publications

ligand's solvent accessible area (SASA) was larger than 0.5, with typical simulations times of 10-20 CPU hours.

Ligand entrance. After convenient preparation of the protein structure an initial simulation was done to create a set of 20 independent initial ligand where ligand's coordinates occupy randomly distributed sites over the protein surface. Having these structures as initial states, free search simulations were performed with runs of 64 independent simulations for 48 CPU hours. Ligand perturbation included equally probable translations of  $3.0\text{\AA}/1.0\text{\AA}$  and rotation of  $0.25/0.05$  radians. The backbone perturbation was the same as in the exit procedure. Ligands displacement direction was randomly updated every 6 steps. Doing so, we guarantee that trajectories explore the entire surface. Furthermore, keeping the perturbation direction for 6 steps is necessary to observe entrance events in difficult cases (buried active sites). We should emphasize that no predetermined search direction is taken and ligand perturbations are completely random. Furthermore, in the entrance protocol all processors search independently (no spawning criteria was used).

## References

- (1) Evans, R.M. The steroid and thyroid hormone receptor superfamily. *Science*. **240**, 889–895 (1988).
- (2) Mangelsdorf, D.J. et al. The nuclear receptor superfamily: the second decade. *Cell* **83**, 835–839 (1995)
- (3) Cole, T.J. Glucocorticoid action and the development of selective glucocorticoid receptor ligands. *Biotechnol. Annu. Rev.* **12**, 269–300 (2006) (prednisolone= GR)
- (4) Gravez, B., Tarjus, A. & Jaisser, F. Mineralocorticoid receptor and cardiac arrhythmia. *Clin. Exp. Pharmacol. Physiol.* **40**, 910–915 (2013)...(Eplerenone = MR)
- (5) Shelley, M., Bennett, C., Nathan, D., & Sartor, O. *Cancer Metastasis-Biology and Treatment* (eds Ablin, R.J. & Mason, M.D.) **13**, 283–307 (Springer Science and Business Media B.V. 2008)...(AR antagonist-Bicalutamid)
- (6) Sitruk-Ware, R. & Nath, A. The use of newer progestins for contraception. *Contraception* **82**, 410–417 (2010)...(Drospirenone = PR)
- (7) Bledsoe, R.K. et al. Crystal structure of the glucocorticoid receptor ligand binding domain reveals a novel mode of receptor dimerization and coactivator recognition. *Cell* **110**, 93–105 (2002). (GR xtal)
- (8) Williams, S.P. & Sigler, P.B. Atomic structure of progesterone complexed with its receptor. *Nature* **393**, 392–396 (1998). (PR xtal)
- (9) Fagart, J. et al. Crystal structure of a mutant mineralocorticoid receptor responsible for hypertension. *Nat. Struct. Mol. Biol.* **12**, 554–555 (2005). (MR xtal structure)
- (10) Matias, P.M. et al. Structural evidence for ligand specificity in the binding domain of the human androgen receptor. Implications for pathogenic gene mutations. *J. Biol. Chem.* **275**, 26164–26171 (2000). (AR x-tal)
- (11) Li, Y., Suino, K., Daugherty, J. & Xu, H.E. Structural and biochemical mechanisms for the specificity of hormone binding and coactivator assembly by mineralocorticoid receptor. *Mol. Cell.* **19**, 367–380 (2005). (outward tilt of the H6-H7 - 17alfa Xtals)
- (12) Borrelli, K.W., Vitalis, A., Alcantara, R. & Guallar, V. Protein energy landscape exploration. A novel Monte Carlo technique. *J. Chem. Theory Comput.* **6**, 1304–1311 (2005).
- (13) Bhabha, G. et al. Divergent evolution of protein conformational dynamics in dihydrofolate reductase. *Nat. Struct. Mol. Biol.* **11**, 1243–1249 (2013).

## Publications

- (14) Burris, T.P. *et al.* Nuclear receptors and their selective pharmacologic modulators. *Pharmacol. Rev.* **65**, 710–778 (2013).
- (15) Bertocchio, J., Warnock, D.G., & Jaisser, F. Mineralocorticoid receptor activation and blockade: an emerging paradigm in chronic kidney disease. *Kidney Int.* **10**, 1051–1060 (2011).
- (17) Dror, R.O. *et al.* Structural basis for modulation of a G-protein-coupled receptor by allosteric drugs. *Nature* **7475**, 295–299 (2013).
- (18) Jensen Ø.M. *et al.* Mechanism of Voltage Gating in Potassium Channels. *Science*, **336**, 229–233 (2012).
- (19) Shan, Y. *et al.* How does a drug molecule find its target binding site? *J. Am. Chem. Soc.* **133**, 9181–9183 (2011).
- (20) Buch, I., Giorgino, T. & De Fabritiis, G. Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proc. Natl Acad. Sci. U.S.A.* **108**, 10184–10189 (2011).
- (21) Madadkar-Sobhani, A. & Guallar, V. PELE web server: atomistic study of biomolecular systems at your fingertips. *Nucleic Acids Res.* **41**, 322–328 (2013).
- (22) Takahashi, R., Gil, V.A. & Guallar, V. Monte carlo free ligand diffusion with markov state model analysis and absolute binding free energy calculations. *J. Chem. Theory Comput.* **10**, 282–288 (2014).
- (24) Li, Y., Suino, K., Daugherty, J. & Xu, H. E. Structural and biochemical mechanisms for the specificity of hormone binding and coactivator assembly by mineralocorticoid receptor. *Mol. Cell* **19**, 367–380 (2005).
- (25) Capelli, A.M., Bruno, A., Guadix, A.E. & Costantino, G. Unbinding pathways from the glucocorticoid receptor shed light on the reduced sensitivity of glucocorticoid ligands to a naturally occurring, clinically relevant mutant receptor. *J. Med. Chem.* **56**, 7003–7014 (2013).
- (27) Hasui, T. *et al.* Identification of benzoxazin-3-one derivatives as novel, potent, and selective nonsteroidal mineralocorticoid receptor antagonists. *J. Med. Chem.* **54**, 8616–8631 (2011)
- (28) Gronemeyer, H., Gustafsson, J.A. & Laudet, V. Principles for modulation of the nuclear receptor superfamily. *Nat. Rev. Drug Discov.* **3**, 950–964 (2004).
- (29) Shibata, S. *et al.* Mineralocorticoid receptor phosphorylation regulates ligand binding and renal response to volume depletion and hyperkalemia. *Cell Metab.* **18**, 660–671 (2013).
- (30) Bridgham, J.T. Carroll S.M. & Thornton, J.W. Evolution of hormone-receptor complexity by molecular exploitation. *Science*, **312**, 97–101 (2006).

- (31) Ortlund, E.A., Bridgham, J.T., Redinbo M.R. & Thornton, J.W. Crystal structure of an ancient protein: evolution by conformational epistasis. *Science*, **317**, 1544–1548 (2007).
- (32) Changeux, J.P. 50 years of allosteric interactions: the twists and turns of the models. *Nat. Rev. Mol. Cell Biol.* **14**, 819–829 (2013).
- (33) Bridgham, J.T., Ortlund, E.A. & Thornton, J. W. An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. *Nature*, **461**, 515–520 (2009).
- (35) Andrieu, T. *et al.* Detection and functional portrayal of a novel class of dihydrotestosterone derived selective progesterone receptor modulators (SPRM). *J. Steroid Biochem. Mol. Biol.* **147**, 111–123 (2015).
- (36) Shan, Y. *et al.* Molecular basis for pseudokinase-dependent autoinhibition of JAK2 tyrosine kinase. *Nat. Struct. Mol.* **21**, 579–584 (2014).
- (37) Moras, D. & Gronemeyer, H. The nuclear receptor ligand-binding domain: structure and function. *Curr. Opin. Cell Biol.* **10**, 384–391 (1998).
- (38) Yen, P. M. Physiological and molecular basis of thyroid hormone action. *Physiol. Rev.* **81**, 1097–1142 (2001).
- (39) Brzozowski, A.M. *et al.* Molecular basis of agonism and antagonism in the oestrogen receptor. *Nature* **389**, 753–758 (1997).
- (40) Batista, M.R. & Martínez, L. Dynamics of nuclear receptor helix-12 switch of transcription activation by modeling time-resolved fluorescence anisotropy decays. *Biophys. J.* **105**, 1670–1680 (2013).
- (41) Sonoda, M.T., Martínez, L. Webb, P., Skaf, M.S., & Polikarpov, I. Ligand dissociation from estrogen receptor is mediated by receptor dimerization: evidence from molecular dynamics simulations. *Mol. Endocrinol.* **22**, 1565–1578 (2008).
- (42) Aci-Sèche, S., Genest, M. & Garnier, N. Ligand entry pathways in the ligand binding domain of PPAR $\gamma$  Receptor. *FEBS. Lett.* **585**, 2599–2603 (2011).
- (43) Li, M.J. *et al.* Structure of estradiol metal chelate and estrogen receptor complex: the basis for designing a new class of selective estrogen receptor modulators. *J Med. Chem.* **54**, 3575–3580 (2011).
- (44) Grad, I. & Picard, D. The glucocorticoid responses are shaped by molecular chaperones. *Mol. Cell Endocrinol.* **275**, 2–12 (2007).
- (45) Ricketson, D., Hostick, U., Fang, L., Yamamoto, K.R. & Darimont, B.D. A conformational switch in the ligand-binding domain regulates the dependence of the glucocorticoid receptor on Hsp90. *Mol. Biol.* **368**, 729–741 (2007).

- (46) Fang, L., Ricketson, D., Getubig, L. & Darimont, B. Unliganded and hormone-bound glucocorticoid receptors interact with distinct hydrophobic sites in the Hsp90 C-terminal domain. *Proc. Natl. Acad. Sci. USA* **103**, 18487–18492 (2006).
- (47) Lorenz, O.R. *et al.* Modulation of the Hsp90 chaperone cycle by a stringent client protein. *Mol. Cell* **53**, 941–953 (2014).
- (48) Caamaño, C.A., Morano, M.I., Dalman, F.C., Pratt, W.B. & Akil, H. A conserved proline in the hsp90 binding region of the glucocorticoid receptor is required for hsp90 heterocomplex stabilization and receptor signaling. *J. Biol. Chem.* **273**, 20473–20480 (1998).
- (49) Cluning, C. *et al.* The helix 1-3 loop in the glucocorticoid receptor LBD is a regulatory element for FKBP cochaperones. *Mol. Endocrinol.* **27**, 1020–1035 (2013).
- (50) Madhavi, S.G., Adzhigirey, M., Day, T., Annabhimoju, R. & Sherman, W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J. Comput. Aided Mol. Des.* **27** 221–234 (2013).
- (51) Benjamin, P., Cossins, B.P., Hosseini, A. & Guallar, V. Exploration of protein conformational change with PELE and meta-dynamics. *J. Chem. Theory Comput.* **8**, 959–965 (2012).
- (52) Bahar, I., Atilgan, A.R., Demirel, M.C. & Erman, B. Vibrational dynamics of folded proteins: significance of slow and fast motions in relation to function and stability. *Phys. Rev. Lett.* **80**, 2733–2736 (1998).
- (53) Alexander, S.P.H. *et al.* The concise guide to pharmacology 2013/14: Nuclear hormone receptors. *Br. J. Pharmacol.* **170**, 1652–1675 (2013).
- (54) Hellal-Levy, C. *et al.* Specific hydroxylations determine selective corticosteroid recognition by human glucocorticoid and mineralocorticoid receptors. *FEBS. Lett.* **464**, 9–13 (1999).
- (56) Nettles, K.W. *et al.* Structural plasticity in the oestrogen receptor ligand-binding domain. *EMBO Rep.* **8**, 563–568 (2007).
- (57) Kabsch W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
- (58) Leslie, A.G.W. & Powell, H.R. *Evolving Methods for Macromolecular Crystallography.* ISBN 978-1-4020-6314-5, **245**, 41–51 (2007)
- (59) Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 72–82 (2006).

## Publications

- (60) Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760–763 (1994).
- (61) McCoy, A.J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
- (62) Bricogne, G. *et al.* BUSTER version 2.11.5 Cambridge, United Kingdom: Global Phasing Ltd. (2011)
- (63) Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
- (64) Waterhouse, R.M., Tegenfeldt, F., Li, J., Zdobnov, E.M. & Kriventseva, E.V. OrthoDB: A hierarchical catalog of animal, fungal and bacterial orthologs. *Nucleic Acids Res.* **41** (Database issue), D358–D365 (2013).
- (65) Larkin, M.A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**, 2947–2948 (2007).

## Acknowledgments

This study was supported by The European Research Council (2009-Adg25027-PELE) to V.G and by the SEV-2011-00067 grant of the Severo Ochoa Program, awarded by the Spanish Government to D.L. We would like to acknowledge our AstraZeneca colleagues R.Unwin, and R.Knöll for helpful discussions and carefully proof-reading the manuscript. We also thank N.Blomberg (ELIXIR) for expert insight into the nuclear hormone target class and for carefully reading the manuscript.

## Author information

### Affiliations

Structural Chemistry, AstraZeneca Mölndal, Mölndal, Sweden.  
Karl Edman

Joint BSC-CRG-IRB Research Program in Computational Biology,  
Barcelona Supercomputing Center, Jordi Girona 29, 08034 Barcelona,  
Spain

Ali Hussein, Daniel Lecina, Ryoji Takahashi & Victor Guallar



## Publications

Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig  
Lluís Companys 23, Barcelona 08010, Spain  
Victor Guallar

Medicinal Chemistry, CVMD, AstraZeneca Mölndal, Mölndal,  
Sweden.  
Anders Hogner

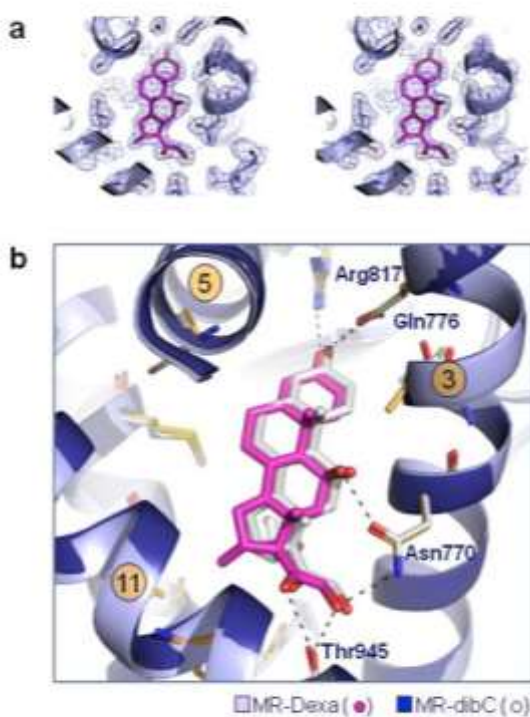
Medicinal Chemistry, RIA, AstraZeneca Mölndal, Mölndal, Sweden.  
Matti Lepistö

R&D Information, AstraZeneca Mölndal, Mölndal, Sweden.  
Magnus Bjursell

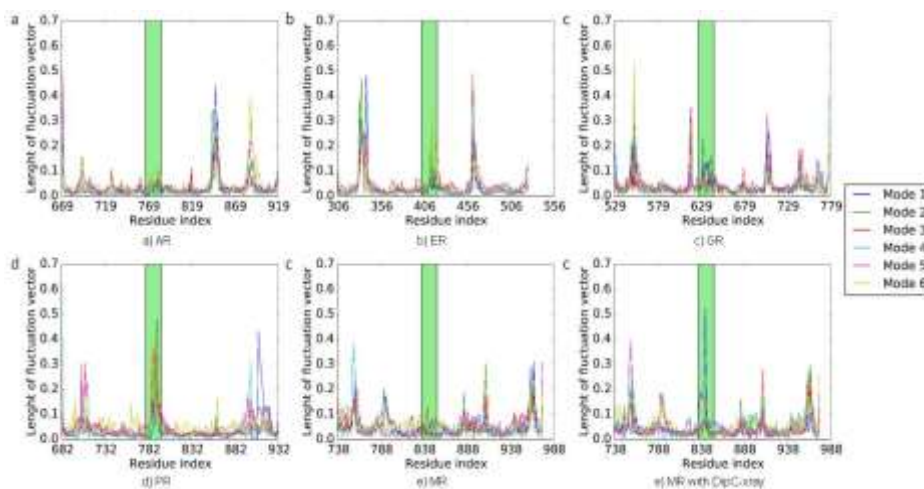
## Supplementary Material

### Ligand recognition in steroid hormone receptors: from conserved plasticity to binding mechanism

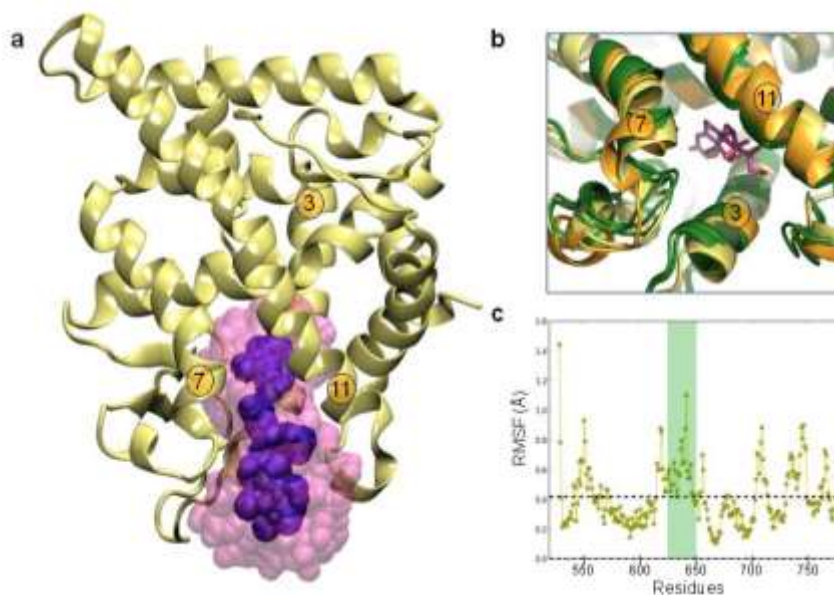
Karl Edman, Anders Hogner, Ali Hussein, Magnus K Bjursell, Anna Aagaard, Stefan Bäckström, Cristian Bodin, Lisa Wissler, Tina Jellesmark-Jensen, Anders Cavallin, Ulla Karlsson, Ewa Nilsson, Daniel Lecina, Ryoji Takahashi, Christoph Grebner, Matti Lepistö & Victor Guallar



**Supplementary Figure 1. Comparison of the complex structures of the MR:Dexa and MR:dibC.** (a) Stereo view of the 2mFo-dFc density map of the MR:Dexa LBP. (b) The structure of MR (light blue) in complex with dexamethasone (magenta) superimposed on the MR structure (dark blue) in complex with dibC (white). The steroid template overlays nearly perfectly (RMSD 0.28Å) with all hydrophilic interactions conserved.

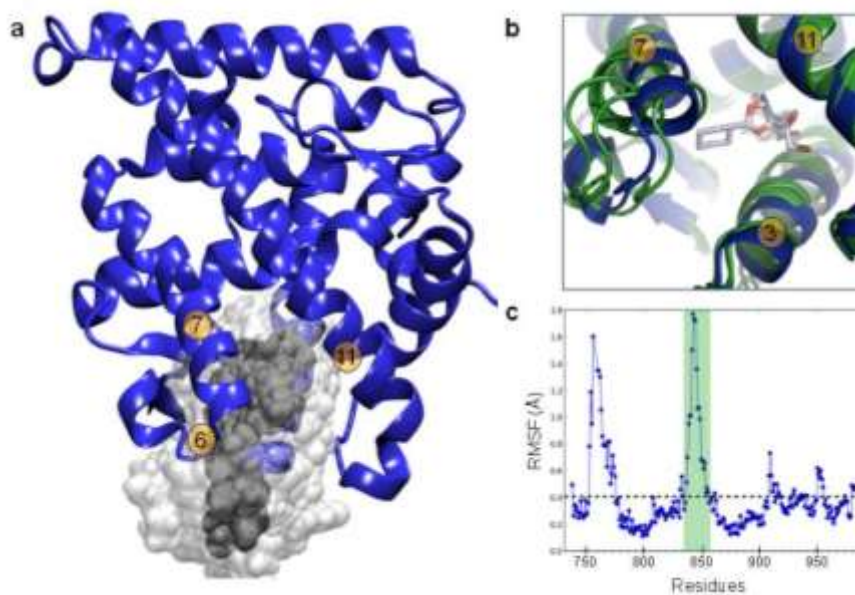


**Supplementary Figure 2. Principle component analysis for all X-ray structures of the steroid hormone receptors (AR, ER, GR, PR, and MR) in the public domain (PDB).** Graphs showing the amplitude of the top six modes from the PCA of the structures in the public domain for AR (a), ER (b), GR (c), PR (d) and MR (e). The H6-H7 sequence is highlighted in green. AR and MR exhibits the smallest variation in the H6-H7 region in the public domain structures.



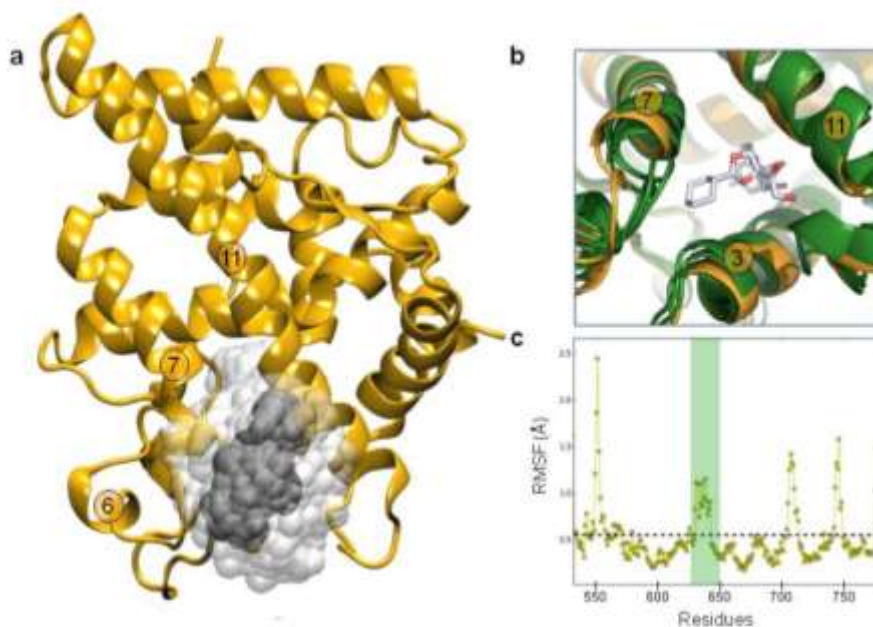
**Supplementary Figure 3. Ligand exit pathway for the GR:Dexa complex.**

(a) The ligand center of mass is highlighted in blue beads all other atoms of the ligand being shown in transparent spacefill. (b) Detail of the backbone rearrangement along the exit pathway. The GR:Dexa and GR:dibC X-ray structures are shown in light yellow and orange, respectively. Three snapshots from the exit simulations are shown in green and a dexamethasone ligand from the exit trajectory is shown in magenta. (c) Residues RMSF fluctuations against the average structure along the GR:Dexa exit pathway where helices 6 and 7 are marked with green shade.



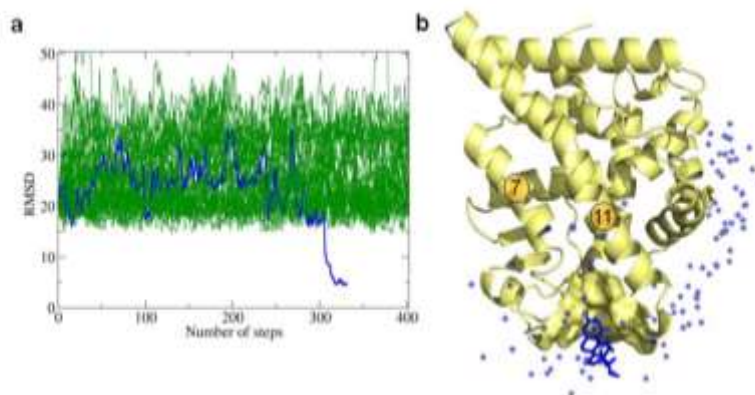
**Supplementary Figure 4. Ligand exit pathway for the MR.dibC complex.**

(a) The ligand center of mass is highlighted in gray beads, all other atoms of the ligand being shown in transparent spacefill. (b) Detail of the backbone rearrangement along the exit pathway. The MR:dibC X-ray structures is shown in dark blue. Three snapshots from the exit simulations are shown in green and a dibC ligand from the exit trajectory is shown in white. (c) Residues RMSF fluctuations against the average structure along the MR.dibC exit pathway where helices 6 and 7 are marked with green shade.

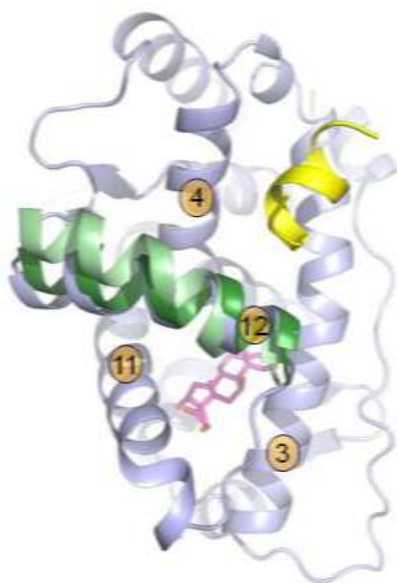


**Supplementary Figure 5. Ligand exit pathway for the GR:dibC complex.**

(a) The ligand center of mass is highlighted in gray beads, all other atoms of the ligand being shown in transparent spacefill. (b) Detail of the backbone rearrangement along the exit pathway. The GR:dibC X-ray structures is shown in orange. Three snapshots from the exit simulations are shown in green and a dibC ligand from the exit trajectory is shown in white. (c) Residues RMSF fluctuations against the average structure along the GR:dibC exit pathway where helices 6 and 7 are marked with green shade.



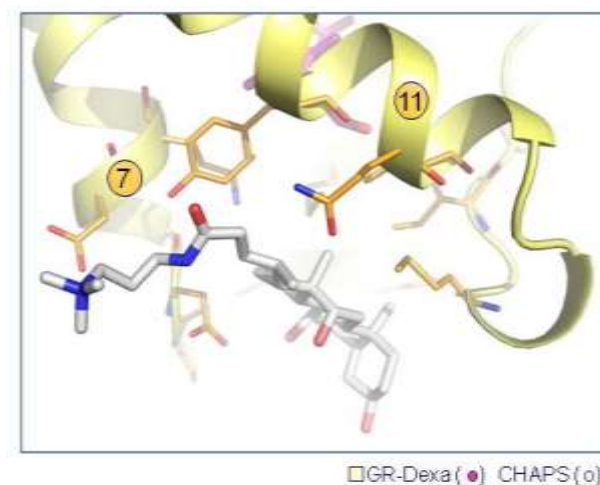
**Supplementary Figure 6.** Unbiased simulation of dexamethasone entering GR. (a) Each line represents the ligand's RMSD (heavy atom) to the bound crystal ligand for the total 64 trajectories. One of the trajectories represented by blue line enter the LBP at step ~330. (b) The ligand's center of mass for the one trajectory that enter the LBP are shown as blue spheres. The region where the ligand enter the LBP is emphasized as a surface with the ligand shown in stick representation.



**Supplementary Figure 7.** Detail of the Helix 12 rearrangement from free simulation dexamethasone entering MR in presence and absence of the co-

## Publications

regulator peptide. The initial structure is shown in light blue and the maximum movement of the helix 12 is shown in dark green (with peptide) and light green (without peptide). The dexamethasone ligand in the LBP is shown in magenta.



**Supplementary Figure 8.** A peripheral binding site about 13Å away from the LBP is revealed in the crystal structure of GR (yellow) in complex with dexamethasone (yellow). This site is occupied by a steroid-like CHAPS molecule (white) that is part of the crystallization conditions.



**Supplementary Table S2.** Data collection and refinement statistics (molecular replacement)

	MR: dexta	MR: dibC	GR:dexta	GR:dibC
<b>Data collection</b>				
Space group	P212121	P41212	P3221	P3221
Cell dimensions				
<i>a</i> , <i>b</i> , <i>c</i> (Å)	73.00, 81.40, 45.23	75.92, 75.92, 117.00	84.66, 84.66, 105.91	87.20, 87.20, 102.89
<i>a</i> , <i>b</i> , <i>g</i> (°)	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 90.00, 120.00	90.00, 90.00, 120.00
Resolution (Å)	40.7-2.03 (2.17-2.03)	48.79-2.36 (2.55-2.36)	31.81-2.50 (2.67-2.50)	40.14-1.80 (1.85-1.80)
<i>R</i> <sub>sym</sub> or <i>R</i> <sub>merge</sub>	0.06(0.50)	0.13(1.30)	0.08(0.55)	0.08(1.05)
<i>I</i> / <i>sI</i>	13.10(2.30)	15.10(1.90)	8.80(1.60)	7.40(0.70)
Completeness (%)	83.9(83.7)	100.0(100.0)	99.6(99.5)	99.9(100.0)
Redundancy	3.3(2.5)	12.6(11.7)	4.1(4.2)	3.5(3.6)
<b>Refinement</b>				
Resolution (Å)	2.03	2.36	2.50	1.80
No. reflections	15085	14672	15559	42339
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	0.185/0.240	0.182/0.218	0.210/0.253	0.213/0.224
No. atoms				
Protein	2080	2118	2133	2184
Ligand/ion	34	49	64	146
Water	101	60	83	250
<i>B</i> -factors (Ask for input)				
Protein	30.14	53.25	49.72	33.25
Ligand/ion	22.12	44.16	34.51	23.55
Water	36.03	56.86	46.23	46.95
R.m.s. deviations				
Bond lengths (Å)	0.010	0.010	0.010	0.010

## Publications

Bond angles (°)	1.01	1.04	1.12	1.06
-----------------	------	------	------	------

---

\*Number of xtals for each structure should be noted in footnote.

\*Values in parentheses are for highest-resolution shell.

[AU: Equations defining various *R*-values are standard and hence are no longer defined in the footnotes.]

[AU: Ramachandran statistics should be in Methods section at the end of Refinement subsection.]

[AU: Wavelength of data collection, temperature and beamline should all be in Methods section.]

## Publications

## **4 Summary of the results and discussion**

## Summary of the results and discussion

## Summary of the results and discussion

A complete understanding of complex formation between proteins and ligands, a crucial matter for pharmacology and, more in general, in biomedicine, requires a detailed knowledge of their static and dynamic atomic interactions. The main objective of this thesis is to test recent developments in conformational sampling techniques in providing such a dynamical view. We aim at developing new protocols and methods for such a study. Moreover, we want to show how its application can aid in addressing existing problems in the biophysics of protein ligand interactions. We provide here a summary of the main results and discussion along this work where we frame the different publication in the overall objectives described above.

### **4.1 Validate our in-house technology PELE (Protein Energy Landscape Exploration) on sampling protein-ligand interactions and induce fit procedure.**

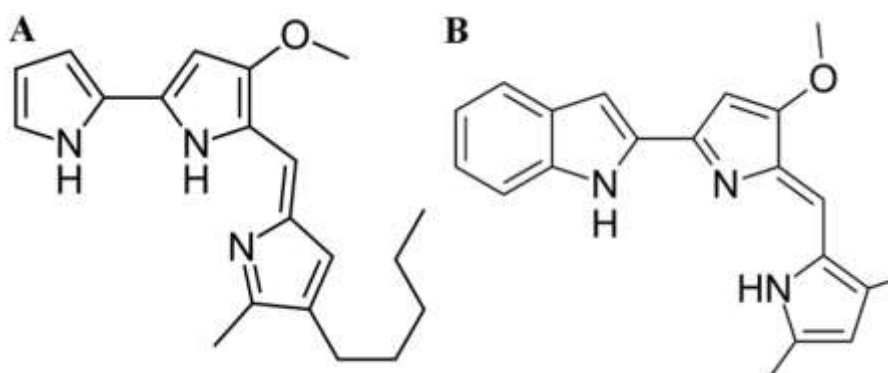
In the introduction of this thesis we have underline the importance of computational techniques in obtaining an atomic detailed and dynamical view of molecular recognition, and, in particular, for protein-ligands interactions. We further highlighted that computational approaches in the majority of the cases are much cheaper and faster than in-vitro and in-vivo experiments; these last methods having difficulties in obtaining atomic detailed information of the binding mechanism. Moreover, we have stated how traditional docking techniques might not be enough for describing the induced fit recognition, and how molecular dynamics, who could describe it, introduce a significant computational cost (computer time).

The initial studies along this thesis aimed at addressing the above difficulties through a series of studies on clinically relevant targets. The PELE algorithm was applied to these systems, which were suggested by our experimental collaborators, to explore the capabilities of a quicker and accurate sampling in protein structure and ligand induced fit docking.

Within the two publications presented in the section 3.1 and 3.2, we used Glide docking scores before and after the PELE run (aimed at modeling the induced fit procedure), to determine if advanced sampling algorithms improve the description of the protein-ligand molecular recognition and binding scores. For each system, involving six know receptors: mTOR, PDK-1, PKC-alpha, MCL1, BCL-xl and BCL-2, we studied the binding of some prodiginines, under investigation by the lab of Prof Ricardo Perez (University of Barcelona), and of some control compounds. These controls involved both positives: crystal structures or know binders, and negative controls: non-

## Summary of the results and discussion

binders as shown by kinase assays. Our results clearly indicate the excellent capabilities of our sampling method. By applying PELE, we could distinguish in every case between binders and non-binders. More importantly, such a distinction is not possible when using standard docking techniques; Glide score after docking but before conformational sampling could not identify binders. Moreover, the results show nice correlation between the amount of the receptor rearrangement, computed as the alpha carbon RMSD along the sampling, and the “need” for induced fit. For example, when a ligand is docked into its crystal structure the induced fit procedure does not significantly change the scores, and only introduces slight changes to the protein–ligand structures in the active site.



**Figure 1.** Image of the two inhibitors used in this work.  
Prodigiosin (A) and Obatoclax (B).

Table 1 summarizes the before and after PELE docking scores, together with the active site RMSD along with the induced fit process, for all systems and ligands which have been used for these two studies.

In the first study, we demonstrate that mTOR is a molecular target of the two prodiginines studied: prodigiosin (PG) and obatoclax (OBX) (Figure 1). This computational study was aimed at finding the molecular targets after noticing that these two drugs were capable of considerably reducing the melanoma cells, a highly drug-resistant cancer model in cellular assays. In addition, *in vitro* binding assays were performed for several kinases (as control). As expected, for PDK1 and PKC-alpha (control systems) we obtained very good docking scores, below -9, for their two crystallographic

## Summary of the results and discussion

residues, LAA and LW4, see table 1. As mentioned before, when starting from the crystal bound structure, no large RMSD along the induced fit is observed. Thus, our (positive) controls indicated good binding scores with low RMSD. Interestingly, the induced fit procedure increased the RMSD but did not substantially improve the affinity for PDK-1 and PKC- $\alpha$  against PG; this (negative) control experiments support the absence of inhibition in PDK1 and PKC observed in the kinase profiling. Regarding mTOR ligand docking, the initial docking score of pp242, PG and OBX is about -7 to -8, similar to the value measured for PG in our in silico control assays. For mTOR, however, the induced fit procedure (the same used for PDK1 and PKC-alpha) introduced significant changes. We observed a clear increase in binding affinity along with a significant active site adjustment, the RMSD increases to 2.1 and 2.3 for PG and OBX, respectively. Thus, our simulation studies support binding of PG to mTOR in a similar fashion of the pp242, a positive control with 0.008  $\mu$ M IC50 in the Presence of 10  $\mu$ M ATP.

This study supported the in vivo cellular assays indicating the capabilities of PG (and its derivatives) for use in some apoptotic targets. Moreover our findings contributed to the understanding of the molecular mechanisms of action of both molecules and provide data about their structural properties that will allow the development of more-effective mTOR inhibitors in the future (the paper has been cited 23 times since 2012, 5 of which reported studies on novel inhibitors).

	PDK-1		PKC-alpha		mTOR			MCL1			BCL-xl			BCL-2		
	LAA	PG	LW4	PG	pp242	PG	OBX	PG	OBX	6	PG	OBX	ABT	PG	OBX	ABT
Initial Score	-9.2	-6.2	-12.0	-7.1	-7.9	-7.4	-8.5	-4.3	-2.9	-6	-7	-3.4	-7.9	-6.3	-6	-8.5
Induced Score	-10.1	-6.6	-11.5	-7.6	-9.7	-10.1	-10.3	-8.6	-8.8	-8.7	-7.4	-8.3	-13.9	-7.9	-8.8	-9.6
RMSD	1.1	1.77	1.3	1.4	1.7	2.1	2.3	3	4	7.4	3	-5.5	1.4	4.8	4.5	4

**Table 1.** Initial and after PELE (induced fit) scores obtained with Glide.

For the second study, we applied a similar protocol to the BCL-2 protein family against PG and OBX (Figure 1), together with two positive control ligands: 6 and ABT. Since there are no crystal structures of this protein family bound to PG, OBX or 6, we expected a significant RMSD change and an improvement in the docking score along the induced fit process. Clearly for all ligands we observed a large RMSD increase, ranging from 3 to 7 and significant improvement in the scores. Interestingly, in MCL1, similar scores ( $\sim$  9) are obtained for the control, ligand 6 with a 0.25  $\mu$ M IC50, and for the two prodigionines studied, OBX and PG.



## Summary of the results and discussion

For BCL-xL the initial crystal structure used to model the target has already the control ligand ABT-737 bound to it. Thus, as expected, we observe the lowest induced fit RMSD for this ligand, 1.4 Å. Additionally, we found good initial and (very good) final scores for ABT-737 a potent inhibitor of BCL-xL from Abbott Laboratories with an IC50 of 0.06 µM. In BCL-2 our model was derived from a peptide (43B) bound crystallographic structure. Accordingly, we observed again significant induced fit RMSD changes and improvements in the scores. For ABT-737 we observe good initial scores and the lower RMSD, possibly as a result of its large size and excellent BH3 helix mimetic properties. Overall this second test confirmed the inhibition of this family of targets by the prodiginin compounds in a similar fashion to other high nanomolar/low micromolar inhibitors (not reaching, however, low nanomolar activity).

These set of initial studies, using PELE as a conformational sampling tool, represent a significant step towards improving the accuracy of modelled protein ligand interactions and demonstrate the necessity of induced fit docking. It showed how computational techniques are mature enough to provide a good protein-ligand recognition mechanism (and binding scores) in ~24 hours of a modest workstation (16 cores) CPU usage.

### **4.2 Besides protocols and software validation, we aim to develop specific application on biomedical and biotechnological relevant systems. Thus, we aim at adding information for contributing to the mechanistic knowledge of important protein-ligand interactions.**

One of the main objectives of this dissertation is to get a deeper understanding of the biophysical mechanism behind protein-ligand interactions. Moreover we aim at performing such study for pharmacologically relevant systems. Besides the above mentioned studies on apoptotic targets inhibition, we have performed additional mechanistic contributions. Possibly, one of the most interesting and challenging (as well as trendy) ones involved the free ligand diffusion coupled to active site search and binding. Such type of studies allow for a full mechanistic exploration of the protein ligand interaction in absence of information of the active site.

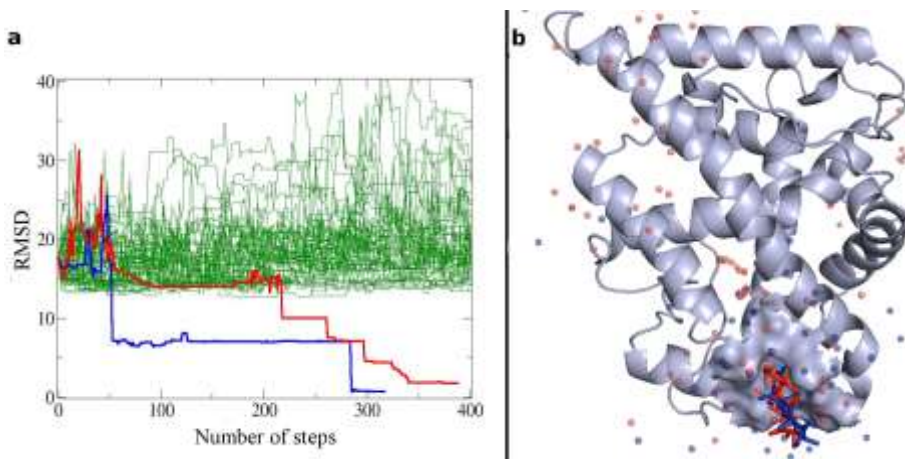
Such a free search has become quite popular since the first molecular dynamics simulation performing such a non-biased sampling by the David

## Summary of the results and discussion

Shaw research group. In their 2011 Journal of American Society paper: “How does a drug molecule find its target binding site?” (already cited 160 times), the authors perform 20 microsecond of non-biased MD where a ligand explores a Src-kinase surface, finding the binding site and identifying the lowest protein-ligand interaction energy as a pose within  $\sim 1\text{\AA}$  of the crystallographic pose. Certainly, studies like this show the potential of biophysical computer simulations in aiding drug design, providing a complete mechanistic knowledge and identifying the correct protein-ligand complex. However, these calculations still use a massive amount of computational time that is not accessible to a normal lab.

Along the thesis, we have developed algorithms to perform such a search with PELE. The same Src kinase study performed by the Shaw group, for example, can be studied analogously with PELE in 24 hours of a modest single processor 32-cores workstation (a commodity machine). Fruit of this study, we have a ready-made script in our server (<https://pele.bsc.es>) “Unconstrained Ligand Exploration and Binding Site Search” which is one of the most popular ones. Some examples are also shown in <https://pele.bsc.es/pele.wt/examples>. Furthermore, PELE’s sampling capabilities allows for such a study in fully occluded (buried) active sites; all MD explorations have been achieved in partially exposed binding sites. Here we show how our technique can be used to scan the binding of dexamethasone (DEX) and dibC to steroid receptors family and perform a blind docking simulation.

In collaboration with AstraZeneca, we performed extensive all atoms unbiased ligand exit and entrance simulations together with structural principal component and bioinformatics analysis for an important pharmacological target: the Nuclear Hormone Receptors (NHRs) family. This study was combined with crystallography and ligand binding assays (scintillation proximity assay) performed in Sweden at the AstraZeneca lab. In this study, presented in section 3.6, ligands were randomly placed outside the receptors and allowed to freely explore the protein surface. Figure 2 shows the unbiased simulation of DEX exploring the mineralocorticoid receptor (MR), where two of the 80 trajectories (running for 24 hours) clearly entered the active site and adopted conformations within  $\sim 1\text{\AA}$  of the crystallographic pose. We should emphasize that no knowledge of the crystal bound complex is used along the simulation.



**Figure 2. Unbiased simulation of dexamethasone entering MR.** (a) Each line represents the ligand's RMSD (heavy atom) to the bound crystal ligand for a different trajectory. Two of the trajectories represented by blue and red lines enter the ligand binding protein (LBP) domain at step 52 and 214, respectively. (b) The ligand center of mass for the two trajectories that enter the LBP are shown using red and blue spheres. The region where the ligands enter the LBP is emphasized as a surface with two ligands shown in stick representation.

Remarkably, such quick performance by PELE allows to run exhaustive sampling in an affordable computational time (using ~600 trajectories in this study), from which absolute binding free energies can be derived by means of Markov State Model techniques. In this study we performed such analysis introducing a 20 Å constraint from the entrance point, obtaining absolute binding energy differences in quantitative agreement to the experimentally measured ones.

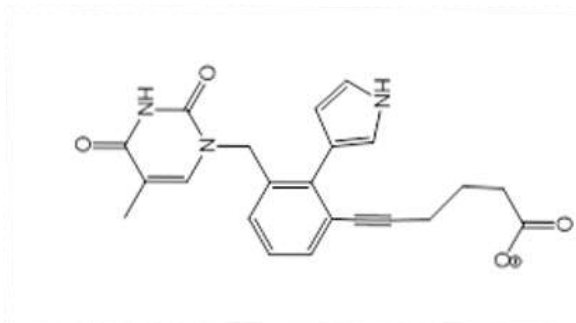
More importantly, such detailed mechanistic study indicated that the large-amplitude protein motion of helix 12, as suggested by apo and holo crystallographic nuclear hormone receptors, is not required for ligand entry. Instead, we show that mid-scale vibrations combined with a structural rearrangement of H6-H7 region are enough to identify an energetically favorable pathway to allow the ligands to diffuse into the LBP. In summary, using PELE we introduced a new structural and dynamic paradigm for ligand binding in this important family of receptors. The discovery may influence (work already being performed at AstraZeneca) future rational design of inhibitors for hormone receptors that exploits this plasticity to generate ligands with differential modes of action. Moreover, the procedure outlined in

this study can be used to challenging problems where huge computational resources are not accessible and a fast and cheap protein-ligand sampling is needed.

### **4.3 Following the previous goal, we aim at the implementation of the atomic detailed knowledge into the rational design of new inhibitors, aiming to enhance specificity and binding strength.**

Motivated by our success with validation studies (applied to several systems for protein-ligand interaction and induce fit procedure) we attempted to design a new inhibitor for a specific target. For doing so, we used the system from our second study: Molecular interactions of prodiginines with the BH3 domain of BCL-2 family members.

Once we had tested the protocol and gained enough information about the protein-ligand structure of our target, we introduced chemical changes on the drugs, working towards a rational design of new inhibitors for Mcl-1 enhancing specificity and binding strength. The strategy was (aimed at) an iterative work between our group and the chemical lab of Dr. Roberto Quesada, from university of Burgos, where different rounds of theoretical prediction and organic synthesis where produced. In this way we aimed towards a highly active, but still easy to synthesize compound. We introduced a novel and in-silico more potent inhibitor of Mcl-1 protein, as compared to Obatoclox (an already known good binder of Bcl-2 family particularly Mcl-1 protein) (Figure 3). As of July 2014, two compounds were produced in silico, but only having about 60% of the original design (due to synthetic limitations and lack of students in the Burgos lab). This compound has been tested in the lab of Professor Ricardo Perez (UB) showing high micromolar binding. Certainly much more resources will be needed to reach the pure compound predicted in silico; due to lack of man-power this objective is, at the present time, in a dead point. Nevertheless, this limited result indicates the potential of this techniques opening new directions in selective drug design (two different projects of drug design are currently running in the lab.)



**Figure 3.** Image of the designed inhibitor for MCL-1 protein.

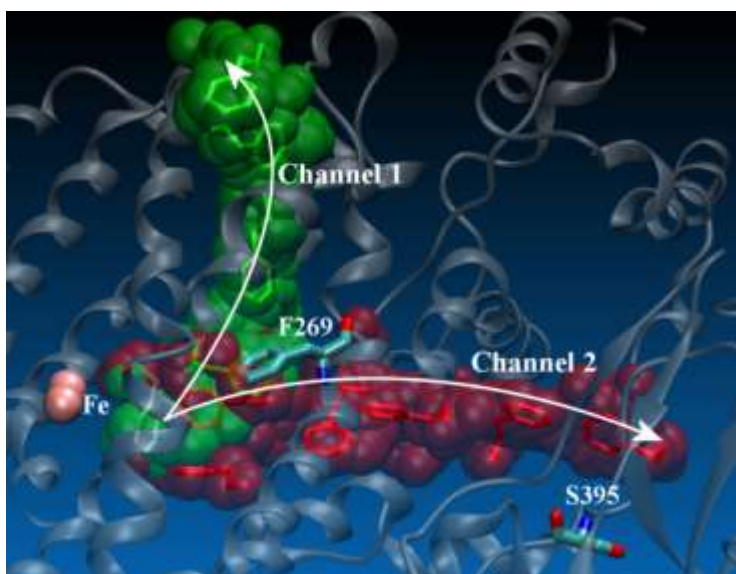
**4.4 An added value of (accurately) describing protein-ligand and protein/substrate interactions at a dynamical level, is being able to map possible changes in ligand/substrate affinities derived from mutations. We aim to develop protocols in PELE for the description of mutational effects in ligand binding. We tested this part on one of the most well studied systems with important mutational effects: HIV-1 protease.**

Recently, due to its implications in drug design and enzyme engineering there has been an increasing interest in applying molecular dynamics techniques for mapping ligand diffusion (entrance and exit), coupled to changes introduced by mutations in them (and in ligand affinities). Within the two publications presented in sections 3.3 and 3.5 we present our efforts in characterizing such mutational effects in ligand binding and enzymatic activity.

In the first study, entering the world of biocatalyst, we present our initial attempts to rationalize the role of single mutations into ligand delivery and product release. In order to understand the atomic detailed mechanism of substrate oxidation in T4MO (a soluble four-component BMM that oxidizes toluene with ~95% regioselectivity at the *para* position) as well as the involvement of S395 in the catalytic rate enhancement, we modelled the ligand migration pathways with all-atom computational techniques. T4MO is of particular interest in industry given the high number of substrates that can be oxidized along with the elevated specificity.

## Summary of the results and discussion

The results indicate that two different tunnels cross T4MO's alpha-subunits in a nearly 90° disposition. As seen in the Figure 4, one channel (hereafter known as channel 1 depicted in green) is a traverse of about 23 Å to the protein's surface which provides the shortest path between the active site and the solvent. The second observed route, for ligand migration, is a long hydrophobic passage (channel 2 in red) with a length of 30–35 Å and exiting close to S395. Channel 2 connects the diiron active site to the solvent, with an overlapping section with channel 1 close to the binding pocket. Both pathways are in good agreement with previous experimental data. Along these pathways, residue F269 plays an important role as a gate keeper for substrate diffusion. Due to its strategic position at the bifurcation point observed in the modeled pathways, and in order to understand the possible influence of the S395C mutation on ligand delivery, we performed the following *in silico* mutations: S395C, F269V and F269W.



**Figure 4.** The two observed expulsion paths obtained with the PELE simulations.

A total of 10 independent runs were produced for each mutant simulation involving: i) ligand exit, where the ligand initially placed in the active site is asked to leave the protein, and ii) ligand entrance, where ligands placed at the solvent and in the vicinity of the exit point are asked to enter the active site. Computational results show that mutations propagate changes to the close-by helices altering the way ligands exit/enter the protein. In the case of the tryptophan, and in contradiction to the larger size of the side chain, the ligand enters more easily the active site, as a result of an increase in the

## Summary of the results and discussion

migration pathway tunnel. This increase of the passageway between the surface of the protein and the active site explains the increase in activity observed in the F269W experiments. In the case of the F269V, we find that very few exits are observed as a consequence of the collapse of the tunnel derived from the smaller side chain (in agreement with the decrease experimental activity). These results clearly show the ability of PELE's technique to connect biophysical response to mutational experiments. Moreover, it represents a pure example of how the simulations might assist in designing improved biocatalysts.

In our second study we performed induce fit docking studies in HIV-1 protease with several numbers of mutations in each of the two chains of this enzyme. Predicting the effects of a mutation in HIV protease has been (and continues to be) a central issue in inhibitor design. Once the virus develops resistance to a drug, as a consequence of a mutation (not largely affecting the protease activity), the patient are forced to change treatment; typical treatments include a cocktail of inhibitors. Thus, it would be ideal to develop methodologies to efficiently predict drug resistance levels, but also to understand the resistance mechanism associated with mutations. Several computational studies have attempted to design such prediction tools, nevertheless still there is no significant success when the number of mutations in each chain is large (on the order of  $>10$ ).

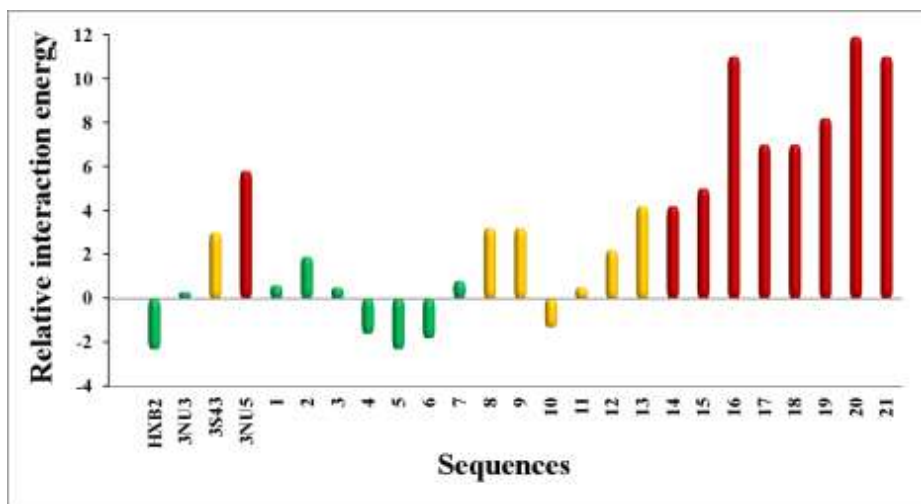
Using PELE together with docking homology modelling, we developed a methodology capable of tackling this problem. Besides checking the prediction capability to a well-known series of mutants and ligands, test case described in Koh et al. (Koh, Nakata et al. 2003), we were challenged by the researchers of the IrsiCaixa AIDS institute to perform "blind resistance predictions" in clinical samples. Thus, we had only access to the sequence but not to the resistance score, which had representatives of three different ranges: sensitive ( $<30$ ), intermediate (between 30 and 60), and high-level resistance ( $>60$ ) based on the resistance scores calculated from expert assessments in HIVdb. Importantly, each of these variants, taken from HIV-infected patients, contained a large number of mutations in each monomer (15-25) when compared to the reference NL4-3 (therefore, in some of the simulated systems the protease bore as many as 50 mutations, representing an incredibly difficult test).

We used the same sampling protocol for all mutants: i) search for the crystal structure with the highest identity to the mutant sequence using BLAST; ii) building our model by replacing each mutant using

## Summary of the results and discussion

Maestro(Sastry, Adzhigirey et al. 2013) ; iii) initial docking of the ligand with Glide; iv) PELE exploration of the induced fit response to the mutations (12 processors x 12 hours); v) computation of the binding energy change (upon mutation) by averaging the interaction energies of all accepted minima (approximately 2000 snapshots) and comparison with the reference (non-resistance) sequence.

Figure 5 shows the results for amprenavir (APV) where we computed PELE's relative binding energies to the reference compound NL4-3. Sensitive, intermediate and resistance HIVdb values for each sequence are shown in green, yellow and red colors, respectively. Clearly our prediction technique could identify all high resistant mutations (with a decrease in binding affinity, turning into an increase of relative interaction energy  $> 5$  kcal/mol); similar accuracy was obtained for Darunavir. This outstanding results indicate how computational techniques are today mature enough to accurately predict the effect of multiple mutations in drug binding (drug resistance). This prediction tool has been added to our public server, <https://pele.bsc.es>.



**Figure 5.** PELE's relative change in APV binding energies (kcal/mol). Sensitive, intermediate and resistance HIVdb values for each sequence are shown in green, yellow and red colors, respectively.



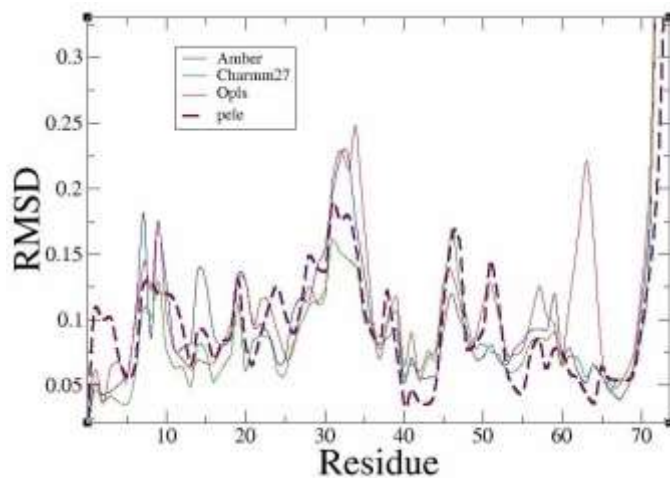
#### **4.5 Beside these main objectives based on methods application, we aim to add methodological improvements derived from the application and validation studies.**

Along with the previous main objectives we performed method development and studied PELE protocols to model long-time protein dynamics by means of normal mode perturbation and constrained minimization.

As we discussed in the introduction section, from the early papers it was clear that PELE needed a better backbone and sampling procedure. Therefore, we introduced a new protein perturbation step based on anisotropic network model methodologies, capable of providing significant backbone motion that was coupled to constraint minimization. Such new perturbation protocol was examined by comparison to microsecond MD on ubiquitin and to metadynamic sampling on T4Lysozyme (Cossins, Hosseini et al. 2012).

Figure 6 shows the comparison of the RMSD from the initial structure for individual residues from the PELE and 3 explicitly solvated 1  $\mu$ s MD simulations of ubiquitin, each of them using a different force field: CHARMM, AMBER and OPLS-AA. While there are some quantitative differences (among PELE and MD but also among different MD simulations) there is an overall good qualitative agreement. The agreement was also present when comparing clustering results and the atomistic forces. For T4Lysozyme, PELE results indicate that we can quickly sample all regions observed in the metadynamics study. Overall, our results indicate that PELE was able to provide protein sampling using a reduced network model eigen problem approach in a fast manner, showing a clear competitive (computational time) advantage over molecular dynamics simulations.

## Summary of the results and discussion



**Figure 6.** Mean residue displacement along the PELE and MD trajectories

## Summary of the results and discussion

Conclusion

## **5 Conclusion**

## Conclusion

## Conclusion

1. The PELE method and its application to proteins and protein complex simulation have been evaluated. The method was tested and compared against available experimental data and designed control test. The technique was used to discover and predict the possible active site and protein-ligand interaction in several proteins, such as mTOR and BCL-2 proteins. Importantly, we demonstrate the critical role of sampling the protein-ligand dynamics in order to improve the docking score. Moreover, the findings reported here clearly shown the capabilities of PG (and its derivatives) for use in particular apoptotic targets.

2. We demonstrated the applicability of the PELE method in solving relevant biophysical problems. In particular, using PELE we introduced a new structural and dynamic paradigm for ligand binding in steroid nuclear receptors.

3. We have shown how PELE can be used in effectively design improved compounds with significant better docking results. We have shown that our designed small molecules bind to the BH3 binding site of MCL1.

4. The protein-ligand landscape exploration of PELE's algorithm allows for an efficient analysis of mutational effects on ligand delivery. Our study on T4MO shows how such an analysis has great possibilities on enzyme engineering. In particular, residue F269 on T4MO, located at the entrance of binding pocket, has a significant role in substrate delivery.

5. We have developed a protocol that is potentially useful in characterizing the effect of multiple mutations on drugs binding to the HIV-1 protease. Using this protocol we introduced a significant advance in predicting the affinity of different drugs against HIV-1 protease with several mutations. This application is fully automated and installed on PELE web server.

6. New backbone perturbation combined with normal modes increased the capability of PELE method to explore local dynamics and large conformational changes.

*Overall, the main message of this thesis is that **an accurate** dynamical view of atomic detailed protein-ligand interaction is necessary... and possible*

## Conclusion

List of publications

## **6 List of publications**



## List of publications

## List of publications

1. M. Espona-Fiedler, V. Soto-Cerrato, **A. Hosseini**, J.M. Lizcano, V. Guallar, R. Quesada, T. Gao, R. Perez-Tomas“ Identification of dual mTORC1 and mTORC2 inhibitors in melanoma cells: Prodigiosin vs. Obatoclox” **Biochem Pharmacol** (2011), doi:10.1016/j.bcp.2011.11.027.
2. Benjamin P. Cossins, **Ali Hosseini**, and Victor Guallar, ”Exploration of protein conformational change with PELE and Meta-dynamics”**J. Chem. Theory Comput.** 10.1021/ct200675g• Publication Date: 27 Jan (2012).
3. **Hosseini A**, Espona-Fiedler M, Soto-Cerrato V, Quesada R, Pérez-Tomás R, et al. Molecular Interactions of Prodiginines with the BH3 Domain of Anti-Apoptotic Bcl-2 Family Members. **PLoS ONE** (2013) 8(2): e57562. doi:10.1371/journal.pone.0057562
4. **Hosseini, A.**; Brouk, M.; Lucas, M. F.; Glaser, F.; Fishman, A.; Guallar, V., Atomic Picture of Ligand Migration in Toluene 4-Monooxygenase. **J. Phys. Chem. B** (2014). DOI 10.1021/jp502509a
5. **Hosseini, A.**; V. Guallar, “Protein-ligand interaction and induce fit docking of different inhibitors for HIV-1 protease” Collaborating with experimental group. **Submitted to The Journal of Medicinal Chemistry.**
6. Karl Edman, **Ali Hosseini**, Anders Hogner, Matti Lepistö, Victor Guallar “Identification of ligand entry pathway in the ligand binding domain of GR and MR receptors” Collaborating with Pharmaceutical company(AstraZeneca). **Submitted to The Journal of Nature structural & molecular biology.**

## List of publications

References

## **7 References**

## References

## References

Adam, G. C., B. F. Cravatt, et al. "Profiling the specific reactivity of the proteome with non-directed activity-based probes." Chemistry & Biology **8**(1): 81-95.

Alberts, B. (1989). Molecular Biology of the Cell, Garland Pub.

Alexandrov, V., U. Lehnert, et al. (2005). "Normal modes for predicting protein motions: A comprehensive database assessment and associated Web tool." Protein Science : A Publication of the Protein Society **14**(3): 633-643.

Ali, A., R. M. Bandaranayake, et al. (2010). "Molecular Basis for Drug Resistance in HIV-1 Protease." Viruses **2**(11): 2509-2535.

Altschul, S. F., W. Gish, et al. (1990). "Basic local alignment search tool." J. Mol. Biol. **215**(3): 403-410.

Atilgan, A. R., S. R. Durell, et al. (2001). "Anisotropy of fluctuation dynamics of proteins with an elastic network model." Biophysical Journal **80**(1): 505-515.

Bahar, I. and A. J. Rader (2005). "Coarse-grained normal mode analysis in structural biology." Current Opinion in Structural Biology **15**(5): 586-592.

Becker, O. M., D. S. Dhanoa, et al. (2006). "An Integrated in Silico 3D Model-Driven Discovery of a Novel, Potent, and Selective Amidosulfonamide 5-HT1A Agonist (PRX-00023) for the Treatment of Anxiety and Depression." Journal of Medicinal Chemistry **49**(11): 3116-3135.

Berendsen, H. J. C. (1988). "Dynamic simulation as an essential tool in molecular modeling." Journal of Computer-Aided Molecular Design **2**(3): 217-221.

Berg, J. M., J. L. Tymoczko, et al. (2010). Biochemistry, W. H. Freeman.

Binder, K. and D. Heermann (2010). Monte Carlo Simulation in Statistical Physics: An Introduction, Springer-Verlag.

Bloomfield, V. A., D. M. Crothers, et al. (2000). Nucleic Acids: Structures, Properties, and Functions, University Science Books.

## References

Boehr, D. D., R. Nussinov, et al. (2009). "The role of dynamic conformational ensembles in biomolecular recognition." Nat Chem Biol **5**(11): 789-796.

Bogoyevitch, M. A., R. K. Barr, et al. (2005). "Peptide inhibitors of protein kinases - discovery, characterisation and use." Biochimica Et Biophysica Acta-Proteins and Proteomics **1754**(1-2): 79-99.

Borrelli, K., A. Vitalis, et al. (2005). "PELE: Protein Energy Landscape Exploration. A Novel Monte Carlo Based Technique." Journal of Chemical Theory and Computation **1**(6): 1304-1311.

Borrelli, K., A. Vitalis, et al. (2005). "PELE: Protein Energy Landscape Exploration. A Novel Monte Carlo Based Technique." J. Chem. Theory Comput. **1**(6): 1304-1311.

Borrelli, K. W., B. Cossins, et al. (2010). "Exploring hierarchical refinement techniques for induced fit docking with protein and ligand flexibility." Journal of Computational Chemistry **31**(6): 1224-1235.

Borrelli, K. W., A. Vitalis, et al. (2005). "PELE: Protein Energy Landscape Exploration. A Novel Monte Carlo Based Technique." Journal of Chemical Theory and Computation **1**(6): 1304-1311.

Bowen, J. P. and N. L. Allinger (2007). Molecular Mechanics: The Art and Science of Parameterization. Reviews in Computational Chemistry, John Wiley & Sons, Inc.: 81-97.

Brocchieri, L. and S. Karlin (2005). "Protein length in eukaryotic and prokaryotic proteomes." Nucleic Acids Research **33**(10): 3390-3400.

Brooijmans, N. and I. D. Kuntz (2003). "Molecular recognition and docking algorithms." Annual Review of Biophysics and Biomolecular Structure **32**: 335-373.

Callaghan, P. T. (1993). Principles of Nuclear Magnetic Resonance Microscopy, Clarendon Press.

Carlson, H. A. and J. A. McCammon (2000). "Accommodating Protein Flexibility in Computational Drug Design." Molecular Pharmacology **57**(2): 213-218.

## References

- Case, D. A. (1994). "Normal mode analysis of protein dynamics." Current Opinion in Structural Biology **4**(2): 285-290.
- Cohen, M. L. (1992). "Epidemiology of Drug-Resistance - Implications for a Postantimicrobial ERA." Science **257**(5073): 1050-1055.
- Condra, J. H., W. A. Schleif, et al. (1995). "IN-VIVO Emergence of HIV-1 Variants Resistant to Multiple Protease Inhibitors." Nature **374**(6522): 569-571.
- Cornell, W. D., P. Cieplak, et al. (1995). "A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules." Journal of the American Chemical Society **117**(19): 5179-5197.
- Cossins, B. P., A. Hosseini, et al. (2012). "Exploration of Protein Conformational Change with PELE and Meta-Dynamics." Journal of Chemical Theory and Computation **8**(3): 959-965.
- Changeux, J.-P. and S. Edelstein (2011). "Conformational selection or induced fit? 50 years of debate resolved." F1000 Biology Reports **3**: 19.
- Cho, A. E., V. Guallar, et al. (2005). "Importance of accurate charges in molecular docking: Quantum mechanical/molecular mechanical (QM/MM) approach." Journal of Computational Chemistry **26**(9): 915-931.
- Damm, W., A. Frontera, et al. (1997). "OPLS all-atom force field for carbohydrates." Journal of Computational Chemistry **18**(16): 1955-1970.
- Doruker, P., A. R. Atilgan, et al. (2000). "Dynamics of proteins predicted by molecular dynamics simulations and analytical approaches: Application to  $\alpha$ -amylase inhibitor." Proteins: Structure, Function, and Bioinformatics **40**(3): 512-524.
- Dror, R. O., A. C. Pan, et al. (2011). "Pathway and mechanism of drug binding to G-protein-coupled receptors." Proceedings of the National Academy of Sciences of the United States of America **108**(32): 13118-13123.
- Espona-Fiedler, M., V. Soto-Cerrato, et al. (2012). "Identification of dual mTORC1 and mTORC2 inhibitors in melanoma cells: Prodigiosin vs. obatoclastax." Biochem. Pharmacol. **83**(4): 489-496.



## References

- Eswar, N., B. Webb, et al. (2002). Comparative Protein Structure Modeling Using Modeller. Curr Protoc Bioinformatics, John Wiley & Sons, Inc.
- Eyal, E., L.-W. Yang, et al. (2006). "Anisotropic network model: systematic evaluation and a new web interface." Bioinformatics **22**(21): 2619-2627.
- Fahy, E., S. Subramaniam, et al. (2009). "Update of the LIPID MAPS comprehensive classification system for lipids." Journal of Lipid Research **50**: S9-S14.
- Fischer, E. (1894). "Einfluss der Configuration auf die Wirkung der Enzyme." Berichte der deutschen chemischen Gesellschaft **27**(3): 2985-2993.
- Fiser, A., R. K. G. Do, et al. (2000). "Modeling of loops in protein structures." Protein Sci. **9**(9): 1753-1773.
- Freire, E., O. L. Mayorga, et al. (1990). "Isothermal titration calorimetry." Analytical Chemistry **62**(18): 950A-959A.
- Ghosh, A. K., Z. L. Dawson, et al. (2007). "Darunavir, a conceptually new HIV-1 protease inhibitor for the treatment of drug-resistant HIV." Bioorganic & Medicinal Chemistry **15**(24): 7576-7580.
- Gilks, C. F., S. Crowley, et al. (2006). "The WHO public-health approach to antiretroviral treatment against HIV in resource-limited settings." Lancet **368**(9534): 505-510.
- Gohlke, H. and G. Klebe (2002). "Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors." Angewandte Chemie (International ed. in English) **41**(15): 2644-2676.
- Gubareva, L. V., R. Bethell, et al. (1996). "Characterization of mutants of influenza A virus selected with the neuraminidase inhibitor 4-guanidino-Neu5Ac2en." J VIROL **70**(3): 1818-1827.
- Gupta, R. K., A. Hill, et al. (2009). "Virological monitoring and resistance to first-line highly active antiretroviral therapy in adults infected with HIV-1 treated under WHO guidelines: a systematic review and meta-analysis." Lancet Infect Dis **9**(7): 409-417.

## References

Halgren, T. A. (1996). "Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94." Journal of Computational Chemistry **17**(5-6): 490-519.

Halgren, T. A., R. B. Murphy, et al. (2004). "Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening." J. Med. Chem. **47**(7): 1750-1759.

Halperin, I., B. Ma, et al. (2002). "Principles of docking: An overview of search algorithms and a guide to scoring functions." Proteins: Structure, Function, and Bioinformatics **47**(4): 409-443.

Hamers, R. L., K. C. E. Sigaloff, et al. (2012). "Patterns of HIV-1 Drug Resistance After First-Line Antiretroviral Therapy (ART) Failure in 6 Sub-Saharan African Countries: Implications for Second-Line ART Strategies." CLIN INFECT DIS **54**(11): 1660-1669.

Hamers, R. L., C. L. Wallis, et al. (2011). "HIV-1 drug resistance in antiretroviral-naive individuals in sub-Saharan Africa after rollout of antiretroviral therapy: a multicentre observational study." Lancet Infect Dis **11**(10): 750-759.

Hogg, R. S., K. V. Heath, et al. (1998). "Improved survival among HIV-infected individuals following initiation of antiretroviral therapy." JAMA-J AM MED ASSOC **279**(6): 450-454.

Hosseini, A., M. Espona-Fiedler, et al. (2013). "Molecular Interactions of Prodiginines with the BH3 Domain of Anti-Apoptotic Bcl-2 Family Members." Plos One **8**(2): 1-8.

Hsu, J. T. A., H. C. Wang, et al. (2006). "Antiviral drug discovery targeting to viral proteases." Current Pharmaceutical Design **12**(11): 1301-1314.

Huang, Z. (2007). Drug Discovery Research: New Frontiers in the Post-Genomic Era, Wiley.

Imanishi, J., Y. Morita, et al. (2011). "Pharmacological profile of FK881(ASP6537), a novel potent and selective cyclooxygenase-1 inhibitor." Biochemical Pharmacology **82**(7): 746-754.

Imming, P., C. Sinning, et al. (2006). "Drugs, their targets and the nature and number of drug targets." Nat Rev Drug Discov **5**(10): 821-834.

## References

- Inc., C. o. G. (2013). "Molecular Operating Environment (MOE)."
- Johnson, V. A., V. Calvez, et al. (2013). "Update of the drug resistance mutations in HIV-1: March 2013." Top Antivir Med. **21**(1): 6-14.
- Jones, G., P. Willett, et al. (1995). "Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation." Journal of Molecular Biology **245**(1): 43-53.
- Jorgensen, W. L., D. S. Maxwell, et al. (1996). "Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids." Journal of the American Chemical Society **118**(45): 11225-11236.
- Kapetanovic, I. M. (2008). "COMPUTER-AIDED DRUG DISCOVERY AND DEVELOPMENT (CADD): in silico-chemico-biological approach." Chemico-biological interactions **171**(2): 165-176.
- Keiser, O., C. Orrell, et al. (2008). "Public-health and individual approaches to antiretroviral therapy: Township South Africa and Switzerland compared." Plos Medicine **5**(7): 1102-1111.
- Kempf, D. J., K. C. Marsh, et al. (1995). "ABT-538 IS A POTENT INHIBITOR OF HUMAN-IMMUNODEFICIENCY-VIRUS PROTEASE AND HAS HIGH ORAL BIOAVAILABILITY IN HUMANS." Proceedings of the National Academy of Sciences of the United States of America **92**(7): 2484-2488.
- Kendrew, J. C., G. Bodo, et al. (1958). "A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis." Nature **181**(4610): 662-666.
- Kim, D. E., D. Chivian, et al. (2004). "Protein structure prediction and analysis using the Robetta server." Nucleic Acids Research **32**(Web Server issue): W526-W531.
- Kim, E. E., C. T. Baker, et al. (1995). "Crystal-structure of HIV-1 Protease in Complex with VX-478, a Potent and Orally Bioavailable Inhibitor of the Enzyme." J. Am. Chem. Soc. **117**(3): 1181-1182.
- Kitchen, D. B., H. Decornez, et al. (2004). "Docking and scoring in virtual screening for drug discovery: methods and applications." Nat Rev Drug Discov **3**(11): 935-949.

## References

- Koh, Y., H. Nakata, et al. (2003). "Novel bis-tetrahydrofuranylurethane-containing nonpeptidic protease inhibitor (PI) UIC-94017 (TMC114) with potent activity against multi-PI-resistant human immunodeficiency virus in vitro." Antimicrob. Agents Chemother. **47**(10): 3123-3129.
- Kollman, P. A., I. Massova, et al. (2000). "Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models." Accounts of Chemical Research **33**(12): 889-897.
- Kumar, S., J. M. Rosenberg, et al. (1992). "THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method." Journal of Computational Chemistry **13**(8): 1011-1021.
- Laio, A. and M. Parrinello (2002). "Escaping free-energy minima." Proceedings of the National Academy of Sciences **99**(20): 12562-12566.
- Lamb, M. L. and W. L. Jorgensen (1997). "Computational approaches to molecular recognition." Current Opinion in Chemical Biology **1**(4): 449-457.
- Larder, B. A. and S. D. Kemp (1989). "Multiple Mutations in HIV-1 Reverse-Transcriptase Confer High-Level Resistance to Zidovudine (AZT)." Science **246**(4934): 1155-1158.
- Lengauer, T. and M. Rarey (1996). "Computational methods for biomolecular docking." Current Opinion in Structural Biology **6**(3): 402-406.
- Lewars, E. G. (2010). Computational Chemistry: Introduction to the Theory and Applications of Molecular and Quantum Mechanics, Springer Netherlands.
- Lewis, E. and K. Murphy (2005). Isothermal Titration Calorimetry. Protein-Ligand Interactions. G. Ulrich Nienhaus, Humana Press. **305**: 1-15.
- Lindberg, J., D. Pyring, et al. (2004). "Symmetric fluoro-substituted diol-based HIV protease inhibitors - Ortho-fluorinated and meta-fluorinated P1/P1'-benzyloxy side groups significantly improve the antiviral activity and preserve binding efficacy." Eur. J. Biochem. **271**(22): 4594-4602.
- Liu, T. F. and R. W. Shafer (2006). "Web resources for HIV type 1 genotypic-resistance test interpretation." CLIN INFECT DIS **42**(11): 1608-1618.

## References

- Lu, I. L., C.-F. Huang, et al. (2006). "Structure-Based Drug Design of a Novel Family of PPAR $\gamma$  Partial Agonists: Virtual Screening, X-ray Crystallography, and in Vitro/in Vivo Biological Activities." Journal of Medicinal Chemistry **49**(9): 2703-2712.
- Lu, Y., Z. Nikolovska-Coleska, et al. (2006). "Discovery of a Nanomolar Inhibitor of the Human Murine Double Minute 2 (MDM2)-p53 Interaction through an Integrated, Virtual Database Screening Strategy." Journal of Medicinal Chemistry **49**(13): 3759-3762.
- Lucas, M. F. and V. Guallar (2013). "Single vs. multiple ligand pathways in globins: A computational view." Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics **1834**(9): 1739-1743.
- Lyne, P. D., P. W. Kenny, et al. (2004). "Identification of Compounds with Nanomolar Binding Affinity for Checkpoint Kinase-1 Using Knowledge-Based Virtual Screening." Journal of Medicinal Chemistry **47**(8): 1962-1968.
- Manetti, F., G. A. Locatelli, et al. (2006). "A Combination of Docking/Dynamics Simulations and Pharmacophoric Modeling To Discover New Dual c-Src/Abl Kinase Inhibitors." Journal of Medicinal Chemistry **49**(11): 3278-3286.
- Mann, S., B. R. Heywood, et al. (1991). Molecular Recognition in Biomineralization. Mechanisms and Phylogeny of Mineralization in Biological Systems. S. Suga and H. Nakahara, Springer Japan: 47-55.
- Mason, J. S., A. C. Good, et al. (2001). "3-D Pharmacophores in Drug Discovery." Current Pharmaceutical Design **7**(7): 567-597.
- Mitra, P., D. Shultis, et al. (2013). "An Evolution-Based Approach to De Novo Protein Design and Case Study on Mycobacterium tuberculosis." PLoS Comput Biol **9**(10): e1003298.
- Mittal, S., R. M. Bandaranayake, et al. (2013). "Structural and Thermodynamic Basis of Amprenavir/Darunavir and Atazanavir Resistance in HIV-1 Protease with Mutations at Residue 50." J. Virol. **87**(8): 4176-4184.
- Mohan, A., C. J. Oldfield, et al. (2006). "Analysis of Molecular Recognition Features (MoRFs)." Journal of Molecular Biology **362**(5): 1043-1059.

## References

Morris, G. M., R. Huey, et al. (2009). "AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility." Journal of Computational Chemistry **30**(16): 2785-2791.

Nienhaus, G. U. (2010). ProteinLigand Interactions: Methods and Applications, Humana Press.

Olsen, L., S. Jost, et al. (2006). "New leads of metallo- $\beta$ -lactamase inhibitors from structure-based pharmacophore design." Bioorganic & Medicinal Chemistry **14**(8): 2627-2635.

Oostenbrink, C., A. Villa, et al. (2004). "A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6." Journal of Computational Chemistry **25**(13): 1656-1676.

Orellana, L., M. Rueda, et al. (2010). "Approaching Elastic Network Models to Molecular Dynamics Flexibility." Journal of Chemical Theory and Computation **6**(9): 2910-2923.

Organization, W. H. (2012). WHO HIV Drug Resistance Report 2012, World Health Organization.

Palella, F. J., K. M. Delaney, et al. (1998). "Declining Morbidity and Mortality among Patients with Advanced Human Immunodeficiency Virus Infection." NEW ENGL J MED **338**(13): 853-860.

Paredes, R., V. C. Marconi, et al. (2013). "Impact of Antiretroviral Drugs in Pregnant Women and Their Children in Africa: HIV Resistance and Treatment Outcomes." J INFECT DIS **207**: S93-S100.

Pattnaik, P. (2005). "Surface plasmon resonance." Applied Biochemistry and Biotechnology **126**(2): 79-92.

Pauling, L., R. B. Corey, et al. (1951). "THE STRUCTURE OF PROTEINS - 2 HYDROGEN-BONDED HELICAL CONFIGURATIONS OF THE POLYPEPTIDE CHAIN." Proceedings of the National Academy of Sciences of the United States of America **37**(4): 205-211.

Peng, H., N. Huang, et al. (2003). "Identification of novel inhibitors of BCR-ABL tyrosine kinase via virtual screening." Bioorganic & Medicinal Chemistry Letters **13**(21): 3693-3699.

## References

Petrillo, E. W. and M. A. Ondetti (1982). "Angiotensin-converting enzyme inhibitors: Medicinal chemistry and biological actions." Medicinal Research Reviews **2**(1): 1-41.

Przybylski, M. and M. O. Glocker (1996). "Electrospray Mass Spectrometry of Biomacromolecular Complexes with Noncovalent Interactions—New Analytical Perspectives for Supramolecular Chemistry and Molecular Recognition Processes." Angewandte Chemie International Edition in English **35**(8): 806-826.

Queneau, Y., A. P. Rauter, et al. (2014). Carbohydrate Chemistry: Chemical and Biological Approaches, Royal Society of Chemistry.

Rapaport, D. C. (2004). The Art of Molecular Dynamics Simulation, Cambridge University Press.

Rathore, N. and J. J. de Pablo (2002). "Monte Carlo simulation of proteins through a random walk in energy space." The Journal of Chemical Physics **116**(16): 7225-7230.

Rhee, S. Y., W. J. Fessel, et al. (2005). "HIV-1 protease and reverse-transcriptase mutations: Correlations with antiretroviral therapy in subtype B isolates and implications for drug-resistance surveillance." J INFECT DIS **192**(3): 456-465.

Rodger, A. J., R. Lodwick, et al. (2013). "Mortality in well controlled HIV in the continuous antiretroviral therapy arms of the SMART and ESPRIT trials compared with the general population." Aids **27**(6): 973-979.

Rosenberg, M. and A. Goldblum (2006). "Computational Protein Design: A Novel Path to Future Protein Drugs." Current Pharmaceutical Design **12**(31): 3973-3997.

Rubinstein, R. Y. and D. P. Kroese (2011). Simulation and the Monte Carlo Method, Wiley.

Ruiz-Carmona, S., D. Alvarez-Garcia, et al. (2014). "rDock: A Fast, Versatile and Open Source Program for Docking Ligands to Proteins and Nucleic Acids." PLoS Comput Biol **10**(4): e1003571.

Sastry, G. M., M. Adzhigirey, et al. (2013). "Protein and Ligand Preparation: Parameters, Protocols, and Influence on Virtual Screening Enrichments." J. Comput. Aided Mol. Des. **27**(3): 221-234.

## References

Sawyer, J. S., B. D. Anderson, et al. (2003). "Synthesis and Activity of New Aryl- and Heteroaryl-Substituted Pyrazole Inhibitors of the Transforming Growth Factor- $\beta$  Type I Receptor Kinase Domain." Journal of Medicinal Chemistry **46**(19): 3953-3956.

Schneider, G., W. Neidhart, et al. (1999). "'Scaffold-Hopping' by Topological Pharmacophore Search: A Contribution to Virtual Screening." Angewandte Chemie International Edition **38**(19): 2894-2896.

Schwede, T., J. Kopp, et al. (2003). "SWISS-MODEL: an automated protein homology-modeling server." Nucleic Acids Research **31**(13): 3381-3385.

Segel, I. H. (1987). "CITATION-CLASSIC - ENZYME-KINETICS - BEHAVIOR AND ANALYSIS OF RAPID EQUILIBRIUM AND STEADY-STATE ENZYME-SYSTEMS." Current Contents/Life Sciences(16): 14-14.

Shen, C.-H., Y.-F. Wang, et al. (2010). "Amprenavir complexes with HIV-1 protease and its drug-resistant mutants altering hydrophobic clusters." FEBS Journal **277**(18): 3699-3714.

Shen, L., J. Shen, et al. (2003). "Steered Molecular Dynamics Simulation on the Binding of NNRTI to HIV-1 RT." Biophysical Journal **84**(6): 3547-3563.

Shenderovich, M. D., R. M. Kagan, et al. (2003). "Structure-based phenotyping predicts HIV-1 protease inhibitor resistance." Prot Sci **12**(8): 1706-1718.

Sigaloff, K. C. E., J. C. J. Calis, et al. (2011). "HIV-1-resistance-associated mutations after failure of first-line antiretroviral treatment among children in resource-poor regions: a systematic review." Lancet Infect Dis **11**(10): 769-779.

Sigaloff, K. C. E., R. L. Hamers, et al. (2012). "Second-Line Antiretroviral Treatment Successfully Resuppresses Drug-Resistant HIV-1 After First-Line Failure: Prospective Cohort in Sub-Saharan Africa." J INFECT DIS **205**(11): 1739-1744.

Singh, J., C. E. Chuaqui, et al. (2003). "Successful shape-Based virtual screening: The discovery of a potent inhibitor of the type I TGF $\beta$  receptor kinase (T $\beta$ RI)." Bioorganic & Medicinal Chemistry Letters **13**(24): 4355-4359.



## References

- Sircar, A., K. A. Sanni, et al. (2011). "Analysis and Modeling of the Variable Region of Camelid Single Domain Antibodies." Journal of immunology (Baltimore, Md. : 1950) **186**(11): 6357-6367.
- Skálová, T., J. Dohnálek, et al. (2006). "HIV-1 Protease Mutations and Inhibitor Modifications Monitored on a Series of Complexes. Structural Basis for the Effect of the A71V Mutation on the Active Site†." J. Med. Chem. **49**(19): 5777-5784.
- Smyth, M. S. and J. H. J. Martin (2000). "x Ray crystallography." Molecular Pathology **53**(1): 8-14.
- Sugita, Y. and Y. Okamoto (1999). "Replica-exchange molecular dynamics method for protein folding." Chemical Physics Letters **314**(1-2): 141-151.
- Takahashi, R., V. A. Gil, et al. (2013). "Monte Carlo Free Ligand Diffusion with Markov State Model Analysis and Absolute Binding Free Energy Calculations." Journal of Chemical Theory and Computation **10**(1): 282-288.
- Tantillo, C., J. P. Ding, et al. (1994). "Locations of ANTI-AIDS Drug-Binding Sites and Resistance Mutations in the 3-Dimensional Structure of HIV-1 Reverse-Transcriptase - Implications for Mechanisms of Drug-Inhibition and Resistance." J. Mol. Biol. **243**(3): 369-387.
- Tie, Y., Y.-F. Wang, et al. (2012). "Critical differences in HIV-1 and HIV-2 protease specificity for clinical inhibitors." Protein Sci. **21**(3): 339-350.
- Tie, Y. F., P. I. Boross, et al. (2004). "High resolution crystal structures of HIV-1 protease with a potent non-peptide inhibitor (UIC-94017) active against multi-drug-resistant clinical strains." J. Mol. Biol. **338**(2): 341-352.
- Tsai, C.-J., R. Norel, et al. (2001). Protein-Ligand Interactions: Energetic Contributions and Shape Complementarity. eLS, John Wiley & Sons, Ltd.
- Turner, S. R., J. W. Strohbach, et al. (1998). "Tipranavir (PNU-140690): A potent, orally bioavailable nonpeptidic HIV protease inhibitor of the 5,6-dihydro-4-hydroxy-2-pyrone sulfonamide class." J. Med. Chem. **41**(18): 3467-3476.

## References

Vangrevelinghe, E., K. Zimmermann, et al. (2003). "Discovery of a Potent and Selective Protein Kinase CK2 Inhibitor by High-Throughput Docking." Journal of Medicinal Chemistry **46**(13): 2656-2662.

Vanommeslaeghe, K., E. Hatcher, et al. (2010). "CHARMM General Force Field: A Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields." Journal of Computational Chemistry **31**(4): 671-690.

Varady, J., X. Wu, et al. (2003). "Molecular Modeling of the Three-Dimensional Structure of Dopamine 3 (D3) Subtype Receptor: Discovery of Novel and Potent D3 Ligands through a Hybrid Pharmacophore- and Structure-Based Database Searching Approach." Journal of Medicinal Chemistry **46**(21): 4377-4392.

Voet, D. and J. G. Voet (2010). Biochemistry, Wiley.

Wang, J., R. M. Wolf, et al. (2004). "Development and testing of a general amber force field." Journal of Computational Chemistry **25**(9): 1157-1174.

Wang, W., O. Donini, et al. (2001). "BIOMOLECULAR SIMULATIONS: Recent Developments in Force Fields, Simulations of Enzyme Catalysis, Protein-Ligand, Protein-Protein, and Protein-Nucleic Acid Noncovalent Interactions." Annual Review of Biophysics and Biomolecular Structure **30**(1): 211-243.

Weiner, S. J., P. A. Kollman, et al. (1986). "An all atom force field for simulations of proteins and nucleic acids." Journal of Computational Chemistry **7**(2): 230-252.

Xu, C. Y., D. Tobi, et al. (2003). "Allosteric changes in protein structure computed by a simple mechanical model: Hemoglobin T  $\leftrightarrow$  R2 transition." Journal of Molecular Biology **333**(1): 153-168.

Yang, S.-Y. (2010). "Pharmacophore modeling and applications in drug discovery: challenges and recent advances." Drug Discovery Today **15**(11-12): 444-450.

Young, C., J. A. Bank, et al. (2009). A 32x32x32, spatially distributed 3D FFT in four microseconds on Anton. High Performance Computing Networking, Storage and Analysis, Proceedings of the Conference on.

## References

Yu, Z., M. P. Jacobson, et al. (2004). "First-Shell Solvation of Ion Pairs: Correction of Systematic Errors in Implicit Solvent Models†." J. Phys. Chem. B **108**(21): 6643-6654.

Zhang, Y. (2008). "Progress and challenges in protein structure prediction." Current Opinion in Structural Biology **18**(3): 342-348.

## References

