BearWorks

Summer 2016

MDM2 Case Study: Computational Protocol Utilizing Protein Flexibility Improves Ligand Binding Mode Predictions

Anthony Thomas Ascone

Follow this and additional works at: https://bearworks.missouristate.edu/theses

Part of the Materials Science and Engineering Commons

Recommended Citation

MDM2 CASE STUDY: COMPUTATIONAL PROTOCOL UTILIZING PROTEIN

FLEXIBILITY IMPROVES LIGAND BINDING MODE PREDICTIONS

A Masters Thesis

Presented to

The Graduate College of

Missouri State University

In Partial Fulfillment

Of the Requirements for the Degree

Master of Science, Materials Science

By

Anthony Ascone

July 2016

MDM2 CASE STUDY: COMPUTATIONAL PROTOCOL UTILIZING PROTEIN FLEXIBILITY IMPROVES LIGAND BINDING MODE PREDICTIONS

Physics, Astronomy and Materials Science

Missouri State University, July 2016

Master of Science

Anthony Ascone

ABSTRACT

Recovery of the P53 tumor suppressor pathway via small molecule inhibitors of onco-protein MDM2 highlights the critical role of computational methodologies in targeted cancer therapies. Molecular docking programs in particular, provide a quantitative ranking of predicted binding geometries based on binding free energy allowing for the screening of large chemical libraries in search of lead compounds for cancer therapeutics. This study found improved binding mode predictions of medicinal compounds to MDM2 using the popular docking programs AutoDock and AutoDock Vina, while adopting a rigid-ligand/flexible-receptor protocol. Crystal structures representing small molecule inhibitors bound to MDM2 were selected and a total of 12 rotatable bonds was supplied to each complex and distributed systematically between the ligand and binding site residues. Docking results were evaluated in terms of the top ranked binding free energy and corresponding RMSD values from the experimentally known binding site. Results show lowest RMSD values coincide with a rigid ligand, while the protein retained the majority of flexibility. This study suggests the future implementation of a rigid-ligand/flexible-receptor protocol may improve accuracy of high throughput screenings of potential cancer drugs targeting the MDM2 protein, while maintaining manageable computational costs.

KEYWORDS: molecular docking, MDM2, autodock, autodock vina, drug discovery

This abstract is approved as to form and content

_____
Dr. Ridwan Sakidja
Chairperson, Advisory Committee
Missouri State University

MDM2 CASE STUDY: COMPUTATIONAL PROTOCOL UTILIZING PROTEIN FLEXIBILITY IMPROVES LIGAND BINDING MODE PREDICTIONS

By

Anthony Ascone

A Masters Thesis
Submitted to the Graduate College
Of Missouri State University
In Partial Fulfillment of the Requirements
For the Degree of Master of Science, Materials Science

July 2016

Approved:

_____
Ridwan Sakidja, PhD

_____
Robert Mayanovic, PhD

_____
Fei Wang, PhD

_____
Julie Masterson, PhD: Dean, Graduate College

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Molecular Docking: A Computational Methodology Aiding Drug Discovery

Our increasing understanding of bio-molecular binding processes has been aided by computational docking programs, which utilize physics-based free energy functions and search algorithms to predict the likely covalent and non-covalent binding sites of small molecules to macromolecules, such as proteins, DNA and RNA, based on the quantitative ranking of the binding free energy [1]. This methodology has become especially relevant in the field of structure-based drug design (SBDD), which uses the known 3D structure of target molecules to design small medicinal compounds as therapeutics [2]. This small molecule referred to as a ligand, is often an inhibitor to some function of the target molecule and must be designed to firmly bind at a specific location on the selected target macromolecule (figure 1) [3]. Consequently, docking programs have become an essential tool in the research and design of cancer therapeutics as they allow the quick and accurate screening of drug candidates early in the drug discovery pipeline, thereby precluding a large investment in time and money [4]. The virtual screening of small molecules from chemical libraries attempts to manage the vast chemical space of all possible compounds available to be optimized as future cancer drugs [5]. These large high throughput screenings (HTS) are able to exclude or include available compounds *in silico,* as drug leads based on desired binding geometry; often referred to as "binding pose" or "binding mode", and binding affinity, which is determined by the binding free energy [6]. Due to the complexity inherent in simulating the dynamics of the molecular binding process, docking programs introduced a time-

independent strategy based on chemical potentials rather than the force calculations

associated with classical molecular dynamics (MD), which use time dependent

Newtonian physics [7][8]. Docking programs use energy evaluations based on

assumptions, estimates and empirical knowledge while estimating, rather than

calculating, binding free energy [9]. Cancer therapeutic research such as targeting onco-

proteins relies on a critical understanding of ligand-protein binding and is subject to the

effectiveness of computational techniques such as molecular docking [10]. Therefore,

Particular attention to understanding the limitations and efficacy of these programs will

ensure proper use and enable further modifications.
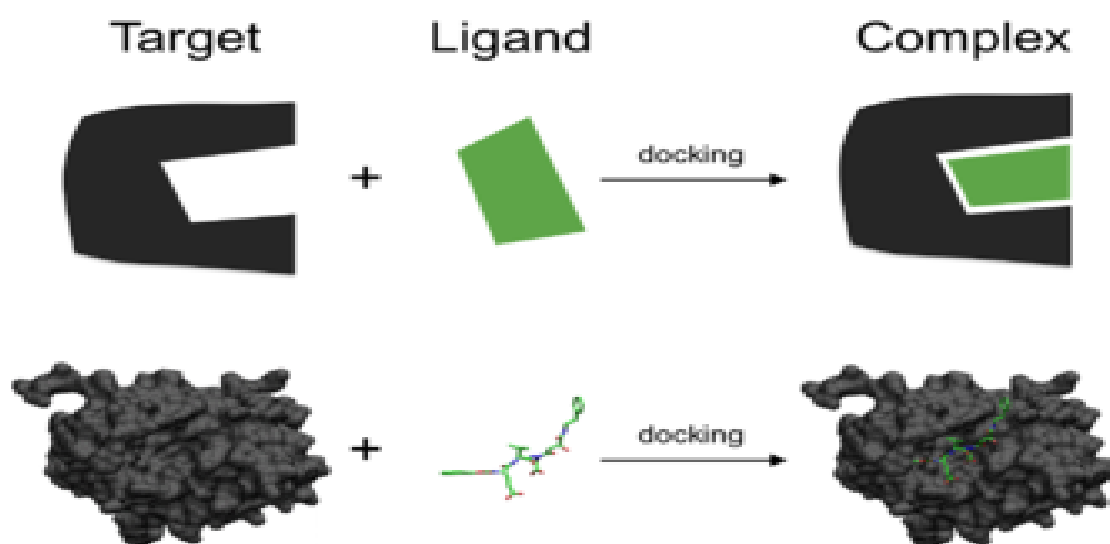
Figure 1. Molecular docking predicts the preferred location and binding affinity of small molecules to proteins, DNA and RNA from an unbound configuration. Taken from [103].

Overview of AutoDock and AutoDock Vina

Currently, there are over 50 docking programs in use worldwide [11]. AutoDock,

introduced in 1990, is an often cited, freely available program maintained by the Scripps Research Institute [12]. Its widespread use can be attributed to its free availability to academic users as well as its versatility and accuracy. AutoDock was the first program to model ligand flexibility, and with the release of AutoDock 4, has incorporated protein side-chain flexibility [13][14]. Although significant advances in this and other programs have been made in the last 25 years, AutoDock suffers from many of the same limitations as other docking programs. In fact, the number of programs reflects these limitations, as it signifies an inability to converge on one scoring function and one search method that would be applicable for a variety of ligand-protein, protein-protein, ligand-DNA or protein-DNA interactions [1][15][16]. Due to its wide use in commercial and academic environments, AutoDock has been the focus of validation studies, which have highlighted efficiencies and certain limitations. For example, AutoDock has been shown to be accurate for ligands with 6 or less rotatable bonds within 2 angstroms of the experimentally known 3D structure, but as the ligands' rotatable bonds increase, performance decreases, and is not reliable after 10 [17][10]. This problem arises from the exponential increase in conformational space that must be evaluated for every additional rotatable bond. This restriction has been the most notable difficulty pertaining to protein-ligand docking and has led to a rigid receptor as the standard methodology because of the computational challenges posed by the numerous degrees of freedom involved in incorporating protein residue flexibility [18]. This lack of allowed flexibility fails to address the proteins' conformational changes during binding. The accurate, computational simulation of protein conformational changes is critical to improved docking studies because it accounts for changes affecting the final binding geometry [19].

It has been shown that, when only a rigid receptor conformation is considered, docking studies predict incorrect binding poses for about 50–70% of all ligands [20]. In response, AutoDock introduced a feature that allowed the incorporation of protein flexibility, which is intended to account for a portion of the conformational changes of the protein upon binding. However, the flexibility of the protein and more importantly, the binding site residues are still subject to the limits imposed by the number of rotatable bonds; therefore, this feature is limited to the ligand and protein having a total of about 10 rotatable bonds.

AutoDock Vina, a faster, more accurate alternative, was released in 2010. (henceforth referred to as Vina). This program was able to improve accuracy, while drastically reducing computation time through effective computer architecture and adopting a "machine learning" approach for the scoring function [21]. Vina was tested using the same complexes evaluated during the development of AutoDock 4 and results show a marked improvement in terms of ligand binding mode accuracy and computational time as compared to AutoDock 4 (figure 2). Vina's combination of speed and accuracy has made it an ideal program for HTS and has been used in several research studies and novel docking approaches [22][23]. Although AutoDock 4 and Vina share similarities in the use of an empirically weighted scoring function and global search optimization algorithm, they differ in their local search strategy and global and scoring function parameters [21].

Improving Binding Mode Predictions

As aforementioned, the two most significant results from a docking experiment
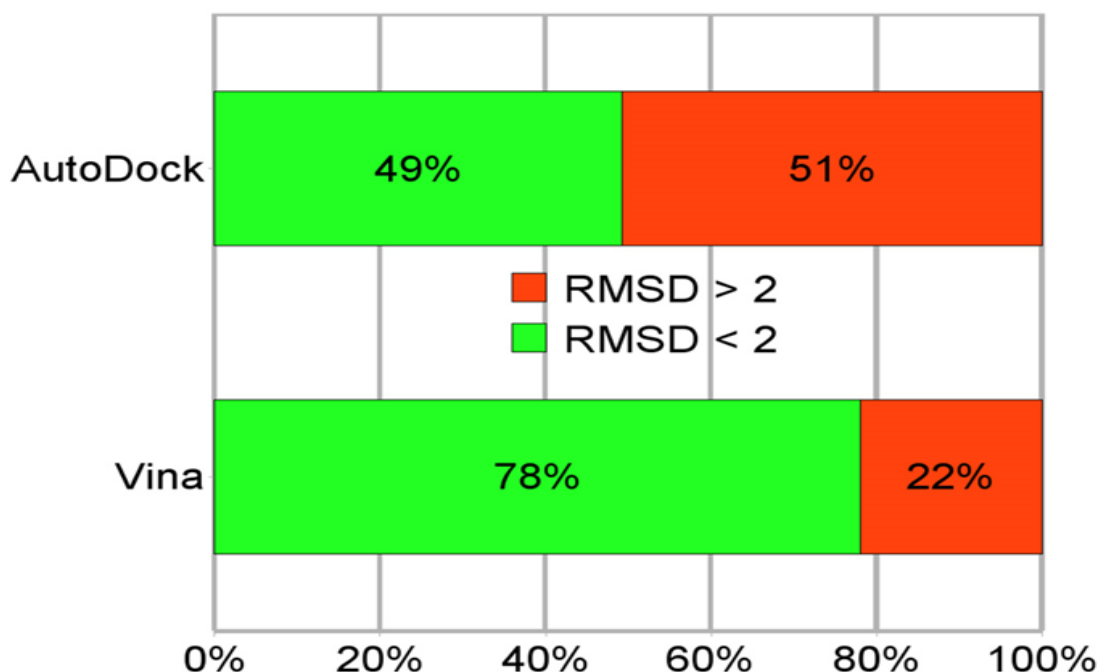
Figure 2. Percentage of the 190 test complexes AutoDock and AutoDock Vina were able to predict within 2 Å of the experimentally known binding site. Taken from [21].

include the binding free energy of a particular bound complex and the ligand binding mode prediction. The focus of this research is improving the ligand binding mode predictions of AutoDock and Vina through selective flexibility of the ligand and receptor by invoking AutoDock's feature of protein residue flexibility, while utilizing the speed and accuracy of Vina to reduce computational time. Few studies have highlighted AutoDock's protein flexibility feature, and no study has been found regarding results from adjusting ligand flexibility, which is another feature available to users. For this study, four structures, all representing small medicinal compounds in complex with the onco-protein MDM2 (murine double minute 2), were selected from the Protein Data Bank (PDB). Docking experiments were performed using all structures with AutoDock and Vina employing the standard protocol. This was followed by docking runs applying a

systematic transfer of flexibility from the ligand to protein for determining the optimal

configuration of flexibility to produce the best binding mode prediction. Analysis of these

results was supplemented by classical MD simulations performed by the Nanoscale

Molecular Dynamics program (NAMD) [24]. MD simulation programs, such as NAMD,

utilize classical Newtonian physics to study the time dependent structure, dynamics, and

thermodynamics of biological molecules. The microscopic properties of atomic positions

and velocities can be translated into macroscopic quantities such as temperature, pressure

and volume using statistical mechanics. This enables the determination of the movement

associated with the flexible binding site residues of the target protein most associated

with the binding interactions [8].

The Cost of Cancer

Cancer is among the leading causes of death worldwide. In 2012, there were 14 million new cases and a reported 8.2 million cancer-related deaths. In 2015, an estimated 1,658,370 new cases of cancer will be diagnosed in the United States and 589,430 people will die from the disease [25]. Pharmaceutical companies are estimated to spend an average of 2.5 billion dollars per every new drug that comes to market because nearly 95% of new drugs tested on humans fail to be effective and safe (figure 3) [26]. A 2012 article in Nature Reviews Drug Discovery reports the number of drugs invented per
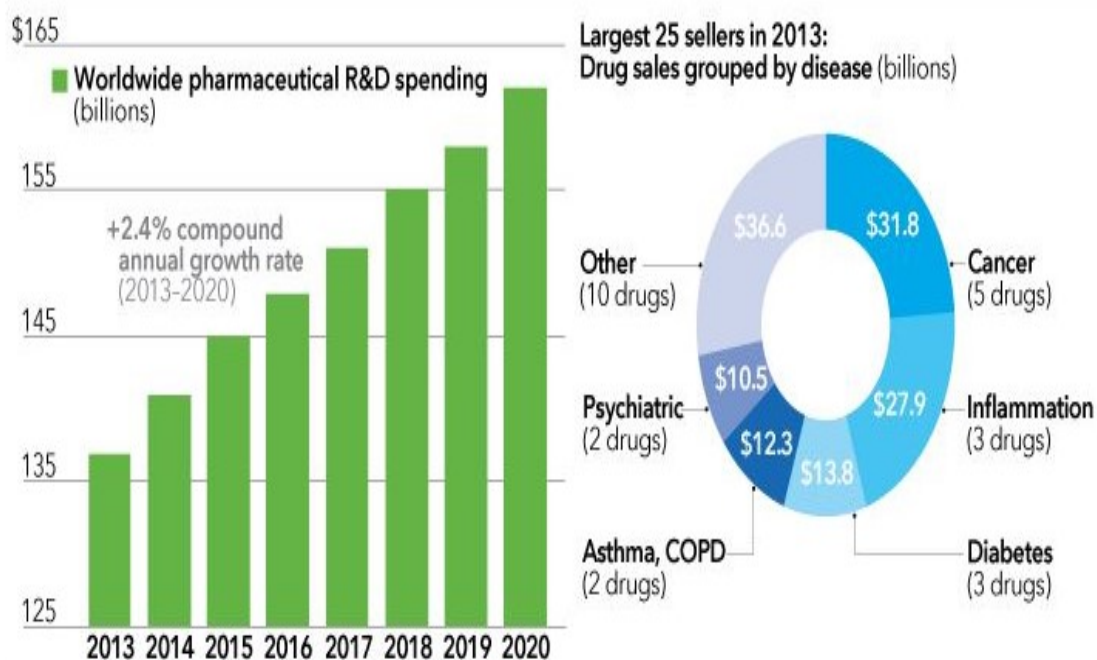
Figure 3. Drug companies spend billions of dollars on developing new drugs, and as this amount rises each year, the cost of Cancer therapeutics and other drugs will continue to rise to offset their investment. Drug companies must spend billions of dollars for each new drug developed that will most likely fail. Taken from [104].

billion dollars of research and development invested has been cut in half every nine years over the past 50 years. In many cases, drug companies spend millions of dollars over the course of several years before the drugs are proven to be ineffective. This enormous investment and large failure rate drives drug costs higher as companies try to recoup their investment when a small amount of medications are successful [27][26]. Drug companies have initiated strategies to lower costs such as abandoning projects earlier that show low probable success rate, partnering with other companies to reduce costs, and targeting rare diseases because they can be more lucrative as the cost per patient is higher [28]. Also, the U.S government has taken steps to help reduce costs by eliminating red tape attempting to fast track drugs to market. In 2014, the Food and Drug Administration (FDA) approved 41 drugs that had not been approved previously for any indication, the most in nearly 20 years. Of these 41 novel drugs, 9 were approved for the treatment of cancer or cancer-related conditions [29].

Targeted Medicine

Over the past ten years, the development of targeted cancer therapies has overtaken standard chemotherapy treatments. Targeted cancer therapies are drugs or other substances that block the growth and spread of cancer by interfering with specific molecules that aide or promote the growth and spread of cancer cells [30]. This strategy reduces toxicity, because unlike chemotherapies, which are chosen for their ability to kill cells, both cancerous and healthy, targeted drugs are designed only to act upon relevant cancerous molecules. These targeted therapies have proved successful as two thirds of cancer patients, most of whom are recipients of such therapies, survive 5

years after diagnosis [31]. However, despite the recent successes and improved diagnosis for cancer patients, targeted therapeutics are more expensive and suffer from certain side-effects and limitations. For example, these drugs mostly target proteins in the cancer cell whose biological function is not reserved for just one activity, and often the function targeted by the small molecule inhibitor is modulated by other proteins. Often the cell will create a new pathway, bypassing the protein the inhibitor was affecting, In other words, the cell acquires a resistance and unfortunately, it can happen quickly [32]. Improved understanding of these molecular pathways has led to the identification of the mechanisms of acquired resistance, which can be mitigated with novel inhibitors. Improving targeted cancer treatments in the future will require expanding our knowledge of resistance mechanisms toward the development of new rationally designed novel inhibitors and translating this information to the clinic to select patients for appropriate therapy [33].

In addition to resistance, another complication associated with targeted drug therapies is off-site binding. This occurs when inhibitors bind to proteins other than the intended target resulting in the unintended disruption of molecular pathways leading to detrimental side-effects. This has become the norm rather than the exception [34]. Research into the extent of drug promiscuity has shown on average each drug binds to 6.3 protein receptors. In fact, the complexity and resiliency of biological systems has led to the failure of most single target, single drug approaches [35]. The problem of off-site binding contributes to the 95% fail rate, as this many times is only discovered in phase 1 or 2 clinical trials [36]. The interactions associated with off-site binding is now being used as a strategy known as polypharmacology, which focuses on searching for drugs that

bind to multiple targets to modify disease associated pathways instead of targeting one protein using a single ligand [34]. This addresses the cells ability to circumvent the modification of certain proteins targeted by small inhibitors, which leads to resistance and continued cancer growth. In 2012, a drug called regorafenib was approved to treat colorectal cancer by blocking six different cancer pathways [37]. Despite this recent success, employing this strategy is a daunting challenge because the effects of binding for the entire biological system must be realized, as well as determining the proper structure of the small molecule to induce high infinity binding. The binding affinity of a ligand to a protein is based on the chemical structure of the ligand and the ligand binding propensity of the protein, and even a weak binding to multiple targets can affect the biological system [38].

Computer-Aided Structure-Based Drug Design

SBDD utilizes the known 3D structure of a target molecule realized through nuclear magnetic resonance spectroscopy, homology modeling or the most common technique, X-ray crystallography, to elucidate the structural components of proteins and ligand-protein complexes [9]. The high resolution associated with these techniques allows for the determination of atomic interactions within the target protein and within small compounds that may be used to bind to the target protein. This allows researchers to locate and focus on a particular area of medicinal interest on the target molecule, while invoking computational methodologies such as docking and MD [39]. This computer-aided approach of SBDD is essential for accurate determination of lead compounds through the virtual screening of chemical libraries early in the drug discovery pipeline as

there are now over one million small molecules in data bases available as possible drug candidates (figure 4) [40].

The structure-based virtual screening procedure performed by docking programs have been used since the early 1980's, tasked with exploring the possible binding modes and binding affinity of a ligand within a known target area. The target area is a specific location on the protein surface active in the modulation or functionality of that protein [41]. Most often, the introduction of a small medicinal compound is used to disrupt oncogenic protein function, rather than attempting to restore a cancerous cell to a healthy one [42]. For example, in healthy cells, the MDM2 protein binds to the tumor suppressor protein P53, marking it for degradation when it has become obsolete after its brief six-minute lifetime. This binding initiates a signaling pathway, alerting other proteins, called ubiquitin, to disassemble P53. In cancerous cells, MDM2 is over expressed, leading to reduced levels of P53. The P53 binding pocket of MDM2 is a medicinal target because it provides an opportunity to block this interaction, which is critical to the levels of these tumor suppression proteins found in cancerous cells [43] . In addition, docking programs perform virtual screenings targeting DNA with small molecule intercalates and minor groove binders designed to disrupt cancerous cell replication leading to cell death [42].

Virtual screenings begin with foreknowledge of a binding target area based on a crystallized structure already developed. The docking results of the tested ligands can then be compared to the experimentally known structure based on position and binding affinity. Location and position, along with binding free energy, are the two key factors in determining a successful docking [3]. The binding affinity is a measure of the estimated energy evaluations based on different molecular interactions. This prediction reflects the
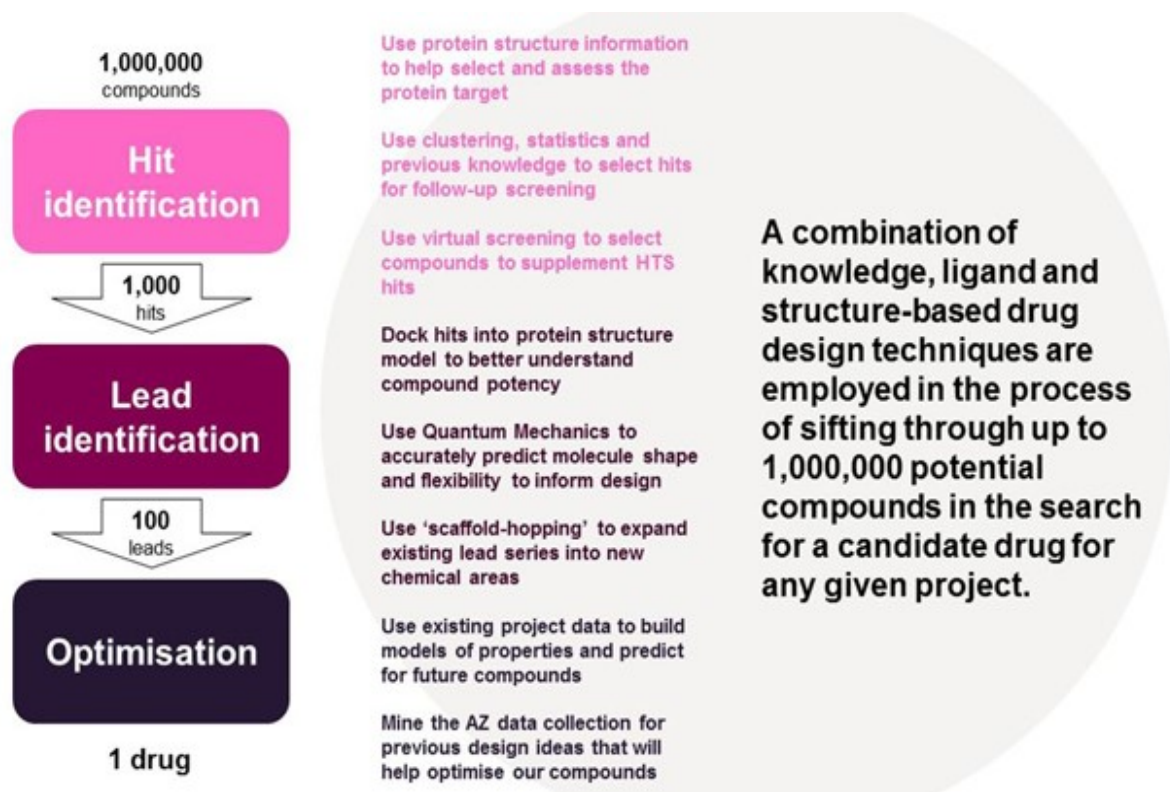
Figure 4. Use of computer-aided drug design in the search for a candidate drug is a time consuming, costly endeavor utilizing numerous methodologies and techniques. Taken from [105].

strength of the non-covalent physical binding between the two molecules [9]. Due to the reliance on estimates and assumptions inherent in the free energy calculations, the binding free energies are generally viewed as un-reliable as a true measure of the free energy. Even when a particular binding mode can be determined, the binding affinity is in doubt due to the complex interactions estimated by the semi-empirical scoring functions often employed by docking programs using simplified free energy models. The problem arises when the experimentally determined binding free energies determined from dissociation constants are quantitatively different from the estimates used for docking

experiments [44]. However, docking methodologies in SBDD are still widely used as they provide a complimentary technique for the discovery and optimization of lead compounds for future cancer drugs. The most common applications of computer methodologies in the drug discovery process are as follows:

1. Due to the large number of small molecules that need to be evaluated, virtual high throughput screenings (VHTS) can include or exclude possible candidates based on the desired binding criteria. This process saves time and money compared to traditionally screenings. The most likely candidates can then be optimized and evaluated more precisely using more expensive methods such as MD simulations [6].

2. When lead compounds are identified, further optimization can be guided by the structure based knowledge from the crystallized complex. MD simulations can be applied to measuring binding free energy and elucidating pharmacological effects, such as absorption, distribution, metabolism, excretion, and the potential for toxicity (ADMET). In addition, quantitative structure-activity relationships (QSAR) analysis is used to reveal the relationship between structure and activity and the relevant molecular parameters [6] [45].

3. Computer techniques allow researchers to build molecules from the ground up, in a process called de novo design, by piecing together functional groups or fragments to design novel compounds. Pharmacophore modeling, a technique that defines the critical features of small compounds that initiate bioactivity with a larger macromolecule, is used to develop different designs that produce the same medicinal outcome [6].

Due to the concurrent emergence of targeted medicine, structure based screening and computer technology, the role of computer aided drug design (CADD) has become an intense focus of medicinal chemistry. In particular, molecular docking programs and classical MD have been used extensively as a preliminary tool for the detection and optimization of small molecules purposed as cancer drugs [46]. Future breakthroughs in cancer treatments will almost certainly rely on improvements related to these methodologies.

THE ROLE OF P53 AND MDM2 IN CANCER THERAPEUTICS

Introduction

P53 is a tumor suppressor protein found in the nucleus of cells, which functions to respond to cellular stress by mediating cell-cycle arrest, senescence or apoptosis in response to DNA damage, oncogene activation, and hypoxia [47]. P53 is one of the most important proteins found in eukaryotic cells because the proper functioning of p53 is critical to tumor suppression. This protein is called "guardian of the genome", because it responds to cell damage by blocking cell division and inducing cell death, thereby stopping malignant cancerous cell division. Inactivation of the p53 pathway is found in the majority of human cancers. In almost 50% of cancers, the TP53 gene that encodes p53 is inactivated through mutation or deletion [48]. In much of the other 50%, functionality of p53 is lost or compromised due to cellular regulatory mechanisms that have been damaged due to cellular stress. In healthy cells, p53 levels are kept low and have a half-life of 6 to 20 minutes, but this half-life doubles as p53 levels increase to promote cell repair through its transcription activity [49]. In 7% of cancers, p53 levels are degraded due to the over expression of the regulatory protein MDM2, which functions as an E3 ubiquitin ligase responsible for the ubiquitination and degradation of p53 (figure5) [50] . This 7% represents a sample of 28 different types of human cancers from almost 4,000 human tumor samples [51]. The over expression of MDM2 is the result of gene amplification mediated by p53. In this way, p53 and MDM2 form an auto-regulatory feedback loop [52]. When this pathway is disabled due to cell damage, the DNA damaged cells may proliferate resulting in the formation of tumors [52][50]. Since
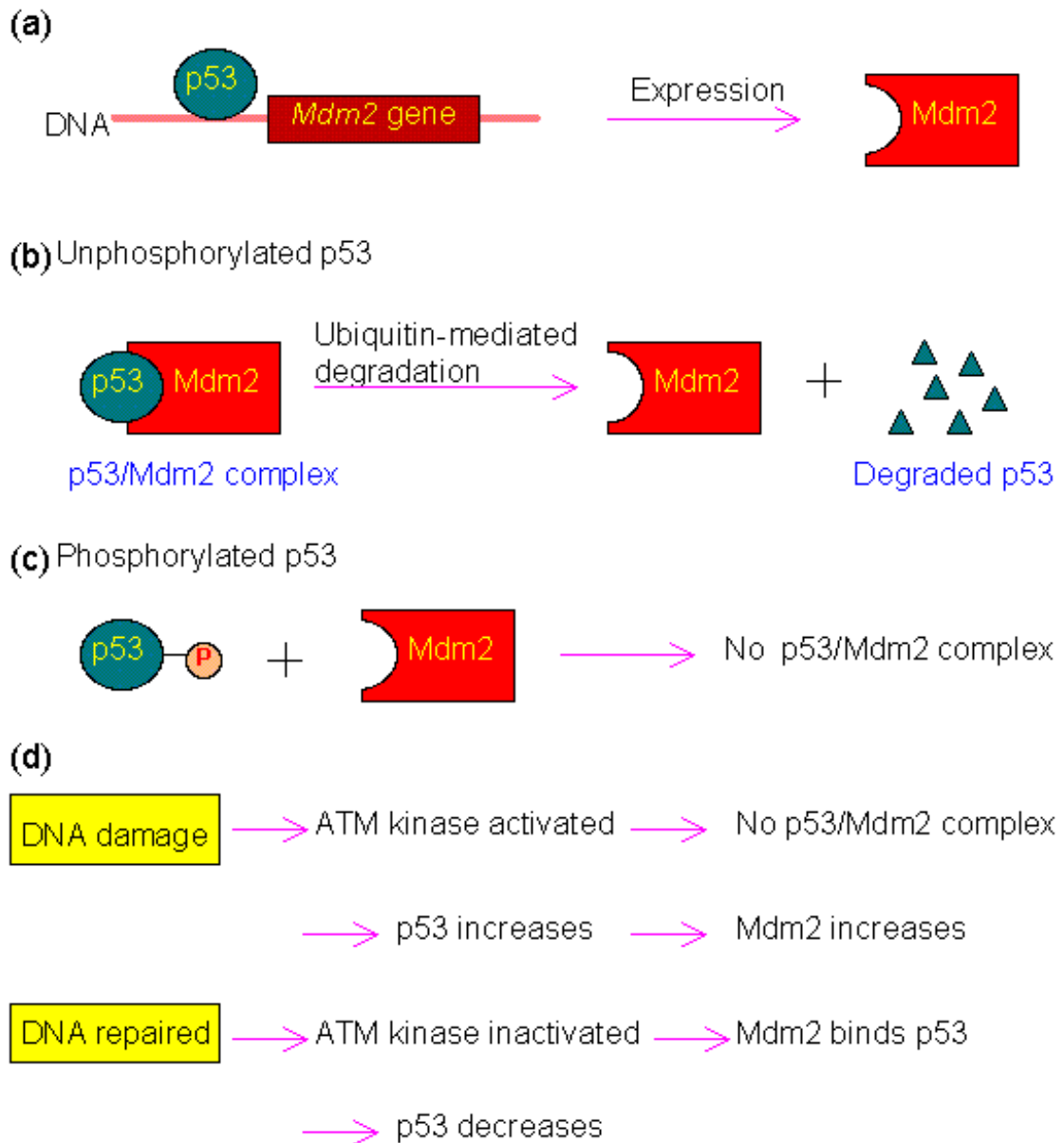
Figure 5. a) P53 activates the production of MDM2. **b**) MDM2 binding to P53 initiates the degradation of P53 via the ubiquitin system. **c**) Phosphorylation of P53 at Ser15, Thr18 or Ser20 will interrupt binding with MDM2. These three residues are not phosphorylated in normal cells keeping P53 at low levels by MDM2. **d**) DNA damage may facilitate the activation of protein kinase to phosphorylate P53 at one of the three residues causing an increase in P53 levels. Due to MDM2 expression being activated by P53, the increase of P53 also increases MDM2, but this has no effect when P53 is phosphorylated. When the DNA damage is repaired, the kinase is no longer active and P53 will quickly be dephosphorylated and have levels reduced by the accumulated MDM2. Taken from [106].

MDM2 affects other cellular processes not associated with p53, and various cellular molecules contribute to regulate the pathway, the full oncogenic role of MDM2 is not fully understood. Studies have shown increased levels of MDM2 may contribute to cancer formation independent of its role as a regulator of p53 [53]. However, there is considerable evidence that increased levels of MDM2 lead to accelerated tumor growth in many types of cancers and in particular those cancers coinciding with low levels of p53 in damaged cells, which has led the p53/MDM2 pathway to become a focus point for cancer therapeutics [54]. In particular, the targeting of the post translational nature of the p53/MDM2 interaction, which can initiate an immediate response towards lowering MDM2 levels [53].

The P53/MDM2 Binding Structure

The binding of MDM2 to p53 is well understood and the biochemical structure of this complex has been established by examining the crystallographic data resulting from the resolved crystal structure. MDM2 binds to a 15 amino acid amphipathic peptide at the NH2 terminus of p53, while the MDM2 binding domain is found in a small hydrophobic pocket located at its NH2 terminus containing residues 25-109 [55]. This binding mediates gene transactivation of p53 through a non-covalent physical binding which blocks the p53 transactional domain. A second and more salient result for our discussion is this binding also marks the protein for degradation by the proteasome, which is a protein designed to break the peptide bonds holding proteins together [56][57][58]. This the mechanism by which p53 levels are lowered and tumor growth is allowed to progress. Chemotherapies attempt to recover the tumor suppression activity of p53 by introducing

small non-peptide molecules into cancerous cells that mimic the critical residues, Phe 19, Trp23 and Leu26, associated with the p53 binding pocket of MDM2. Levels of p53 should increase given that small molecules can be designed to bind to MDM2 in these critical domains before the ubiquitination process occurs [55]. The fortuitous nature of the structure of MDM2 has been exploited in the development of such small molecules. Human MDM2 is a 491- amino acid long phosphoprotein, whose p53 binding domain is found in a well-defined hydrophobic surface pocket encompassing a relatively small area of the protein. The binding pocket measures only 18 Å along the long edge; the size of a typical small molecule [59]. The MDM2 cleft is formed by amino acids 26–108, and consists of two structurally similar portions that fold up into a deep groove lined by 14 hydrophobic and aromatic residues [55][60].

Small Molecule Inhibitors of MDM2

The physical and chemical structure of MDM2 has enabled many small molecule inhibitors, such as nutlins, spiroxindoles and isolindones, and chalcone derivatives, to be designed from lead compounds discovered via the experimental and computational screening of large chemical libraries (figure 6) [60]. Pharmacological optimization can then be carried out to produce the desired physicochemical properties, such as essential high binding affinities and high specificity over other proteins. The nutlins were the first group of optimized small molecule inhibitors discovered at Hoffmann-La Roche in 2004 and advanced to phase 1 clinical trials in 2012, and a second generation of inhibitors has been designed with greater potency and improved selectivity, which are currently in phase 1 clinical trials [61]. Results from the first trial, using an optimized nutlin
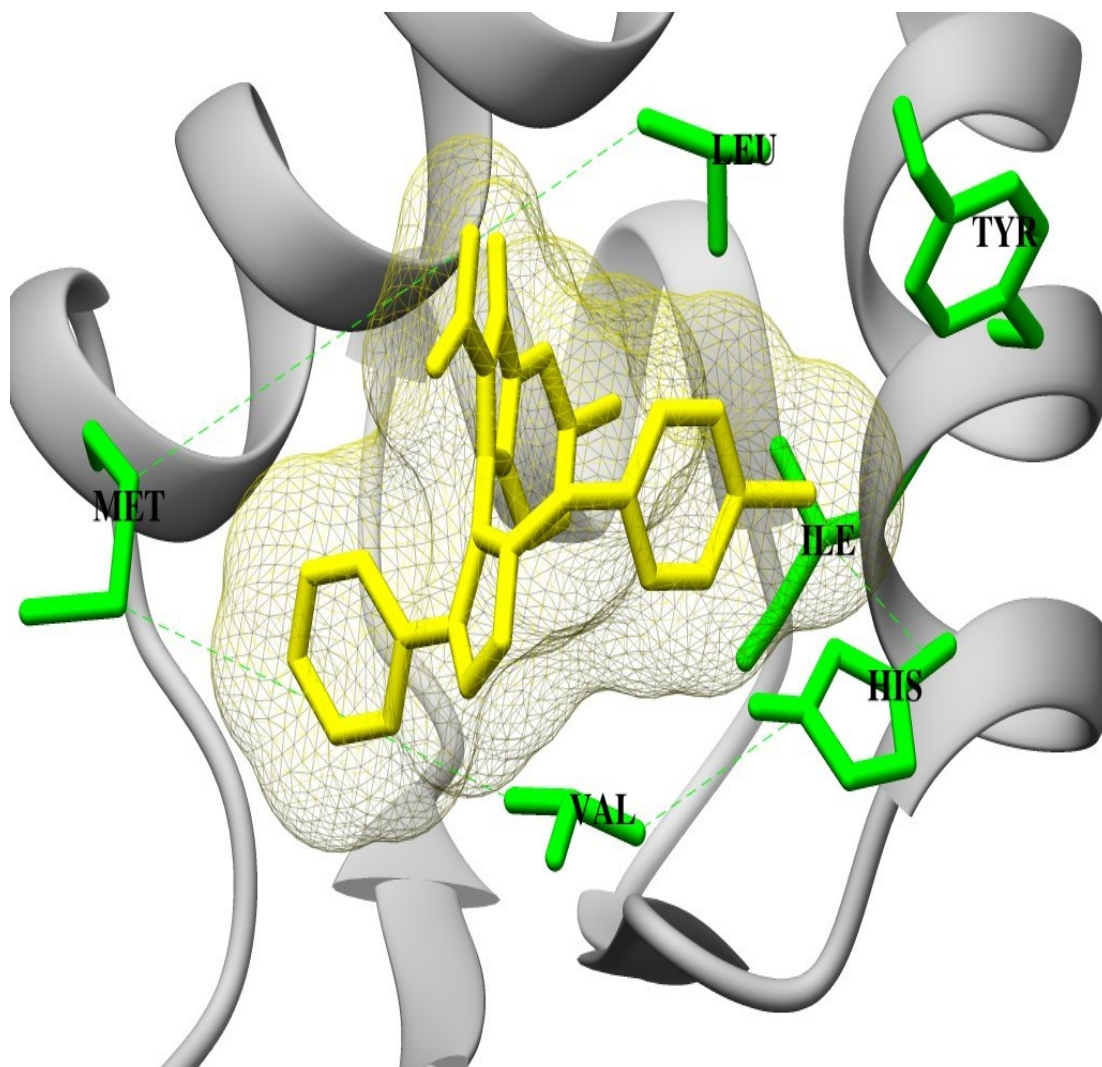
Figure 6. Small molecule inhibitors must be designed to firmly attach to the P53 binding site of MDM2 in order to block binding and subsequent degradation of tumor suppressor P53 in damaged cells. Knowledge and simulation of binding site residues shown in green, enable accurate predictions by docking programs and lead to a more efficient drug discovery pipeline. The efficient utilization and future optimization of these programs are essential for hastening drug discovery and reducing costs of therapeutic development. This illustration was produced using Chimera software during this research.

compound administered to 20 patients, have been mixed. The drug known as RG7112,

was administered to patients with lipsarcoma, a form of malignancy of fat cells. The post

clinical data showed while some tumors were kept from growing; only one was reduced and several patients showed toxicity in the form of thrombocytopenia, a condition in which your immune system attacks your own platelets. However, although most of the focus has been to develop drugs targeting the p53/MDM2 pathway, a new class of MDM2 inhibitors has been developed for cancers that have no functional p53, including pancreatic, prostate, and advanced breast cancer. These drugs bind directly to MDM2 promoting self-degradation and inhibiting transcription in a p53 independent and MDM2 dependent manner [62][63].

COMPUTATIONAL METHODOLOGIES AIDING STRUCTURE BASED

DRUG DESIGN

Molecular Docking

The computational evaluation of bio-molecular interactions is a challenging task, responsible for the development of over 50 programs dedicated to predicting the likely location and binding affinity of small molecules to macro-molecule receptors, such as DNA, RNA and proteins. The two principle components of these programs include conformational search methods and binding free energy evaluations. In practice, a balance must be struck between computational cost and accuracy, for medicinal chemistry, this trade-off becomes increasingly important, as thousands of compounds can be tested focusing on one macromolecule target [3]. Considering the receptor and ligand each possess six degrees of translational and rotational freedom as well as conformational degrees of freedom, there exists a tremendous number of possible binding modes. In an attempt to manage this complexity, molecular docking programs utilize various conformational search strategies and scoring functions (figure 7) [64].

Conformational Search Strategies. The first docking programs developed in the early 1980's used a simplified model of ligand-protein interaction called "lock and key" theory, which held both the ligand and receptor rigid. Later, the "induced fit" theory stated that the active site of the protein is continually reshaped by interactions with the ligand, and accounting for this movement would increase the accuracy of binding mode predictions [65]. However, as stated earlier, this is a daunting computational challenge and has led to the use of a flexible ligand-rigid receptor protocol, which still remains the
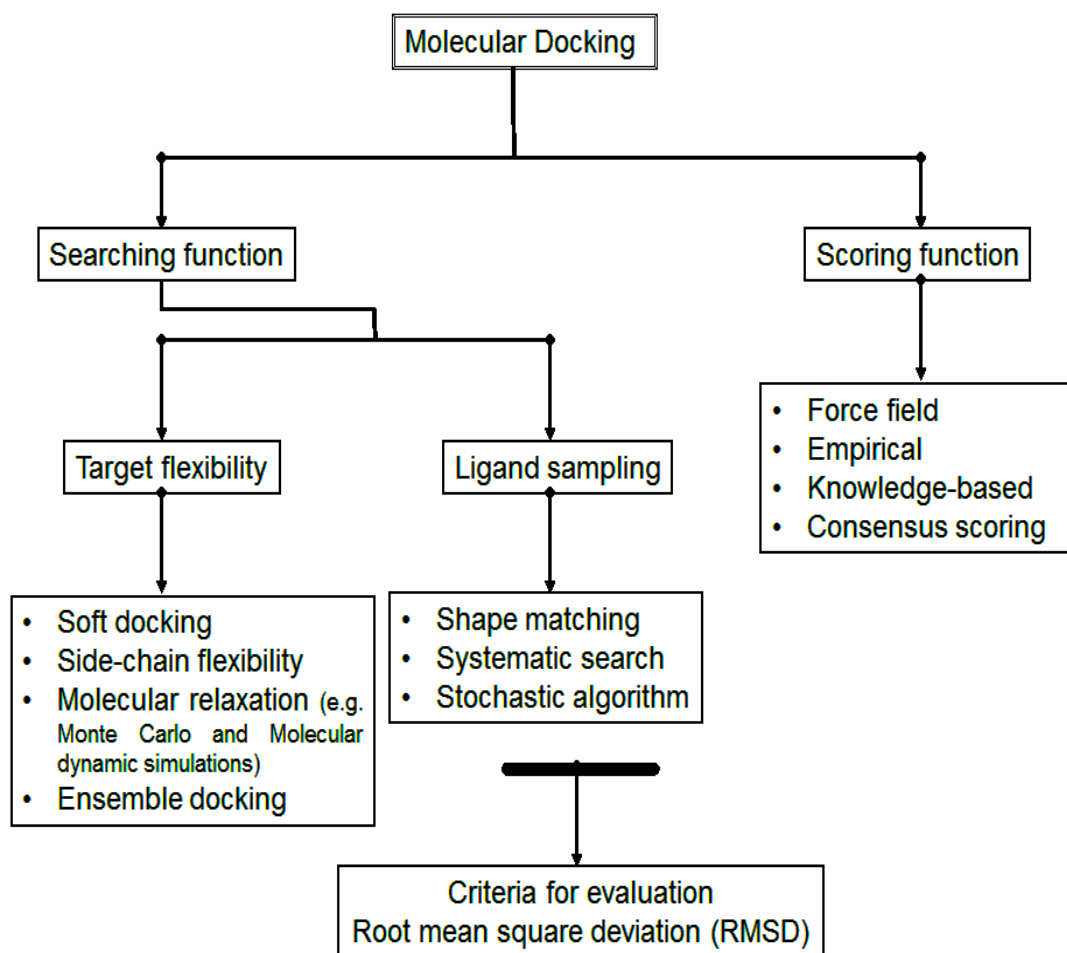
Figure 7. The two main facets of a molecular docking program are the search method and scoring function. Different methods are used for both and have been adopted by 50 different docking programs throughout the world. This chart reflects the flow of the most important aspects of these programs culminating in evaluation of the RMSD from the either known experimental site or targeted location. Molecular docking represents an early evaluation and precedes more expensive time consuming evaluations and trials. Taken from [107].

most popular method in use [20]. Using this protocol reduces the amount of possible

binding orientations thereby reducing computational time and cost. Contemporary

docking programs have adopted various sampling algorithms to survey the user defined

conformational binding space.

The two methods most commonly used can be classified as systematic and

stochastic [66]. Both apply an incremental change to the structural parameters of the ligand, including the torsional, translational and rotational degrees of freedom. The systematic approach methodically changes the structural orientation of the ligand, thereby inducing different conformational states and binding modes [67]. The binding energies are calculated for each docking during numerous evaluation cycles in search of the ligand-protein orientation producing the lowest binding free energy and therefore, the most likely binding mode [68] . This structural manipulation attempts to simulate the changes to the ligand facilitated by its interaction with the protein [69][67]. This method is conducive to a thorough exploration of the search space, but is subject to settling on a local, rather than global energy minimum. It also encounters a combinatorial explosion due to all the possible ligand orientations [66]. To counter this, an incremental algorithm is employed to evaluate fragments of the ligand separately as they are placed within the either desired, or suspected protein binding site. A fragment is selected and evaluated until the binding mode corresponding to the minimum energy is realized, than another fragment is added, and the process is repeated until the entire ligand is constructed and evaluated ending in a final binding orientation with minimum binding energy [66][9].

The Conformational space can also be surveyed through Stochastic methods by randomly modifying a ligand conformation or a population of ligands. Monte Carlo (MC) and genetic algorithms (GA) are typical algorithms representing this technique. The basic MC search has the ligand perform a random sampling of different locations around the protein. At each location, the ligand orientation is randomly altered and then evaluated in terms of acceptance criteria. If the energy is lower than the previous step, it proceeds from that orientation. If the energy is higher or the same, a probabilistic approach is used

to dictate the most advantageous adjustment. This will continue until a user defined number of conformations is gathered in search of the global minimum energy. A modification of this technique is simulated annealing MC. After a defined number of steps, the system temperature is lowered, producing conformations less likely to have increased energy, thereby moving the system toward energy minima [70] [9].

Similar to MC, the GA also adopts a randomness approach, which is based on Darwin's theory of evolution and natural selection [71][72]. First, the structural parameters of the ligand are translated or encoded into a chromosome. From this chromosome, which actually represents the pose of the ligand, the search algorithm randomly produces a population of chromosomes representing a substantial portion of the energy landscape. This population's fitness is evaluated and the most adapted chromosomes, or those with the lowest energy values, are selected as templates to parturitate the next population. Then, mutation makes random alterations to the genes, such as crossover exchanging of genes between two chromosomes, producing a new ligand pose with new structural parameters. This cyclic process reduces the average energy of the chromosome ensemble by transferring an optimal set of values describing the translational, orientation, and conformation of the ligand as it binds to the protein, thereby reducing the conformational space to be explored [73][72]. The search performed by the GA is iterative, and after a large, yet manageable number of cycles, converges on a conformation corresponding to the global energy minimum [66]. The main advantage of the GA and MC strategies is their ability to sample a large amount, rather than all of the possible binding orientations in a reasonable amount of time, while still locating the minimum binding energy. In contrast, the systematic approach aided by

incremental procedures searches all possible conformational space, making it less practical for larger ligands.

      Energy Evaluations. Although scoring functions used in molecular docking for determining the binding free energy are still considered inaccurate, docking programs provide a guide to the realization of pharmaceuticals based on the cost effective evaluation of drug candidates [16][2]. This can be accomplished without an exact determination of the binding free energy by establishing binding location and orientation before more expensive, time consuming techniques are utilized such as linear activation energy (LAE) and free-energy perturbation methods (FEP) [74]. Given current computing speeds, a more robust, accurate scoring function would be impractical due to the computational time necessitated by each small molecule being screened. Again, this is due to the combinatorial explosion of conformational states accompanied by degrees of freedom issued to the ligand or protein. In addition, properly accounting for surrounding water molecules and ions would further add to the computational costs. This would eliminate the cost and time benefit allotted by *in silico* techniques. Hence, the scoring functions adopted by docking programs provide a trade-off between time and accuracy via best estimates and assumptions of binding interactions using implicit solvent and fixed entropic contributions [15]. The most basic assumption is molecules will seek the most stable binding orientation characterized by a deep potential energy well indicated by the amount of energy released upon binding. The energy functions attempt to estimate this energy and therefore, the most likely binding mode. The scoring functions most commonly used to estimate binding affinity are classified into force-field based, empirical based and knowledge-based scoring functions (figure 8) [1][6].
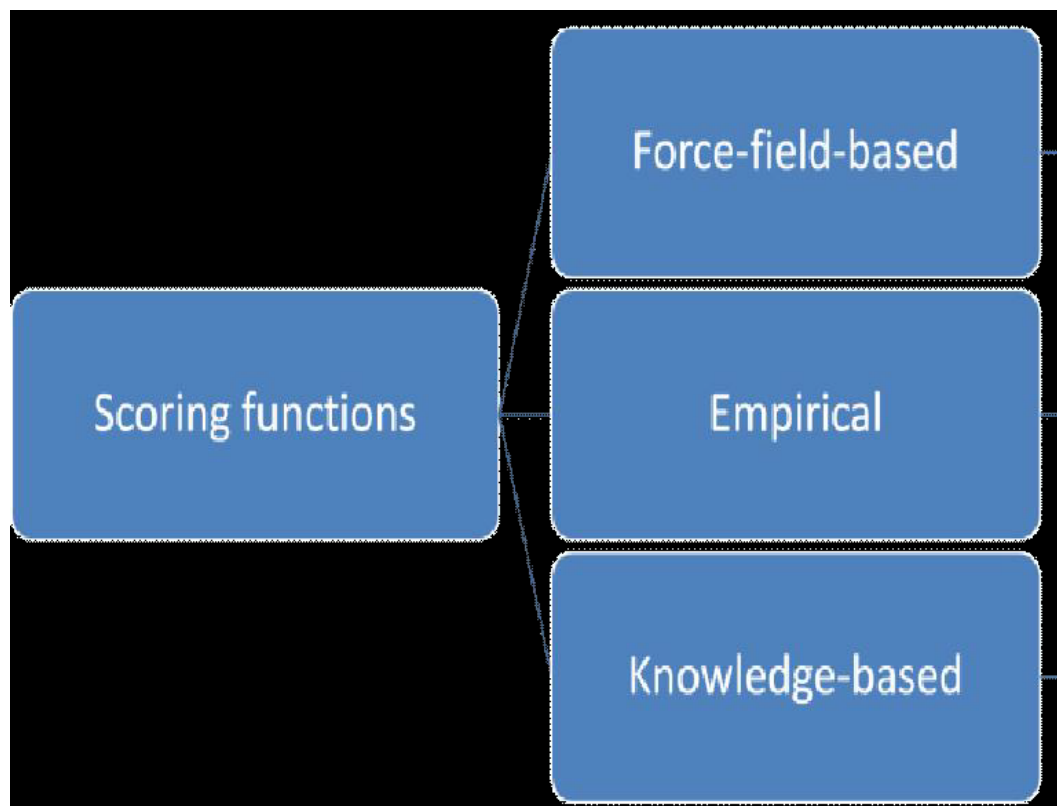
Figure 8. Scoring functions are divided into three main categories although many, such as AutoDock, may incorporate different features of each. AutoDock uses force-field based calculations and empirical quantities during energy evaluations.

Force-field based scoring methods utilize chemical potentials to calculate binding energy by summing the non-bonded, distant-dependent, atomic pairwise van der Waals and electrostatic interactions. The van der Waals terms are calculated from a Lennard-Jones potential function. This distance-dependent function can be adjusted by setting different parameters for the Lennard-Jones potential such as controlling how close a contact between protein and ligand atoms will be accepted as contributing to the total energy. Therefore, reducing this distance will lead to diminished accuracy of long-range effects contributing to binding. Additional terms are often added to these scoring

functions considering hydrogen bonding, solvation and entropy contributions as well [9][66][7]. The particular scoring function employed by a docking program varies by the terms used and the values adopted by empirical testing when installing the hydrogen, solvation, and entropy terms. Therefore, the values attributed to these terms will depend on the test set selected for calibration. A large set of diverse ligand-receptor complexes including ligand-protein, ligand-DNA, protein-protein and ligand-RNA will be more applicable, but less accurate, while only selecting for instance, ligand-protein complexes will be accurate, but less adaptable [75]. The advantage of force-field based scoring functions is that they provide the most realistic model of binding interactions with acceptable computational costs. The problem is their inaccuracy in estimating entropic contributions. This limitation is due to the lack of a reasonable physical model to describe this phenomenon. Furthermore, the solvent is not explicitly considered, hindering the estimation of desolvation energies [66].

A popular form of the force-field based potentials is currently used with the AMBER molecular dynamics software package and has been adopted for use in docking programs. The AMBER force fields are represented by:

$$E_{amber} = E_{angle} + E_{bond} + E_{dihedral} + E_{non-bonded} \qquad (1)$$

E_{angle} and E_{bond} are harmonic approximations of the bond angle and strain energies, respectively, and $E_{dihedral}$ is an energy term associated with the dihedral angles of linearly-bonded sets of four atoms. The term $E_{non-bonded}$ aggregates the non-bonded interactions: a Lennard-Jones 6-12 potential which approximates the van der Waals

attraction, and Pauli repulsion, and an electrostatic potential term. The ff94 force field,
which uses the AMBER functional form, has been very popular for simulating proteins.
Also, subsequent versions including AMBER 99SB force field and the general AMBER
force field (GAFF) contain parameters suitable for modeling the interactions of small
organic molecules associated with drug design [15].

Empirical scoring functions use a simplified calculation that allows for high-speed
computation for virtual database screening. Empirical functions provide efficient
performance for groups of proteins or compounds that are similar to complexes within
the training set, but performance decreases when applied to different protein families
[66]. The foundation of empirical scoring functions is based on structural features and is
often guided by physical interactions. However, these functions are not derived from first
principles, but are directly calibrated with a set of protein-ligand complexes with
experimentally determined structures and binding affinities based on multiple linear
regression analysis or machine learning methods in an attempt to measure the favorability
of an interaction rather than fully elucidate the underlying physics of the event. It is a
simplified approach, attempting to save computational time, while accurately predicting
binding interactions [76]. The terms in the master equation represent a particular physical
process involved in the formation of a ligand-receptor complex. The functions combine
features such as hydrophobic contacts, hydrophilic contacts, number of rotatable bonds or
number of hydrogen bonds, and parameterize these features using a diverse training set of
ligand-receptor complexes. This establishes the coefficients of the scoring function which
is multiplied by each component and summed to produce a final score [77].

Knowledge-based scoring functions use pairwise energy potentials taken from

known ligand-receptor complexes to create a general function. These potentials are developed by accounting for the frequency two distinct atoms are found within a given distance in the structural dataset based on the assumption that more favorable interactions occur with greater frequency. The frequency distributions are then converted into pairwise atom-type potentials. The different types of interactions observed in the dataset are classified and weighted according to their frequency of occurrence and the sum of these individual interactions is given by the final docking score [66][78]. The inverse-Boltzmann equation provides the relationship between the frequency of features and the energy that is assigned to those features. For protein-ligand interactions, the energy assigned to the interaction between ligand atom type i and protein atom type j, at a distance of r_k (the distance of the kth bin), can be computed as follows:

$$U_{\rho mf,ij}(r_k) = -k_B T \, \ln\left[\frac{\rho_{ij}(r_k)}{\rho_{ij,ref}}\right] \qquad (2)$$

The quantity $(\rho ij(rk))/\rho ij,ref)$ is the relative radial density for atom pair type ij within the training set and is a function of the bond distance r_k. The density $\rho_{ij,ref}$ associated with the reference state may be computed using an ideal gas approximation or other approaches. Since knowledge-based functions are not reliant on the calculation of binding affinities as with force-field methods, nor do they attempt to reproduce binding affinities, as with empirical functions, they offer a good balance between accuracy and speed. They also perform well with ligand-receptor complexes outside their structural dataset [15].

AutoDock

   Scoring Function. AutoDock 4 employs a semi-empirical free energy force field

based on a linear regression analysis and the amber force field described in the previous

section. The force field parameters were chosen from using a large number of protein-

inhibitor complexes for which both structure and inhibition constants, or K_i , are known.

Improvements installed in this newest version allows for the incorporation of protein side

chain flexibility due to the calculation of bound and unbound inter and intra-molecular

energies (figure 9) [14]. The total binding free energy is the difference between the

unbound states of the ligand and protein and the total bound energy of the complex. Most

simply, the energy of the ligand and protein separately is always greater than the bound

complex, the difference is the binding free energy, it is always negative, and determines

the depth of the energy well and the amount of energy needed to break apart the complex

[7]. These calculations are augmented by a second program called Autogrid, which

performs pre-calculations of binding affinity's and stores this information for the actual

docking run. Autogrid works with AutoDock and must be implemented before all

docking runs. The force field evaluates binding in two steps, while the protein and ligand

begin in an unbound conformation. First, the intramolecular energetics are estimated for

the transition from these unbound states to the conformation of the ligand and protein in

the bound state. Then, the free energy function evaluates intermolecular energetics of

combining the ligand and protein in their bound conformation:
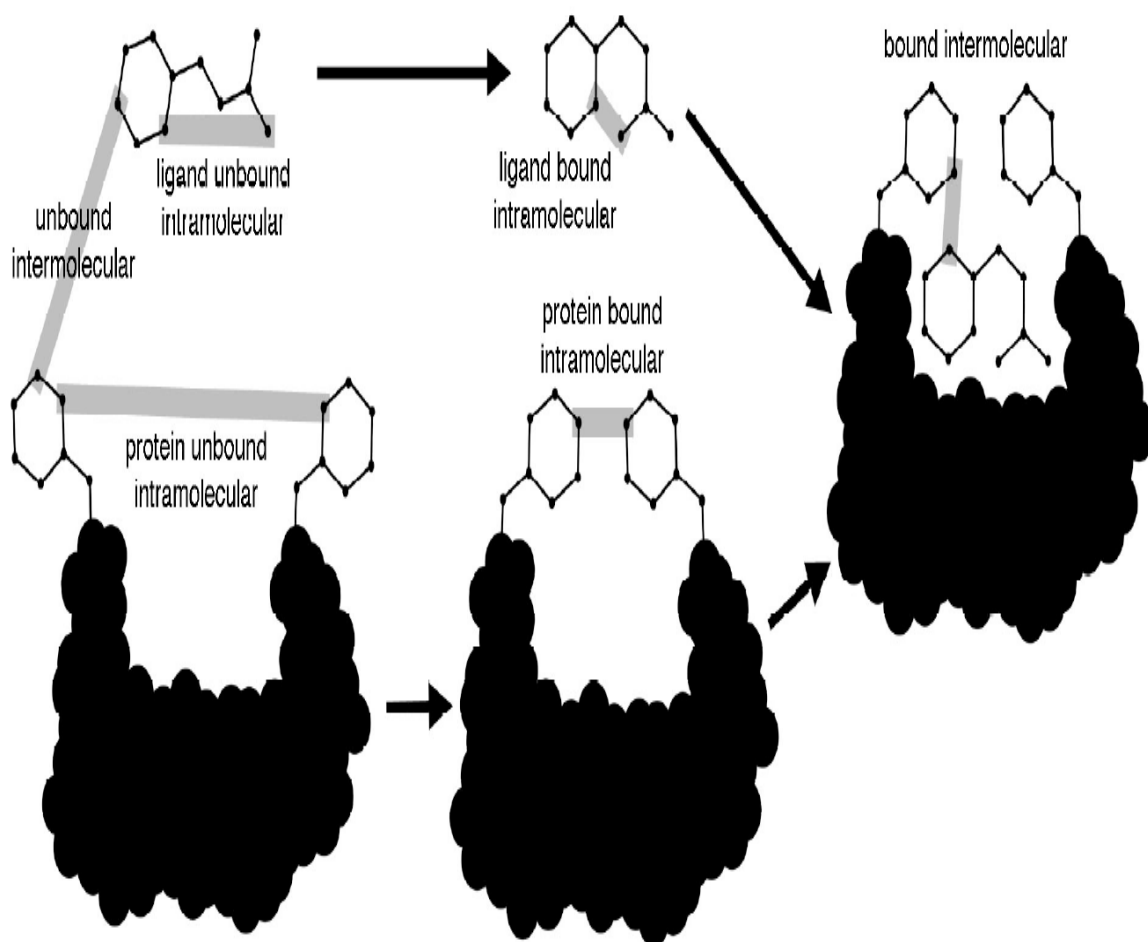
Figure 9. The force field used by AutoDock evaluates binding in two steps. The first step evaluates the energetics of the ligand and protein separately by determining the intramolecular energies, which is the energy within the ligand or protein rather than between the two, associated with the transition from the unbound state to the bound state they will adopt when they form the complex. Then, the second step evaluates the intermolecular energies, or the energy between the ligand and protein associated with the binding. The difference between the unbound and bound energies represents the binding free energy. It is the energy released during the formation of the complex in an attempt by the ligand and protein to reach the lowest energy possible. Molecules will settle in a location and geometry with the lowest possible negative binding energy to ensure a secure conformation. Docking programs use this tendency to evaluate this energy and determine likely binding geometries. Taken from [7].

$$G = (V_{bound}^{L-L} - V_{unbound}^{L-L}) + (V_{bound}^{P-P} - V_{bound}^{P-P}) + (V_{bound}^{P-L}) \qquad (3)$$

The force field includes six pair-wise evaluations (V) and includes a term for the conformational entropy lost upon binding, ΔS_{conf}. Where 'L' and 'P' refer to the ligand and protein. The terms included in the first two parentheses calculate the intramolecular energies for the bound and unbound states of the protein and ligand separately. The third parentheses estimates the intermolecular energy from the bound and unbound state. The total conformational entropy lost upon binding is calculated by multiplying a constant (W) by the total number of rotatable bonds (N). The default constant used for AutoDock 4 is .314 Kcal/mol.

$$\Delta s_{conf} = W_{conf} N_{tor} \qquad (4)$$

Each pair-wise energy evaluation includes terms for dispersion/repulsion, hydrogen bonding, electrostatics and desolvation (figure 10):

$$\Delta G = W_{vdw} \sum_{ij} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right) + W_{hbond} \sum_{ij} E(t) \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right) + W_{elec} \sum_{ij} \frac{q_i \, q_j}{\varepsilon(r_{ij})r_{ij}}$$

$$+ W_{sol} \sum_{ij} \left(S_i V_j + S_j V_i \, \exp\left(\frac{-r_{ij}^2}{2\sigma^2} \right) \right) \qquad (5)$$

All terms are preceded by a weighting constant (W) determined by a set of empirically determined binding constants. The summations are calculated for all pairs of ligand

atoms, i, and protein atoms, j, as well as all pairs of atoms comprising the ligand separated by three or more bonds. The first term is a 6/12 potential representing dispersion/repulsion interactions. The second, is a directional H-bond term based on a 10/12 potential. The parameters C and D provide a well depth of 5 kcal/mol at 1.9Å for hydrogen bonds with oxygen and nitrogen, and a well depth of 1 kcal/mol at 2.5Å for hydrogen bonds with sulfur. The function E (t) provides directionality based on the angle from ideal H-bonding geometry. The third term is a screened Coulomb potential for electrostatics. The desolvation potential is based on the volume of atoms (V) that surround a given atom weighted by a solvation parameter (S) and an exponential term with distance-weighting factor σ=3.5Å. AutoDock4 introduced an improved desolvation model that includes terms for more atom types. The atomic solvation parameter (S_i) is a function of the atomic charge (q_i):

$$S_i = (ASP_i + QASP \times |q_i|) \tag{6}$$

Where ASP_i is the atomic solvation parameter of atom i and QASP is the charged-based atomic solvation parameter (QASP = 0.01097, stderror = 0.0063) [7][79].

Autogrid. AutoGrid is a program responsible for the pre-calculation of grid maps reflecting the interaction energies for various atom types, such as aliphatic carbons, aromatic carbons, hydrogen bonding oxygens, and so on, with a macromolecule such as a protein, DNA or RNA. Autogrid also produces an electrostatics potential map and a desolvation map [79]. A probe atom from each atom type of the ligand is placed at different points on a grid designated by the user as for the size and spacing. The energy of

AutoDock 4.1 Forcefield

Interatomic Distance (Angstroms)
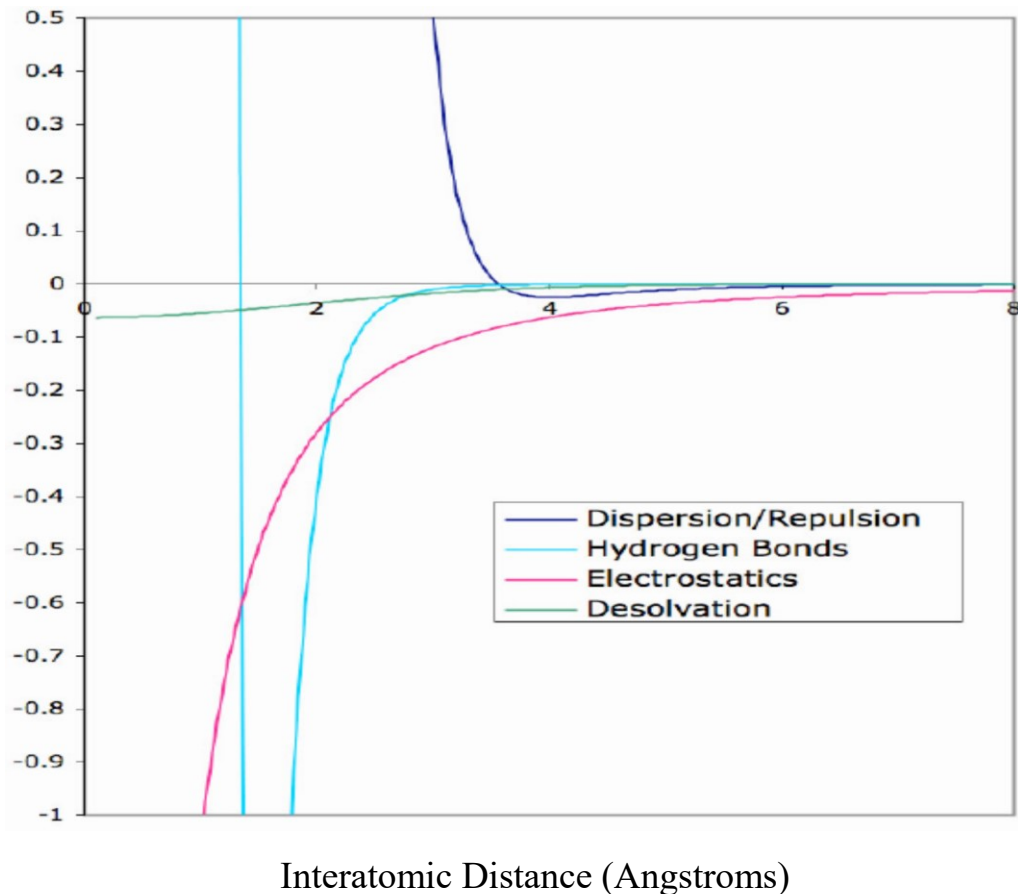
Figure 10. Interatomic forces used for the AutoDock force field. These terms are used together with weighting constants which are optimized based on experimentally characterized complexes. This reduces computation time, but also reduces accuracy of the binding free energy calculation. Taken from [7].

interaction of this single atom with the protein is assigned to the grid point. The default

spacing for AutoDock is .375 Å, which is about a quarter of the length of a carbon-carbon

bond [79]. The maximum grid box size is 126 points in each direction. The distance

between each point correlates to one grid spacing which can be increased up to 1 Å when

creating the grid box. These grid maps are then used by AutoDock to determine the total

interaction energy for a ligand when docked with a macromolecule. This pre-calculation

of the atomic affinity potentials for each atom type performed by Autogrid saves time during the docking because the non-bonded interactions are stored and used during the calculation. Also, what was a calculation with order N-squared complexity, is reduced to one that is order N, where N is the number of atoms interacting [13].

      Protein Side-Chain Flexibility. AutoDock 4 allows for flexibility of protein residue rotamers to reflect, at least to a small extent, conformational changes of the protein when binding with a ligand (figure 11). Accounting for protein conformational changes has long been a goal of molecular docking programs for accurate binding predictions [80][81]. There have long been reported examples of differences in the proteins bound and un-bound state ranging from local side-chain rearrangements, to large-scale conformational changes, such as domain motions, hinge bending motions, or un-folding of entire segments [82]. Although different strategies have been designed to reflect protein flexibility including soft docking, selective docking, ensemble docking, four dimensional docking, on-the-fly docking, hybrid docking, and composite structure docking, AutoDock side-chain flexibility is treated explicitly to simulate the induced fit mechanism of the side-chain residues as the ligand enters the binding site. This technique is suitable for side-chain residues and is not applicable for reflecting substantial protein changes associated with protein domain movements [83]. Due to the exponential increase in search space for each rotatable bond granted to the ligand or protein, only a small number of bonds can be selected in addition to the ligands rotatable bonds before the added search space prevents convergence on a local minima. Those residues most related with binding interactions and that are not embedded within the protein are ideal for flexibility (figure 12). Binding site residue information is usually available from the

crystallographic data associated with the complex. Visualization software, such as Chimera, visual molecular dynamics (VMD) or AutoDock tools (ADT), just to mention a few, can provide further insight into the binding site interface.

Conformational Search Method. AutoDock provides a choice of search methods including simulated annealing and a genetic algorithm without a local search, but for most applications, the Lamarckian Genetic Algorithm (LGA) produces the best results. An initial population of conformations is established and then mutated during the course of successive generations. Future generations inherent conformational parameters, and those "individuals" possessing local minima binding free energy, pass these "traits" to the next generation, mimicking evolutionary natural selection. The LGA search method involves several steps. Initially, all the variables and evaluations are defined. The binding geometry of a ligand and a protein is defined by values representing the translation, orientation, and conformation of the ligand with respect to the protein. These values are the ligands state variables and each corresponds to a gene. The ligand's state corresponds to the genotype and its atomic coordinates correspond to the phenotype. A genotype is described by one chromosome. The chromosome is composed of a string of real valued genes: three Cartesian coordinates for the ligand translation; four variables defining a quaternion specifying the ligand orientation; and one real–value for each ligand torsion. In evolutionary terms, the fitness of an individual determines which traits are advanced to the next generation. The fitness variable in the LGA is the total interaction energy of the ligand with the protein, and is evaluated using the free energy function. Random individuals are mated using a process of crossover, where each new individual inherits genes from either parent. Also, some offspring mutate randomly when
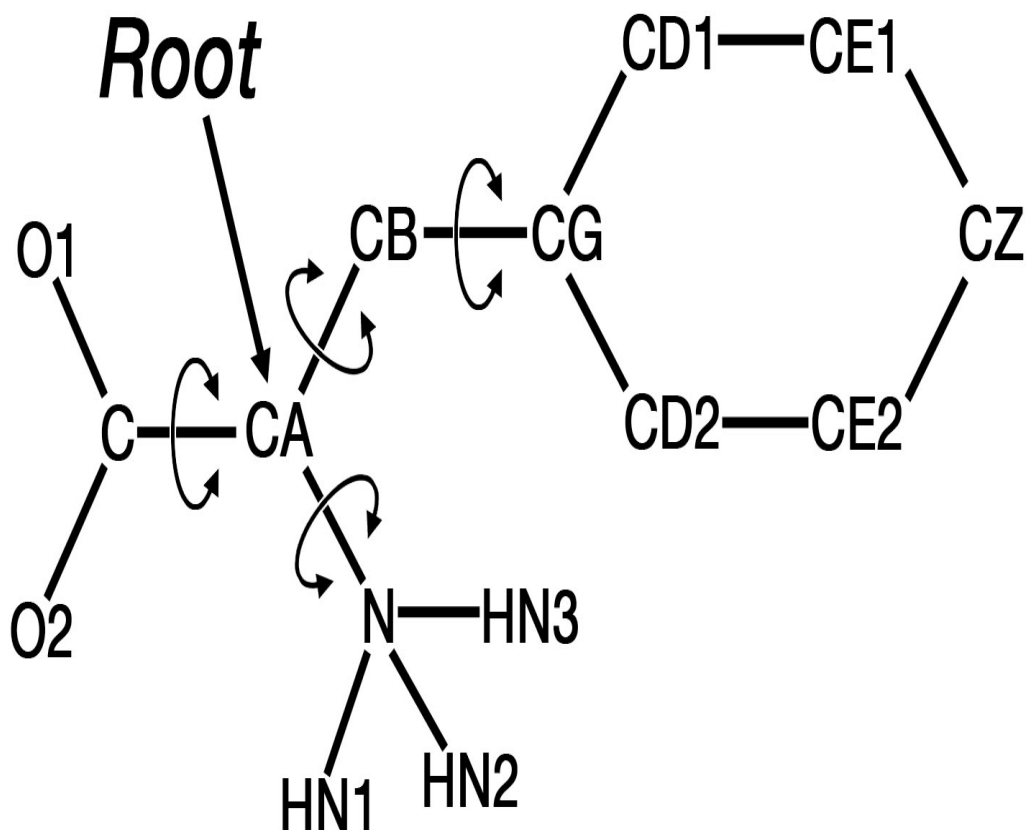
Figure 11. AutoDock uses the same method to model flexibility in both the ligand and side-chain residue. A root atom for the ligand can either be chosen by the user or the AutoDock tools module when preparing the ligand for docking. The atom linking the amino acid to the protein, which will remain in a fixed position during the simulation, is included as the root for protein flexibility and is selected by AutoDock tools automatically when preparing the flexible residue file. Flexibility in the case of the ligand and protein only refers to rotation, the bond angles and bond lengths remain fixed. Taken from [79].

one gene changes by a random amount [71].

After each generation, a local search is performed using a defined portion of the population with an algorithm developed by Solis and Wets, which surveys the genotypic space rather than the phenotypic space. This explicit local search allows

```
BEGIN_RES ARG A 364
REMARK  4 active torsions:
REMARK  status: ('A' for Active; 'I' for Inactive)
REMARK    1  A    between atoms: CA   and  CB
REMARK    2  A    between atoms: CB   and  CG
REMARK    3  A    between atoms: CG   and  CD
REMARK    4  A    between atoms: CD   and  NE
REMARK      I    between atoms: NE   and  CZ
REMARK      I    between atoms: CZ   and  NH1
REMARK      I    between atoms: CZ   and  NH2
ROOT
ATOM     1 CA  ARG A 364     15.543 -3.599 11.058 1.00 20.66    0.176 C
ENDROOT
BRANCH  1  2
ATOM     2 CB  ARG A 364     14.218 -2.825 10.902 1.00 23.02     0.036 C
BRANCH  2  3
ATOM     3 CG  ARG A 364     13.215 -3.537  9.968 1.00 25.74     0.023 C
BRANCH  3  4
ATOM     4 CD  ARG A 364     11.968 -2.720  9.626 1.00 29.66     0.138 C
BRANCH  4  5
ATOM     5 NE  ARG A 364     10.929 -2.819 10.645 1.00 32.22    -0.227 N
ATOM     6 HE  ARG A 364     11.111 -2.383 11.537 1.00  0.00     0.177 HD
ATOM     7 CZ  ARG A 364      9.757 -3.440 10.493 1.00 31.86     0.665 C
ATOM     8 NH1 ARG A 364      9.435 -4.047  9.353 1.00 31.44    -0.235 N
ATOM     9 NH2 ARG A 364      8.886 -3.427 11.490 1.00 29.11    -0.235 N
ATOM    10 1HH1 ARG A 364     8.538 -4.501  9.259 1.00  0.00     0.174 HD
ATOM    11 2HH1 ARG A 364    10.088 -4.054  8.582 1.00  0.00     0.174 HD
ATOM    12 2HH2 ARG A 364     9.116 -2.960 12.356 1.00  0.00     0.174 HD
ATOM    13 1HH2 ARG A 364      7.991 -3.884 11.385 1.00  0.00     0.174 HD
ENDBRANCH  4  5
END_RES ARG A 364
```

Figure 12. Example of AutoDock's flexible residue file describing parameters for flexible protein residues. This file is separate from both the ligand and rigid portion of the protein so that each residue can be modeled explicitly. The user decides the bonds made active (A). All other bonds within the selected residue will remain rigid (I). The maximum number of rotatable bonds is 32, but results begin to decline after 10 depending on the structure. Vina has reported success with up to 20 rotatable bonds in terms of RMSD values. The Root Atom labeled 'CA' is a carbon atom connecting the amino acid to the residue ARG364.

for the inclusion of small, refined changes without the need for switching from the genotypic space to the phenotypic space as is the case for most algorithms. This local search method enables individual conformations to search their local conformational space, finding local minima, and then pass these parameters to later generations. The individual's fitness determined by the binding free energy will determine survival to the next generation [71]. As with all computational methodologies, a balance must be struck between an exhaustive calculation and computational cost. Although a rigorous searching algorithm would go through all possible binding modes between the two molecules, this search would be impractical due to the size of the search space and amount of time it might take to complete. As a consequence, only a small amount of the total conformational space can be sampled. Theoretically, given enough computational power and time, every possible rotational and translational orientation of the ligand relative to the protein could be explored.

AutoDock Vina

     Scoring Function. The developers of Vina adopted a machine learning approach instead of using a traditional physics-based scoring function. The general functional form of the conformation-dependent part of the scoring function Vina is designed to work with is:

$$c = \sum_{i<j} f t_i t_j \left(r_{ij} \right) \tag{7}$$

where the summation is over all of the pairs of atoms that can move relative to each other, normally excluding atoms separated by 3 consecutive covalent bonds. Here, each

atom i is assigned a type t_j and a symmetric set of interaction functions $ft_i ft_j$ of the interatomic distance r_{ij} should be defined. This value can be seen as a sum of intermolecular and intramolecular contributions:

$$c = c_{inter} + c_{intra} \tag{8}$$

The optimization algorithm is tasked with finding and ranking the global minimum of c and additional low-scoring conformations. The predicted binding free energy is calculated from the intermolecular portion of the lowest-scoring conformation, designated as 1:

$$s_1 = g(c_1 - c_{intra1}) = g(c_{inter1}) \tag{9}$$

where the function g can be an arbitrary strictly increasing smooth possibly nonlinear function. In the output, other low-scoring conformations are also formally given s values, but, to preserve the ranking, using c_{intra} of the best binding mode:

$$s_i = g(c_{intra1}) \tag{10}$$

For modularity reasons, much of the program does not rely on any particular functional form of $ft_i t_j$ interactions or g. Basically, these functions are passed as a parameter for the rest of the code. Also, this program is designed so that alternative atom typing schemes could be used, such as the AutoDock 4 atom typing. The AutoDock Vina scoring

function is a combination of knowledge-based potentials and empirical scoring functions, which uses information from both the conformational preferences of the receptor-ligand complexes and known affinity measurements. The interaction functions $ft_i t_j$ are defined relative to the surface distance,

$$d_{ij} = r_{ij} - Rt_i - Rt_j \qquad (11)$$

Where R_t is the van der Waals radius of atom type t and $ht_i t_j$ is a weighted sum of

$$f_{t_i t_j}(r_{ij}) \equiv h_{t_i t_j}(d_{ij}) \qquad (12)$$

steric interactions identical for all atom pairs, hydrophobic interaction between hydrophobic atoms, and, where applicable, hydrogen bonding. The hydrophobic term equals 1, when $d < 0.5$Å; and 0, when $d > 1.5$Å, and is linearly interpolated between these values. The hydrogen bonding term equals 1, when $d < -0.7$Å; and 0, when $d > 0$, and is similarly linearly interpolated in between. In this implementation, all interaction functions $ft_i t_j$ are cut off at $r_{ij} = 8$Å. The conformation-independent function g was chosen to be:

$$g(c_{inter}) = \frac{c_{inter}}{1 + w N_{rot}} \qquad (13)$$

where N_{rot} is the number of active rotatable bonds between heavy atoms in the ligand, and w is the associated weight. The speed and accuracy of Vina was tested using the

same 190 receptor-ligand training set sampled by AutoDock 4. Results showed a two orders of magnitude improvement in speed and significant improvement of binding mode predictions. Vina can achieve near-ideal speed-up by utilizing multiple CPU cores through multithreading (figures 13 and 14). Most computer code is serialized, resulting in tasks performed by the program to be carried out in a sequential manner, one after another. The multithreaded architecture divides tasks and runs them concurrently, leading to a substantial decrease in computation time [21].

Conformational Search Method. The search method used by Vina is an iterated local search global optimizer. In this algorithm, a succession of steps consisting of a mutation and a local optimization are taken, with each step being accepted according to the Metropolis criterion. Local optimization is carried out using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method, which is a quasi-Newton method. BFGS uses the value of the scoring function, in addition to its gradient, by taking the derivatives of the value of the scoring function with respect to its arguments. These arguments are the position and orientation of the ligand, in addition to the values of the torsions for the active rotatable bonds in the ligand and flexible residues, if any. These derivatives would have a simple mechanical interpretation, if the scoring function were an energy. The derivatives with respect to the position, orientation and torsions would be the negative total force acting on the ligand, the negative total torque and the negative torque projections, respectively, where the projections refer to the torque applied to the branch "moved" by the torsion, projected on its rotation axis. While it may take longer to evaluate the gradient in addition to the value of the scoring function, using the gradient can speed up the optimization significantly. Several runs starting from random
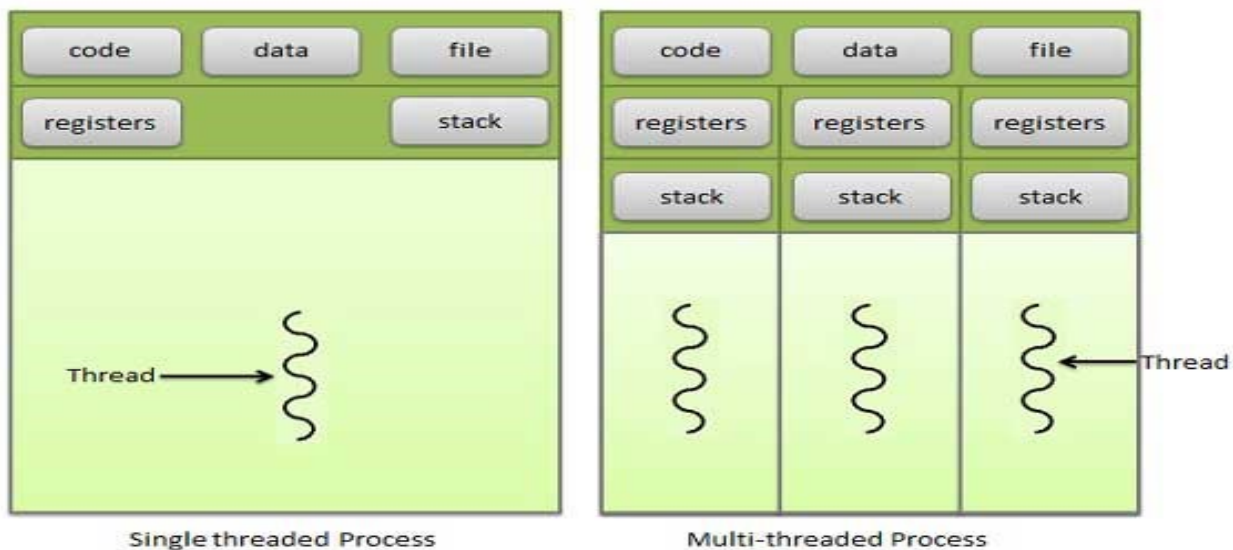
Figure 13. Contrast between single and multi-threaded computer architecture. Taken from [108].

minutes
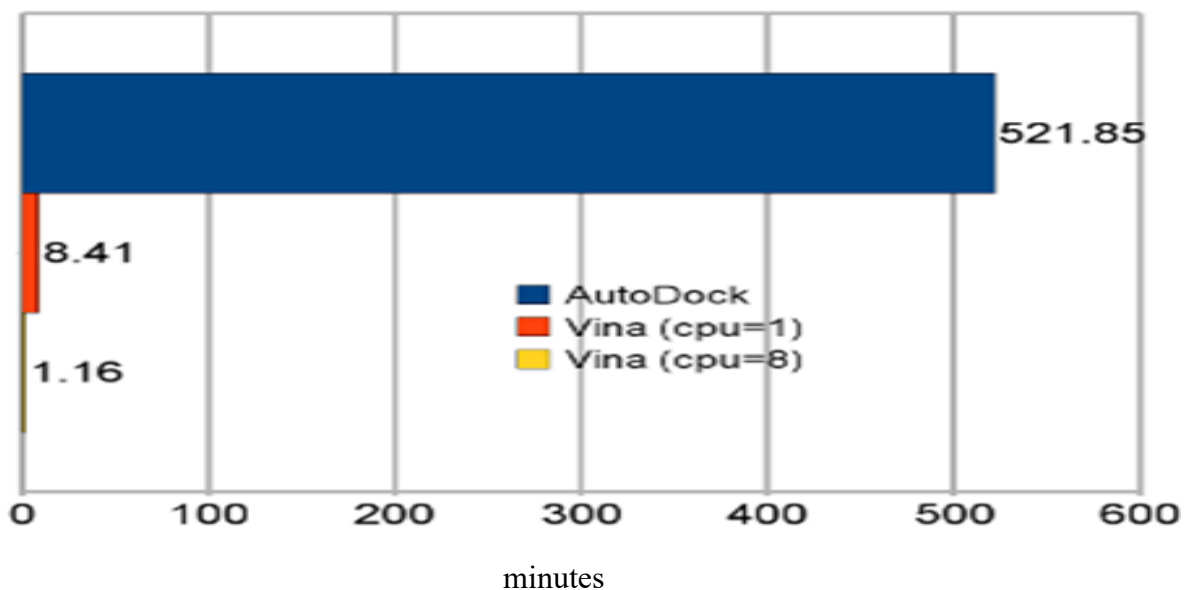
Figure 14. Vina has achieved a drastic reduction in computation time along with improved accuracy. Taken from [21].

conformations are performed. These runs can be performed concurrently using the

multithreading computer strategy built into Vina resulting in substantially reduced

computational time. The optimization algorithm maintains a set of diverse significant minima found that are then combined from the separate runs and used during the structure refinement and clustering stage [21].

Molecular Dynamics

In addition to molecular docking, MD simulations have become an essential computational tool used in the drug design process. In contrast to docking experiments, MD simulations are time dependent. This allows for understanding the dynamic behavior of proteins at different timescales, from fast internal motions to slow conformational changes or even protein folding processes. It is also possible to study the effect of explicit solvent molecules on protein structure and stability to obtain time-averaged properties of the biomolecular system, such as density, conductivity, and dipolar moment, as well as different thermodynamic parameters, including interactions energies and entropies [46]. Although MD is often used as an umbrella term, there are two related forms of this technique. In the classical mechanics approach to MD simulations, molecules are treated as classical objects, resembling very much the 'ball and stick' model. Atoms correspond to soft balls and elastic sticks correspond to bonds. The dynamics of the system using this technique is governed by classical mechanics and Newtonian physics. The quantum or 'first-principles' MD simulations, take explicitly into account the quantum nature of the chemical bond. The electron density function for the valence electrons that determine bonding in the system is computed using quantum equations, whereas the dynamics of ions (nuclei with their inner electrons) is followed classically. This strategy provides a detailed, quantum mechanical insight toward a number of biological systems, but is

impractical for systems with thousands of atoms with timescales of processes as long as a nano-second [84]. The precise calculations of quantum methods are applicable to smaller more defined systems.

MD simulations use the iterative numerical calculation of instantaneous forces present in a system and the subsequent movements in that system (figure 15). The system consists of a set of particles that move in response to their interactions according to the equations of motion defined in classical Newtonian mechanics. As with molecular docking, classical MD is much more efficient than might be expected from full consideration of the physics of biomolecular systems due to the number of substantial approximations. Most notably, each atom is treated as a point mass. This approximation is justified in terms of the Born-Oppenheimer approximation, which states only the nuclear displacements need to be considered while ignoring the electronic configuration [8] . With this assumption, Newton's second law can be used and the calculation becomes deterministic in nature. Knowing the starting positions and velocities of the atoms allows us to apply Newton's second law and determine the future or past state of the system. The mathematical description of this method is as follows:

$$F_i = m_i a_i \qquad (14)$$

where F_i is the force exerted on particle i, m_i is the mass of particle i, and a_i is the acceleration of particle i. The net force, F_i, exerted on the atom i by the remainder of the system is given by the negative gradient of the potential-energy function with respect to the position of atom i,

$$F_i = -\frac{dV}{dr_i} \tag{15}$$
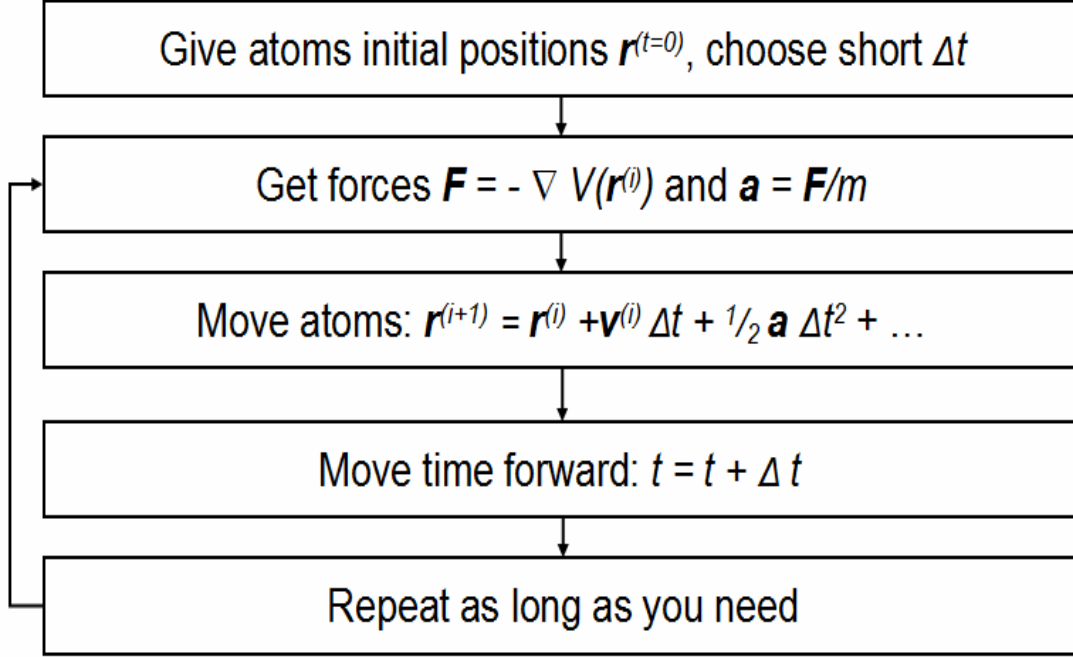
Figure 15. MD calculations are iterative. Form initial positions and velocities future and past positions can be calculated. Taken from Wikimedia-Knordlun.

The Newtonian equation of motion for atom i is,

$$\frac{dp_i}{dt} = F_i \tag{16}$$

Given the position with respect to a single component of vector r_i, at a specific time, t, then the position after a short and finite interval, denoted Δt, is given by a standard Taylor series:

$$x(t + \Delta t) = x(t) + \frac{dx(t)}{dt}\Delta t + \frac{d^2x(t)}{dt^2}\frac{\Delta t^2}{2} + \ldots \tag{17}$$

45

The position $x(t)$, the velocity $dx(t)/dt$, and the acceleration $d^2x(t)/dt^2$ are sufficient for numerical solution to the equations of motion if some approximation to account for higher order terms in the Taylor series can be made. For this single dimension, Newton's second law describes the acceleration,

$$\frac{d^2x(t)}{dt^2} = \frac{F_x}{m} \tag{18}$$

where F_x is the component of the net force acting on the atom parallel to the direction of x. Determination of future acceleration, velocity and positions are calculated by integration of the equations of motion using integration algorithms. The Verlet algorithm uses the atomic positions and accelerations at time t and the positions from the prior step, $x(t - \Delta t)$, to determine the new positions at $t + \Delta t$.

$$x(t + \Delta t) = 2x(t) - x(t - \Delta t) + \frac{d^2x(t)}{dt^2}\Delta t^2 \tag{19}$$

A slight modification of this, known as the leapfrog algorithm, is popular. The leapfrogalgorithm uses the positions at time t and the velocities at time $t - (\Delta t/2)$ for the update of both positions and velocities via the calculated forces, $F(t)$, acting on the atoms at time t,

$$x(t + \Delta) = x(t) + \frac{dx(t)}{dt}(t + \frac{\Delta t}{2})\Delta t \tag{20}$$

$$\frac{dx(t)}{dt}\left(t + \frac{\Delta t}{2}\right) = \frac{dx(t)}{dt}\left(t - \frac{\Delta t}{2}\right) + \frac{d^2x(t)}{dt^2}\Delta t \tag{21}$$

These iterative calculations are performed every one to two femto-seconds for a set period determined usually by computational time and expense. Other parameters are determining factors during MD simulations, including temperature control, constraints, periodic boundary conditions, using implicit or explicit solvent, and pressure control [8].

Trajectory Analysis. A basic MD simulation is a minimization and equilibration run designed to collect data of a system in equilibrium. Information on thermodynamic variables is provided by the NAMD log file including temperature, bond energy, total energy, pressure, volume and kinetic energy. These variables give insight into the thermodynamic state of the entire system, but we also want to know if the system is conformationally stable. For this, we can calculate the RMSD of the protein backbone as a measure of the stability of the protein. If the RMSD is still increasing at the end of the run, the protein is still searching for a lower energy state, and indicates the system is not equilibrated. The RMSD is a numerical difference between two structures. It is defined as:

$$RMSD(t) = \sqrt{\frac{\sum_{i=0}^{N} |\vec{r}_i(t) - r_i^{-0}|^2}{N}} \qquad (22)$$

Where N is the set of atoms at time t, with respect to the reference conformation. $|\vec{r}_i(t) - r_i^{-0}|^2$ is the displacement of the ith atom at time t from the reference position r_i^0. After the run, the RMSD values are plotted over time and although the protein and RMSD values will always fluctuate, we expect the data line to flatten out before the end of the simulation [85][86].

MDM2 CASE STUDY: COMPUTATIONAL PROTOCOL UTILIZING PROTEIN FLEXIBILITY IMPROVES LIGAND BINDING MODE PREDICTIONS

Introduction

P53 is a tumor suppressor protein found in the nucleus of cells, which functions to respond to cellular stress by mediating cell-cycle arrest, senescence, or apoptosis in response to DNA damage, oncogene activation, and hypoxia [48][87]. Inactivation of the P53 pathway is found in the majority of human cancers and is facilitated by mutation or deletion of the TP53 gene or damage to cellular regulatory mechanisms [88][89]. The primary regulator of P53 is MDM2, an E3 ubiquitan ligase protein, which binds to P53 marking it for degradation. In damaged cells, over-expression of MDM2 results in reduced levels of P53, initiating the onset of oncogenesis [62][90]. Chemotherapies attempt to block this interaction and recover the tumor suppression activity of P53 by introducing small non-peptide molecules designed to target and bind to the P53 binding domain of MDM2 [59]. Several small molecule inhibitors have been designed from lead compounds discovered via the structure-based virtual screening of chemical libraries performed by docking programs and many have entered and completed Phase 1 cancer drug clinical trials [47][60]. HTS evaluate thousands of small molecules and are a cost effective approach designed to rely on fast, accurate predictions, intended to isolate a small number of molecules as possible cancer therapeutics [15].

Molecular docking programs represent a critical tool in the early stages of structure-based drug design while providing important insights into molecular binding processes [91]. The focus on MDM2 as a target for cancer therapeutics and subsequent

literature underscores the importance of docking programs such as AutoDock and the more efficient, reliable alternative, Vina, for the quick and accurate screening of cancer drug candidates [92]. Many scientific articles have cited these programs and the programs themselves have been the focus of validation studies highlighting efficiencies and certain limitations. AutoDock has been shown to be reliable within 2 angstroms of the experimentally known binding site for smaller ligands targeting DNA and proteins, but struggles with larger ligands having 10 or more rotatable bonds due to the increased conformational search space [69][16][93]. This restriction has led to the use of a rigid receptor/flexible ligand protocol as standard methodology due to the computational challenges posed by incorporating protein flexibility [81][80]. This methodology fails to account for side-chain residue movement at binding site interfaces, resulting in a less reliable prediction of the ligands' docked binding mode [94][9]. Invoking AutoDock's side-chain flexibility feature accounts for some protein movement, but this adds to the conformational search space associated with a flexible ligand, which can reduce accuracy and increase computational costs [95].

Many parameters can be adjusted by the user for any particular experiment using AutoDock. This analysis focuses on improving ligand binding mode predictions through selective flexibility of the ligand and receptor utilizing AutoDock's feature of protein side-chain flexibility, while employing the speed and accuracy of Vina to reduce computational time. However, it's important to realize, AutoDock and Vina are two very different programs as they use a different scoring function and search method [21]. The results must be compared separately and it is for the user to determine the program best suited for their needs. Using Vina will determine if accuracy of the results is shared by

both, therefore, the user can take advantage of Vina's computational speed and accuracy or AutoDock's flexibility and large data output if desired.

The protocol followed for this study transfers flexibility from the ligand to the protein binding site residues resulting in improved ligand binding mode predictions of small medicinal compounds to the P53 binding site of MDM2, based on RMSD from the experimentally determined binding site (figure 16). Lowest RMSD values coincide with a majority of the rotatable bond flexibility transferred to the protein. This determination resulted from data collected from 90 docking runs using AutoDock and Vina. Further, a MD simulation revealed the minimized and equilibrated MDM2 protein shows considerable movement of binding site residues made flexible during docking calculations. The docking and MD results from this study highlights the importance of modeling protein flexibility for the determination of accurate binding mode predictions of small molecules to MDM2 in contrast to traditionally rigid docking approaches. This protocol could be especially useful for HTS of potential cancer drugs and other therapeutics targeting proteins and DNA.

An additional brief docking study of a cancer drug target suggests this protocol may be applicable to other structures. The study evaluated a cancer drug complexed with a protein enzyme and DNA (figure 17). Camptothecin is a substance, poison to human cells, it is produced by a tree native to China as a natural defense mechanism. In humans, it binds to topoisomerase, a critical enzyme responsible for binding to DNA in order to relax the helical structure during replication [96]. Camptothecin binds to the covalent topoisomerase -DNA complex and stabilizes the helical structure leading to DNA strand breaks that lead to the apoptosis of drug-treated cells. This drug is used to treat colorectal
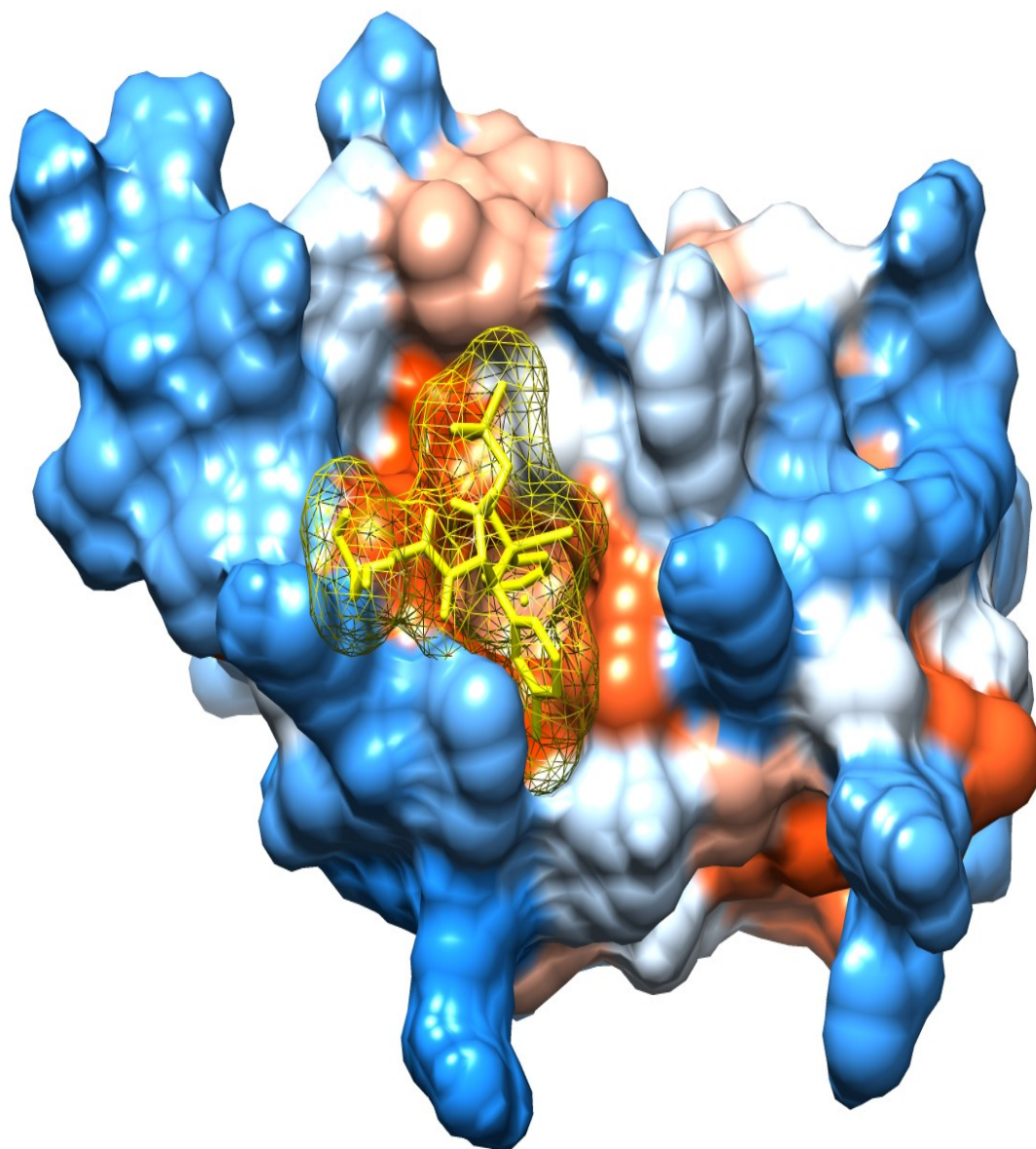
Figure 16. Snapshot of PDB structure 4JRG. The small molecule inhibitor shown in yellow is bound to the P53 binding domain of MDM2. The mesh surrounding the molecule is the surface rendering designed to highlight the full extent of the interactions with the protein. The re-docking experiments performed during this study uses the known binding site pictured here against AutoDock and Vina's best predictions.

cancer, ovarian and small cell lung cancers. While effective for about 30% of patients,

the effects are temporary, due to topoisomerase resistance mutations [97]. Research has
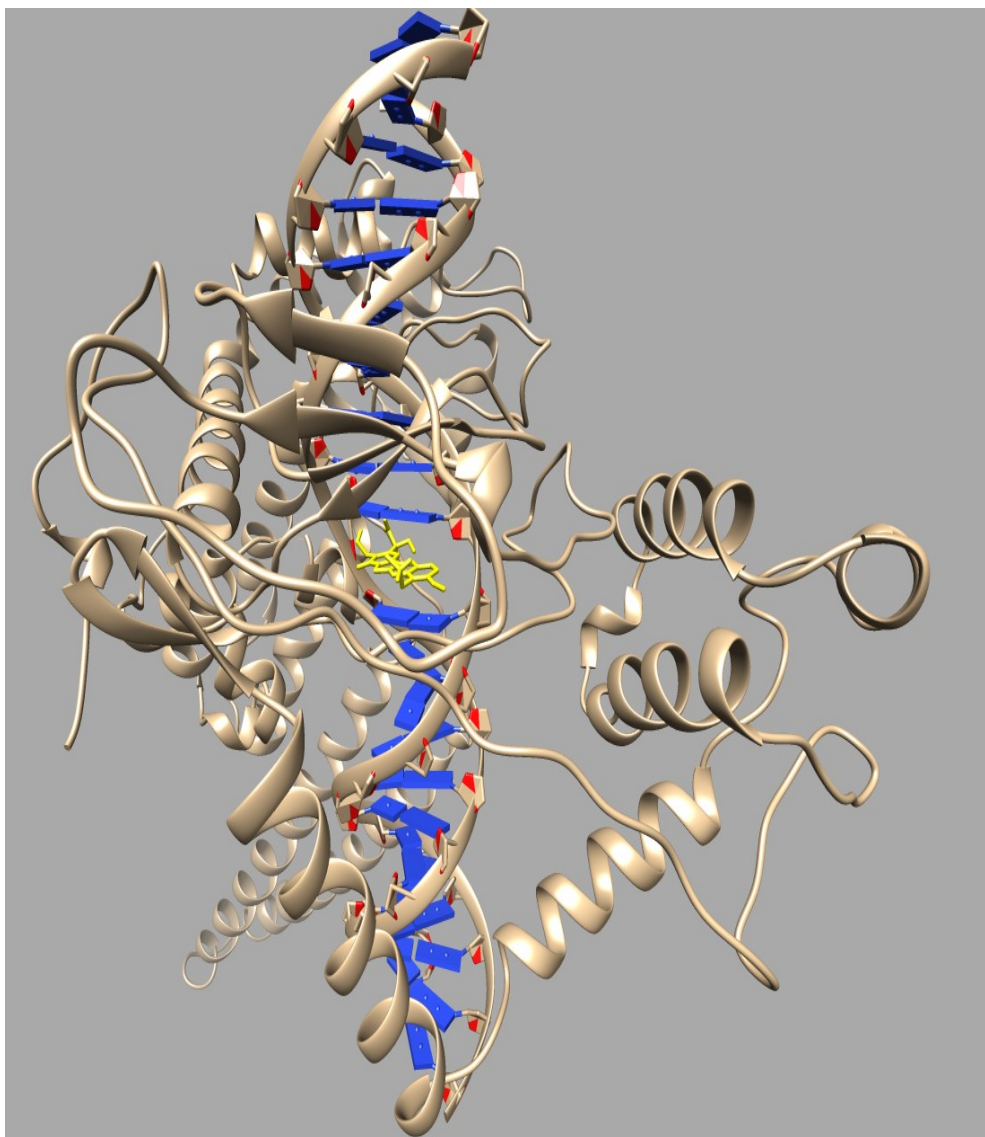
Figure 17. Ligand TTG (yellow) is shown intercalated between DNA base pairs (blue) while bound to the topoisomerase enzyme (ribbon). This forms a ternary complex designed to stabilize the DNA in cancerous cells and disrupt replication during cell division. Research is now focused on topoisomerase mutations undermining the drug effectiveness possibly due to over-accumulation of the drug in tumor cells or adoptive responses to DNA disruption.

focused on not only understanding the mutations, but modifying the chemical structure of

camptothecin to improve efficacy [98]. CADD utilizing docking methods can aide this

effort by modifying and testing possible changes to the drug *in silico,* saving time and money during development.

Methods and Materials

Standard docking experiments employ a rigid receptor/flexible ligand protocol while exploring conformational space within a specified grid box designated by the user. Results are assessed according to the binding free energy, which represents a measure of how securely the small molecule is attached to the desired area of a target molecule. The binding mode is then analyzed to ensure the small molecule's correct location and geometry accompanies the top ranked binding energy. A successful re-docking will be within 2 angstroms of the experimentally known binding site and correspond to one of the top ranked binding energies [99][100]. In short, the sum of the energy of ligand and receptor separately is greater than the total energy when bound together. The difference is the binding free energy. A higher negative energy indicates a deeper potential energy well, a more stable complex, and more likely binding mode [7]. For this study, only the top ranked binding energy and corresponding RMSD from the known binding site was considered as a data point.

Experimental details. A set of four structures, representing small molecule inhibitors in complex with MDM2 was retrieved from the PDB (table 1). PDB codes: 4JRG, 3LBK, 4IPF and 4ZYI. For each complex, the implemented protocol systematically distributed a total of 12 rotatable bonds between the ligand and receptor until all combinations were tested beginning with 0 flexible bonds for the ligand and 12 for the receptor, and then 1 flexible bond for the ligand and 11 for the receptor and so on,

53

Table 1. Ligand and receptor parameters of structures tested.

PDB Structure	Ligand		Receptor
	Ligand Identifier	# of Rotatable Bonds	Binding Site Residues
4JRG	I09	12	ILE57 ILE95 TYR HIS69 PHE LEU HIS92 MET VAL LYS
4IPF	1F0	10	HIS69 HIS92 GLY LEU50 LEU53 MET VAL ILE57 GLN
4ZYI	4TH	9	PHE55 PHE86 LEU GLN24 GLN72 HIS TYR67 TYR100 MET
3LBK	K23	5	ILE61 LEU PHE ILE99 GLY HIS VAL MET TYR100

using the notation (0,12) and (1,11) respectively. When the maximum number of rotatable bonds was reached inherent in each ligand, the remainder was transferred to the protein. Docking parameters for all calculations using AutoDock version 4.2 were adjusted to 100 runs with 2×10^7 energy evaluations and a grid box size of 60 Å 62 Å 62 Å centered on the ligand with .375 Å grid spacing (figure 18). The grid box for Vina version 1.1.2 was set to 27 Å 27 Å 27 Å centered on the ligand with a 1 Å grid spacing and the exhaustiveness was set to 12. The grid box created for AutoDock and Vina runs are the same size, but the spacing is different. This accounts for the apparent smaller grid box of Vina whose spacing is 1 Å. The LGA was designated as the AutoDock search method, while all other settings for both programs were kept at default parameters.

The structure representing camptothecin in complex with topoisomerase and DNA was retrieved from the PDB, (PDB code: 1RRJ). The ligand (camptothecin) identifier used for this study is TTG, which is the hydrolyzed version (1 extra Oxygen and 2 extra Hydrogen) of another form of the drug TTC. The ligand has 9 inherent rotatable bonds instead of the 5 found with TTC. All settings were kept consistent with those used for MDM2 docking runs apart from the search grid box. The grid box used for this docking was 60 Å 72 Å 60 Å (figure 19). The 72 Å reflects the unique geometry associated with a protein bound to DNA. This translates to a Vina grid box of 22.5 Å 27 Å 22.5 Å. Instead of a systematic evaluation of each configuration of ligand and protein flexibility, the docking study for 1RRJ only contained two configurations. One, using the standard protocol of a flexible ligand and rigid protein, and the second, using the rigid ligand protocol transferring 12 rotatable bonds to the binding site residues of the protein. The notation is (9,0) and (0,12) respectively. This is sufficient to evaluate the merits of a

rigid ligand protocol in contrast to the rigid protein protocol currently employed by HTS

without exploring all combinations of flexibility because it explores the only other

pragmatic option for large virtual screenings. When evaluating thousands of small

molecules, it becomes impractical to select the optimal number of rotatable bonds for

each ligand. Alternatively, a target protein could be prepared with selected flexible
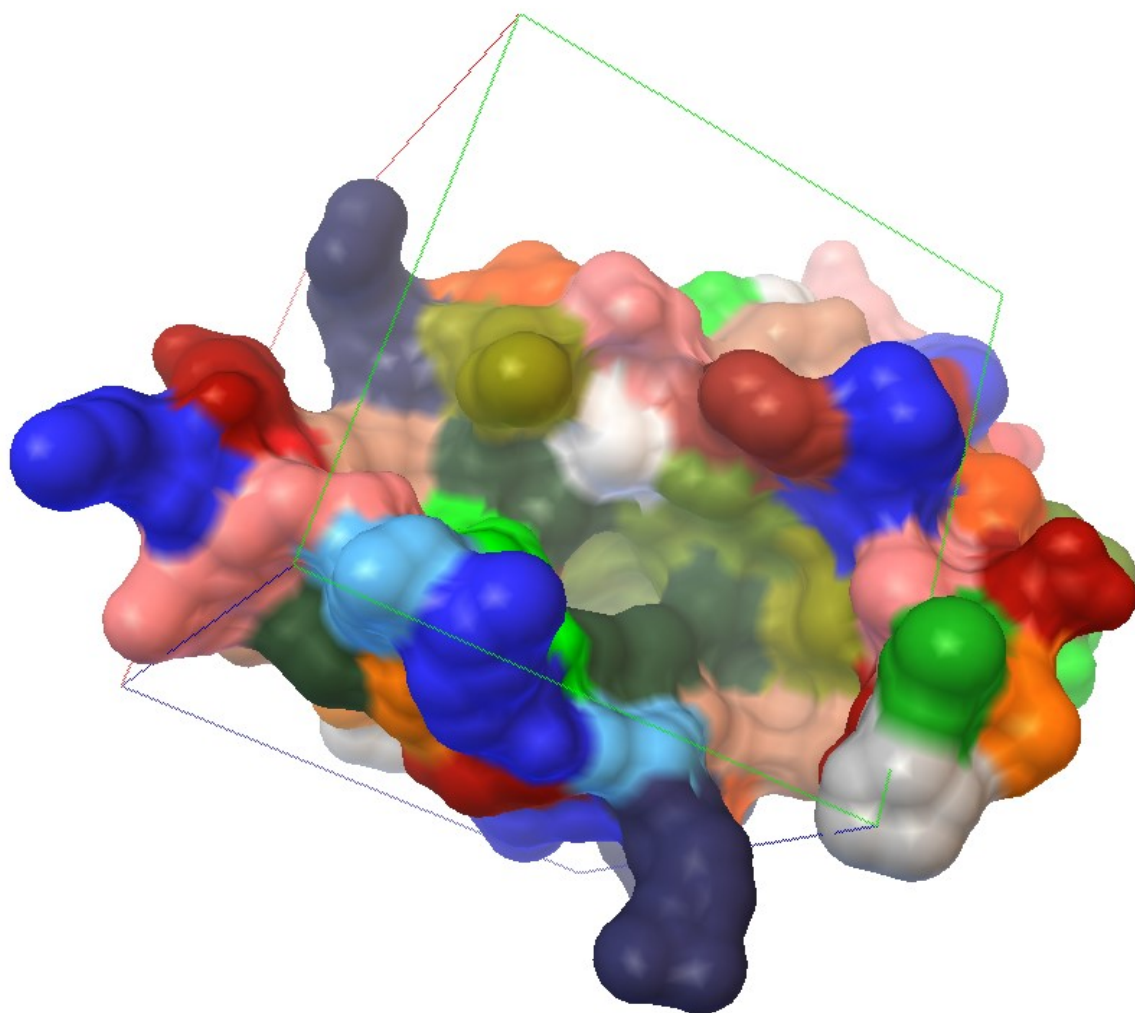
Figure 18. Conformational search grid box for all MDM2 docking runs as rendered
by AutoDock tools. The protein residues are labeled by color.
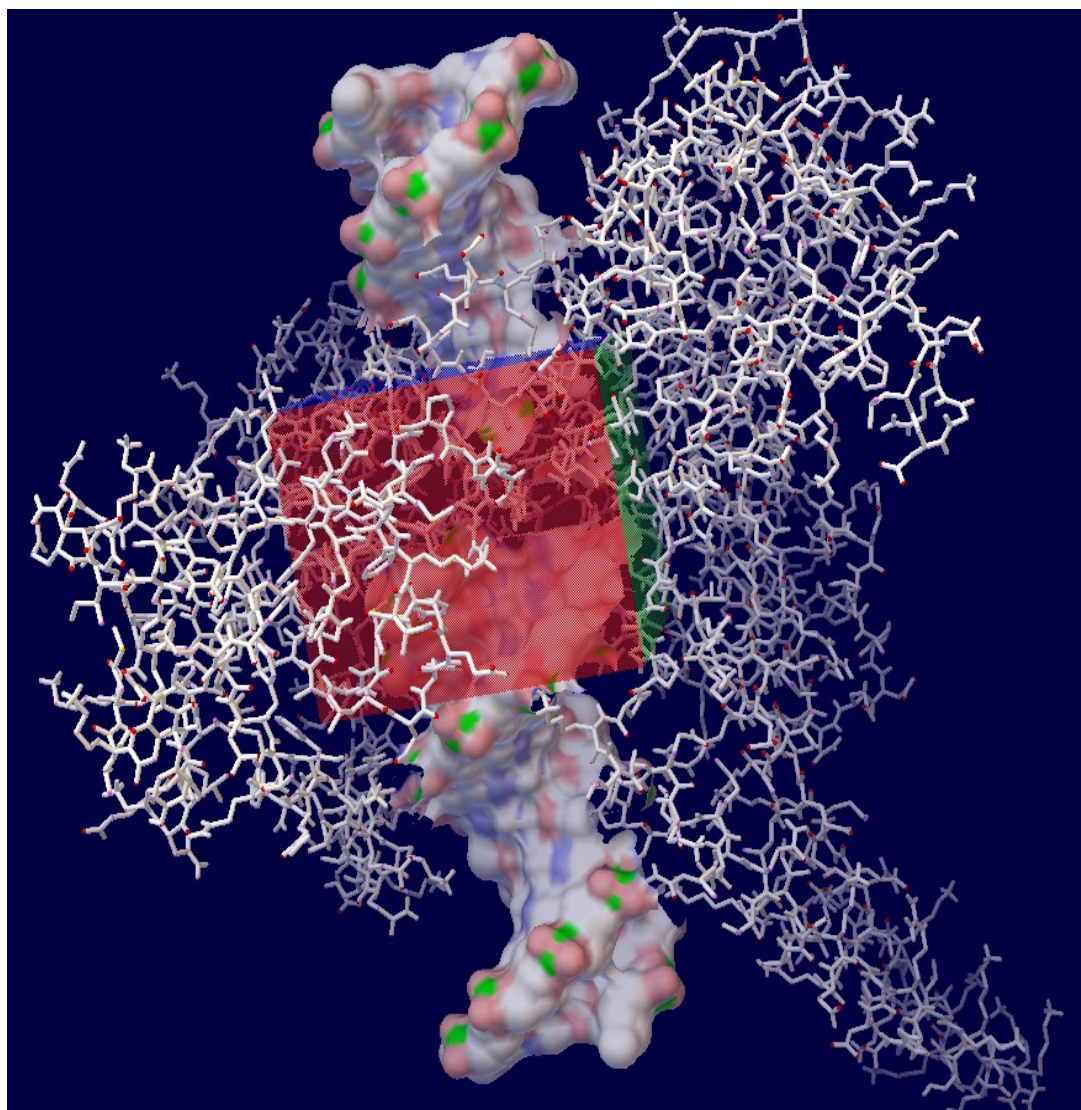
Figure 19. Grid box for two docking runs of structure 1RRJ. The box must be large enough to encompass the binding site flexible residues of the protein and the experimentally known ligand binding site. Failure to make the box large enough will result in positive binding energy.

binding site residues, while keeping all ligands rigid without disruption of the screening.

This would allow for the modeling, at least to a small extent, of binding site flexibility.

The residues and number of rotatable bonds for the (0,12) configuration of 1RRJ include:

Arg364 (4), Asp533 (2), Lys532 (5) and Glu356 (1).

Ligand/Protein Preparation. All structures were retrieved from the PDB and initially prepared for docking using Chimera software [101] . The ligand was separated from the protein and a short energy minimization was applied to each structure for a duration of 10 steps. Hydrogen atoms were added, water molecules removed, and Gasteiger charges were automatically added to the ligand and protein. The necessary files for docking were prepared in AutoDock tools (ADT). When files are imported into ADT, they are checked for polar hydrogens, water molecules and proper charges. The rotatable bonds of the ligand were altered using the 'choose torsions' option. Here, the initially flexible bonds of the ligand can be adjusted for docking and saved. A flexible residue file was created for the rotatable bonds of the selected protein binding site residues in addition to a separate rigid protein file. The identical grid box was created for all structures tested using AutoDock. The grid box values for Vina dockings are placed directly into the Vina configuration file (See appendix A for example input files).

AutoDock/Vina Docking Output. AutoDock results are ranked according to the highest negative binding free energies and corresponding RMSD values from the experimentally determined binding site. Each of the configurations tested produced 100 predicted binding sites that were clustered and ranked. A cluster refers to dockings within 2 Å of each other and they are also ranked according to the binding free energy. The number of runs parameter set by AutoDock users instructs the program to conduct, in this case 100 separate searches beginning from a random location inside the designated grid box. Each run ends when the set number of energy evaluations have been reached in an attempt to find the global minimum energy. Vina's results also rank dockings according to the binding free energy. However, Vina's top ranked binding energy always

58

corresponds to a 0 RMSD regardless if the predicted binding mode matches the experimental structure. The subsequent RMSD values are in relation to this top ranked pose. Determining if Vina and AutoDock can converge on a similar binding mode can be accomplished utilizing visualization software, which can directly compare the experimentally known structure to both programs best prediction. AutoDock and Vina share functional commonalities including the global optimization of the scoring function, pre-calculation of grid maps, and the pre-calculation of distant dependent pair-wise energetics between each atom type. However, they employ a different scoring function and algorithms to obtain binding free energies. Conveniently, both programs utilize the same ligand, receptor, and flexible residue files. This allows for a seamless transition between the two programs when the input files are prepared.

Results and Discussion

The MDM2 oncoprotein is a current target for cancer drug development in the form of small molecule inhibitors designed to firmly attach to its P53 binding domain, thus blocking P53/MDM2 interaction. The physical and chemical structure of MDM2 has been widely studied and a variety of crystallized structures have been developed. Many small molecule inhibitors have been crystallized complexed with the protein including many cancer drugs. This has enabled a thorough understanding of the binding site domain of MDM2 inhibitors. This domain contains 9 or 10 binding site residues depending on the crystal structure, each with between 1 and 4 rotatable bonds. This research sought to improve binding mode predictions using the popular docking programs AutoDock and Vina through the selective flexibility of both ligand and binding site residues using four

crystallized structures obtained from the PDB.

The docking results indicate best binding mode predictions correspond to ligand rigidity and protein flexibility for all structures tested using either program (table 2). Further, configurations of complete ligand rigidity (0,12), produced smaller RMSD values compared to the standard protocol for all structures. Surprisingly, supplying additional rotatable bonds well past AutoDock's usual accuracy threshold improves binding site predictions when this additional flexibility is transferred to the protein. This was not expected, since AutoDock usually experiences reduced accuracy as the number of rotatable bonds approach and exceed 10. The PDB structure 3LBK represents a small molecule in complex with MDM2 containing only 5 inherent rotatable bonds, but AutoDock's most likely docked pose has it 2.20 Å from the experimentally known site using the standard protocol. This improves to .51 Å when 12 rotatable bonds are supplied to selected binding site residues (figure 20). A snapshot of each of these two binding modes highlights the contrast between AutoDock's best prediction and the experimentally known binding geometry. From figures 21 and 22, we can see the juxtaposition of predicted and experimental geometries of configuration (5,0) representing the standard protocol and (0,12), the rigid ligand docking. The geometry and proximity of the standard protocol docking is not nearly as precise as (0,12), shown by the 2.20 Å RMSD as compared to .51 Å. from the experimentally known site using the standard protocol. This improves to .51 Å when 12 rotatable bonds are supplied to selected binding site residues (figure 20). A snapshot of each of these two binding modes highlights the contrast between AutoDock's best prediction and the experimentally known binding geometry. From figures 21 and 22, we can see the juxtaposition of

Table 2. AutoDock results from rigid ligand-flexible residue configuration (green) contrasted with the standard protocol methodology (red).

PDB Structure	RMSD(Å)		Binding Free Energies (Kcal/mol)		Binding Site Flexible Residues (# of rotatable bonds)		
4JRG	.66	3.84	-10.25	-9.63	LEU(2)	ILE57(2)	HIS(2)
					VAL(1)	ILE95(2)	LYS(3)
4IPF	.64	1.68	-14.28	-12.29	LEU50(1)	LEU53(2)	ILE(2)
					HIS69(2)	HIS92(2)	VAL(1)
					TYR(2)		
4ZYI	.44	.77	-20.58	-12.63	GLN24(1)	GLN72(2)	LEU(2)
					TYR67(2)	TYR100(2)	PHE(1)
					HIS(2)		
3LBK	1.02	3.52	-11.14	-10.32	LEU(2)	MET(3)	VAL(1)
					HIS(2)	ILE(2)	TYR(2)

predicted and experimental geometries of configuration (5,0) representing the standard

protocol and (0,12), the rigid ligand docking. The geometry and proximity of the standard

protocol docking is not nearly as precise as (0,12), shown by the 2.20 Å RMSD as

compared to .51 Å.

The ligand bound to MDM2 in structure 4JRG contains 12 rotatable bonds, which

is well above AutoDock's validated limit for a successful docking. Using the standard

protocol (12,0), AutoDock's best prediction is 2.83 Å from the experimentally

determined binding mode, while a rigid ligand protocol (0,12), yields a prediction within

.58 Å (figure 23). The trendline shown is indicative of all MDM2 structures tested. As

The ligand takes on more flexibility, the RMSD values rise until, most notably in this

example, spikes at 2.83Å with a rigid protein. The RMSD values of configurations

between the two extremes fluctuate, but it is difficult to infer from these results the reason

for the anamolies of certain configurations showing smaller RMSD values as with (7,5),

next to an almost doubling of RMSD (6,6). Due to the randomness componet of the

search method, we do expect slight differences in results to accompany identical docking

trials. The same input files run on the same system using the same version of the program

is not likely to produce identical results. This does not account for the consistent patern of

increased RMSD values found in these docking calculations. The pattern we see of

increased fluctuation, especially of those configurations with almost equal number of

rotatable bonds attributed to the ligand and protein, is probably due to the selection of

rotatable   bonds (figure 24). When a ligand with 12 inherent rotatable bonds has 6 held

rigid as with (6,6), and is docked to a protein with 6 rotatable bonds within the binding

site residues, we can anticipate how the result may change when we simply reverse the

rigid and flexible bonds of the ligand. Due to the randomness of the search and the

enourmous amount of possible combinations, we can only infer from the pattern we see

here a likely strategy for success. In the case of MDM2 at least, the diversity associtated
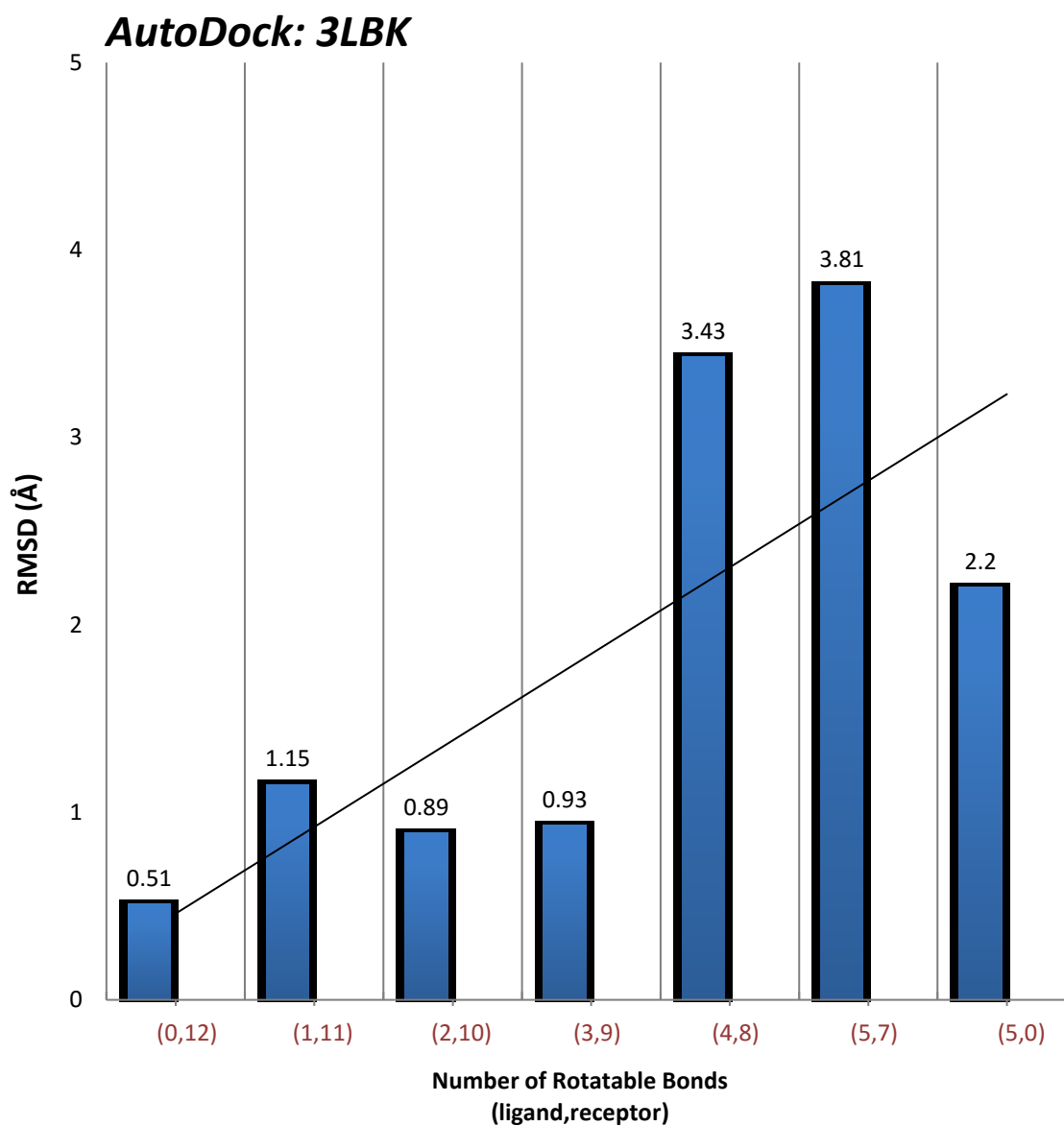
Figure 20. RMSD values corresponding to the top ranked binding energies for all configurations indicate a total rigid ligand (0,12) has the lowest RMSD value of .51 Å. As the number of rotatable bonds become more evenly distributed, binding mode accuracy declines.
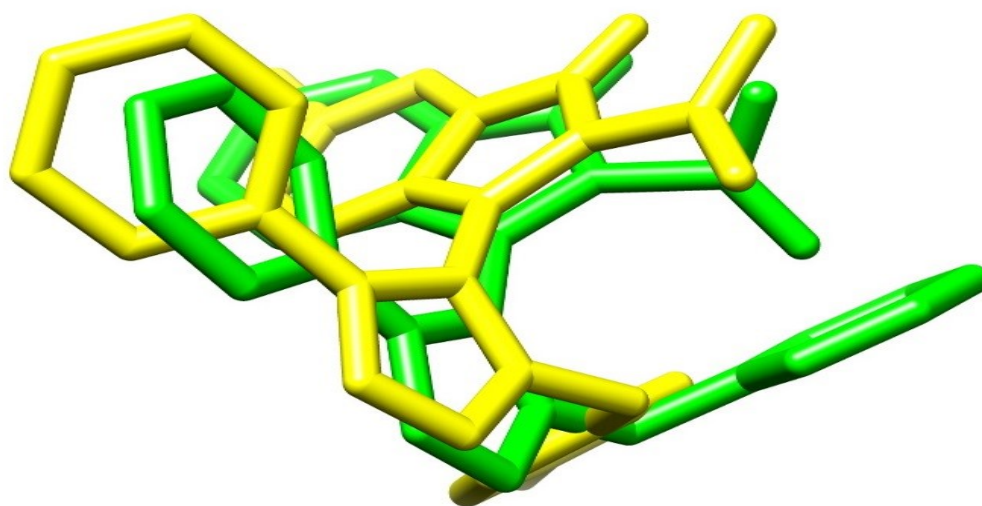
Figure 21. The standard protocol docking pose for 3LBK (5,0) using AutoDock with an RMSD of 2.20 Å (green). In contrast with the experimentally determined structure (yellow).
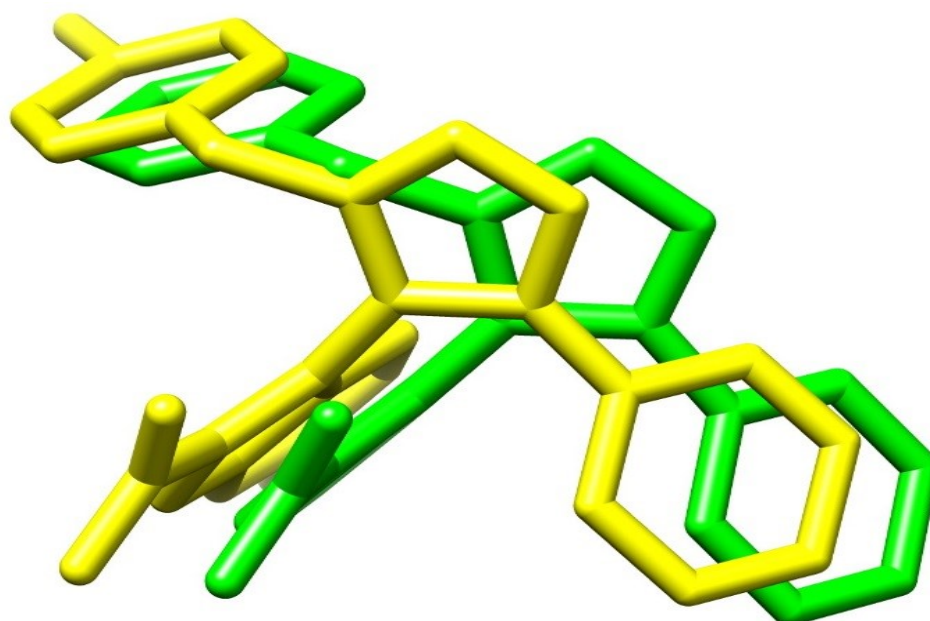
Figure 22. 3LBK rigid ligand configuration (0,12) using AutoDock with RMSD of .51 Å (green). In contrast to the experimentally known structure (yellow).

AutoDock: 4JRG
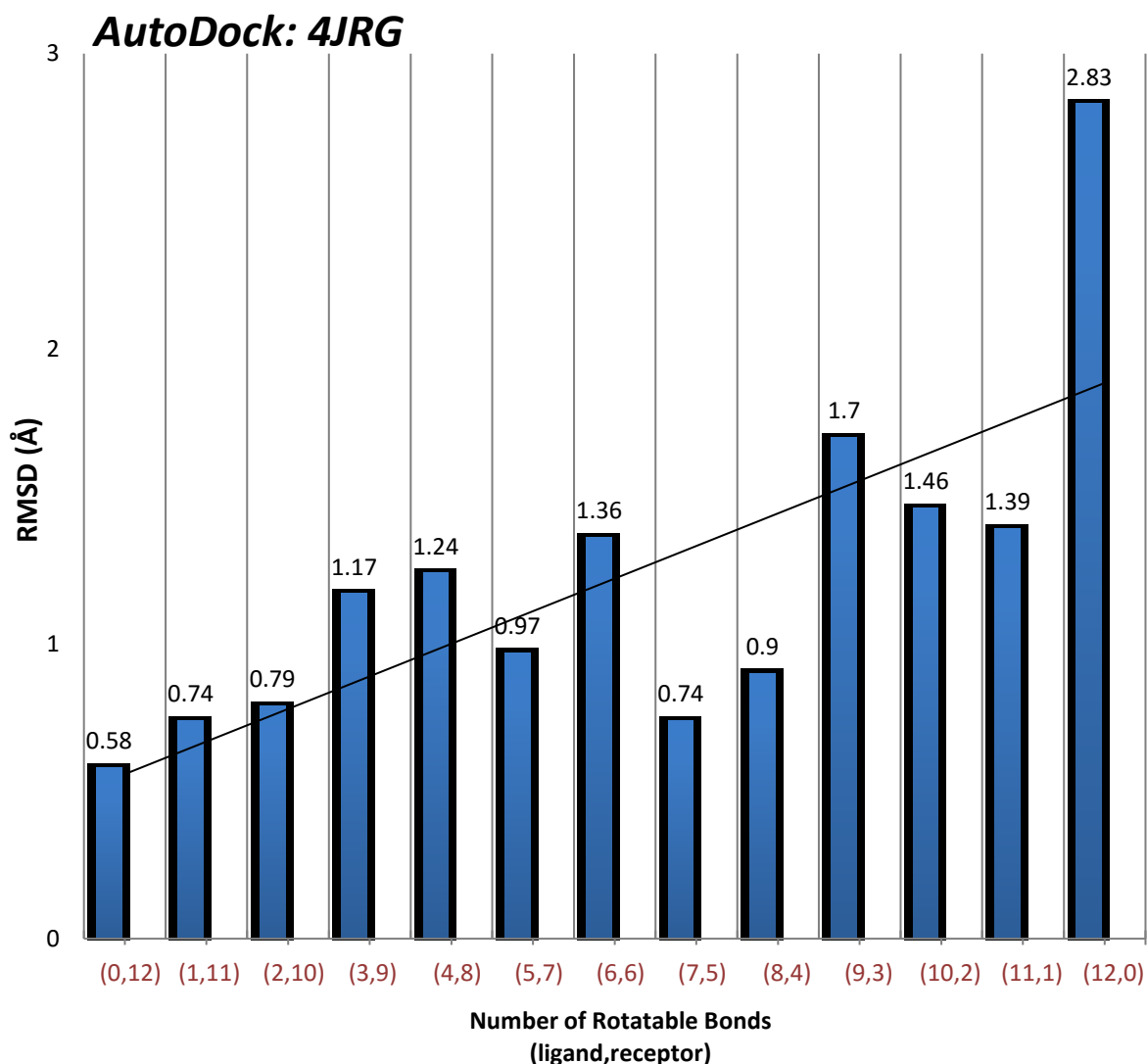
Figure 23**.** The standard protocol (12,0) configuration shows an RMSD of 2.83 Å with the ligand having 12 rotatable bonds in contrast to a rigid ligand and all 12 rotatable bonds transferred to the MDM2 protein represented by configuration (0,12) with an RMSD of .58 Å.

with equal flexibility of ligand and protein reduces accuracy. We would expect (6,6) to

have the lowest RMSD values if just modeling flexibilty was essential for improved

binding mode predictions.

Another notable result from this study is the large RMSD values for those
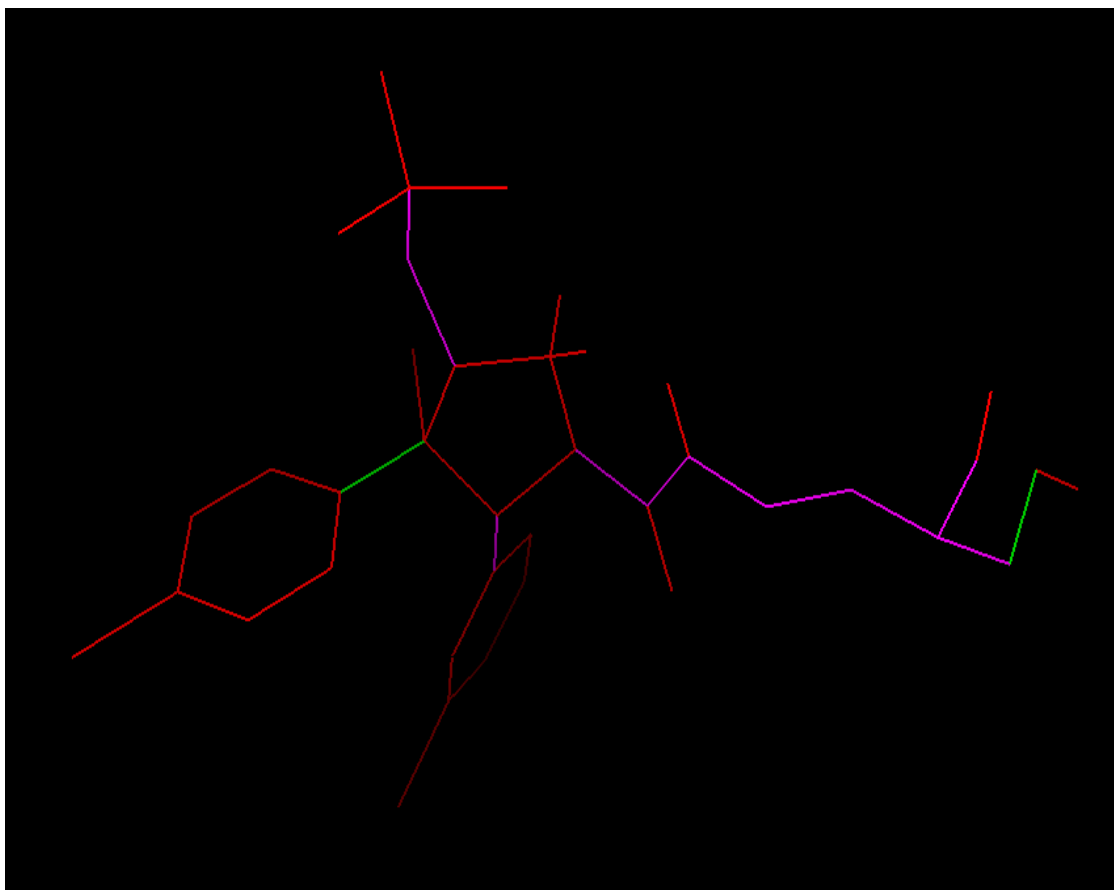
Figure 24. Snapshot of prepared ligand for structure 4JRG, configuration (2,10). The red lines indicate the fixed portions of the small molecule. The magenta lines indicate bonds that could be selected as rotatable, while the green lines are selected rotatable bonds. The selection of rotatable bonds for all ligands was based on allowing the outer-most bonds to become flexible first and then move inward as flexibility increased. If two entirely different bonds were chosen, we can see how the docking calculation may have changed. This is also true for protein flexibility where the same general approach was used to select flexibility.

configurations representing a completely flexible ligand with limited protein binding site

flexibility. For 3LBK, this is docking run (5,7). Although this would seem the ideal

distribution of flexibility as it incorporates the inherent flexibility of the ligand and

binding protein residues, the RMSD value is 3.81 Å, far above the rigid ligand docking of

.51 Å. The same is true of 4JRG, with the (10,2) configuration showing an RMSD of 1.46

Å compared to .58 Å. This is important because the (5,7) and (10,2) configuration of 3LBK and 4JRG respectively would be applicable to HTS as it does not require ligand modification or a change in protein flexibility once the screening starts, and would simulate in part, both protein and ligand binding dynamics more accurately than the standard protocol. This study finds a fully flexible ligand in combination with selective protein binding site flexibility fails to optimize the binding mode predictions. Docking results for structures 4IPF and 4ZYI, also show higher RMSD values for this distribution, with values of 1.64 Å and 3.75 Å respectively (figure 25 and 26). An explanation for the poor docking results could be the consequence of re-docking a rigid ligand, and the built-in bias of the small molecule to assume the same binding mode associated with the experimentally known structure. However, the same bias is present using a rigid protein protocol, as the protein's binding site residues are essentially frozen in the optimal binding mode as is the rigid ligand. When the ligand and protein both have flexibility, this bias is partially removed and predicting binding modes close to the experimental structure becomes more difficult. In fact, all structures excluding 4JRG, produced higher RMSD values compared to both the standard and rigid ligand protocols when a fully flexible ligand was in combination with selected binding site flexibility.

A more exhaustive examination of possible combinations of ligand and protein flexibility may, for any particular complex, improve the ligand binding mode prediction and improve RMSD values. The results from this study only indicates applying a rigid ligand protocol will produce lower RMSD values compared to the standard protocol for MDM2. This is of some importance considering the focus of medicinal interest on MDM2 and considering the amount of MDM2 targeted inhibitors developed as cancer

AutoDock: 4IPF
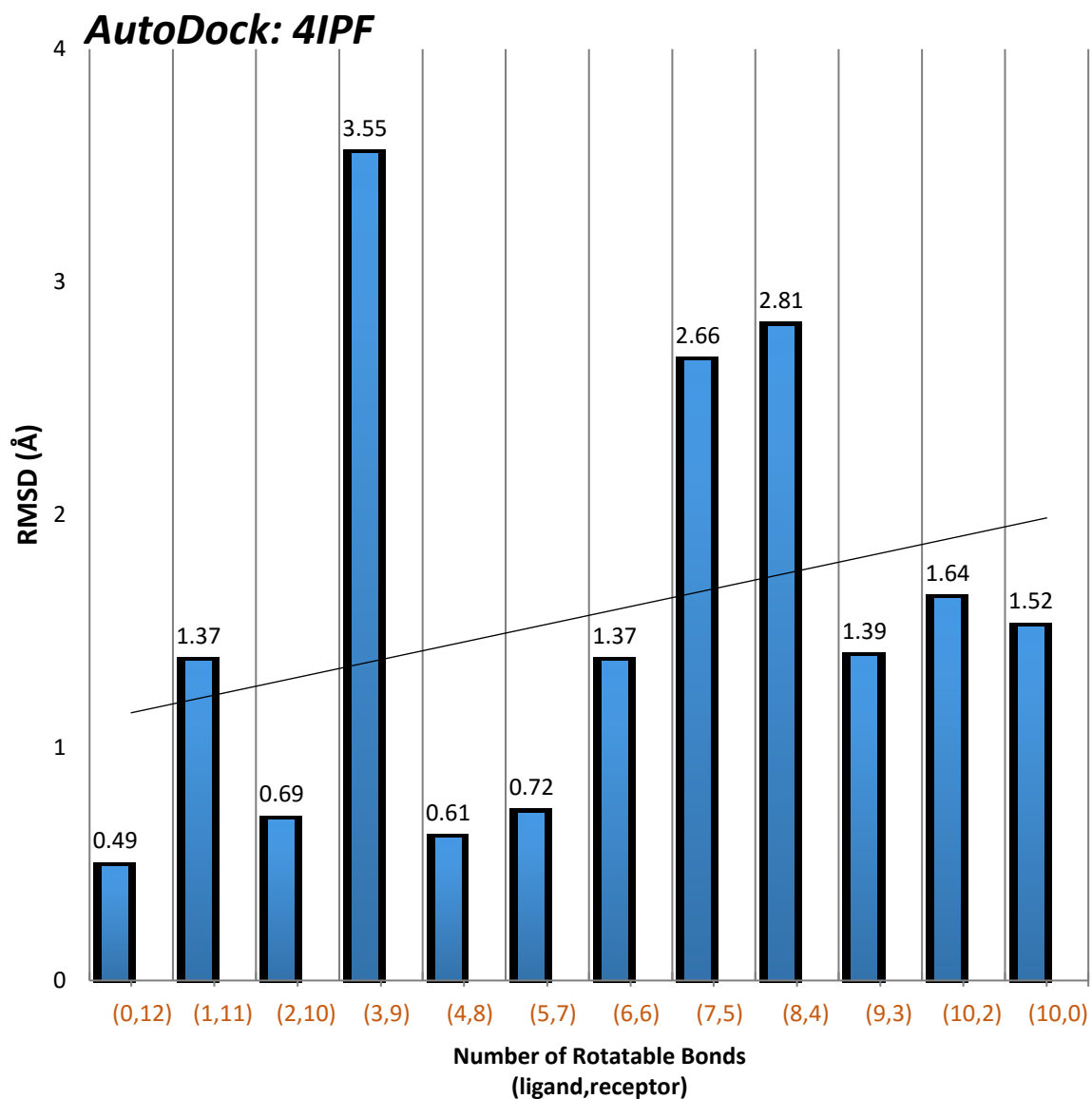
Figure 25. A fully flexible ligand with 2 flexible binding site rotatable bonds (10,2) yields a larger RMSD value of 1.64Å compared to both the standard (10,0) and rigid ligand protocol (0,12).

therapeutics. For HTS, an exhaustive examination of all possible flexibility is not a practical strategy as thousands of drug candidates are screened and evaluated. A large assortment of small molecules is evaluated with varying chemical and physical properties, including flexibility. There is little doubt more robust scoring functions

AutoDock: 4ZYI
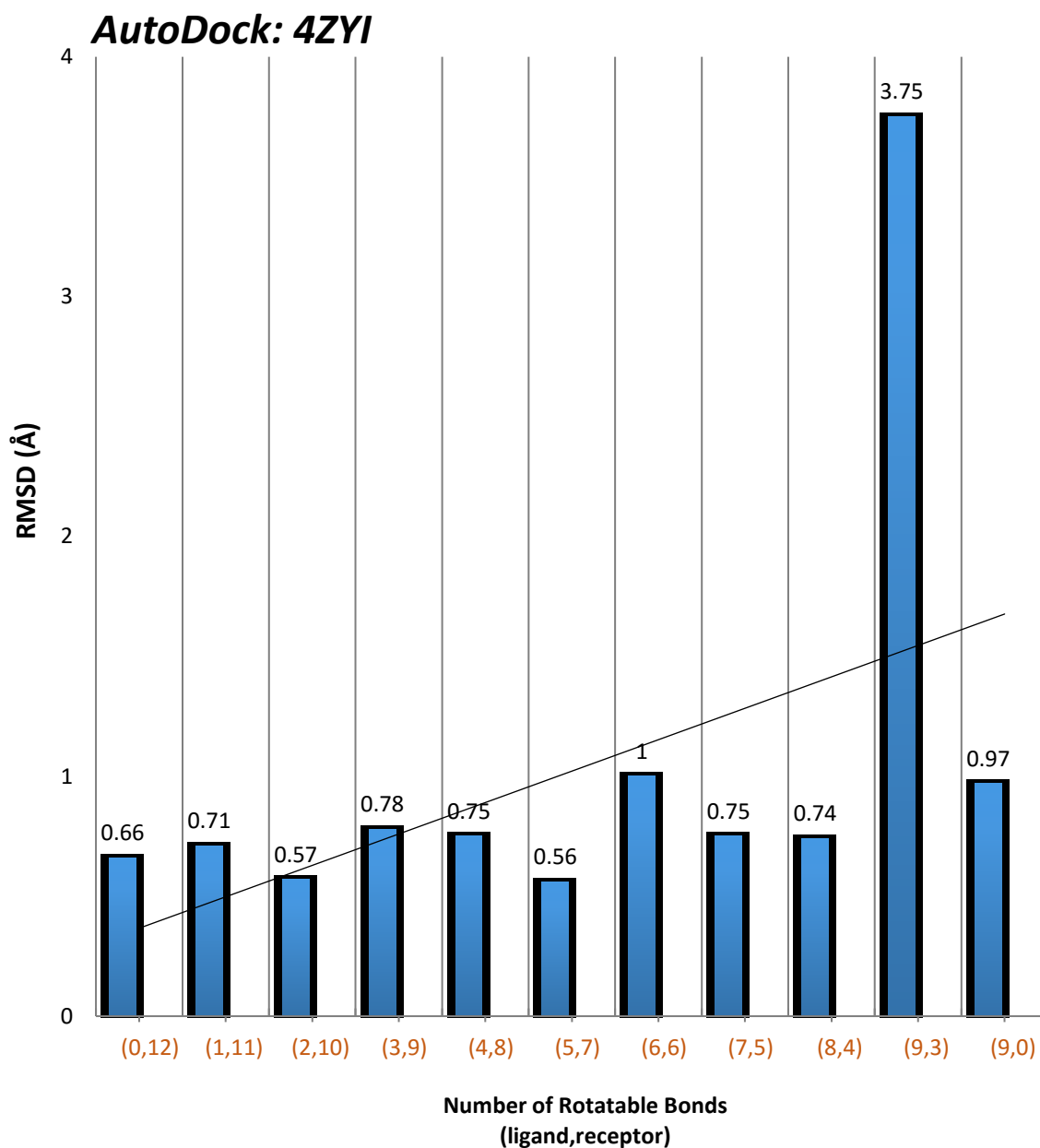
Figure 26. The flexible ligand configuration (9,3) produces an RMSD value much larger than all other combinations. This docking result is the exception as the rigid ligand (0,12) does not produce the lowest RMSD value, but is still well within the successful docking range of 2Å.

and search methods, along with increased computational speed from advancing computer architecture, will improve docking results.

The new scoring function and search method employed by Vina has improved

speed and binding mode predictions and is probably more suited for HTS then AutoDock. However, AutoDock does hold some advantages over Vina in terms of changing docking parameters and output. Vina's docking results are not nearly as extensive as AutoDock. Vina supplies the top ten binding energies but does not associate a RMSD value to the top ranked binding free energy (figure 27). Consequently, Vina results are not charted for this study in terms of RMSD as with AutoDock. Confirmation of a successful docking and comparison to AutoDock must be accomplished using renderings of the binding mode in direct contrast to the experimentally determined structure. Comparing binding free energies is not applicable for two reasons. As discussed previously, AutoDock and Vina use different methods to determine binding free energy with both using many assumptions and estimates. This becomes especially problematic when adding flexibility as the loss of entropy is merely a weighted constant and not calculated. If this constant is in error by just .20 kcal/mol for structure 3LBK, a configuration such as (0,12) will have a widely different free energy than configuration (5,0). It would not be valid to compare the two binding energies or to compare the binding energies between the two programs. The ranking of the free energy is only truly reliable within the same docking run. Secondly, the AutoDock lowest binding energy does not always correspond to the binding mode most closely associated with the experimental structure. A higher binding energy may correspond to a slightly better RMSD value close to the top ranked energy possessing a higher RMSD, but with an almost equally accurate prediction.

An important advantage of using Vina along with a rigid ligand protocol for MDM2 is that it mitigates concerns of additional computational time and costs while allowing for a flexible protein and more accurate binding prediction. Vina's

70

multithreaded configuration can drastically reduce computational time while providing

accurate results when docking ligands, as in this study, with 12 rotatable bonds. Vina's

binding prediction for structure 4JRG mirrors AutoDock's best result of .58 Å as shown

in figures 28 and 29. However, Vina's calculation took less than 2 minutes, while the

AutoDock calculation lasted 15 hours. The increased speed is not surprising, as this has

been shown previously, but while transferring all flexibility to the protein, Vina also

produced a noticeable improvement of binding mode predictions as compared to the

standard protocol with no increase in computational time [102]. Applying a rigid ligand

protocol when using Vina for HTS concerning MDM2 affords the consideration of larger

ligands without reduced accuracy or increased computational costs.

The improved determination of binding geometries using a rigid ligand protocol is

probably best explained by the physical structure of the P53 binding domain of MDM2.

The binding domain as seen clearly in figure 16 sits inside a pocket flanked by residues

not embedded within the protein. These residues can fluctuate during the binding process,

allowing a rigid ligand to enter while the protein conforms to the small molecule

according to the energy landscape determined by its chemical structure as evidenced by

the before and after snapshots of figure 32. The modeling of movement of critical

residues in the same way a flexible ligand would be modeled may explain how the rigid

ligand can enter the binding site and settle in the preferred binding mode. It is important

to remember the program moves both the ligand's and protein's bonds randomly, while

searching for the local energy minima. Modeling the protein movement may be more

essential in finding the global minimum and facilitating changes to ligand geometry to

more accurately match the true binding orientation.
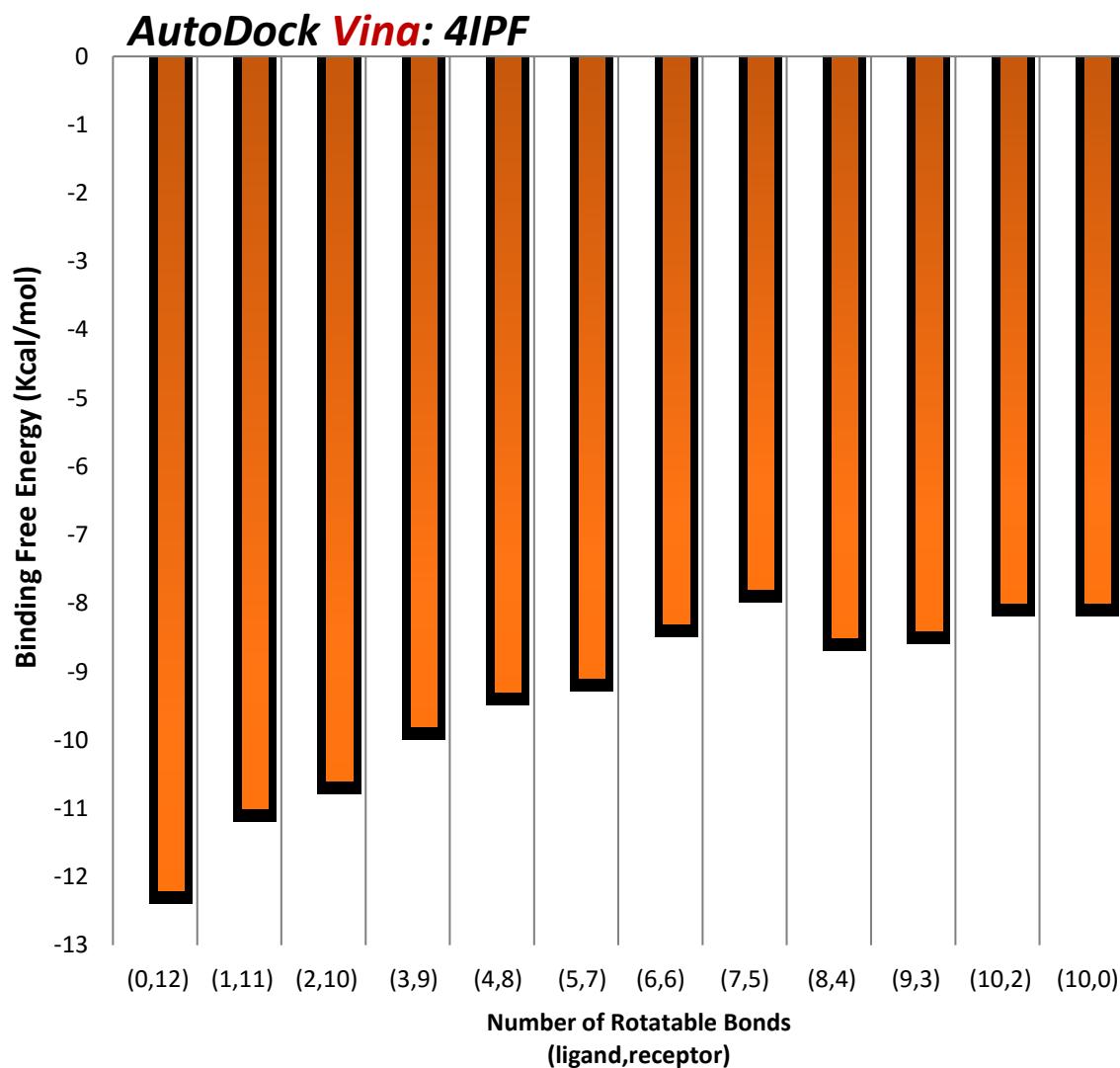
Figure 27. Binding free energy results from 4IPF using Vina shows a similar pattern to AutoDock. The rigid ligand docking has a much higher negative free energy compared to the standard protocol indicating a more likely binding mode. Also, the combinations with a more equal distribution of rotatable bonds show a higher fluctuation in binding energies. This is important because both programs reach this result using different scoring and search methods.

Molecular Dynamics Analysis. A time dependent evaluation of binding site residue movement can be conducted with a classical MD simulation of the MDM2 protein. Using the crystallized protein structure of the 4IPF complex, a simulation
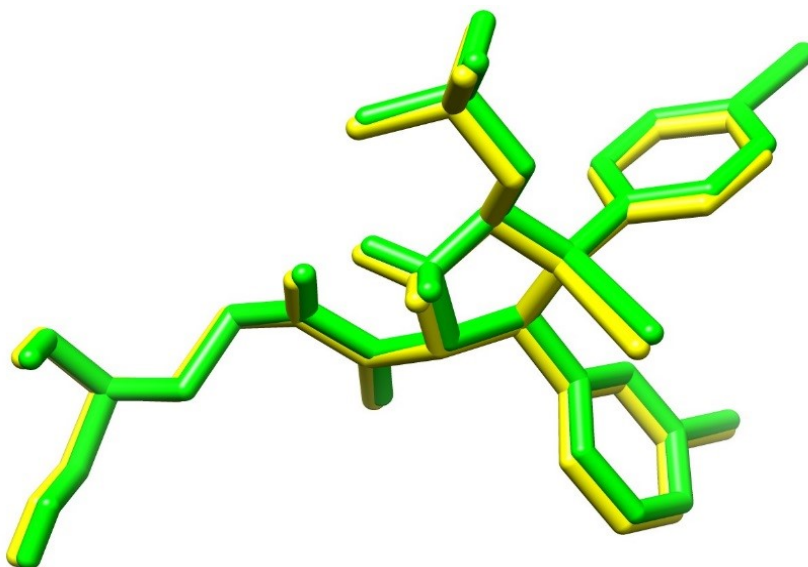
Figure 28. Vina's prediction of rigid ligand docking of 4JRG (0,12). This binding mode prediction (green) is slightly closer to the experimentally known pose found by AutoDock in figure 29.
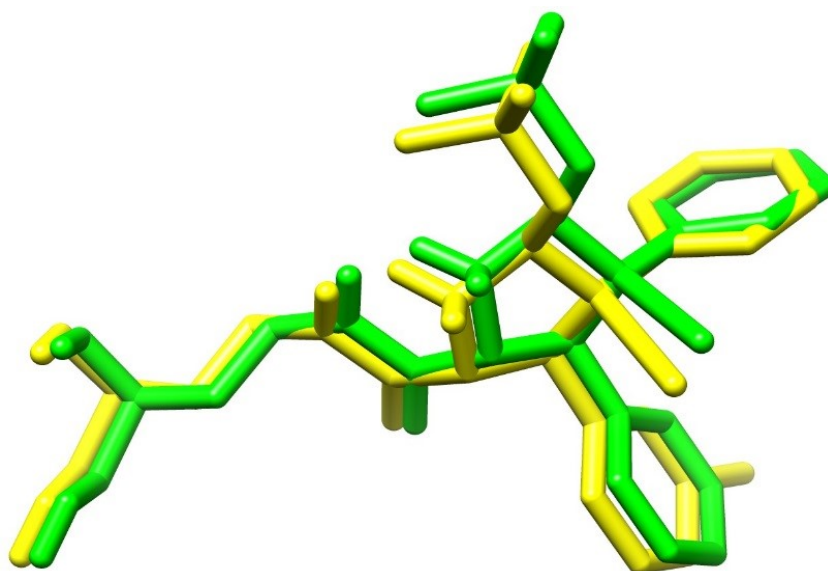
Figure 29. AutoDock's prediction of rigid ligand docking of 4JRG (0,12) shown in green. This docked pose is .58 Å from the experimentally known docked ligand (yellow).

was performed using NAMD. The protein is prepared and necessary files created using VMD (visual molecular dynamics), the graphical user interface (GUI) designed to work with NAMD for the preparation, evaluation and visualization of MD simulations (appendix B). A minimization and equilibration simulation allows for the determination of residue mobility before interaction with the ligand from average RMSD values calculated during equilibration. The data provides valuble insights toward understanding the protein residue dynamics the ligand encounters as it enters the MDM2 binding site and searches for the preferred binding mode. The simulation lasted .5 ns with a 1 fs time step at a constant temperature of 310K in an explicit solvent (figure 30). The long run time ensures the protein will be able to reach equilibrium providing reliable data for the residue RMSD calculation. The use of an explicit solvent accounts for the aqueous enviroment and the desolvation energy associated with removing water molecules from binding site interfaces. Results show the residues made flexible for docking run (0,12) fluctuate between 1.4 and 3.2 Å, with binding site residues HIS 69, HIS 92 and TYR 96 all moving an average of 3 Å (figure 31). The RMSD values serve to quantify movement of the protein in equilibrium while affirming the importance of modeling residue flexibility during docking calculations (figure 31). Although flexibility modeled by docking is just the limited rotation of bonds, with bond lengths and angles kept constant, the change observed from the binding site residues can be pronounced. The before and after snapshot of binding site residues from figure 32 highlights the change in orientation necessary to accommodate a successful docking within .49 Å of the known structures ligand binding mode. This suggests ligand flexibility may not be as essential as modeling protein conformation changes that accommodate a small molecule during the binding

process. Analysis using Ramachandran plots confirm the protein residues settle

in allowed regions according to possible φ and ψ dihedral angles. A selection of these

**RMSD of Minimization and Equilibration of MDM2
(PDB: 4IPF)**
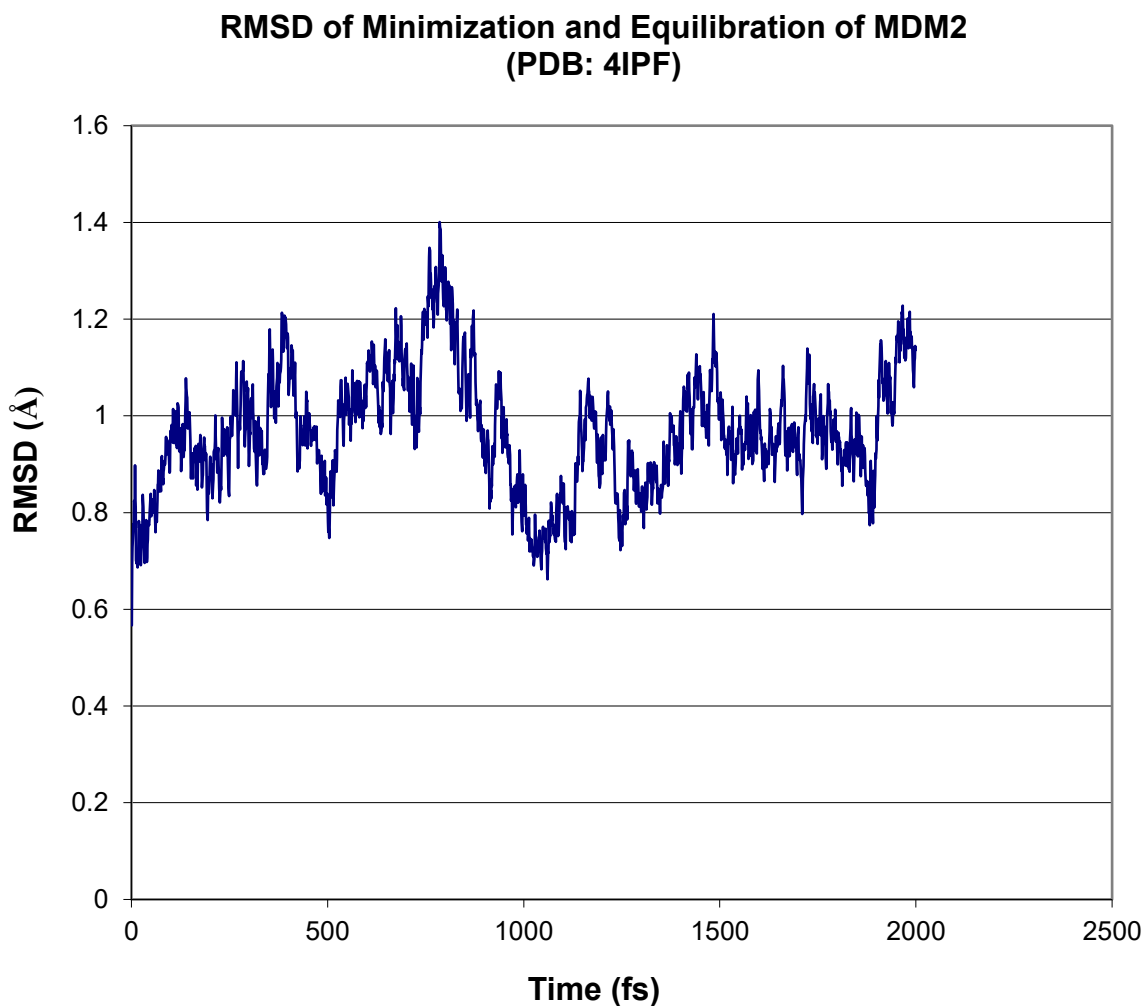
Figure 30. The MDM2 protein as represented by crystal structure 4IPF was used to evaluate fluctuation of specific binding site residues. The equilibration simulation should run long enough as to let the system stabilize according to RMSD fluctuation. Any data collected before that will be artificially altered by the heat energy placed into the system. The graph above shows flattening out, indicating the system is in equilibrium. This allows data to be collected for individual residues in terms of RMSD from their equilibrium position providing a better understanding of the binding site dynamics.

RMSD of Residues for the Minimization and Equilibration of MDM2 (4IPF)
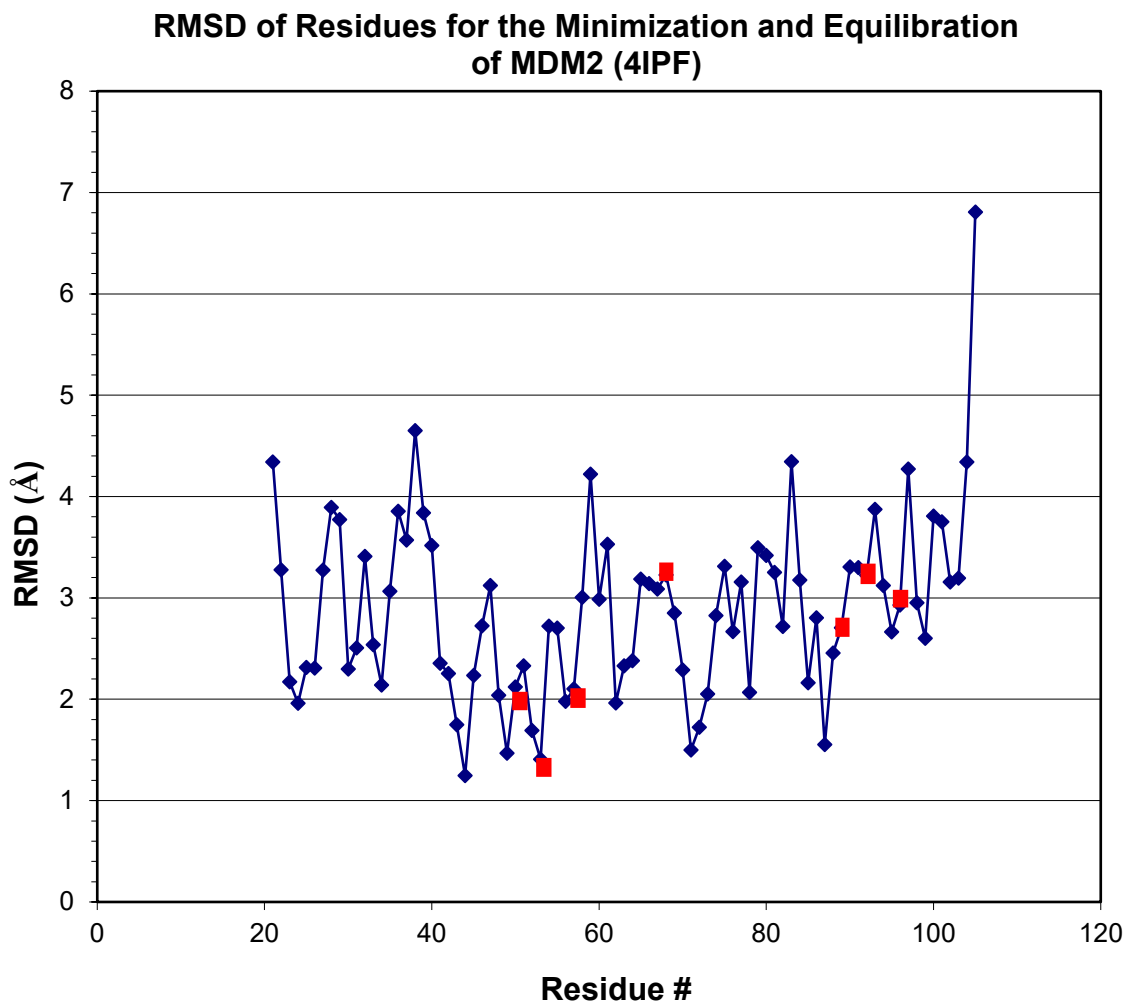
Figure 31. A Molecular dynamics simulation was performed using NAMD to establish residue fluctuation at equilibrium. The red markers indicate binding site residues made flexible for the 4IPF (0,12) docking run. The missing data results from the loss of the first and last twenty residues while crystallizing MDM2, which is typical for this structure.

plots can be found in appendices C through E. The Ramachandran plots ensure the conformation changes predicted by the programs are theorhetically possible and therefore a valid representation of the biological system.

Data Mining Analysis. Given the large number of potential rotatable bond combinations between the ligand and protein, a 'a priori' knowledge of the optimal

distribution of ligand/protein flexibility would be ideal considering testing all possible

considerations would be impractical. For example, 3lbk configuration (5,7) contains 792

possible combinations. As a guide, we can employ a data mining approach where each of

the potential bonds is assigned as one of the parameters to be turned off or turned on (0 or

1). We can then seek the weight factor of each of these parameters by supposing the total
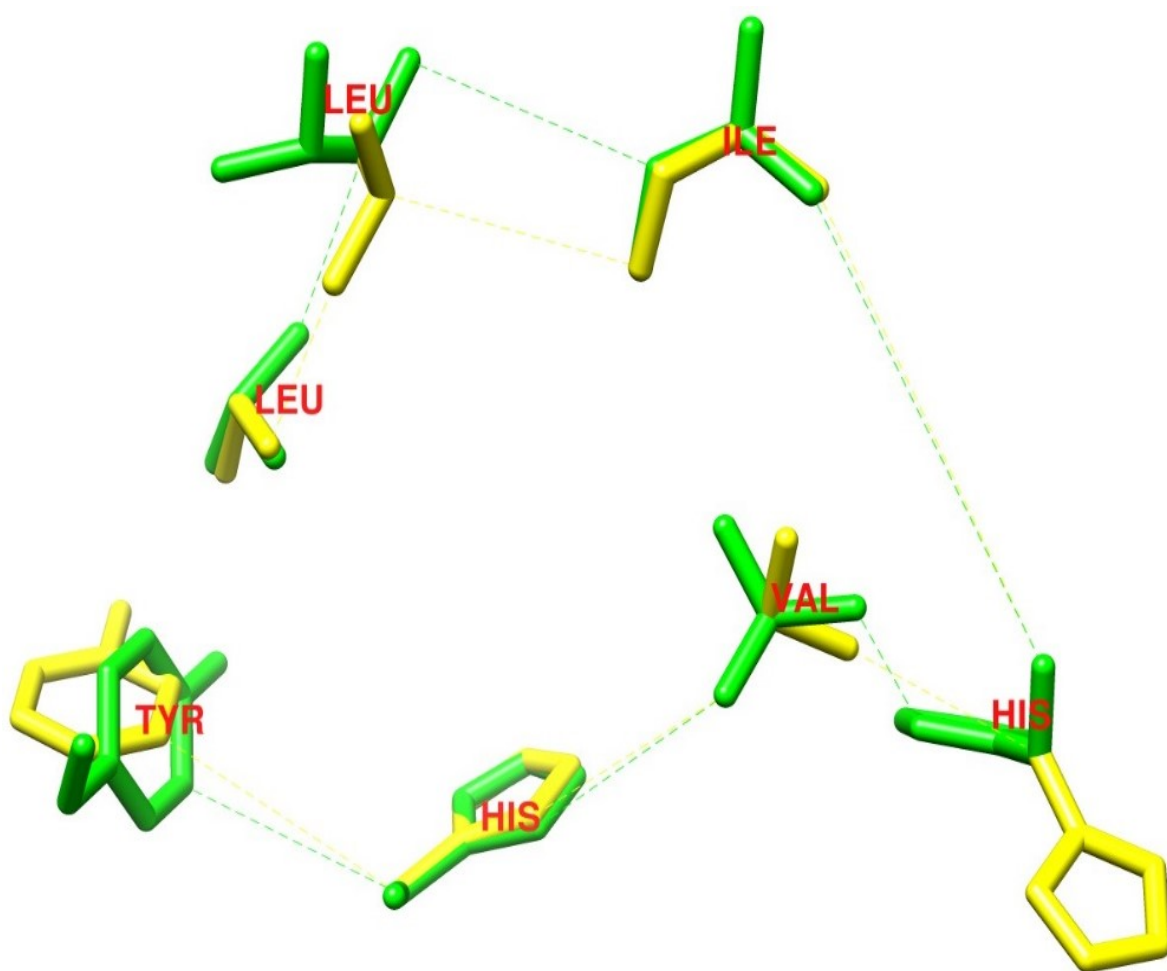
Figure 32. Final positions of flexible binding site residues (green) for 4IPF (0,12) indicates residues LEU, TYR, VAL and HIS have shifted considerable from the experimentally determined geometry (yellow). Due to the time independent nature of docking, we can only capture before and after states of the complex. The protein and ligand have been removed for clarity.

cohesive energy as well as RMSD can be quantitatively linked to a linear regression as a superposition of all the parameters:

$$E = \sum_{i=1}^{N} W_{Li} L_i + \sum_{i=1}^{M} W_{pi} P_i \qquad (23)$$

$$RMSD = \sum_{i=1}^{N} W_{si} S_i + \sum_{i=1}^{M} W_{Ti} T_i \qquad (24)$$

Where:

$L_i = 0$ or 1 (a binary for each rotatable bond in the ligand) for energy

W_{Li} = weight factors of each L_i for energy

$P_i = 0$ or 1 (a binary for each rotatable bond in the residues) for energy

W_{Pi} = weight factors of each P_i for energy

$S_i = 0$ or 1 (a binary for each rotatable bond in the ligand) for RMSD

W_{Si} = weight factors of each L_i for RMSD

$T_i = 0$ or 1 (a binary for each rotatable bond in the residues) for RMSD

W_{Ti} = weight factors of each T_i for RMSD

The number of parameters will depend on the total number of rotatable bonds between the ligand and protein. Each parameter can be toggled on or off corresponding to a flexible or rigid bond. Data mining calculations using the energy, RMSD and distribution of flexibility data as shown in table 3 yielded promising results as particular ligand and protein rotatable bonds showed strong weighted constants which match closely with low RMSD values and high negative binding energies. This simplified method certainly may not generate fully recoverable linear superposition for a complex docking process, but at

Table 3. Structure 3lbk bond toggle displays the activation status of each bond for each configuration in addition to the correspond energy and RMSD values.

(L,P)	L1	L2	L3	L4	L5	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	Energy	RMSD
(0,12)	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	-10.03	.51
(1,11)	0	0	1	0	0	1	1	0	1	1	1	1	1	1	1	1	1	-11.50	1.15
(2,10)	0	0	1	0	1	0	1	1	1	1	1	1	1	1	1	0	1	-11.39	.89
(3,9)	1	0	1	0	1	1	0	1	1	1	0	1	0	1	1	1	1	-13.05	.93
(4,8)	1	0	1	1	1	0	1	1	1	1	0	1	0	1	1	1	0	-14.77	3.43
(5,7)	1	1	1	1	1	1	1	0	0	0	1	1	1	1	1	0	0	-8.36	1.20
(5,0)	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	-10.32	3.52
(5,7)	1	1	1	1	1	0	1	1	0	1	0	0	0	1	1	1	1	-13.69	3.81
(5,7)	1	1	1	1	1	0	0	0	0	1	1	1	1	0	1	1	1	-8.41	2.19
(4,8)	1	1	1	1	0	1	1	0	0	0	0	1	1	1	1	1	1	-8.55	2.30
(4,8)	1	1	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0	-7.95	3.14
(4,8)	0	1	1	1	1	1	1	1	1	0	0	1	1	0	0	1	1	-8.97	2.04

least it does provide a guidance, based on a given number of allotted rotatable bonds, as to which rotatable bonds that are preferred to be "activated". The weight functions may also indicate the relative level of importance quantitatively on activating a specific set of rotatable bonds. Further, we can probably identify if there is a potential negative effect in

activating certain rotatable bonds toward the total energy and/or RMSD through the

formation of negative weight factors. Using this technique, we can locate particular bonds

within binding site residues to activate that will most likely lead to a successful docking.

The linear regression analysis of structure 3lbk indicated activation of bonds P4 and P9

play a significant role in energy optimization while P8 improves RMSD values. Bonds

L4 and L8 play a significant role, both negative and positive, for RMSD. Figure 33 shows

a small nutlin derivative and MDM2 with each ligand and protein labeled for a clear

representation of the 3D structure.
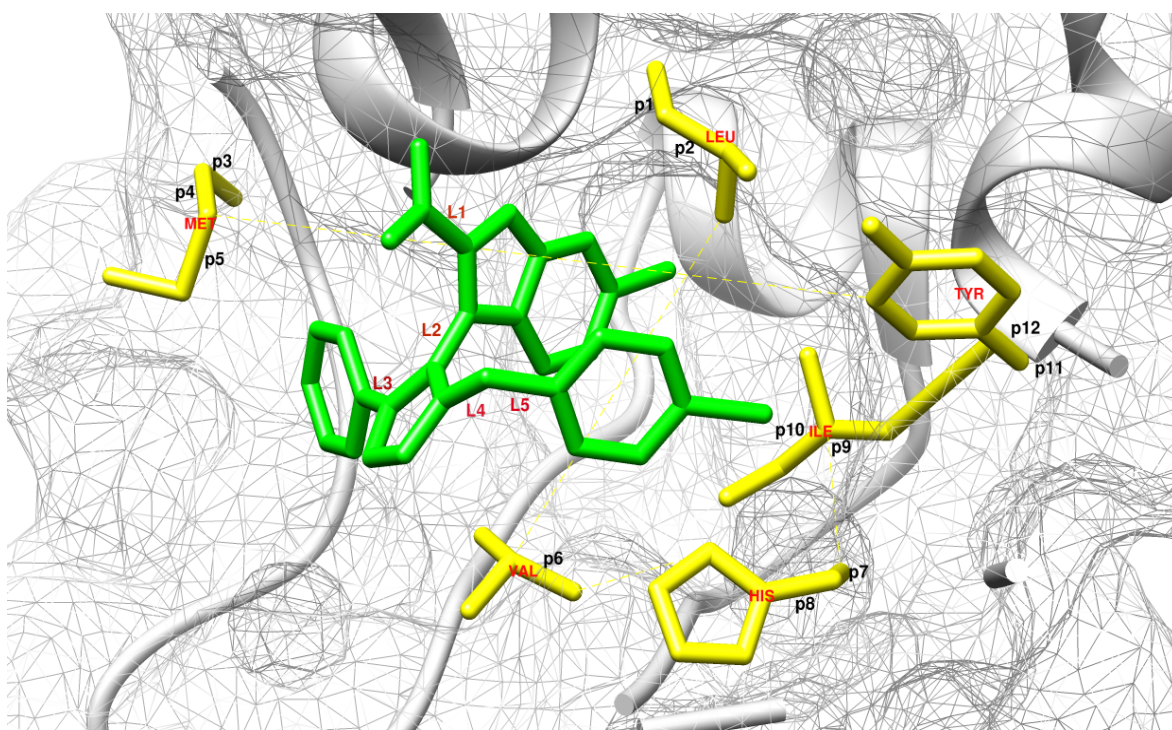
Figure 33. Structure 3lbk with protein and ligand labels together with data mining analysis allows for a clear 3D assessment of the residues and rotatable bonds most critical for a successful docking.

Analysis of Cancer Drug Camptothecin. A brief study of a rigid ligand

protocol in contrast to the standard protocol was conducted testing AutoDock and Vina's

binding mode predictions of a small molecule complexed with DNA and a protein enzyme. This forms a ternary binding since the ligand is bound to the protein while the protein is bound to DNA. This complex is distinct from MDM2 and provides a suitable test case in order to evaluate if the successful rigid ligand docking regarding MDM2 was a special case or could be applicable to other systems. The standard protocol was applied using both programs for the initial docking allowing all 9 of the ligands' rotatable bonds to be flexible (9,0), followed by a rigid ligand docking, with 12 rotatable bonds being assigned to selected protein residues.

The top ranked binding mode for (9,0) using AutoDock was 1.14 Å from the experimentally determined binding site, while the rigid ligand docking was within .74 Å. In this case, both re-docking calculations would be considered successful as they are both within 2 Å of the known binding area, but as with MDM2, the rigid ligand docking produces a noticeable improvement (figures 33 and 34). Vina results also show a more precise orientation utilizing protein flexibility evident from the comparison of figures 35 and 36. The noticeable exception of this study is the fact AutoDock's best binding mode prediction is actually closer in orientation to the known structure as compared to Vina in contrast to all MDM2 calculations. This is surprising because Vina, as mentioned previously, out performs AutoDock in both speed and accuracy. As the parameters for both programs were not changed, the difference in scoring and search method in addition to the chemical and physical structure of the complex may explain AutoDock's superior prediction. Given this result, along with the more comprehensive findings from this research, it further serves to illustrate the importance of testing and validating these programs for optimum performance and results.
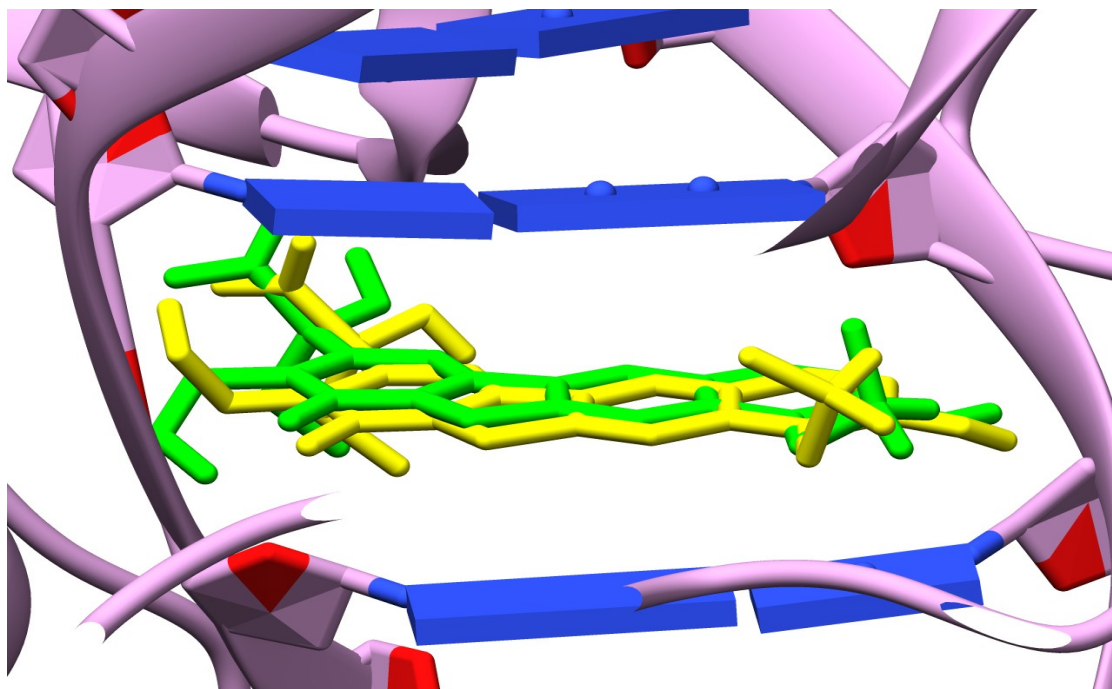
Figure 34. AutoDock standard protocol using a flexible ligand shows the contrast between the experimental orientation (yellow) and predicted binding mode (green).
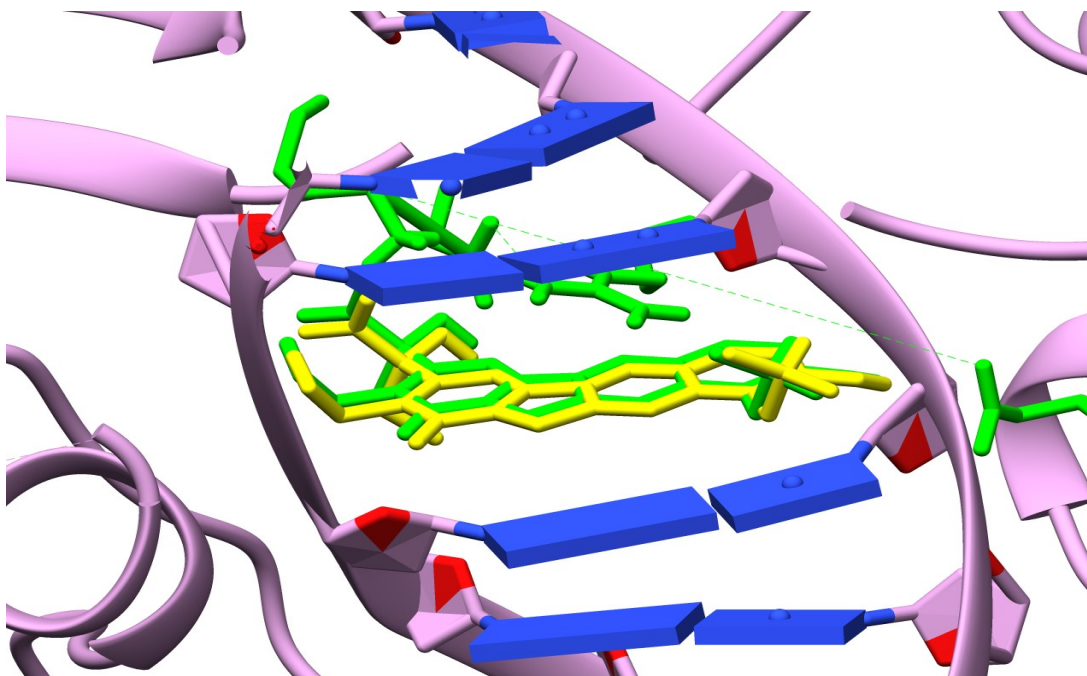
Figure 35. AutoDock utilizing protein flexibility improves binding mode accuracy. Protein flexible residues are shown in green.
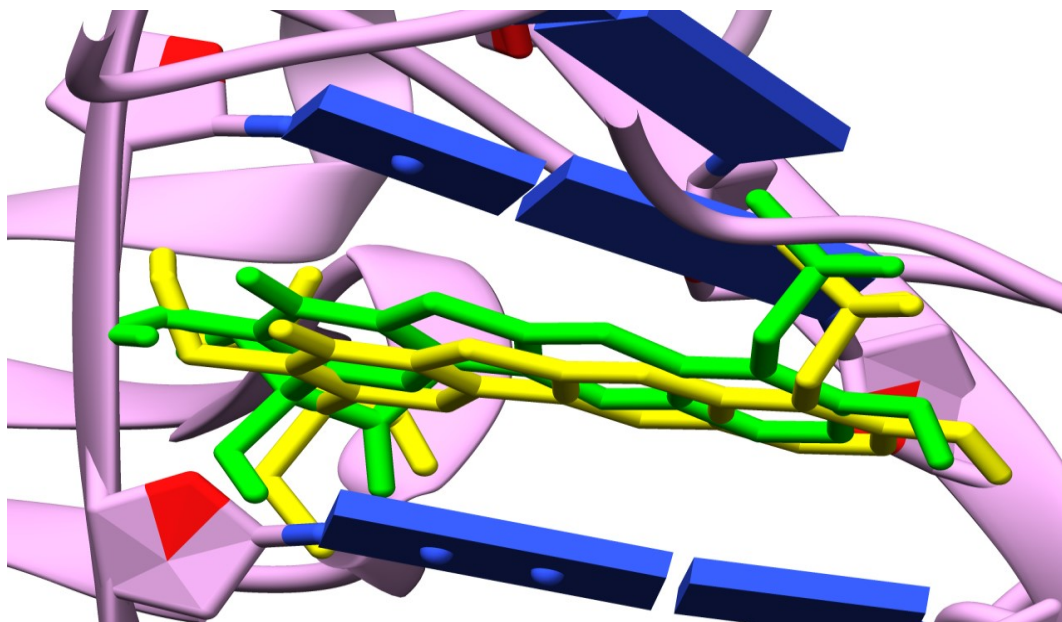
Figure 36. Vina's top ranked pose for (9,0) appears to be well within 2 Å of known structure, but struggles to match orientation at the far left end.
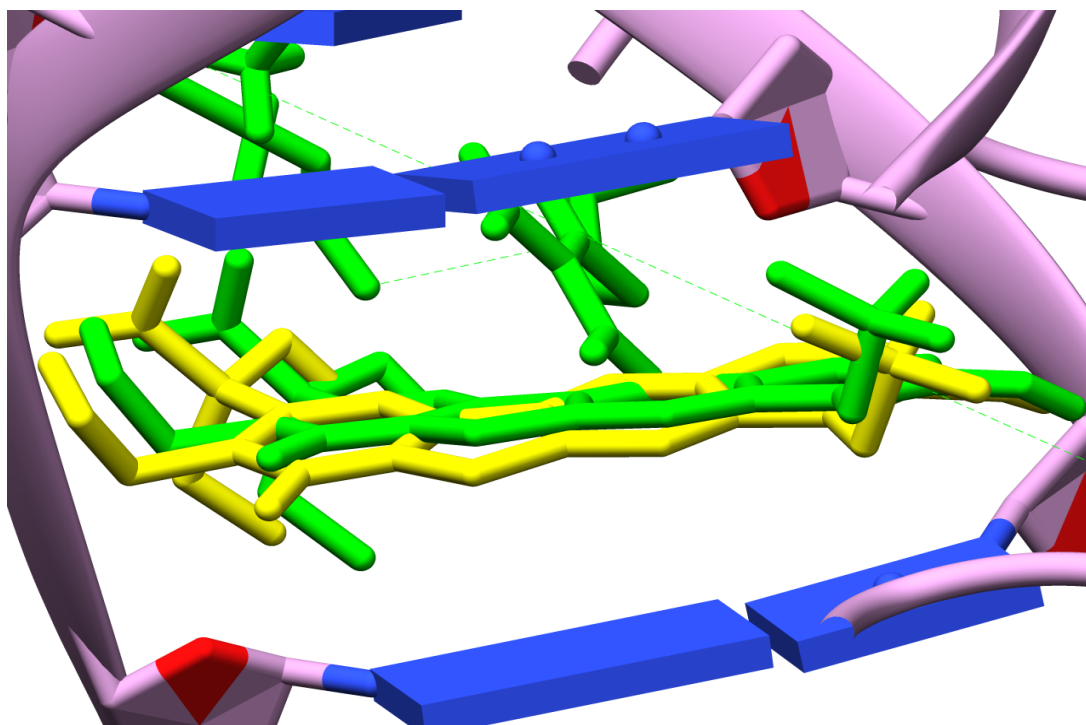
Figure 37. Vina utilizing protein flexibility improves orientation of ligand due to ligand rigidity within the binding domain.

CONCLUSION

Recovery of the P53 tumor suppressor pathway via small molecule inhibitors of onco-protein MDM2 highlights the critical role of computational methodologies in targeted cancer therapies. Improved binding mode predictions of small molecule inhibitors targeting the MDM2 protein was achieved with AutoDock and Vina using a systematic distribution of 12 rotatable bonds between the ligand and protein. This study found a rigid ligand in combination with flexible binding site residues produced lower RMSD values from the known binding site when compared to standard rigid receptor docking. Further analysis of the flexible binding site residues found considerable movement illustrated by a MD simulation and examination of the predicted final positions and experimentally determined positions of the selected residues. An additional evaluation of a DNA targeted cancer drug also produced improved ligand binding mode results incorporating the rigid ligand protocol. The modeling of residues located within the binding site were found to be more determinate than modeling ligand flexibility. Docking runs simulating a fully flexible ligand, in addition to 7,3 and 2 rotatable bonds transferred to MDM2, failed to come as close as a rigid ligand to the experimentally known structure. Adopting a rigid ligand-flexible protein docking protocol for future medicinal studies of MDM2 using Vina in particular, will enable accurate, fast predictions of MDM2 binding modes of small molecules with 12 and possibly more rotatable bonds.

Future studies will determine if this protocol may be applied to other proteins or DNA targets of medicinal interest. The emergence of molecular docking as an essential

tool for SBDD has fostered many clinical studies as well as exhaustive validation studies. The continued evaluation and optimization of these programs complimented by advanced computer architecture will aide in reducing the cost of cancer drug development, as well as foster new insights into bio-molecular binding processes.

REFERENCES

[1]   S.-Y. Huang and X. Zou, "Advances and challenges in protein-ligand docking.," *Int. J. Mol. Sci.*, vol. 11, no. 8, pp. 3016–34, Jan. 2010.

[2]   L. Chen, J. K. Morrow, H. T. Tran, S. S. Phatak, L. Du-Cuny, and S. Zhang, "From laptop to benchtop to bedside: structure-based drug design on protein targets.," *Curr. Pharm. Des.*, vol. 18, no. 9, pp. 1217–39, Jan. 2012.

[3]   S. Cosconati, S. Forli, A. L. Perryman, R. Harris, D. S. Goodsell, and A. J. Olson, "Virtual Screening with AutoDock: Theory and Practice.," *Expert Opin. Drug Discov.*, vol. 5, no. 6, pp. 597–607, Jun. 2010.

[4]   P. Szymański, M. Markowicz, and E. Mikiciuk-Olasik, "Adaptation of high-throughput screening in drug discovery-toxicological screening tests.," *Int. J. Mol. Sci.*, vol. 13, no. 1, pp. 427–52, Jan. 2012.

[5]   J. L. Medina-Franco, K. Martínez-Mayorga, M. A. Giulianotti, R. A. Houghten, and C. Pinilla, "Visualization of the Chemical Space in Drug Discovery," *Curr. Comput. Aided. Drug Des.*, vol. 4, pp. 322–333, 2008.

[6]   G. Sliwoski, S. Kothiwale, J. Meiler, and E. W. Lowe, "Computational methods in drug discovery.," *Pharmacol. Rev.*, vol. 66, no. 1, pp. 334–95, Jan. 2014.

[7]   R. Huey, G. M. Morris, A. J. Olson, and D. S. Goodsell, "A semiempirical free energy force field with charge-based desolvation.," *J. Comput. Chem.*, vol. 28, no. 6, pp. 1145–52, Apr. 2007.

[8]   S. A. Adcock and J. A. McCammon, "Molecular dynamics: survey of methods for simulating the activity of proteins.," *Chem. Rev.*, vol. 106, no. 5, pp. 1589–615, May 2006.

[9]   X.-Y. Meng, H.-X. Zhang, M. Mezei, and M. Cui, "Molecular docking: a powerful approach for structure-based drug discovery.," *Curr. Comput. Aided. Drug Des.*, vol. 7, no. 2, pp. 146–57, Jun. 2011.

[10]   A. Dhanik, J. S. McMurray, and L. E. Kavraki, "DINC: A new AutoDock-based protocol for docking large ligands," *BMC Struct. Biol.*, vol. 13, no. Suppl 1, p. S11, 2013.

[11]   "Directory of computer-aided Drug Design tools," *Swiss Inst. Bioinforma.*, 2014.

[12]   A. P. Norgan, P. K. Coffman, J.-P. A. Kocher, D. J. Katzmann, and C. P. Sosa, "Multilevel Parallelization of AutoDock 4.2.," *J. Cheminform.*, vol. 3, no. 1, p. 12,

Jan. 2011.

[13]   G. M. MORRIS, "AutoDock-Resources," *Scripps Inst.*, 2007.

[14]   G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell, and A. J. Olson, "AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility.," *J. Comput. Chem.*, vol. 30, no. 16, pp. 2785–91, Dec. 2009.

[15]   S. Z. Grinter and X. Zou, "Challenges, applications, and recent advances of protein-ligand docking in structure-based drug design.," *Molecules*, vol. 19, no. 7, pp. 10150–76, Jan. 2014.

[16]   D. Plewczynski, M. Łaźniewski, R. Augustyniak, and K. Ginalski, "Can we trust docking results? Evaluation of seven commonly used programs on PDBbind database.," *J. Comput. Chem.*, vol. 32, no. 4, pp. 742–55, Mar. 2011.

[17]   C. Deligkaris, A. T. Ascone, K. J. Sweeney, and A. J. Q. Greene, "Validation of a computational docking methodology to identify the non-covalent binding site of ligands to DNA.," *Mol. Biosyst.*, vol. 10, no. 8, pp. 2106–25, Aug. 2014.

[18]   M. L. Teodoro, G. N. Phillips, and L. E. Kavraki, "Molecular docking: a problem with thousands of degrees of freedom," in *Proceedings 2001 ICRA. IEEE International Conference on Robotics and Automation (Cat. No.01CH37164)*, 2002, vol. 1, pp. 960–965.

[19]   N. Plattner and F. Noé, "Protein conformational plasticity and complex ligand-binding kinetics explored by atomistic simulations and Markov models.," *Nat. Commun.*, vol. 6, p. 7653, Jan. 2015.

[20]   M. Totrov and R. Abagyan, "Flexible ligand docking to multiple receptor conformations: a practical alternative.," *Curr. Opin. Struct. Biol.*, vol. 18, no. 2, pp. 178–84, Apr. 2008.

[21]   O. Trott and A. J. Olson, "AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading.," *J. Comput. Chem.*, vol. 31, no. 2, pp. 455–61, Jan. 2010.

[22]   S. R. Ellingson, J. C. Smith, and J. Baudry, "VinaMPI: facilitating multiple receptor high-throughput virtual docking on high-performance computers.," *J. Comput. Chem.*, vol. 34, no. 25, pp. 2212–21, Sep. 2013.

[23]   H. I. Januar, A. S. Dewi, E. Marraskuranto, and T. Wikanta, "In silico study of fucoxanthin as a tumor cytotoxic agent.," *J. Pharm. Bioallied Sci.*, vol. 4, no. 1, pp. 56–9, Jan. 2012.

[24]   J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, and K. Schulten, "Scalable molecular dynamics with NAMD.," *J. Comput. Chem.*, vol. 26, no. 16, pp. 1781–802, Dec. 2005.

[25]   "Cancer Statistics - National Cancer Institute." [Online]. Available: http://www.cancer.gov/about-cancer/what-is-cancer/statistics. [Accessed: 03-Apr-2016].

[26]   S. Desmond-Hellmann, "The Cost Of Creating A New Drug Now $5 Billion, Pushing Big Pharma To Change Title," *newdrugapprovals.org*. [Online]. Available: https://newdrugapprovals.org/2013/09/01/the-cost-of-creating-a-new-drug-now-5-billion-pushing-big-pharma-to-change/.

[27]   M. Siddiqui and S. V. Rajkumar, "The high cost of cancer drugs and what we can do about it.," *Mayo Clin. Proc.*, vol. 87, no. 10, pp. 935–43, Oct. 2012.

[28]   C. for D. E. and Research, "Information for Consumers (Drugs) - The FDA's Drug Review Process: Ensuring Drugs Are Safe and Effective." Center for Drug Evaluation and Research.

[29]   "Novel New Drugs 2014 Summary," *U.S. Food Drug Adm. Cent. Drug Eval. Res.*, 2014.

[30]   "Targeted Cancer Therapies Fact Sheet - National Cancer Institute." [Online]. Available: http://www.cancer.gov/about-cancer/treatment/types/targeted-therapies/targeted-therapies-fact-sheet. [Accessed: 03-Apr-2016].

[31]   "Costs mount for targeted cancer therapies | Science Life on WordPress.com." [Online]. Available: http://sciencelife.uchospitals.edu/2015/05/19/costs-mount-for-targeted-cancer-therapies/. [Accessed: 03-Apr-2016].

[32]   M. M. Gottesman, "Mechanisms of cancer drug resistance.," *Annu. Rev. Med.*, vol. 53, pp. 615–27, Jan. 2002.

[33]   L. M. Ellis and D. J. Hicklin, "Resistance to Targeted Therapies: Refining Anticancer Therapy in the Era of Molecular Oncology.," *Clin. Cancer Res.*, vol. 15, no. 24, pp. 7471–7478, Dec. 2009.

[34]   L. Xie, L. Xie, and P. E. Bourne, "Structure-based systems biology for analyzing off-target binding.," *Curr. Opin. Struct. Biol.*, vol. 21, no. 2, pp. 189–99, Apr. 2011.

[35]   J. Mestres, E. Gregori-Puigjané, S. Valverde, and R. V Solé, "Data completeness--the Achilles heel of drug-target networks.," *Nat. Biotechnol.*, vol. 26, no. 9, pp. 983–4, Sep. 2008.

[36]   K. A. O'Connor and B. L. Roth, "Finding new tricks for old drugs: an efficient

route for public-sector drug discovery.," *Nat. Rev. Drug Discov.*, vol. 4, no. 12, pp. 1005–14, Dec. 2005.

[37]   "HOW FAR WE'VE COME: A DECADE IN REVIEW," *cancerprogress.net*, 2016. [Online]. Available: http://www.cancerprogress.net/cca/how-far-weve-come-decade-review.

[38]   V. Ágoston, P. Csermeyl, and S. Pongor, "Multiple, weak hits confuse complex systems: A transcriptional regulatory network as an example," *Phys. Rev*, vol. 71, 2005.

[39]   U. Debnath, "Structure-Based Drug Design (SBDD) : Positive Approach For Drug Discovery," *Netaji Subhas Chandra Bose Institute of Pharmacy*, 2012. [Online]. Available: http://www.nscbip.org/?p=459.

[40]   M. K. Gilson, T. Liu, M. Baitaluk, G. Nicola, L. Hwang, and J. Chong, "BindingDB in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology.," *Nucleic Acids Res.*, vol. 44, no. D1, pp. D1045–53, Jan. 2016.

[41]   E. Krovat, T. Steindl, and T. Langer, "Recent Advances in Docking and Scoring," *Curr. Comput. Aided-Drug Des.*, vol. 1, no. 1, pp. 93–102, Jan. 2005.

[42]   P. A. Holt, J. B. Chaires, and J. O. Trent, "Molecular docking of intercalators and groove-binders to nucleic acids using Autodock and Surflex.," *J. Chem. Inf. Model.*, vol. 48, no. 8, pp. 1602–15, Aug. 2008.

[43]   S. Nag, J. Qin, K. S. Srivenugopal, M. Wang, and R. Zhang, "The MDM2-p53 pathway revisited.," *J. Biomed. Res.*, vol. 27, no. 4, pp. 254–71, Jul. 2013.

[44]   V. Lounnas, T. Ritschel, J. Kelder, R. McGuire, R. P. Bywater, and N. Foloppe, "Current progress in Structure-Based Rational Drug Design marks a new mindset in drug discovery.," *Comput. Struct. Biotechnol. J.*, vol. 5, no. 6, p. e201302011, Jan. 2013.

[45]   A. Cherkasov, E. N. Muratov, D. Fourches, A. Varnek, I. I. Baskin, M. Cronin, J. Dearden, P. Gramatica, Y. C. Martin, R. Todeschini, V. Consonni, V. E. Kuz'min, R. Cramer, R. Benigni, C. Yang, J. Rathman, L. Terfloth, J. Gasteiger, A. Richard, and A. Tropsha, "QSAR modeling: where have you been? Where are you going to?," *J. Med. Chem.*, vol. 57, no. 12, pp. 4977–5010, Jun. 2014.

[46]   H. Alonso, A. A. Bliznyuk, and J. E. Gready, "Combining Docking and Molecular Dynamic Simulations in Drug Design," *Wiley Intersci.*, vol. 26, no. 5, pp. 531–568, 2006.

[47]   N. Bharatham, K. Bharatham, A. A. Shelat, and D. Bashford, "Ligand binding mode prediction by docking: mdm2/mdmx inhibitors as a case study.," *J. Chem.*

Inf. Model., vol. 54, no. 2, pp. 648–59, Feb. 2014.

[48] K. H. Vousden and C. Prives, "Blinded by the Light: The Growing Complexity of p53.," *Cell*, vol. 137, no. 3, pp. 413–31, May 2009.

[49] Zhang and H. Wang, "MDM2 oncogene as a novel target for human cancer therapy.," *Curr. Pharm. Des.*, vol. 6, no. 4, pp. 393–416, Mar. 2000.

[50] J. Momand, D. Jung, S. Wilczynski, and J. Niland, "The MDM2 gene amplification database.," *Nucleic Acids Res.*, vol. 26, no. 15, pp. 3453–9, Aug. 1998.

[51] S. Wang, Y. Zhao, D. Bernard, and A. Aguilar, "Protein-Protein Interactions," vol. 8, pp. 57–80, 2012.

[52] X. Wu, J. H. Bayle, D. Olson, and A. J. Levine, "The p53-mdm-2 autoregulatory feedback loop.," *Genes Dev.*, vol. 7, no. 7A, pp. 1126–32, Jul. 1993.

[53] M. Wade, Y.-C. Li, and G. M. Wahl, "MDM2, MDMX and p53 in oncogenesis and cancer therapy.," *Nat. Rev. Cancer*, vol. 13, no. 2, pp. 83–96, Feb. 2013.

[54] K. Onel and C. Cordon-Cardo, "MDM2 and prognosis.," *Mol. Cancer Res.*, vol. 2, no. 1, pp. 1–8, Jan. 2004.

[55] U. M. Moll and O. Petrenko, "The MDM2-p53 Interaction," *Mol. Cancer Res.*, vol. 1, no. 14, pp. 1001–1008, Dec. 2003.

[56] J. D. Oliner, J. A. Pietenpol, S. Thiagalingam, J. Gyuris, K. W. Kinzler, and B. Vogelstein, "Oncoprotein MDM2 conceals the activation domain of tumour suppressor p53.," *Nature*, vol. 362, no. 6423, pp. 857–60, Apr. 1993.

[57] A. J. Levine and M. Oren, "The first 30 years of p53: growing ever more complex," *Nat. Rev. Cancer*, vol. 9, no. 10, pp. 749–758, Oct. 2009.

[58] P. H. Kussie, S. Gorina, V. Marechal, B. Elenbaas, J. Moreau, A. J. Levine, and N. P. Pavletich, "Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain.," *Science*, vol. 274, no. 5289, pp. 948–53, Nov. 1996.

[59] K. Khoury and A. Dömling, "P53 mdm2 inhibitors.," *Curr. Pharm. Des.*, vol. 18, no. 30, pp. 4668–78, Jan. 2012.

[60] S. Wang, Y. Zhao, D. Bernard, A. Aguilar, and S. Kumar, *Protein-Protein Interactions*, vol. 8. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012.

[61] L. T. Vassilev, B. T. Vu, B. Graves, D. Carvajal, F. Podlaski, Z. Filipovic, N.

Kong, U. Kammlott, C. Lukacs, C. Klein, N. Fotouhi, and E. A. Liu, "In vivo activation of the p53 pathway by small-molecule antagonists of MDM2.," *Science*, vol. 303, no. 5659, pp. 844–8, Feb. 2004.

[62] H. Wang, D. Yu, S. Agrawal, and R. Zhang, "Experimental therapy of human prostate cancer by inhibiting MDM2 expression with novel mixed-backbone antisense oligonucleotides: In vitro and in vivo activities and mechanisms," *Prostate*, vol. 54, no. 3, pp. 194–205, Feb. 2003.

[63] J.-J. Qin, W. Wang, S. Voruganti, H. Wang, W.-D. Zhang, and R. Zhang, "Identification of a new class of natural product MDM2 inhibitor: In vitro and in vivo anti-breast cancer activities and target validation.," *Oncotarget*, vol. 6, no. 5, pp. 2623–40, Feb. 2015.

[64] H. Li, H. Zhang, M. Zheng, J. Luo, L. Kang, X. Liu, X. Wang, and H. Jiang, "An effective docking strategy for virtual screening based on multi-objective optimization algorithm.," *BMC Bioinformatics*, vol. 10, no. 1, p. 58, Jan. 2009.

[65] G. M. Whitesides and V. M. Krishnamurthy, "Designing ligands to bind proteins.," *Q. Rev. Biophys.*, vol. 38, no. 4, pp. 385–95, Nov. 2005.

[66] L. Ferreira, R. dos Santos, G. Oliva, and A. Andricopulo, "Molecular Docking and Structure-Based Drug Design Strategies," *Molecules*, vol. 20, no. 7, pp. 13384–13421, Jul. 2015.

[67] V. Tran, M. M. Delage, and A. Bul�on, "A systematic docking approach. Application to the ?-cyclodextrin/phenyl-ethanol complex," *J. Incl. Phenom. Mol. Recognit. Chem.*, vol. 14, no. 3–4, pp. 271–284, 1992.

[68] R. T. Kroemer, "Structure-Based Drug Design: Docking and Scoring," *Curr. Protein Pept. Sci.*, vol. 8, pp. 312–328, 2007.

[69] *In Silico Drug Discovery and Design: Theory, Methods, Challenges, and Applications*. CRC Press, 2015.

[70] B. Qian, S. Raman, R. Das, P. Bradley, A. J. McCoy, R. J. Read, and D. Baker, "High-resolution structure prediction and the crystallographic phase problem.," *Nature*, vol. 450, no. 7167, pp. 259–64, Nov. 2007.

[71] G. M. M. D. S. G. R. S. H. R. H. W. E. H. R. K. B. A. J. O. A. J. Autodock, "Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function."

[72] E. Atilgan and J. Hu, "Improving Protein Docking Using Sustainable Genetic Algorithms," *Int. J. Comput. Inf. Syst. Ind. Manag. Appl.*, vol. 3, pp. 248–255, 2011.

[73] C. S. de Magalhães, H. J. C. Barbosa, and L. E. Dardenne, "A genetic algorithm for the ligand-protein docking problem," *Genet. Mol. Biol.*, vol. 27, no. 4, pp. 605–610, 2004.

[74] J. Wang, Y. Deng, and B. Roux, "Absolute binding free energy calculations using molecular dynamics simulations with restraining potentials.," *Biophys. J.*, vol. 91, no. 8, pp. 2798–814, Oct. 2006.

[75] S. F. Sousa, A. J. M. Ribeiro, J. T. S. Coimbra, R. P. P. Neves, S. A. Martins, N. S. H. N. Moorthy, P. A. Fernandes, and M. J. Ramos, "Protein-Ligand Docking in the New Millennium – A Retrospective of 10 Years in the Field," *Curr. Med. Chem.*, vol. 20, pp. 2296–2234, 2013.

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

[77] C. SOTRIFFER and G. KLEBE, *Docking and Scoring Functions Virtual Screening*. 2003.

[78] H. Gohlke, M. Hendlich, and G. Klebe, "Knowledge-based Scoring Function to Predict Protein-Ligand Interactions," *J. Mol. Biol*, no. 295, pp. 337–356, 2000.

[79] G. M. Morris, D. S. Goodsell, M. E. Pique, W. "Lindy" Lindstrom, R. Huey, Stefano, and Forli, "AutoDock 4.2 User Guide." pp. 1–66, 2009.

[80] F. Spyrakis, A. BidonChanal, X. Barril, and F. J. Luque, "Protein flexibility and ligand recognition: challenges for molecular modeling.," *Curr. Top. Med. Chem.*, vol. 11, no. 2, pp. 192–210, Jan. 2011.

[81] R. M. V. Abreu, H. J. C. Froufe, M.-J. R. P. Queiroz, and I. C. F. R. Ferreira, "Selective Flexibility of Side-Chain Residues Improves VEGFR-2 Docking Score using AutoDock Vina," *Chem. Biol. Drug Des.*, vol. 79, no. 4, pp. 530–534, Apr. 2012.

[82] B. Ma, M. Shatsky, H. J. Wolfson, and R. Nussinov, "Multiple diverse ligands binding at a single protein site: a matter of pre-existing populations.," *Protein Sci.*, vol. 11, no. 2, pp. 184–97, Feb. 2002.

[83] D. A. Antunes, D. Devaurs, and L. E. Kavraki, "Understanding the Challenges of Protein Flexibility in Drug Design," *Expert Opin. Drug Discov.*, vol. 10, no. 12, pp. 1–17, 2015.

[84] J. Meller, "Molecular Dynamics," *Encycl. LIFE Sci.*, 2001.

[85] T. Isgro, J. Phillips, M. Sotomayor, and E. Villa, "NAMD TUTORIAL." Beckman Institute Computational Biophysics Workshop, pp. 31–37, 2007.

[86] F. E. Herrera, "Computational approaches to the investigation of proteins involved in Parkinson's Disease," 2008.

[87] Y. Zhao, D. Bernard, and S. Wang, "Small Molecule Inhibitors of MDM2-p53 and MDMX-p53 Interactions as New Cancer Therapeutics," *Biodiscovery*, no. 8, p. 4, 2013.

[88] C. Klein and L. T. Vassilev, "Targeting the p53-MDM2 interaction to treat cancer.," *Br. J. Cancer*, vol. 91, no. 8, pp. 1415–9, Oct. 2004.

[89] A. C. Joerger and A. R. Fersht, "The Tumor Suppressor p53: From Structures to Drug Discovery," *Cold Spring Harb. Perspect. Biol.*, vol. 2, no. 6, pp. a000919–a000919, Feb. 2010.

[90] H. Wang, L. Nan, D. Yu, S. Agrawal, and R. Zhang, "Antisense anti-MDM2 oligonucleotides as a novel therapeutic approach to human breast cancer: in vitro and in vivo activities and mechanisms.," *Clin. Cancer Res.*, vol. 7, no. 11, pp. 3613–24, Nov. 2001.

[91] W. A. Warner, R. Sanchez, A. Dawoodian, E. Li, and J. Momand, "Identification of FDA-approved Drugs that Computationally Bind to MDM2," *Chem. Biol. Drug Des.*, vol. 80, no. 4, pp. 631–637, Oct. 2012.

[92] D. R. Houston, L.-H. Yen, S. Pettit, and M. D. Walkinshaw, "Structure- and ligand-based virtual screening identifies new scaffolds for inhibitors of the oncoprotein MDM2.," *PLoS One*, vol. 10, no. 4, p. e0121424, Jan. 2015.

[93] Z. Huang, *Drug Discovery Research: New Frontiers in the Post-Genomic Era*. John Wiley & Sons, 2007.

[94] C. A. Sotriffer, "Accounting for induced-fit effects in docking: what is possible and what is not?," *Curr. Top. Med. Chem.*, vol. 11, no. 2, pp. 179–91, Jan. 2011.

[95] E. Perola, W. P. Walters, and P. S. Charifson, "A Detailed Comparison of Current Docking and Scoring Methods on Systems of Pharmaceutical Relevance," *PROTEINS Struct. Funct. Bioinforma.*, vol. 56, pp. 235–249, 2004.

[96] Y. Arakawa, K. Ozaki, Y. Okawa, and H. Yamada, "Three missense mutations of DNA topoisomerase I in highly camptothecin-resistant colon cancer cell sublines.," *Oncol. Rep.*, vol. 30, no. 3, pp. 1053–8, Sep. 2013.

[97] S. Sirikantaramas, M. Yamazaki, and K. Saito, "Mutations in topoisomerase I as a self-resistance mechanism coevolved with the production of the anticancer alkaloid camptothecin in plants.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 105, no. 18, pp. 6782–6, May 2008.

[98] S. Singh, T. Das, M. Awasthi, V. P. Pandey, B. Pandey, and U. N. Dwivedi, "DNA topoisomerase-directed anticancerous alkaloids: ADMET-based screening, molecular docking, and dynamics simulation.," *Biotechnol. Appl. Biochem.*, vol. 63, no. 1, pp. 125–37, Feb. 2016.

[99] S.-Y. Huang and X. Zou, "Efficient molecular docking of NMR structures: Application to HIV-1 protease," *Protein Sci.*, vol. 16, no. 1, pp. 43–51, Nov. 2006.

[100] Z. Bikadi and E. Hazai, "Application of the PM6 semi-empirical method to modeling proteins enhances docking accuracy of AutoDock.," *J. Cheminform.*, vol. 1, no. 1, p. 15, Jan. 2009.

[101] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin, "UCSF Chimera--a visualization system for exploratory research and analysis.," *J. Comput. Chem.*, vol. 25, no. 13, pp. 1605–12, Oct. 2004.

[102] M. W. Chang, C. Ayeni, S. Breuer, and B. E. Torbett, "Virtual Screening for HIV Protease Inhibitors: A Comparison of AutoDock 4 and Vina," *PLoS One*, vol. 5, no. 8, p. e11955, Aug. 2010.

[103] "Docking (molecular)," *Wikipedia*. [Online]. Available: https://en.wikipedia.org/wiki/Docking_(molecular).

[104] "Innovation Opening New Fronts in the War on Cancer," *Capital Group*. [Online]. Available: https://www.thecapitalgroup.com/us/insights/investment-insights/war-on-cancer.html.

[105] G. Robb, "Computer-aided drug design – an integral part of the discovery process," *AstraZeneca*, 2013. [Online]. Available: http://www.labtalk.astrazeneca.com/oncology/computer-aided-drug-design-an-integral-part-of-the-discovery-process/.

[106] "Primary information of p53 gene," *bioinformatics.org*. [Online]. Available: http://www.bioinformatics.org/p53/introduction.html.

[107] A. Y. T.-B. V. A. H. V.-C. and C. M.-B. Alejandra Hernández-Santoyo, *Protein Engineering - Technology and Application*. InTech, 2013.

[108] "Operating System - Multi-Threading," *tutorialspoint.com*. [Online]. Available: http://www.tutorialspoint.com/operating_system/os_multi_threading.htm.

Appendix A

Sample of Input Files For Docking Experiments

Ligand File

```
REMARK  0 active torsions:
REMARK  status: ('A' for Active; 'I' for Inactive)
REMARK       I     between atoms: N1_1   and   C10_23
REMARK       I     between atoms: C5_2   and   C51_4
REMARK       I     between atoms: C4_5   and   C41_7
REMARK       I     between atoms: C2_9   and   C21_10
REMARK       I     between atoms: C10_23  and   N10_25
REMARK       I     between atoms: N12_28  and   C15_49
REMARK       I     between atoms: C24_33  and   C30_34
REMARK       I     between atoms: C26_38  and   O27_39
REMARK       I     between atoms: O27_39  and   C28_40
REMARK       I     between atoms: C16_43  and   C17_44
REMARK       I     between atoms: C16_43  and   C15_49
REMARK       I     between atoms: C17_44  and   S17_45
ROOT
ATOM      1  N1  1F0 A         12.567   8.335  35.668  0.00
0.00    -0.226 N
ATOM      2  C5  1F0 A         11.695   9.079  34.712  0.00
0.00     0.096 C
ATOM      3  C57 1F0 A         10.610   8.150  34.083  0.00
0.00     0.041 C
ATOM      4  C51 1F0 A         12.522   9.750  33.579  0.00
0.00    -0.021 A
ATOM      5  C4  1F0 A         11.040  10.172  35.653  0.00
0.00     0.111 C
ATOM      6  C47 1F0 A          9.767   9.671  36.400  0.00
0.00     0.045 C
ATOM      7  C41 1F0 A         10.778  11.519  34.951  0.00
0.00    -0.018 A
ATOM      8  N3  1F0 A         12.042  10.275  36.747  0.00
0.00    -0.255 NA
ATOM      9  C2  1F0 A         12.808   9.256  36.742  0.00
0.00     0.134 C
ATOM     10  C21 1F0 A         13.874   9.070  37.782  0.00
0.00     0.045 A
ATOM     11  C56 1F0 A         13.883  10.041  33.743  0.00
0.00     0.007 A
ATOM     12  C55 1F0 A         14.610  10.691  32.750  0.00
0.00     0.021 A
ATOM     13  C54 1F0 A         13.989  11.076  31.567  0.00
```

```
            0.00      0.041 A
ATOM     14 CL5  1F0 A        14.895  11.887  30.337  0.00
  0.00     -0.084 Cl
ATOM     15  C53 1F0 A        12.635  10.815  31.385  0.00
  0.00      0.021 A
ATOM     16  C52 1F0 A        11.910  10.159  32.380  0.00
  0.00      0.007 A
ATOM     17  C42 1F0 A         9.556  11.785  34.315  0.00
  0.00      0.008 A
ATOM     18  C43 1F0 A         9.354  12.981  33.629  0.00
  0.00      0.021 A
ATOM     19  C44 1F0 A        10.379  13.922  33.552  0.00
  0.00      0.041 A
ATOM     20 CL4  1F0 A        10.137  15.393  32.673  0.00
  0.00     -0.084 Cl
ATOM     21  C45 1F0 A        11.600  13.672  34.172  0.00
  0.00      0.021 A
ATOM     22  C46 1F0 A        11.794  12.482  34.871  0.00
  0.00      0.008 A
ATOM     23  C10 1F0 A        12.978   7.059  35.425  0.00
  0.00      0.310 C
ATOM     24  O10 1F0 A        13.183   6.667  34.294  0.00
  0.00     -0.253 OA
ATOM     25  N10 1F0 A        13.169   6.220  36.487  0.00
  0.00     -0.275 N
ATOM     26  C14 1F0 A        13.880   4.941  36.289  0.00
  0.00      0.160 C
ATOM     27  C13 1F0 A        14.711   4.566  37.526  0.00
  0.00      0.148 C
ATOM     28  N12 1F0 A        13.928   4.572  38.814  0.00
  0.00      0.249 N
ATOM     29  C12 1F0 A        13.249   5.917  38.965  0.00
  0.00      0.148 C
ATOM     30  C11 1F0 A        12.377   6.267  37.741  0.00
  0.00      0.160 C
ATOM     31  C22 1F0 A        13.559   9.415  39.098  0.00
  0.00      0.015 A
ATOM     32  C23 1F0 A        14.477   9.252  40.134  0.00
  0.00      0.005 A
ATOM     33  C24 1F0 A        15.758   8.750  39.887  0.00
  0.00     -0.040 A
ATOM     34  C30 1F0 A        16.797   8.569  41.017  0.00
  0.00     -0.013 C
ATOM     35  C33 1F0 A        17.792   9.754  40.991  0.00
  0.00      0.016 C
ATOM     36  C31 1F0 A        17.589   7.246  40.846  0.00
  0.00      0.016 C
ATOM     37  C25 1F0 A        16.098   8.462  38.559  0.00
  0.00      0.038 A
ATOM     38  C26 1F0 A        15.172   8.596  37.515  0.00
  0.00      0.086 A
ATOM     39  O27 1F0 A        15.527   8.251  36.223  0.00
```

```
0.00     -0.346 OA
ATOM     40  C28 1F0 A        16.855    7.848   35.817  0.00
0.00      0.164 C
ATOM     41  C29 1F0 A        16.837    7.543   34.313  0.00
0.00      0.035 C
ATOM     42  C32 1F0 A        16.163    8.516   42.435  0.00
0.00      0.016 C
ATOM     43  C16 1F0 A        14.280    4.213   41.357  0.00
0.00      0.176 C
ATOM     44  C17 1F0 A        15.260    3.679   42.435  0.00
0.00      0.530 C
ATOM     45  S17 1F0 A        14.568    4.133   43.989  0.00
0.00     -0.280 S
ATOM     46  C19 1F0 A        15.491    3.473   45.350  0.00
0.00      0.666 C
ATOM     47  O17 1F0 A        13.278    3.527   44.023  0.00
0.00     -0.525 OA
ATOM     48  O18 1F0 A        14.693    5.552   44.050  0.00
0.00     -0.525 OA
ATOM     49  C15 1F0 A        14.886    4.237   39.941  0.00
0.00      0.141 C
ATOM     50 2HN1 1F0 A        13.190    3.846   38.764  0.00
0.00      0.200 HD
ENDROOT
TORSDOF 10
```

Flexible Residue File

```
BEGIN_RES LEU A  50
REMARK  2 active torsions:
REMARK  status: ('A' for Active; 'I' for Inactive)
REMARK    1  A    between atoms: CA   and  CB
REMARK    2  A    between atoms: CB   and  CG
ROOT
ATOM      1  CA  LEU A  50       9.533    9.929   29.079  1.00
10.27     0.177 C
ENDROOT
BRANCH   1   2
ATOM      2  CB  LEU A  50       8.870   10.875   30.081  1.00
10.02     0.038 C
BRANCH   2   3
ATOM      3  CG  LEU A  50       8.727   12.295   29.523  1.00
13.36    -0.020 C
ATOM      4  CD2 LEU A  50      10.076   12.999   29.587  1.00
11.01     0.009 C
ATOM      5  CD1 LEU A  50       7.673   13.063   30.296  1.00
11.28     0.009 C
ENDBRANCH   2    3
ENDBRANCH   1    2
END_RES LEU A  50
BEGIN_RES LEU A  53
```

```
REMARK  1 active torsions:
REMARK  status: ('A' for Active; 'I' for Inactive)
REMARK    3  A    between atoms: CA   and  CB
REMARK       I    between atoms: CB   and  CG
ROOT
ATOM      6  CA  LEU A  53      13.763   7.673  27.601  1.00
10.12    0.177 C
ENDROOT
BRANCH   6   7
ATOM      7  CB  LEU A  53      13.561   9.190  27.484  1.00
10.09    0.038 C
ATOM      8  CG  LEU A  53      14.846  10.013  27.342  1.00
14.17   -0.020 C
ATOM      9  CD2 LEU A  53      14.500  11.493  27.234  1.00
15.42    0.009 C
ATOM     10  CD1 LEU A  53      15.623   9.551  26.114  1.00
13.92    0.009 C
ENDBRANCH   6   7
END_RES LEU A  53
BEGIN_RES ILE A  57
REMARK  1 active torsions:
REMARK  status: ('A' for Active; 'I' for Inactive)
REMARK       I    between atoms: CA   and  CB
REMARK    4  A    between atoms: CB   and  CG1
ROOT
ATOM     11  CB  ILE A  57      18.694   7.357  31.036  1.00
8.44     0.013 C
ENDROOT
BRANCH  11  12
ATOM     12  CG1 ILE A  57      18.441   8.505  30.053  1.00
10.51    0.002 C
ATOM     13  CD1 ILE A  57      17.724   9.691  30.650  1.00
10.63    0.005 C
ENDBRANCH  11  12
END_RES ILE A  57
BEGIN_RES HIS A  69
REMARK  2 active torsions:
REMARK  status: ('A' for Active; 'I' for Inactive)
REMARK    5  A    between atoms: CA   and  CB
REMARK    6  A    between atoms: CB   and  CG
ROOT
ATOM     14  CA  HIS A  69      19.751  12.132  39.250  1.00
10.97    0.182 C
ENDROOT
BRANCH  14  15
ATOM     15  CB  HIS A  69      19.087  13.136  40.206  1.00
13.11    0.093 C
BRANCH  15  16
ATOM     16  CG  HIS A  69      20.025  14.161  40.764  1.00
13.45    0.030 A
ATOM     17  CD2 HIS A  69      20.759  14.170  41.902  1.00
17.53    0.143 A
```

98

```
ATOM        18  NE2 HIS A  69        21.452  15.357  41.926  1.00
17.19     -0.254 NA
ATOM        19  CE1 HIS A  69        21.155  16.050  40.842  1.00
18.20      0.207 A
ATOM        20  ND1 HIS A  69        20.296  15.350  40.123  1.00
16.07     -0.353 N
ATOM        21  HD1 HIS A  69        19.900  15.641  39.241  1.00
0.00       0.166 HD
ENDBRANCH  15  16
ENDBRANCH  14  15
END_RES HIS A  69
BEGIN_RES VAL A  89
REMARK  1 active torsions:
REMARK  status: ('A' for Active; 'I' for Inactive)
REMARK     7  A    between atoms: CA   and  CB
ROOT
ATOM        22  CA  VAL A  89        16.027  14.124  35.948  1.00
10.52      0.180 C
ENDROOT
BRANCH  22  23
ATOM        23  CB  VAL A  89        15.913  12.637  36.388  1.00
10.21      0.009 C
ATOM        24  CG1 VAL A  89        16.875  11.774  35.569  1.00
11.26      0.012 C
ATOM        25  CG2 VAL A  89        16.187  12.499  37.875  1.00
10.50      0.012 C
ENDBRANCH  22  23
END_RES VAL A  89
BEGIN_RES HIS A  92
REMARK  2 active torsions:
REMARK  status: ('A' for Active; 'I' for Inactive)
REMARK     8  A    between atoms: CA   and  CB
REMARK     9  A    between atoms: CB   and  CG
ROOT
ATOM        26  CA  HIS A  92        11.157  18.396  34.121  1.00
10.57      0.182 C
ENDROOT
BRANCH  26  27
ATOM        27  CB  HIS A  92         9.916  18.053  34.950  1.00
11.85      0.095 C
BRANCH  27  28
ATOM        28  CG  HIS A  92        10.067  16.802  35.761  1.00
11.71      0.053 A
ATOM        29  ND1 HIS A  92         8.991  16.050  36.181  1.00
13.27     -0.247 NA
ATOM        30  CE1 HIS A  92         9.426  15.005  36.863  1.00
13.40      0.207 A
ATOM        31  NE2 HIS A  92        10.745  15.054  36.903  1.00
14.08     -0.359 N
ATOM        32  HE2 HIS A  92        11.344  14.381  37.360  1.00
0.00       0.166 HD
ATOM        33  CD2 HIS A  92        11.171  16.168  36.223  1.00
```

```
13.21     0.116 A
ENDBRANCH  27  28
ENDBRANCH  26  27
END_RES HIS A  92
BEGIN_RES TYR A  96
REMARK  3 active torsions:
REMARK  status: ('A' for Active; 'I' for Inactive)
REMARK   10  A    between atoms: CA   and  CB
REMARK   11  A    between atoms: CB   and  CG
REMARK   12  A    between atoms: CZ   and  OH
ROOT
ATOM     34  CA  TYR A  96       9.955  18.557  28.206  1.00
9.98      0.180 C
ENDROOT
BRANCH   34  35
ATOM     35  CB  TYR A  96       8.759  19.165  28.947  1.00
10.95     0.073 C
BRANCH   35  36
ATOM     36  CG  TYR A  96       7.758  18.146  29.436  1.00
10.20    -0.056 A
ATOM     37  CD2 TYR A  96       6.561  17.944  28.758  1.00
12.34     0.010 A
ATOM     38  CE2 TYR A  96       5.646  16.998  29.188  1.00
10.72     0.037 A
ATOM     39  CZ  TYR A  96       5.924  16.244  30.307  1.00
12.51     0.065 A
ATOM     40  CE1 TYR A  96       7.105  16.426  31.000  1.00
12.43     0.037 A
ATOM     41  CD1 TYR A  96       8.014  17.373  30.562  1.00
10.18     0.010 A
BRANCH   39  42
ATOM     42  OH  TYR A  96       5.018  15.302  30.733  1.00
14.86    -0.361 OA
ATOM     43  HH  TYR A  96       4.184  15.730  30.939  1.00
0.00      0.217 HD
ENDBRANCH  39  42
ENDBRANCH  35  36
ENDBRANCH  34  35
END_RES TYR A  96
```

Grid Parameter File

```
npts 60 62 62                          # num.grid points in xyz
gridfld 4ipf_rigid.maps.fld            # grid_data_file
spacing 0.375                          # spacing(A)
receptor_types A C NA OA N SA HD       # receptor atom types
ligand_types A C Cl NA OA N S HD       # ligand atom types
receptor 4ipf_rigid.pdbqt             # macromolecule
gridcenter 13.988 9.562 35.855         # xyz-coordinates or auto
smooth 0.5                             # store minimum energy w/in
rad(A)
```

```
map 4ipf_rigid.A.map                    # atom-specific affinity map
map 4ipf_rigid.C.map                    # atom-specific affinity map
map 4ipf_rigid.Cl.map                   # atom-specific affinity map
map 4ipf_rigid.NA.map                   # atom-specific affinity map
map 4ipf_rigid.OA.map                   # atom-specific affinity map
map 4ipf_rigid.N.map                    # atom-specific affinity map
map 4ipf_rigid.S.map                    # atom-specific affinity map
map 4ipf_rigid.HD.map                   # atom-specific affinity map
elecmap 4ipf_rigid.e.map                # electrostatic potential
map
dsolvmap 4ipf_rigid.d.map                # desolvation potential
map
dielectric -0.1465                      # <0, AD4 distance-
dep.diel;>0, constant
```

Docking Parameter File

```
autodock_parameter_version 4.2         # used by autodock to
validate parameter set
outlev 1                               # diagnostic output level
intelec                                # calculate internal
electrostatics
seed pid time                          # seeds for random generator
ligand_types A C Cl NA OA N S HD       # atoms types in ligand
fld 4ipf_rigid.maps.fld                # grid_data_file
map 4ipf_rigid.A.map                   # atom-specific affinity map
map 4ipf_rigid.C.map                   # atom-specific affinity map
map 4ipf_rigid.Cl.map                  # atom-specific affinity map
map 4ipf_rigid.NA.map                  # atom-specific affinity map
map 4ipf_rigid.OA.map                  # atom-specific affinity map
map 4ipf_rigid.N.map                   # atom-specific affinity map
map 4ipf_rigid.S.map                   # atom-specific affinity map
map 4ipf_rigid.HD.map                  # atom-specific affinity map
elecmap 4ipf_rigid.e.map               # electrostatics map
desolvmap 4ipf_rigid.d.map             # desolvation map
move 4ipf_lig.pdbqt                    # small molecule
 flexres 4ipf_flex.pdbqt                  # file containing
flexible residues
about 13.5638 8.4251 37.0794           # small molecule center
tran0 random                           # initial coordinates/A or
random
quaternion0 random                     # initial orientation
dihe0 random                           # initial dihedrals
(relative) or random
torsdof 10                              # torsional degrees of
freedom
rmstol 2.0                             # cluster_tolerance/A
extnrg 1000.0                          # external grid energy
e0max 0.0 10000                        # max initial energy; max
number of retries
ga_pop_size 150                        # number of individuals in
```

```
population
ga_num_evals 20000000                    # maximum number of energy
evaluations
ga_num_generations 27000                 # maximum number of
generations
ga_elitism 1                             # number of top individuals
to survive to next generation
ga_mutation_rate 0.02                    # rate of gene mutation
ga_crossover_rate 0.8                    # rate of crossover
ga_window_size 10                        #
ga_cauchy_alpha 0.0                      # Alpha parameter of Cauchy
distribution
ga_cauchy_beta 1.0                       # Beta parameter Cauchy
distribution
set_ga                                   # set the above parameters
for GA or LGA
sw_max_its 300                           # iterations of Solis & Wets
local search
sw_max_succ 4                            # consecutive successes
before changing rho
sw_max_fail 4                            # consecutive failures
before changing rho
sw_rho 1.0                               # size of local search space
to sample
sw_lb_rho 0.01                           # lower bound on rho
ls_search_freq 0.06                      # probability of performing
local search on individual
set_psw1                                 # set the above pseudo-Solis
& Wets parameters
unbound_model bound                      # state of unbound ligand
ga_run 100                                # do this many hybrid GA-LS
runs
analysis                                 # perform a ranked cluster
analysis
```

Vina Configuration File

```
receptor = 4ipf_rigid.pdbqt
flex = 4ipf_flex.pdbqt
ligand = 4ipf_lig.pdbqt

out = out.4ipf.pdbqt

center_x = 13.988
center_y = 9.562
center_z = 35.855

size_x = 27
size_y = 27
size_z = 27
```

```
exhaustiveness = 12


Ligand Input File

REMARK   10 active torsions:
REMARK   status: ('A' for Active; 'I' for Inactive)
REMARK      I    between atoms: N1_1  and  C10_23
REMARK    1 A    between atoms: C5_2  and  C51_4
REMARK    2 A    between atoms: C4_5  and  C41_7
REMARK    3 A    between atoms: C2_9  and  C21_10
REMARK      I    between atoms: C10_23  and  N10_25
REMARK    4 A    between atoms: N12_28  and  C15_49
REMARK    5 A    between atoms: C24_33  and  C30_34
REMARK    6 A    between atoms: C26_38  and  O27_39
REMARK    7 A    between atoms: O27_39  and  C28_40
REMARK    8 A    between atoms: C16_43  and  C17_44
REMARK    9 A    between atoms: C16_43  and  C15_49
REMARK   10 A    between atoms: C17_44  and  S17_45
ROOT
ATOM      1  N1  1F0 A       12.567   8.335  35.668  0.00
0.00    -0.226 N
ATOM      2  C5  1F0 A       11.695   9.079  34.712  0.00
0.00     0.096 C
ATOM      3  C57 1F0 A       10.610   8.150  34.083  0.00
0.00     0.041 C
ATOM      4  C4  1F0 A       11.040  10.172  35.653  0.00
0.00     0.111 C
ATOM      5  C47 1F0 A        9.767   9.671  36.400  0.00
0.00     0.045 C
ATOM      6  N3  1F0 A       12.042  10.275  36.747  0.00
0.00    -0.255 NA
ATOM      7  C2  1F0 A       12.808   9.256  36.742  0.00
0.00     0.134 C
ATOM      8  C10 1F0 A       12.978   7.059  35.425  0.00
0.00     0.310 C
ATOM      9  O10 1F0 A       13.183   6.667  34.294  0.00
0.00    -0.253 OA
ATOM     10  N10 1F0 A       13.169   6.220  36.487  0.00
0.00    -0.275 N
ATOM     11  C14 1F0 A       13.880   4.941  36.289  0.00
0.00     0.160 C
ATOM     12  C13 1F0 A       14.711   4.566  37.526  0.00
0.00     0.148 C
ATOM     13  N12 1F0 A       13.928   4.572  38.814  0.00
0.00     0.249 N
ATOM     14  C12 1F0 A       13.249   5.917  38.965  0.00
0.00     0.148 C
ATOM     15  C11 1F0 A       12.377   6.267  37.741  0.00
0.00     0.160 C
ATOM     16 2HN1 1F0 A       13.190   3.846  38.764  0.00
```

```
0.00     0.200 HD
ENDROOT
BRANCH   7  17
ATOM      17  C21 1F0 A         13.874    9.070   37.782   0.00
0.00     0.045 A
ATOM      18  C22 1F0 A         13.559    9.415   39.098   0.00
0.00     0.015 A
ATOM      19  C23 1F0 A         14.477    9.252   40.134   0.00
0.00     0.005 A
ATOM      20  C24 1F0 A         15.758    8.750   39.887   0.00
0.00    -0.040 A
ATOM      21  C25 1F0 A         16.098    8.462   38.559   0.00
0.00     0.038 A
ATOM      22  C26 1F0 A         15.172    8.596   37.515   0.00
0.00     0.086 A
BRANCH  20  23
ATOM      23  C30 1F0 A         16.797    8.569   41.017   0.00
0.00    -0.013 C
ATOM      24  C33 1F0 A         17.792    9.754   40.991   0.00
0.00     0.016 C
ATOM      25  C31 1F0 A         17.589    7.246   40.846   0.00
0.00     0.016 C
ATOM      26  C32 1F0 A         16.163    8.516   42.435   0.00
0.00     0.016 C
ENDBRANCH  20  23
BRANCH  22  27
ATOM      27  O27 1F0 A         15.527    8.251   36.223   0.00
0.00    -0.346 OA
BRANCH  27  28
ATOM      28  C28 1F0 A         16.855    7.848   35.817   0.00
0.00     0.164 C
ATOM      29  C29 1F0 A         16.837    7.543   34.313   0.00
0.00     0.035 C
ENDBRANCH  27  28
ENDBRANCH  22  27
ENDBRANCH   7  17
BRANCH   2  30
ATOM      30  C51 1F0 A         12.522    9.750   33.579   0.00
0.00    -0.021 A
ATOM      31  C56 1F0 A         13.883   10.041   33.743   0.00
0.00     0.007 A
ATOM      32  C55 1F0 A         14.610   10.691   32.750   0.00
0.00     0.021 A
ATOM      33  C54 1F0 A         13.989   11.076   31.567   0.00
0.00     0.041 A
ATOM      34  CL5 1F0 A         14.895   11.887   30.337   0.00
0.00    -0.084 Cl
ATOM      35  C53 1F0 A         12.635   10.815   31.385   0.00
0.00     0.021 A
ATOM      36  C52 1F0 A         11.910   10.159   32.380   0.00
0.00     0.007 A
ENDBRANCH   2  30
```

```
BRANCH    4  37
ATOM       37  C41 1F0 A         10.778  11.519  34.951  0.00
0.00    -0.018 A
ATOM       38  C42 1F0 A          9.556  11.785  34.315  0.00
0.00     0.008 A
ATOM       39  C43 1F0 A          9.354  12.981  33.629  0.00
0.00     0.021 A
ATOM       40  C44 1F0 A         10.379  13.922  33.552  0.00
0.00     0.041 A
ATOM       41  CL4 1F0 A         10.137  15.393  32.673  0.00
0.00    -0.084 Cl
ATOM       42  C45 1F0 A         11.600  13.672  34.172  0.00
0.00     0.021 A
ATOM       43  C46 1F0 A         11.794  12.482  34.871  0.00
0.00     0.008 A
ENDBRANCH    4  37
BRANCH   13  44
ATOM       44  C15 1F0 A         14.886   4.237  39.941  0.00
0.00     0.141 C
BRANCH   44  45
ATOM       45  C16 1F0 A         14.280   4.213  41.357  0.00
0.00     0.176 C
BRANCH   45  46
ATOM       46  C17 1F0 A         15.260   3.679  42.435  0.00
0.00     0.530 C
BRANCH   46  47
ATOM       47  S17 1F0 A         14.568   4.133  43.989  0.00
0.00    -0.280 S
ATOM       48  C19 1F0 A         15.491   3.473  45.350  0.00
0.00     0.666 C
ATOM       49  O17 1F0 A         13.278   3.527  44.023  0.00
0.00    -0.525 OA
ATOM       50  O18 1F0 A         14.693   5.552  44.050  0.00
0.00    -0.525 OA
ENDBRANCH   46  47
ENDBRANCH   45  46
ENDBRANCH   44  45
ENDBRANCH   13  44
TORSDOF 10
```

Grid Parameter File

```
npts 60 62 62                          # num.grid points in xyz
gridfld 4ipf_rigid.maps.fld            # grid_data_file
spacing 0.375                          # spacing(A)
receptor_types A C HD N NA OA SA       # receptor atom types
ligand_types A C Cl NA OA N S HD       # ligand atom types
receptor 4ipf_rigid.pdbqt             # macromolecule
gridcenter 13.988 9.562 35.855         # xyz-coordinates or auto
smooth 0.5                             # store minimum energy w/in
rad(A)
```

```
map 4ipf_rigid.A.map               # atom-specific affinity map
map 4ipf_rigid.C.map               # atom-specific affinity map
map 4ipf_rigid.Cl.map              # atom-specific affinity map
map 4ipf_rigid.NA.map              # atom-specific affinity map
map 4ipf_rigid.OA.map              # atom-specific affinity map
map 4ipf_rigid.N.map               # atom-specific affinity map
map 4ipf_rigid.S.map               # atom-specific affinity map
map 4ipf_rigid.HD.map              # atom-specific affinity map
elecmap 4ipf_rigid.e.map           # electrostatic potential
map
dsolvmap 4ipf_rigid.d.map           # desolvation potential
map
dielectric -0.1465                 # <0, AD4 distance-
dep.diel;>0, constant
```

Docking Parameter File

```
autodock_parameter_version 4.2     # used by autodock to
validate parameter set
outlev 1                           # diagnostic output level
intelec                            # calculate internal
electrostatics
seed pid time                      # seeds for random generator
ligand_types A C Cl NA OA N S HD   # atoms types in ligand
fld 4ipf_rigid.maps.fld            # grid_data_file
map 4ipf_rigid.A.map               # atom-specific affinity map
map 4ipf_rigid.C.map               # atom-specific affinity map
map 4ipf_rigid.Cl.map              # atom-specific affinity map
map 4ipf_rigid.NA.map              # atom-specific affinity map
map 4ipf_rigid.OA.map              # atom-specific affinity map
map 4ipf_rigid.N.map               # atom-specific affinity map
map 4ipf_rigid.S.map               # atom-specific affinity map
map 4ipf_rigid.HD.map              # atom-specific affinity map
elecmap 4ipf_rigid.e.map           # electrostatics map
desolvmap 4ipf_rigid.d.map         # desolvation map
move 4ipf_lig.pdbqt                # small molecule
about 13.5638 8.4251 37.0794       # small molecule center
tran0 random                       # initial coordinates/A or
random
quaternion0 random                 # initial orientation
dihe0 random                       # initial dihedrals
(relative) or random
torsdof 10                          # torsional degrees of
freedom
rmstol 2.0                         # cluster_tolerance/A
extnrg 1000.0                      # external grid energy
e0max 0.0 10000                    # max initial energy; max
number of retries
ga_pop_size 150                    # number of individuals in
population
ga_num_evals 20000000               # maximum number of energy
```

```
evaluations
ga_num_generations 27000              # maximum number of
generations
ga_elitism 1                          # number of top individuals
to survive to next generation
ga_mutation_rate 0.02                 # rate of gene mutation
ga_crossover_rate 0.8                 # rate of crossover
ga_window_size 10                     #
ga_cauchy_alpha 0.0                   # Alpha parameter of Cauchy
distribution
ga_cauchy_beta 1.0                    # Beta parameter Cauchy
distribution
set_ga                                # set the above parameters
for GA or LGA
sw_max_its 300                        # iterations of Solis & Wets
local search
sw_max_succ 4                         # consecutive successes
before changing rho
sw_max_fail 4                         # consecutive failures
before changing rho
sw_rho 1.0                            # size of local search space
to sample
sw_lb_rho 0.01                        # lower bound on rho
ls_search_freq 0.06                   # probability of performing
local search on individual
set_psw1                              # set the above pseudo-Solis
& Wets parameters
unbound_model bound                   # state of unbound ligand
ga_run 100                             # do this many hybrid GA-LS
runs
analysis                              # perform a ranked cluster
analysis
```

Vina Configuration File

```
receptor = 4ipf.rec.pdbqt
ligand = 4ipf.lig.pdbqt

out = out.4ipf.pdbqt

center_x = 13.988
center_y = 9.562
center_z = 35.855

size_x = 27
size_y = 27
size_z = 27

exhaustiveness = 12
```

Ligand Input File

```
REMARK   6 active torsions:
REMARK   status: ('A' for Active; 'I' for Inactive)
REMARK       I    between atoms: N1_1   and  C10_23
REMARK    1  A    between atoms: C5_2   and  C51_4
REMARK    2  A    between atoms: C4_5   and  C41_7
REMARK    3  A    between atoms: C2_9   and  C21_10
REMARK       I    between atoms: C10_23  and  N10_25
REMARK       I    between atoms: N12_28  and  C15_49
REMARK    4  A    between atoms: C24_33  and  C30_34
REMARK       I    between atoms: C26_38  and  O27_39
REMARK       I    between atoms: O27_39  and  C28_40
REMARK    5  A    between atoms: C16_43  and  C17_44
REMARK       I    between atoms: C16_43  and  C15_49
REMARK    6  A    between atoms: C17_44  and  S17_45
ROOT
ATOM      1  N1  1F0 A        12.593   8.448  35.680  0.00
0.00    -0.226 N
ATOM      2  C5  1F0 A        11.693   9.088  34.737  0.00
0.00     0.096 C
ATOM      3  C57 1F0 A        10.648   8.109  34.173  0.00
0.00     0.041 C
ATOM      4  C4  1F0 A        11.040  10.162  35.660  0.00
0.00     0.111 C
ATOM      5  C47 1F0 A         9.715   9.646  36.222  0.00
0.00     0.045 C
ATOM      6  N3  1F0 A        11.997  10.317  36.772  0.00
0.00    -0.255 NA
ATOM      7  C2  1F0 A        12.817   9.323  36.800  0.00
0.00     0.134 C
ATOM      8  C10 1F0 A        13.066   7.097  35.478  0.00
0.00     0.310 C
ATOM      9  O10 1F0 A        13.331   6.708  34.334  0.00
0.00    -0.253 OA
ATOM     10  N10 1F0 A        13.241   6.247  36.580  0.00
0.00    -0.275 N
ATOM     11  C14 1F0 A        13.894   4.961  36.360  0.00
0.00     0.160 C
ATOM     12  C13 1F0 A        14.773   4.636  37.554  0.00
0.00     0.148 C
ATOM     13  N12 1F0 A        13.986   4.650  38.811  0.00
0.00     0.249 N
ATOM     14  C12 1F0 A        13.338   5.980  39.031  0.00
0.00     0.148 C
ATOM     15  C11 1F0 A        12.468   6.395  37.819  0.00
0.00     0.160 C
ATOM     16  C16 1F0 A        14.105   4.266  41.285  0.00
0.00     0.176 C
ATOM     17  C15 1F0 A        14.837   4.223  39.942  0.00
0.00     0.141 C
```

```
ATOM     18 2HN1 1F0 A        13.248   3.967  38.713  0.00
0.00      0.200 HD
ENDROOT
BRANCH   2  19
ATOM     19  C51 1F0 A        12.519   9.725  33.633  0.00
0.00     -0.021 A
ATOM     20  C56 1F0 A        13.895   9.911  33.814  0.00
0.00      0.007 A
ATOM     21  C55 1F0 A        14.651  10.585  32.847  0.00
0.00      0.021 A
ATOM     22  C54 1F0 A        14.023  11.069  31.699  0.00
0.00      0.041 A
ATOM     23  CL5 1F0 A        14.949  12.020  30.533  0.00
0.00     -0.084 Cl
ATOM     24  C53 1F0 A        12.664  10.823  31.476  0.00
0.00      0.021 A
ATOM     25  C52 1F0 A        11.913  10.151  32.443  0.00
0.00      0.007 A
ENDBRANCH  2  19
BRANCH   7  26
ATOM     26  C21 1F0 A        13.848   9.094  37.832  0.00
0.00      0.045 A
ATOM     27  C22 1F0 A        13.564   9.406  39.163  0.00
0.00      0.015 A
ATOM     28  C23 1F0 A        14.532   9.237  40.156  0.00
0.00      0.005 A
ATOM     29  C24 1F0 A        15.800   8.743  39.831  0.00
0.00     -0.040 A
ATOM     30  C25 1F0 A        16.109   8.427  38.504  0.00
0.00      0.038 A
ATOM     31  C26 1F0 A        15.127   8.612  37.519  0.00
0.00      0.086 A
ATOM     32  O27 1F0 A        15.422   8.340  36.211  0.00
0.00     -0.346 OA
ATOM     33  C28 1F0 A        16.739   7.881  35.949  0.00
0.00      0.164 C
ATOM     34  C29 1F0 A        16.882   7.542  34.459  0.00
0.00      0.035 C
BRANCH  29  35
ATOM     35  C30 1F0 A        16.824   8.563  40.922  0.00
0.00     -0.013 C
ATOM     36  C33 1F0 A        17.797   9.747  40.911  0.00
0.00      0.016 C
ATOM     37  C31 1F0 A        17.604   7.264  40.697  0.00
0.00      0.016 C
ATOM     38  C32 1F0 A        16.117   8.501  42.279  0.00
0.00      0.016 C
ENDBRANCH  29  35
ENDBRANCH  7  26
BRANCH   4  39
ATOM     39  C41 1F0 A        10.803  11.463  34.943  0.00
0.00     -0.018 A
```

```
ATOM     40  C42 1F0 A        9.543  11.722  34.394  0.00
0.00     0.008 A
ATOM     41  C43 1F0 A        9.314  12.907  33.701  0.00
0.00     0.021 A
ATOM     42  C44 1F0 A       10.353  13.825  33.559  0.00
0.00     0.041 A
ATOM     43  CL4 1F0 A       10.088  15.305  32.647  0.00
0.00    -0.084 Cl
ATOM     44  C45 1F0 A       11.608  13.583  34.122  0.00
0.00     0.021 A
ATOM     45  C46 1F0 A       11.828  12.400  34.819  0.00
0.00     0.008 A
ENDBRANCH   4  39
BRANCH  16  46
ATOM     46  C17 1F0 A       15.068   3.786  42.382  0.00
0.00     0.530 C
BRANCH  46  47
ATOM     47  S17 1F0 A       14.470   4.217  43.996  0.00
0.00    -0.280 S
ATOM     48  C19 1F0 A       15.569   3.418  45.143  0.00
0.00     0.666 C
ATOM     49  O17 1F0 A       13.138   3.656  44.163  0.00
0.00    -0.525 OA
ATOM     50  O18 1F0 A       14.594   5.666  44.158  0.00
0.00    -0.525 OA
ENDBRANCH  46  47
ENDBRANCH  16  46
TORSDOF 10
```

Flexible Residue File

```
BEGIN_RES LEU A  50
REMARK  2 active torsions:
REMARK  status: ('A' for Active; 'I' for Inactive)
REMARK   1  A    between atoms: CA   and  CB
REMARK   2  A    between atoms: CB   and  CG
ROOT
ATOM      1  CA  LEU A  50     9.533   9.929  29.079  1.00
10.27    0.186 C
ENDROOT
BRANCH   1   2
ATOM      2  CB  LEU A  50     8.870  10.875  30.081  1.00
10.02    0.034 C
BRANCH   2   3
ATOM      3  CG  LEU A  50     8.727  12.295  29.523  1.00
13.36    0.002 C
ATOM      4  CD1 LEU A  50     7.673  13.063  30.296  1.00
11.28    0.000 C
ATOM      5  CD2 LEU A  50    10.076  12.999  29.587  1.00
11.01    0.000 C
ENDBRANCH   2   3
```

```
ENDBRANCH   1    2
END_RES LEU A   50
BEGIN_RES LEU A   53
REMARK  2 active torsions:
REMARK  status: ('A' for Active; 'I' for Inactive)
REMARK    3  A     between atoms: CA   and   CB
REMARK    4  A     between atoms: CB   and   CG
ROOT
ATOM      6  CA  LEU A  53      13.763   7.673  27.601  1.00
10.12     0.186 C
ENDROOT
BRANCH   6    7
ATOM      7  CB  LEU A  53      13.561   9.190  27.484  1.00
10.09     0.034 C
BRANCH   7    8
ATOM      8  CG  LEU A  53      14.846  10.013  27.342  1.00
14.17     0.002 C
ATOM      9  CD2 LEU A  53      14.500  11.493  27.234  1.00
15.42     0.000 C
ATOM     10  CD1 LEU A  53      15.623   9.551  26.114  1.00
13.92     0.000 C
ENDBRANCH   7    8
ENDBRANCH   6    7
END_RES LEU A   53
BEGIN_RES HIS A   69
REMARK  2 active torsions:
REMARK  status: ('A' for Active; 'I' for Inactive)
REMARK    5  A     between atoms: CA   and   CB
REMARK    6  A     between atoms: CB   and   CG
ROOT
ATOM     11  CA  HIS A  69      19.751  12.132  39.250  1.00
10.97     0.192 C
ENDROOT
BRANCH  11  12
ATOM     12  CB  HIS A  69      19.087  13.136  40.206  1.00
13.11     0.082 C
BRANCH  12  13
ATOM     13  CG  HIS A  69      20.025  14.161  40.764  1.00
13.45     0.082 A
ATOM     14  ND1 HIS A  69      20.296  15.350  40.123  1.00
16.07    -0.227 NA
ATOM     15  CE1 HIS A  69      21.155  16.050  40.842  1.00
18.20     0.199 A
ATOM     16  NE2 HIS A  69      21.452  15.357  41.926  1.00
17.19    -0.221 NA
ATOM     17  CD2 HIS A  69      20.759  14.170  41.902  1.00
17.53     0.116 A
ENDBRANCH  12  13
ENDBRANCH  11  12
END_RES HIS A   69
```

Grid Parameter File

```
npts 60 62 62                          # num.grid points in xyz
gridfld 4ipf_rigid.maps.fld            # grid_data_file
spacing 0.375                          # spacing(A)
receptor_types A C N NA OA SA HD        # receptor atom types
ligand_types A C Cl NA OA N S HD       # ligand atom types
receptor 4ipf_rigid.pdbqt              # macromolecule
gridcenter 13.988 9.562 35.855         # xyz-coordinates or auto
smooth 0.5                             # store minimum energy w/in
rad(A)
map 4ipf_rigid.A.map                   # atom-specific affinity map
map 4ipf_rigid.C.map                   # atom-specific affinity map
map 4ipf_rigid.Cl.map                  # atom-specific affinity map
map 4ipf_rigid.NA.map                  # atom-specific affinity map
map 4ipf_rigid.OA.map                  # atom-specific affinity map
map 4ipf_rigid.N.map                   # atom-specific affinity map
map 4ipf_rigid.S.map                   # atom-specific affinity map
map 4ipf_rigid.HD.map                  # atom-specific affinity map
elecmap 4ipf_rigid.e.map               # electrostatic potential
map
dsolvmap 4ipf_rigid.d.map              # desolvation potential
map
dielectric -0.1465                     # <0, AD4 distance-
dep.diel;>0, constant
```

Docking Parameter File

```
autodock_parameter_version 4.2         # used by autodock to
validate parameter set
outlev 1                               # diagnostic output level
intelec                                # calculate internal
electrostatics
seed pid time                          # seeds for random generator
ligand_types A C Cl NA OA N S HD       # atoms types in ligand
fld 4ipf_rigid.maps.fld                # grid_data_file
map 4ipf_rigid.A.map                   # atom-specific affinity map
map 4ipf_rigid.C.map                   # atom-specific affinity map
map 4ipf_rigid.Cl.map                  # atom-specific affinity map
map 4ipf_rigid.NA.map                  # atom-specific affinity map
map 4ipf_rigid.OA.map                  # atom-specific affinity map
map 4ipf_rigid.N.map                   # atom-specific affinity map
map 4ipf_rigid.S.map                   # atom-specific affinity map
map 4ipf_rigid.HD.map                  # atom-specific affinity map
elecmap 4ipf_rigid.e.map               # electrostatics map
desolvmap 4ipf_rigid.d.map             # desolvation map
move 4ipf_lig.pdbqt                    # small molecule
flexres 4ipf_flex.pdbqt                # file containing
flexible residues
about 13.5629 8.4362 37.0971           # small molecule center
tran0 random                           # initial coordinates/A or
```

```
random
quaternion0 random                           # initial orientation
dihe0 random                                  # initial dihedrals
(relative) or random
torsdof 10                                     # torsional degrees of
freedom
rmstol 2.0                                    # cluster_tolerance/A
extnrg 1000.0                                 # external grid energy
e0max 0.0 10000                               # max initial energy; max
number of retries
ga_pop_size 150                               # number of individuals in
population
ga_num_evals 20000000                          # maximum number of energy
evaluations
ga_num_generations 27000                      # maximum number of
generations
ga_elitism 1                                  # number of top individuals
to survive to next generation
ga_mutation_rate 0.02                         # rate of gene mutation
ga_crossover_rate 0.8                         # rate of crossover
ga_window_size 10                             #
ga_cauchy_alpha 0.0                           # Alpha parameter of Cauchy
distribution
ga_cauchy_beta 1.0                            # Beta parameter Cauchy
distribution
set_ga                                        # set the above parameters
for GA or LGA
sw_max_its 300                                # iterations of Solis & Wets
local search
sw_max_succ 4                                 # consecutive successes
before changing rho
sw_max_fail 4                                 # consecutive failures
before changing rho
sw_rho 1.0                                    # size of local search space
to sample
sw_lb_rho 0.01                                # lower bound on rho
ls_search_freq 0.06                           # probability of performing
local search on individual
set_psw1                                      # set the above pseudo-Solis
& Wets parameters
unbound_model bound                           # state of unbound ligand
ga_run 100                                     # do this many hybrid GA-LS
runs
analysis                                      # perform a ranked cluster
analysis
```

Vina Configuration File

```
receptor = 4ipf_rigid.pdbqt
flex = 4ipf_flex.pdbqt
ligand = 4ipf_lig.pdbqt
```

```
out = out.4ipf.pdbqt

center_x = 13.988
center_y = 9.562
center_z = 35.855

size_x = 27
size_y = 27
size_z = 27

exhaustiveness = 12
```

Ligand Input File

```
REMARK   0 active torsions:
REMARK   status: ('A' for Active; 'I' for Inactive)
REMARK       I    between atoms: C1_1   and  C10_2
REMARK       I    between atoms: C10_2  and  N11_3
REMARK       I    between atoms: N11_3  and  C12_4
REMARK       I    between atoms: C12_4  and  C13_5
REMARK       I    between atoms: C13_5  and  C14_6
REMARK       I    between atoms: C14_6  and  O14_7
REMARK       I    between atoms: C14_6  and  C15_8
REMARK       I    between atoms: C15_8  and  O15_9
REMARK       I    between atoms: C2_12  and  C21_13
REMARK       I    between atoms: C3_20  and  C31_23
REMARK       I    between atoms: C4_30  and  C41_31
REMARK       I    between atoms: C41_31  and  C42_32
ROOT
ATOM      1  C1  I09 A         11.781  10.738  -1.043  0.00
0.00    0.288 C
ATOM      2  C10 I09 A         12.414  11.957  -0.431  0.00
0.00    0.249 C
ATOM      3  N11 I09 A         11.556  12.817   0.220  0.00
0.00   -0.359 N
ATOM      4  C12 I09 A         12.033  14.023   0.890  0.00
0.00    0.125 C
ATOM      5  C13 I09 A         12.241  15.149  -0.123  0.00
0.00    0.053 C
ATOM      6  C14 I09 A         12.646  16.423   0.616  0.00
0.00    0.143 C
ATOM      7  O14 I09 A         11.500  17.054   1.182  0.00
0.00   -0.393 OA
ATOM      8  C15 I09 A         13.403  17.385  -0.310  0.00
0.00    0.195 C
ATOM      9  O15 I09 A         12.564  17.825  -1.369  0.00
0.00   -0.398 OA
ATOM     10  O10 I09 A         13.634  12.192  -0.479  0.00
0.00   -0.271 OA
```

```
ATOM     11  N1  I09 A        10.350  10.632  -0.589  0.00
0.00    -0.055 N
ATOM     12  C2  I09 A        12.467   9.466  -0.549  0.00
0.00     0.088 C
ATOM     13  C21 I09 A        13.595   9.007  -1.435  0.00
0.00    -0.048 A
ATOM     14  C26 I09 A        13.601   9.292  -2.802  0.00
0.00     0.007 A
ATOM     15  C25 I09 A        14.641   8.817  -3.613  0.00
0.00     0.002 A
ATOM     16  C22 I09 A        14.630   8.269  -0.877  0.00
0.00     0.031 A
ATOM     17  C23 I09 A        15.669   7.801  -1.685  0.00
0.00     0.034 A
ATOM     18 CL2  I09 A        16.985   6.894  -0.955  0.00
0.00    -0.084 Cl
ATOM     19  C24 I09 A        15.678   8.063  -3.063  0.00
0.00     0.024 A
ATOM     20  C3  I09 A        11.330   8.456  -0.438  0.00
0.00     0.114 C
ATOM     21  C37 I09 A        10.965   7.977  -1.780  0.00
0.00     0.077 C
ATOM     22  N37 I09 A        10.689   7.604  -2.820  0.00
0.00    -0.196 NA
ATOM     23  C31 I09 A        11.662   7.293   0.469  0.00
0.00    -0.033 A
ATOM     24  C32 I09 A        12.218   7.516   1.745  0.00
0.00     0.011 A
ATOM     25  C33 I09 A        12.538   6.429   2.561  0.00
0.00     0.024 A
ATOM     26  C36 I09 A        11.426   5.988   0.024  0.00
0.00     0.011 A
ATOM     27  C35 I09 A        11.742   4.901   0.841  0.00
0.00     0.024 A
ATOM     28  C34 I09 A        12.299   5.122   2.104  0.00
0.00     0.034 A
ATOM     29 CL3  I09 A        12.718   3.745   3.119  0.00
0.00    -0.084 Cl
ATOM     30  C4  I09 A        10.187   9.303   0.095  0.00
0.00     0.215 C
ATOM     31  C41 I09 A         8.816   8.670  -0.208  0.00
0.00     0.043 C
ATOM     32  C42 I09 A         7.720   9.186   0.758  0.00
0.00    -0.049 C
ATOM     33  C45 I09 A         7.803  10.724   0.897  0.00
0.00     0.014 C
ATOM     34  C44 I09 A         7.901   8.544   2.142  0.00
0.00     0.014 C
ATOM     35  C43 I09 A         6.336   8.796   0.210  0.00
0.00     0.014 C
ATOM     36  H11 I09 A        10.568  12.607   0.234  0.00
0.00     0.162 HD
```

```
ATOM      37 2HN1 I09 A           9.731  10.688  -1.385  0.00
0.00      0.278 HD
ATOM      38 3HN1 I09 A          10.141  11.378   0.059  0.00
0.00      0.278 HD
ATOM      39 4HO1 I09 A          10.882  17.281   0.483  0.00
0.00      0.210 HD
ATOM      40 5HO1 I09 A          13.053  18.425  -1.937  0.00
0.00      0.209 HD
ENDROOT
TORSDOF 11
```

Flexible Residue File

```
BEGIN_RES LEU A  50
REMARK  2 active torsions:
REMARK  status: ('A' for Active; 'I' for Inactive)
REMARK    1  A    between atoms: CA   and   CB
REMARK    2  A    between atoms: CB   and   CG
ROOT
ATOM       1  CA  LEU A  50       12.169   2.858  -2.898  1.00
29.74     0.177 C
ENDROOT
BRANCH   1    2
ATOM       2  CB  LEU A  50       13.101   3.947  -3.433  1.00
28.76     0.038 C
BRANCH   2    3
ATOM       3  CG  LEU A  50       14.583   3.735  -3.102  1.00
30.22    -0.020 C
ATOM       4  CD1 LEU A  50       15.429   4.648  -3.971  1.00
30.37     0.009 C
ATOM       5  CD2 LEU A  50       14.833   4.003  -1.629  1.00
24.61     0.009 C
ENDBRANCH   2    3
ENDBRANCH   1    2
END_RES LEU A  50
BEGIN_RES ILE A  57
REMARK  2 active torsions:
REMARK  status: ('A' for Active; 'I' for Inactive)
REMARK    3  A    between atoms: CA   and   CB
REMARK    4  A    between atoms: CB   and   CG1
ROOT
ATOM       6  CA  ILE A  57        5.611   4.348   4.893  1.00
31.92     0.180 C
ENDROOT
BRANCH   6    7
ATOM       7  CB  ILE A  57        6.798   5.276   4.579  1.00
30.37     0.013 C
ATOM       8  CG2 ILE A  57        6.745   6.518   5.452  1.00
32.04     0.012 C
BRANCH   7    9
ATOM       9  CG1 ILE A  57        8.107   4.518   4.810  1.00
```

```
32.69     0.002 C
ATOM     10  CD1 ILE A  57         9.258   5.022   3.961  1.00
30.75     0.005 C
ENDBRANCH   7    9
ENDBRANCH   6    7
END_RES ILE A  57
BEGIN_RES VAL A  89
REMARK  1 active torsions:
REMARK  status: ('A' for Active; 'I' for Inactive)
REMARK    5  A    between atoms: CA   and  CB
ROOT
ATOM     11  CA  VAL A  89        13.479  10.288   4.280  1.00
30.94     0.180 C
ENDROOT
BRANCH  11  12
ATOM     12  CB  VAL A  89        12.045  10.614   3.779  1.00
27.73     0.009 C
ATOM     13  CG1 VAL A  89        11.030   9.813   4.563  1.00
28.65     0.012 C
ATOM     14  CG2 VAL A  89        11.769  12.096   3.902  1.00
29.10     0.012 C
ENDBRANCH  11  12
END_RES VAL A  89
BEGIN_RES ILE A  95
REMARK  2 active torsions:
REMARK  status: ('A' for Active; 'I' for Inactive)
REMARK    6  A    between atoms: CA   and  CB
REMARK    7  A    between atoms: CB   and  CG1
ROOT
ATOM     15  CA  ILE A  95        18.516   4.625   3.211  1.00
26.56     0.180 C
ENDROOT
BRANCH  15  16
ATOM     16  CB  ILE A  95        17.419   5.493   2.567  1.00
27.98     0.013 C
ATOM     17  CG2 ILE A  95        16.365   4.619   1.878  1.00
24.30     0.012 C
BRANCH  16  18
ATOM     18  CG1 ILE A  95        16.779   6.376   3.646  1.00
27.40     0.002 C
ATOM     19  CD1 ILE A  95        15.909   7.506   3.097  1.00
29.20     0.005 C
ENDBRANCH  16  18
ENDBRANCH  15  16
END_RES ILE A  95
BEGIN_RES LYS A  90
REMARK  5 active torsions:
REMARK  status: ('A' for Active; 'I' for Inactive)
REMARK    8  A    between atoms: CA   and  CB
REMARK    9  A    between atoms: CB   and  CG
REMARK   10  A    between atoms: CG   and  CD
REMARK   11  A    between atoms: CD   and  CE
```

117

```
REMARK    12  A    between atoms: CE    and  NZ
ROOT
ATOM     20  CA  LYS A  90      16.135  12.789   3.183  1.00
34.59     0.176 C
ENDROOT
BRANCH  20  21
ATOM     21  CB  LYS A  90      16.176  14.231   3.703  1.00
37.01     0.035 C
BRANCH  21  22
ATOM     22  CG  LYS A  90      15.016  15.094   3.270  1.00
38.28     0.004 C
BRANCH  22  23
ATOM     23  CD  LYS A  90      15.117  16.472   3.909  1.00
44.74     0.027 C
BRANCH  23  24
ATOM     24  CE  LYS A  90      14.329  17.506   3.125  1.00
48.47     0.229 C
BRANCH  24  25
ATOM     25  NZ  LYS A  90      14.085  18.742   3.928  1.00
52.71    -0.079 N
ATOM     26  HZ2 LYS A  90      13.124  19.031   3.817  1.00
0.00      0.274 HD
ATOM     27  HZ3 LYS A  90      14.268  18.552   4.903  1.00
0.00      0.274 HD
ATOM     28  HZ1 LYS A  90      14.697  19.479   3.608  1.00
0.00      0.274 HD
ENDBRANCH  24   25
ENDBRANCH  23   24
ENDBRANCH  22   23
ENDBRANCH  21   22
ENDBRANCH  20   21
END_RES LYS A  90
```

Grid Parameter File

```
npts 60 62 62                        # num.grid points in xyz
gridfld 4jrg_rigid.maps.fld          # grid_data_file
spacing 0.375                        # spacing(A)
receptor_types A C HD N NA OA SA     # receptor atom types
ligand_types A C Cl NA OA N HD       # ligand atom types
receptor 4jrg_rigid.pdbqt            # macromolecule
gridcenter 11.493 10.5 2.759         # xyz-coordinates or auto
smooth 0.5                           # store minimum energy w/in
rad(A)
map 4jrg_rigid.A.map                 # atom-specific affinity map
map 4jrg_rigid.C.map                 # atom-specific affinity map
map 4jrg_rigid.Cl.map                # atom-specific affinity map
map 4jrg_rigid.NA.map                # atom-specific affinity map
map 4jrg_rigid.OA.map                # atom-specific affinity map
map 4jrg_rigid.N.map                 # atom-specific affinity map
map 4jrg_rigid.HD.map                # atom-specific affinity map
```

```
elecmap 4jrg_rigid.e.map              # electrostatic potential
map
dsolvmap 4jrg_rigid.d.map               # desolvation potential
map
dielectric -0.1465                    # <0, AD4 distance-
dep.diel;>0, constant
```

Docking Parameter File

```
autodock_parameter_version 4.2        # used by autodock to
validate parameter set
outlev 1                              # diagnostic output level
intelec                               # calculate internal
electrostatics
seed pid time                         # seeds for random generator
ligand_types A C Cl NA OA N HD        # atoms types in ligand
fld 4jrg_rigid.maps.fld               # grid_data_file
map 4jrg_rigid.A.map                  # atom-specific affinity map
map 4jrg_rigid.C.map                  # atom-specific affinity map
map 4jrg_rigid.Cl.map                 # atom-specific affinity map
map 4jrg_rigid.NA.map                 # atom-specific affinity map
map 4jrg_rigid.OA.map                 # atom-specific affinity map
map 4jrg_rigid.N.map                  # atom-specific affinity map
map 4jrg_rigid.HD.map                 # atom-specific affinity map
elecmap 4jrg_rigid.e.map              # electrostatics map
desolvmap 4jrg_rigid.d.map            # desolvation map
move 4jrg_lig.pdbqt                   # small molecule
flexres 4jrg_flex.pdbqt               # file containing flexible
residues
about 11.8028 10.3609 -0.2311         # small molecule center
tran0 random                          # initial coordinates/A or
random
quaternion0 random                    # initial orientation
dihe0 random                          # initial dihedrals
(relative) or random
torsdof 11                             # torsional degrees of
freedom
rmstol 2.0                            # cluster_tolerance/A
extnrg 1000.0                         # external grid energy
e0max 0.0 10000                       # max initial energy; max
number of retries
ga_pop_size 150                       # number of individuals in
population
ga_num_evals 20000000                 # maximum number of energy
evaluations
ga_num_generations 27000              # maximum number of
generations
ga_elitism 1                          # number of top individuals
to survive to next generation
ga_mutation_rate 0.02                 # rate of gene mutation
```

```
ga_crossover_rate 0.8                    # rate of crossover
ga_window_size 10                        #
ga_cauchy_alpha 0.0                      # Alpha parameter of Cauchy
distribution
ga_cauchy_beta 1.0                       # Beta parameter Cauchy
distribution
set_ga                                   # set the above parameters
for GA or LGA
sw_max_its 300                           # iterations of Solis & Wets
local search
sw_max_succ 4                            # consecutive successes
before changing rho
sw_max_fail 4                            # consecutive failures
before changing rho
sw_rho 1.0                               # size of local search space
to sample
sw_lb_rho 0.01                           # lower bound on rho
ls_search_freq 0.06                      # probability of performing
local search on individual
set_psw1                                 # set the above pseudo-Solis
& Wets parameters
unbound_model bound                      # state of unbound ligand
ga_run 100                                # do this many hybrid GA-LS
runs
analysis                                 # perform a ranked cluster
analysis
```

Vina Configuration File

```
receptor = 4jrg_rigid.pdbqt
flex = 4jrg_flex.pdbqt
ligand = 4jrg_lig.pdbqt

out = out.4jrg.pdbqt

center_x = 11.493
center_y = 10.5
center_z = 2.759

size_x = 27
size_y = 27
size_z = 27

exhaustiveness = 12
```

Ligand Input File

```
REMARK  12 active torsions:
REMARK  status: ('A' for Active; 'I' for Inactive)
REMARK    1  A    between atoms: C1_1  and  C10_2
```

```
REMARK    2  A    between atoms: C10_2   and  N11_3
REMARK    3  A    between atoms: N11_3   and  C12_4
REMARK    4  A    between atoms: C12_4   and  C13_5
REMARK    5  A    between atoms: C13_5   and  C14_6
REMARK    6  A    between atoms: C14_6   and  O14_7
REMARK    7  A    between atoms: C14_6   and  C15_8
REMARK    8  A    between atoms: C15_8   and  O15_9
REMARK    9  A    between atoms: C2_12   and  C21_13
REMARK   10  A    between atoms: C3_20   and  C31_23
REMARK   11  A    between atoms: C4_30   and  C41_31
REMARK   12  A    between atoms: C41_31  and  C42_32
ROOT
ATOM      1  C10 I09 A         12.414  11.957  -0.431  0.00
0.00     0.249 C
ATOM      2  O10 I09 A         13.634  12.192  -0.479  0.00
0.00    -0.271 OA
ENDROOT
BRANCH   1   3
ATOM      3  C1  I09 A         11.781  10.738  -1.043  0.00
0.00     0.288 C
ATOM      4  N1  I09 A         10.350  10.632  -0.589  0.00
0.00    -0.055 N
ATOM      5  C4  I09 A         10.187   9.303   0.095  0.00
0.00     0.215 C
ATOM      6 2HN1 I09 A          9.731  10.688  -1.385  0.00
0.00     0.278 HD
ATOM      7 3HN1 I09 A         10.141  11.378   0.059  0.00
0.00     0.278 HD
ATOM      8  C3  I09 A         11.330   8.456  -0.438  0.00
0.00     0.114 C
ATOM      9  C2  I09 A         12.467   9.466  -0.549  0.00
0.00     0.088 C
ATOM     10  C37 I09 A         10.965   7.977  -1.780  0.00
0.00     0.077 C
ATOM     11  N37 I09 A         10.689   7.604  -2.820  0.00
0.00    -0.196 NA
BRANCH   5  12
ATOM     12  C41 I09 A          8.816   8.670  -0.208  0.00
0.00     0.043 C
BRANCH  12  13
ATOM     13  C42 I09 A          7.720   9.186   0.758  0.00
0.00    -0.049 C
ATOM     14  C45 I09 A          7.803  10.724   0.897  0.00
0.00     0.014 C
ATOM     15  C44 I09 A          7.901   8.544   2.142  0.00
0.00     0.014 C
ATOM     16  C43 I09 A          6.336   8.796   0.210  0.00
0.00     0.014 C
ENDBRANCH  12  13
ENDBRANCH   5  12
BRANCH   8  17
ATOM     17  C31 I09 A         11.662   7.293   0.469  0.00
```

```
0.00     -0.033 A
ATOM      18  C32 I09 A         12.218   7.516   1.745  0.00
0.00      0.011 A
ATOM      19  C33 I09 A         12.538   6.429   2.561  0.00
0.00      0.024 A
ATOM      20  C34 I09 A         12.299   5.122   2.104  0.00
0.00      0.034 A
ATOM      21  C35 I09 A         11.742   4.901   0.841  0.00
0.00      0.024 A
ATOM      22 CL3  I09 A         12.718   3.745   3.119  0.00
0.00     -0.084 Cl
ATOM      23  C36 I09 A         11.426   5.988   0.024  0.00
0.00      0.011 A
ENDBRANCH   8  17
BRANCH    9  24
ATOM      24  C21 I09 A         13.595   9.007  -1.435  0.00
0.00     -0.048 A
ATOM      25  C26 I09 A         13.601   9.292  -2.802  0.00
0.00      0.007 A
ATOM      26  C25 I09 A         14.641   8.817  -3.613  0.00
0.00      0.002 A
ATOM      27  C24 I09 A         15.678   8.063  -3.063  0.00
0.00      0.024 A
ATOM      28  C23 I09 A         15.669   7.801  -1.685  0.00
0.00      0.034 A
ATOM      29  C22 I09 A         14.630   8.269  -0.877  0.00
0.00      0.031 A
ATOM      30 CL2  I09 A         16.985   6.894  -0.955  0.00
0.00     -0.084 Cl
ENDBRANCH   9  24
ENDBRANCH   1   3
BRANCH    1  31
ATOM      31  N11 I09 A         11.556  12.817   0.220  0.00
0.00     -0.359 N
ATOM      32  H11 I09 A         10.568  12.607   0.234  0.00
0.00      0.162 HD
BRANCH   31  33
ATOM      33  C12 I09 A         12.033  14.023   0.890  0.00
0.00      0.125 C
BRANCH   33  34
ATOM      34  C13 I09 A         12.241  15.149  -0.123  0.00
0.00      0.053 C
BRANCH   34  35
ATOM      35  C14 I09 A         12.646  16.423   0.616  0.00
0.00      0.143 C
BRANCH   35  36
ATOM      36  O14 I09 A         11.500  17.054   1.182  0.00
0.00     -0.393 OA
ATOM      37 4HO1 I09 A         10.882  17.281   0.483  0.00
0.00      0.210 HD
ENDBRANCH  35  36
BRANCH   35  38
```

```
ATOM      38  C15 I09 A          13.403  17.385  -0.310  0.00
0.00      0.195 C
BRANCH   38  39
ATOM      39  O15 I09 A          12.564  17.825  -1.369  0.00
0.00     -0.398 OA
ATOM      40 5HO1 I09 A          13.053  18.425  -1.937  0.00
0.00      0.209 HD
ENDBRANCH  38  39
ENDBRANCH  35  38
ENDBRANCH  34  35
ENDBRANCH  33  34
ENDBRANCH  31  33
ENDBRANCH   1  31
TORSDOF 11
```

Grid Parameter File

```
npts 60 62 62                        # num.grid points in xyz
gridfld 4jrg_rigid.maps.fld          # grid_data_file
spacing 0.375                        # spacing(A)
receptor_types A C H HD N NA OA SA   # receptor atom types
ligand_types A C Cl NA OA N HD       # ligand atom types
receptor 4jrg_rigid.pdbqt            # macromolecule
gridcenter 11.493 10.5 2.759         # xyz-coordinates or auto
smooth 0.5                           # store minimum energy w/in
rad(A)
map 4jrg_rigid.A.map                 # atom-specific affinity map
map 4jrg_rigid.C.map                 # atom-specific affinity map
map 4jrg_rigid.Cl.map                # atom-specific affinity map
map 4jrg_rigid.NA.map                # atom-specific affinity map
map 4jrg_rigid.OA.map                # atom-specific affinity map
map 4jrg_rigid.N.map                 # atom-specific affinity map
map 4jrg_rigid.HD.map                # atom-specific affinity map
elecmap 4jrg_rigid.e.map             # electrostatic potential
map
dsolvmap 4jrg_rigid.d.map              # desolvation potential
map
dielectric -0.1465                   # <0, AD4 distance-
dep.diel;>0, constant
```

Docking Parameter File

```
autodock_parameter_version 4.2       # used by autodock to
validate parameter set
outlev 1                             # diagnostic output level
intelec                              # calculate internal
electrostatics
seed pid time                        # seeds for random generator
ligand_types A C Cl NA OA N HD       # atoms types in ligand
fld 4jrg_rigid.maps.fld              # grid_data_file
map 4jrg_rigid.A.map                 # atom-specific affinity map
```

```
map 4jrg_rigid.C.map                      # atom-specific affinity map
map 4jrg_rigid.Cl.map                     # atom-specific affinity map
map 4jrg_rigid.NA.map                     # atom-specific affinity map
map 4jrg_rigid.OA.map                     # atom-specific affinity map
map 4jrg_rigid.N.map                      # atom-specific affinity map
map 4jrg_rigid.HD.map                     # atom-specific affinity map
elecmap 4jrg_rigid.e.map                  # electrostatics map
desolvmap 4jrg_rigid.d.map                # desolvation map
move 4jrg_lig.pdbqt                       # small molecule
about 11.8028 10.3609 -0.2311             # small molecule center
tran0 random                              # initial coordinates/A or
random
quaternion0 random                        # initial orientation
dihe0 random                              # initial dihedrals
(relative) or random
torsdof 12                                 # torsional degrees of
freedom
rmstol 2.0                                # cluster_tolerance/A
extnrg 1000.0                             # external grid energy
e0max 0.0 10000                           # max initial energy; max
number of retries
ga_pop_size 150                           # number of individuals in
population
ga_num_evals 20000000                      # maximum number of energy
evaluations
ga_num_generations 27000                  # maximum number of
generations
ga_elitism 1                              # number of top individuals
to survive to next generation
ga_mutation_rate 0.02                     # rate of gene mutation
ga_crossover_rate 0.8                     # rate of crossover
ga_window_size 10                         #
ga_cauchy_alpha 0.0                       # Alpha parameter of Cauchy
distribution
ga_cauchy_beta 1.0                        # Beta parameter Cauchy
distribution
set_ga                                    # set the above parameters
for GA or LGA
sw_max_its 300                            # iterations of Solis & Wets
local search
sw_max_succ 4                             # consecutive successes
before_changing rho
sw_max_fail 4                             # consecutive failures
before changing rho
sw_rho 1.0                                # size of local search space
to sample
sw_lb_rho 0.01                            # lower bound on rho
ls_search_freq 0.06                       # probability of performing
local search on individual
set_psw1                                  # set the above pseudo-Solis
& Wets parameters
unbound_model bound                       # state of unbound ligand
```

```
ga_run 100                                    # do this many hybrid GA-LS
runs
analysis                                      # perform a ranked cluster
analysis
```

Vina Configuration File

```
receptor = 4jrg_rec.pdbqt
ligand = 4jrg_lig.pdbqt

out = out.4jrg.pdbqt

center_x = 11.493
center_y = 10.5
center_z = 2.759

size_x = 27
size_y = 27
size_z = 27

exhaustiveness = 12
```

Ligand Input File

```
REMARK  6 active torsions:
REMARK  status: ('A' for Active; 'I' for Inactive)
REMARK      I    between atoms: C1_1   and  C10_2
REMARK      I    between atoms: C10_2  and  N11_3
REMARK      I    between atoms: N11_3  and  C12_4
REMARK      I    between atoms: C12_4  and  C13_5
REMARK      I    between atoms: C13_5  and  C14_6
REMARK   1  A    between atoms: C14_6  and  O14_37
REMARK   2  A    between atoms: C14_6  and  C15_7
REMARK   3  A    between atoms: C15_7  and  O15_39
REMARK   4  A    between atoms: C2_10  and  C21_19
REMARK   5  A    between atoms: C3_11  and  C31_26
REMARK      I    between atoms: C4_14  and  C41_15
REMARK   6  A    between atoms: C41_15  and  C42_33
ROOT
ATOM      1  C1  I09 A       11.781  10.738  -1.043  0.00
0.00     0.159 C
ATOM      2  C10 I09 A       12.414  11.957  -0.431  0.00
0.00     0.260 C
ATOM      3  N11 I09 A       11.556  12.817   0.220  0.00
0.00    -0.311 N
ATOM      4  C12 I09 A       12.033  14.023   0.890  0.00
0.00     0.107 C
ATOM      5  C13 I09 A       12.241  15.149  -0.123  0.00
0.00     0.053 C
ATOM      6  C14 I09 A       12.646  16.423   0.616  0.00
0.00     0.141 C
```

```
ATOM          7  O10 I09 A          13.634  12.192  -0.479  0.00
0.00     -0.271 OA
ATOM          8  N1  I09 A          10.350  10.632  -0.589  0.00
0.00      0.252 N
ATOM          9  C2  I09 A          12.467   9.466  -0.549  0.00
0.00      0.107 C
ATOM         10  C3  I09 A          11.330   8.456  -0.438  0.00
0.00      0.136 C
ATOM         11  C37 I09 A          10.965   7.977  -1.780  0.00
0.00      0.081 C
ATOM         12  N37 I09 A          10.689   7.604  -2.820  0.00
0.00     -0.197 NA
ATOM         13  C4  I09 A          10.187   9.303   0.095  0.00
0.00      0.087 C
ATOM         14  C41 I09 A           8.816   8.670  -0.208  0.00
0.00      0.062 C
ATOM         15  H11 I09 A          10.568  12.607   0.234  0.00
0.00      0.150 HD
ATOM         16 2HN1 I09 A          10.137  11.380   0.056  0.00
0.00      0.199 HD
ATOM         17 3HN1 I09 A           9.733  10.684  -1.387  0.00
0.00      0.199 HD
ENDROOT
BRANCH   9  18
ATOM         18  C21 I09 A          13.595   9.007  -1.435  0.00
0.00     -0.035 A
ATOM         19  C22 I09 A          14.630   8.269  -0.877  0.00
0.00      0.024 A
ATOM         20  C23 I09 A          15.669   7.801  -1.685  0.00
0.00      0.041 A
ATOM         21  CL2 I09 A          16.985   6.894  -0.955  0.00
0.00     -0.084 Cl
ATOM         22  C24 I09 A          15.678   8.063  -3.063  0.00
0.00      0.020 A
ATOM         23  C25 I09 A          14.641   8.817  -3.613  0.00
0.00      0.002 A
ATOM         24  C26 I09 A          13.601   9.292  -2.802  0.00
0.00      0.004 A
ENDBRANCH   9  18
BRANCH  10  25
ATOM         25  C31 I09 A          11.662   7.293   0.469  0.00
0.00     -0.021 A
ATOM         26  C36 I09 A          11.426   5.988   0.024  0.00
0.00      0.007 A
ATOM         27  C35 I09 A          11.742   4.901   0.841  0.00
0.00      0.021 A
ATOM         28  C34 I09 A          12.299   5.122   2.104  0.00
0.00      0.041 A
ATOM         29  C33 I09 A          12.538   6.429   2.561  0.00
0.00      0.021 A
ATOM         30  CL3 I09 A          12.718   3.745   3.119  0.00
0.00     -0.084 Cl
```

126

```
ATOM      31  C32 I09 A           12.218   7.516   1.745  0.00
0.00      0.007 A
ENDBRANCH  10  25
BRANCH  14  32
ATOM      32  C42 I09 A            7.720   9.186   0.758  0.00
0.00     -0.033 C
ATOM      33  C43 I09 A            6.336   8.796   0.210  0.00
0.00      0.011 C
ATOM      34  C44 I09 A            7.901   8.544   2.142  0.00
0.00      0.011 C
ATOM      35  C45 I09 A            7.803  10.724   0.897  0.00
0.00      0.011 C
ENDBRANCH  14  32
BRANCH   6  36
ATOM      36  O14 I09 A           11.500  17.054   1.182  0.00
0.00     -0.391 OA
ATOM      37 14HO I09 A           10.882  17.281   0.483  0.00
0.00      0.211 HD
ENDBRANCH   6  36
BRANCH   6  38
ATOM      38  C15 I09 A           13.403  17.385  -0.310  0.00
0.00      0.187 C
BRANCH  38  39
ATOM      39  O15 I09 A           12.564  17.825  -1.369  0.00
0.00     -0.394 OA
ATOM      40 15HO I09 A           13.053  18.425  -1.937  0.00
0.00      0.210 HD
ENDBRANCH  38  39
ENDBRANCH   6  38
TORSDOF 11
```

Flexible Residue File

```
BEGIN_RES HIS A   69
REMARK  2 active torsions:
REMARK  status: ('A' for Active; 'I' for Inactive)
REMARK    1  A    between atoms: CA   and  CB
REMARK    2  A    between atoms: CB   and  CG
ROOT
ATOM      1  CA  HIS A  69      10.614  13.888   7.010  0.70
35.51     0.182 C
ENDROOT
BRANCH   1   2
ATOM      2  CB  HIS A  69      11.813  14.799   6.727  0.70
36.82     0.095 C
BRANCH   2   3
ATOM      3  CG  HIS A  69      12.379  15.449   7.950  0.70
38.82     0.053 A
ATOM      4  ND1 HIS A  69      13.531  15.004   8.564  0.70
39.45    -0.247 NA
ATOM      5  CE1 HIS A  69      13.801  15.779   9.601  0.70
```

```
38.18      0.207 A
ATOM      6  NE2 HIS A  69        12.868  16.707   9.684  0.70
38.06     -0.359 N
ATOM      7  HE2 HIS A  69        12.824  17.434  10.384  0.70
0.00      0.166 HD
ATOM      8  CD2 HIS A  69        11.965  16.525   8.665  0.70
37.74      0.116 A
ENDBRANCH   2    3
ENDBRANCH   1    2
END_RES HIS A   69
BEGIN_RES ILE A   57
REMARK  2 active torsions:
REMARK  status: ('A' for Active; 'I' for Inactive)
REMARK    3  A     between atoms: CA    and   CB
REMARK    4  A     between atoms: CB    and   CG1
ROOT
ATOM      9  CA  ILE A  57         5.611   4.348   4.893  1.00
31.92      0.180 C
ENDROOT
BRANCH   9   10
ATOM     10  CB  ILE A  57         6.798   5.276   4.579  1.00
30.37      0.013 C
ATOM     11  CG2 ILE A  57         6.745   6.518   5.452  1.00
32.04      0.012 C
BRANCH  10   12
ATOM     12  CG1 ILE A  57         8.107   4.518   4.810  1.00
32.69      0.002 C
ATOM     13  CD1 ILE A  57         9.258   5.022   3.961  1.00
30.75      0.005 C
ENDBRANCH  10   12
ENDBRANCH   9   10
END_RES ILE A   57
BEGIN_RES LEU A   50
REMARK  2 active torsions:
REMARK  status: ('A' for Active; 'I' for Inactive)
REMARK    5  A     between atoms: CA    and   CB
REMARK    6  A     between atoms: CB    and   CG
ROOT
ATOM     14  CA  LEU A  50        12.169   2.858  -2.898  1.00
29.74      0.177 C
ENDROOT
BRANCH  14   15
ATOM     15  CB  LEU A  50        13.101   3.947  -3.433  1.00
28.76      0.038 C
BRANCH  15   16
ATOM     16  CG  LEU A  50        14.583   3.735  -3.102  1.00
30.22     -0.020 C
ATOM     17  CD1 LEU A  50        15.429   4.648  -3.971  1.00
30.37      0.009 C
ATOM     18  CD2 LEU A  50        14.833   4.003  -1.629  1.00
24.61      0.009 C
ENDBRANCH  15   16
```

```
ENDBRANCH  14  15
END_RES LEU A  50
```


Grid Parameter File

```
npts 60 62 62                        # num.grid points in xyz
gridfld 4jrg_rigid.maps.fld          # grid_data_file
spacing 0.375                        # spacing(A)
receptor_types A C NA OA N SA HD     # receptor atom types
ligand_types A C Cl NA OA N HD       # ligand atom types
receptor 4jrg_rigid.pdbqt            # macromolecule
gridcenter 11.493 10.5 2.759        # xyz-coordinates or auto
smooth 0.5                           # store minimum energy w/in
rad(A)
map 4jrg_rigid.A.map                 # atom-specific affinity map
map 4jrg_rigid.C.map                 # atom-specific affinity map
map 4jrg_rigid.Cl.map                # atom-specific affinity map
map 4jrg_rigid.NA.map                # atom-specific affinity map
map 4jrg_rigid.OA.map                # atom-specific affinity map
map 4jrg_rigid.N.map                 # atom-specific affinity map
map 4jrg_rigid.HD.map                # atom-specific affinity map
elecmap 4jrg_rigid.e.map             # electrostatic potential
map
dsolvmap 4jrg_rigid.d.map            # desolvation potential
map
dielectric -0.1465                   # <0, AD4 distance-
dep.diel;>0, constant
```

Docking Parameter File

```
autodock_parameter_version 4.2       # used by autodock to
validate parameter set
outlev 1                             # diagnostic output level
intelec                              # calculate internal
electrostatics
seed pid time                        # seeds for random generator
ligand_types A C Cl NA OA N HD       # atoms types in ligand
fld 4jrg_rigid.maps.fld              # grid_data_file
map 4jrg_rigid.A.map                 # atom-specific affinity map
map 4jrg_rigid.C.map                 # atom-specific affinity map
map 4jrg_rigid.Cl.map                # atom-specific affinity map
map 4jrg_rigid.NA.map                # atom-specific affinity map
map 4jrg_rigid.OA.map                # atom-specific affinity map
map 4jrg_rigid.N.map                 # atom-specific affinity map
map 4jrg_rigid.HD.map                # atom-specific affinity map
elecmap 4jrg_rigid.e.map             # electrostatics map
desolvmap 4jrg_rigid.d.map           # desolvation map
move 4jrg_lig.pdbqt                  # small molecule
flexres 4jrg_flex.pdbqt              # file containing
flexible residues
```

```
about 11.8028 10.3609 -0.2312          # small molecule center
tran0 random                           # initial coordinates/A or
random
quaternion0 random                     # initial orientation
dihe0 random                           # initial dihedrals
(relative) or random
torsdof 11                              # torsional degrees of
freedom
rmstol 2.0                             # cluster_tolerance/A
extnrg 1000.0                          # external grid energy
e0max 0.0 10000                        # max initial energy; max
number of retries
ga_pop_size 150                        # number of individuals in
population
ga_num_evals 20000000                   # maximum number of energy
evaluations
ga_num_generations 27000               # maximum number of
generations
ga_elitism 1                           # number of top individuals
to survive to next generation
ga_mutation_rate 0.02                  # rate of gene mutation
ga_crossover_rate 0.8                  # rate of crossover
ga_window_size 10                      #
ga_cauchy_alpha 0.0                    # Alpha parameter of Cauchy
distribution
ga_cauchy_beta 1.0                     # Beta parameter Cauchy
distribution
set_ga                                 # set the above parameters
for GA or LGA
sw_max_its 300                         # iterations of Solis & Wets
local search
sw_max_succ 4                          # consecutive successes
before changing rho
sw_max_fail 4                          # consecutive failures
before changing rho
sw_rho 1.0                             # size of local search space
to sample
sw_lb_rho 0.01                         # lower bound on rho
ls_search_freq 0.06                    # probability of performing
local search on individual
set_psw1                               # set the above pseudo-Solis
& Wets parameters
unbound_model bound                    # state of unbound ligand
ga_run 100                 # do this many hybrid GA-LS runs
analysis                               # perform a ranked cluster
analysis
```

Vina Configuration File

```
receptor = 4jrg_rigid.pdbqt
flex = 4jrg_flex.pdbqt
```

```
ligand = 4jrg_lig.pdbqt

out = out.4jrg.pdbqt

center_x = 11.493
center_y = 10.5
center_z = 2.759

size_x = 27
size_y = 27
size_z = 27

exhaustiveness = 12
```

Appendix B

NAMD Configuration File: 4ipf

```
#############################################################
## JOB DESCRIPTION                                         ##
#############################################################

# Minimization and Equilibration of
# MDM2 in a Water Sphere


#############################################################
## ADJUSTABLE PARAMETERS                                   ##
#############################################################

structure          4ipf_ws.psf
coordinates        4ipf_ws.pdb

set temperature    310
set outputname     4ipf_ws_eq

firsttimestep      0


#############################################################
## SIMULATION PARAMETERS                                   ##
#############################################################

# Input
paraTypeCharmm        on
parameters            par_all27_prot_lipid.inp
temperature           $temperature
```

```
# Force-Field Parameters
exclude             scaled1-4
1-4scaling          1.0
cutoff              12.0
switching           on
switchdist          10.0
pairlistdist        14.0


# Integrator Parameters
timestep            1.0  ;# 1fs/step
rigidBonds          all  ;# needed for 2fs steps
nonbondedFreq       1
fullElectFrequency  2
stepspercycle       10


# Constant Temperature Control
langevin            on    ;# do langevin dynamics
langevinDamping     1     ;# damping coefficient (gamma) of 1/ps
langevinTemp        $temperature
langevinHydrogen    off   ;# don't couple langevin bath to
hydrogens


# Output
outputName          $outputname

restartfreq         500       ;# 500steps = every 1ps
dcdfreq             250
outputEnergies      100
outputPressure      100


############################################################
## EXTRA PARAMETERS                                       ##
############################################################

# Spherical boundary conditions
sphericalBC         on
sphericalBCcenter   30.3081743413, 28.8049907121, 15.353994423
sphericalBCr1       26.0
sphericalBCk1       10
sphericalBCexp1     2


############################################################
## EXECUTION SCRIPT                                       ##
############################################################

# Minimization
minimize            100
```

```
reinitvels          $temperature

run 5000000 ;# .5 ns
```

NAMD PGN File

```
package require psfgen
topology top_all27_prot_lipid.inp
pdbalias residue HIS HSE
segment U {pdb 4ipfp.pdb}
coordpdb 4ipfp.pdb U
guesscoord
writepdb 4ipf.pdb
writepsf 4ipf.psf
```

NAMD Script for Water Sphere

```
### Script to immerse MDM2 in a sphere of water just large enough
### to cover it

set molname 4ipf

mol new ${molname}.psf
mol addfile ${molname}.pdb

### Determine the center of mass of the molecule and store the
coordinates
set cen [measure center [atomselect top all] weight mass]
set x1 [lindex $cen 0]
set y1 [lindex $cen 1]
set z1 [lindex $cen 2]
set max 0

### Determine the distance of the farthest atom from the center
of mass
foreach atom [[atomselect top all] get index] {
  set pos [lindex [[atomselect top "index $atom"] get {x y z}] 0]
  set x2 [lindex $pos 0]
  set y2 [lindex $pos 1]
  set z2 [lindex $pos 2]
  set dist [expr pow(($x2-$x1)*($x2-$x1) + ($y2-$y1)*($y2-$y1) +
($z2-$z1)*($z2-$z1),0.5)]
  if {$dist > $max} {set max $dist}
  }

mol delete top

### Solvate the molecule in a water box with enough padding (15
A).
```

```
### One could alternatively align the molecule such that the
vector
### from the center of mass to the farthest atom is aligned with
an axis,
### and then use no padding
package require solvate
solvate ${molname}.psf ${molname}.pdb -t 20 -o del_water

resetpsf
package require psfgen
mol new del_water.psf
mol addfile del_water.pdb
readpsf del_water.psf
coordpdb del_water.pdb

### Determine which water molecules need to be deleted and use a
for loop
### to delete them
set wat [atomselect top "same residue as {water and ((x-$x1)*(x-
$x1) + (y-$y1)*(y-$y1) + (z-$z1)*(z-$z1))<($max*$max)}"]
set del [atomselect top "water and not same residue as {water and
((x-$x1)*(x-$x1) + (y-$y1)*(y-$y1) + (z-$z1)*(z-
$z1))<($max*$max)}"]
set seg [$del get segid]
set res [$del get resid]
set name [$del get name]
for {set i 0} {$i < [llength $seg]} {incr i} {
  delatom [lindex $seg $i] [lindex $res $i] [lindex $name $i]
  }
writepsf ${molname}_ws.psf
writepdb ${molname}_ws.pdb

mol delete top

mol new ${molname}_ws.psf
mol addfile ${molname}_ws.pdb
puts "CENTER OF MASS OF SPHERE IS: [measure center [atomselect
top all] weight mass]"
puts "RADIUS OF SPHERE IS: $max"
mol delete top
```
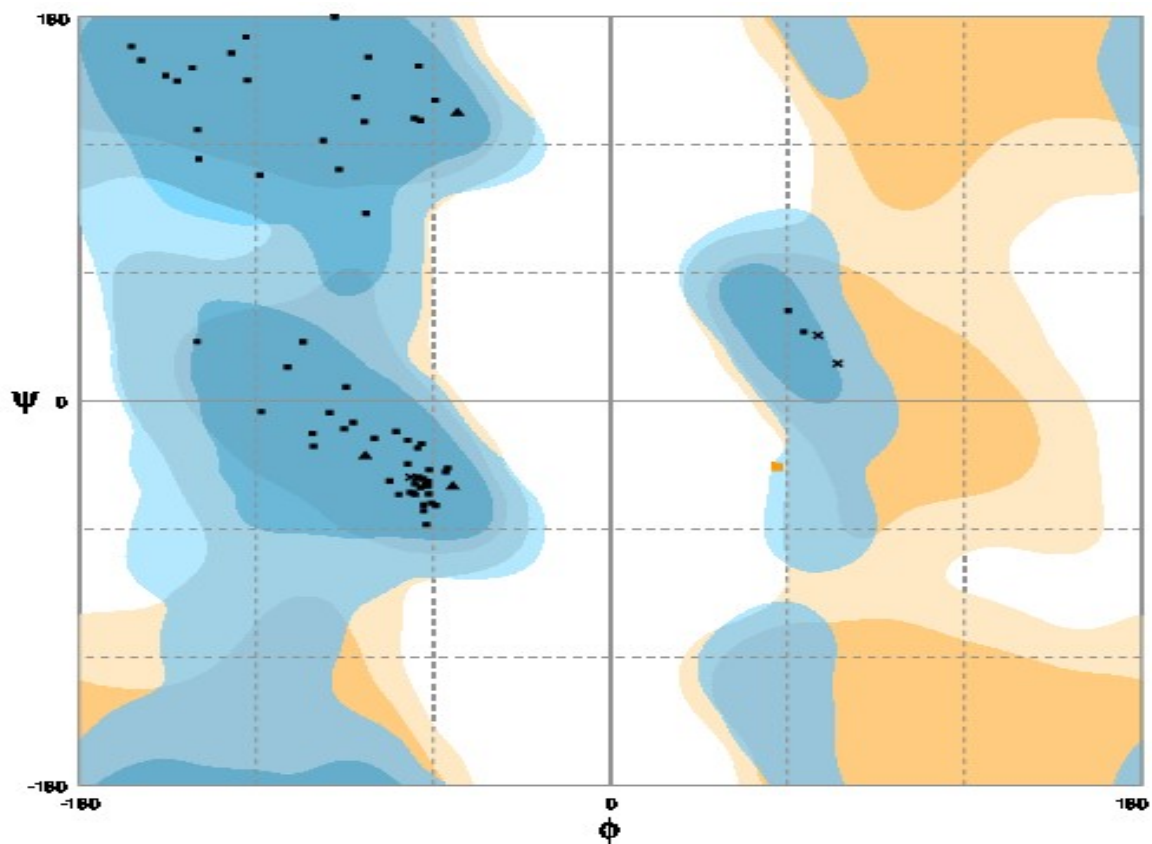
Appendix C

Rama Plot 3lbk (0,12)

Evaluation of residues
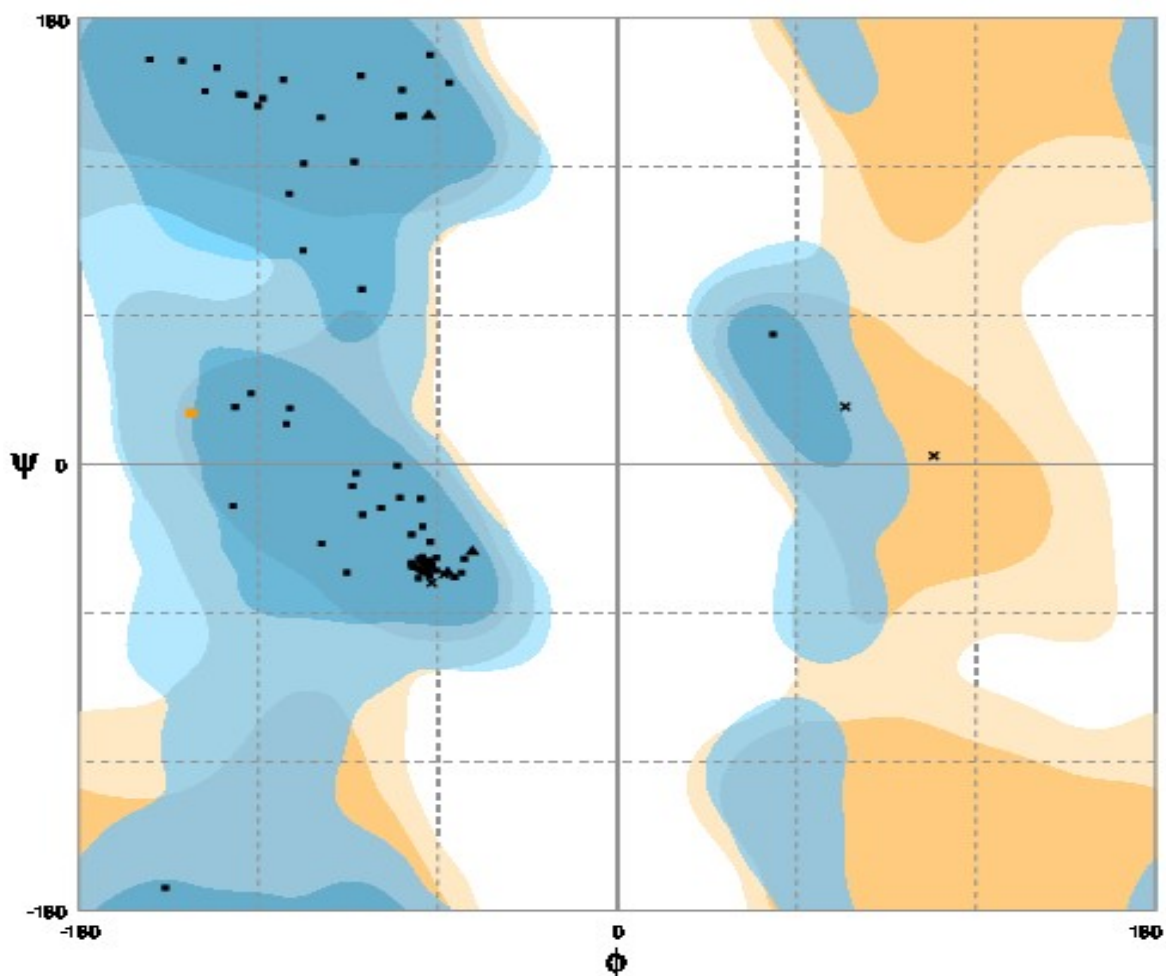Residue [A 72 :GLN] (56.75, -30.86) in Allowed region
Number of residues in favored region   (~98.0% expected) :   69 (98.6%)
Number of residues in allowed region   (~2.0% expected) :   1 (1.4%)
Number of residues in outlier region   :   0 (0.0%)

Rama Plot 4jrg (6,6)

Evaluation of residues

Residue [A 73 :CYS] (-142.26, 20.56) in Allowed region
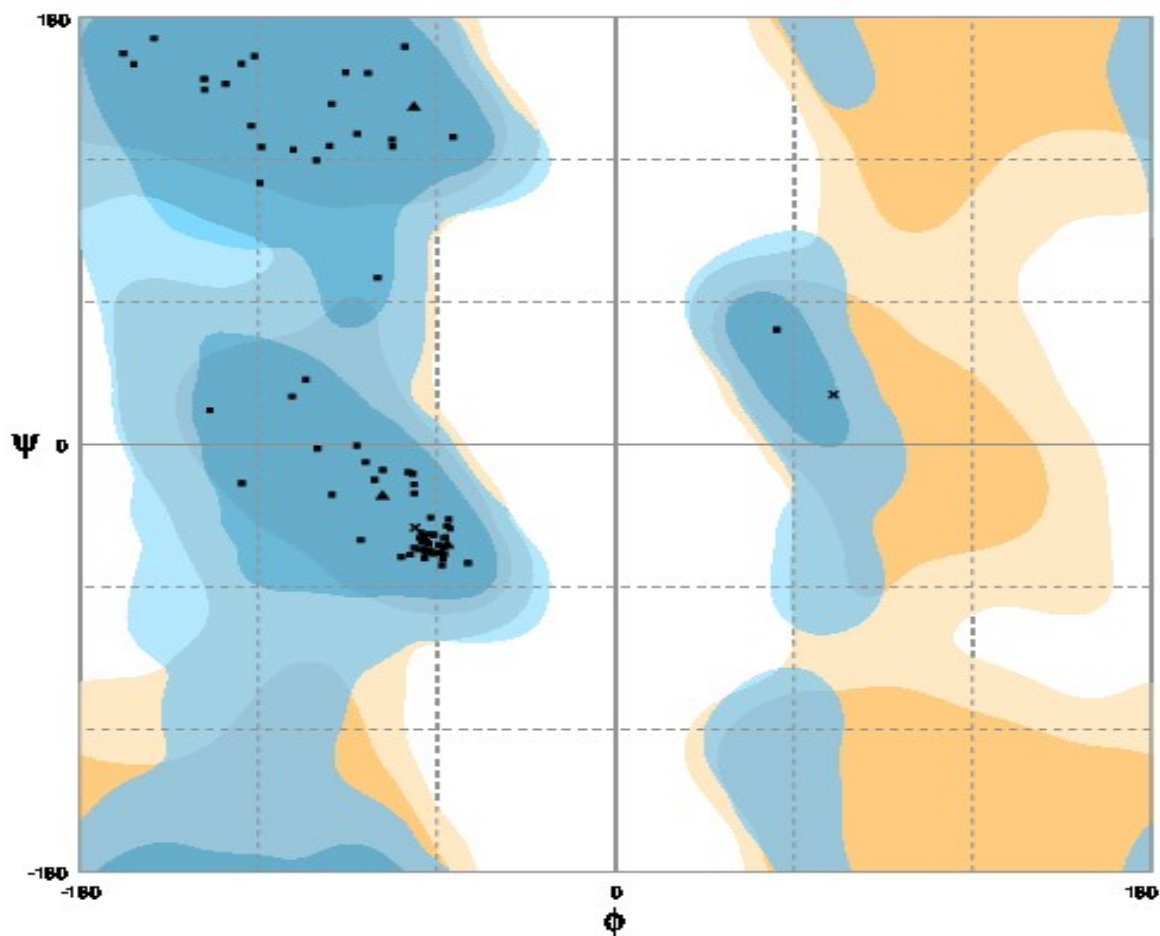Number of residues in favoured region   (~98.0% expected) :   72 (98.6%)
Number of residues in allowed region   (~2.0% expected) :   1 (1.4%)
Number of residues in outlier region   :   0 (0.0%)

Appendix E

Rama Plot 4zyi (0,12)

Evaluation of residues
Number of residues in favoured region   (~98.0% expected) :   77 (100.0%)
Number of residues in allowed region   (~2.0% expected) :   0 (0.0%)
Number of residues in outlier region   :   0 (0.0%)