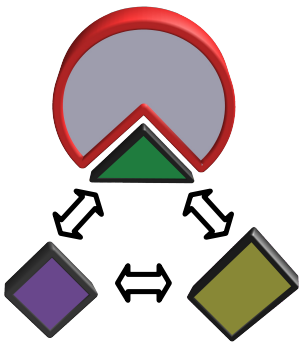
University of Southampton
Faculty of Natural and Environmental Sciences

Computing Free Energy, Binding and Competition within Fragment Based Drug Discovery

Ana Isabel Cabedo Martinez

PhD Thesis, May 2016



‘Calling the bravest hearts’

We Can Change **Heather Peace**

Abstract

The development of JAFS, a new computational method to study the binding geometries of small fragment molecules to protein cavities, estimate their binding affinities and analyse how they compete for a common protein binding site, all in the context of Fragment Based Drug Discovery, is presented in this thesis.

Fragment Based Drug Discovery is an approach to drug development which studies the binding of small ligands (fragments) forming high quality interactions with their target. Further optimization of these fragments into drug-like molecules, adding functionalities to increase affinity while controlling other relevant properties such as toxicity and absorption then takes place. JAFS studies the binding of fragments to their target proteins.

The JAFS method consists of the execution and analysis of Monte Carlo simulations of fragments (and waters) in the binding cavities of proteins with an added degree of freedom which accounts for the scaling of the interaction energy of the fragment (and water). Sampling of states at very low interaction energies gives a boost in fragment configurational sampling while competition between different fragments to remain at unscaled (high) interaction energies at a given binding site provides information on their relative binding affinities. JAFS is built on the JAWS formulation for water binding to protein cavities.

The performance of the JAFS method on a range of different test cases (T4 Lyzozyme, Major Urinary Protein I, Cyclin Dependent Kinase 2 and Heat Shock Protein 90) was studied. JAFS is divided in two protocols to rank fragments by affinity and locate binding geometries, respectively. The ranking of fragments by affinity to a common protein target was satisfactory (as compared to experimental data) for the simpler systems (T4 Lyzozyme and Major Urinary Protein I). However, more demanding systems proved problematic, where the ranking of nine different ligands to the binding site of Cyclin Dependent Kinase 2 provided results unrelated to experimental binding affinities.

Studying pose generation in sets of five repeats per simulation, the crystal binding geometry of every fragment studied was found in at least one of the repeats, without providing any previous information on the system (such as the presence or location of water mediated interactions or the hydration state of the cavity). Consistency between repeats was however found to be problematic and no method is currently able to select the optimal binding geometry among all the generated poses. Suggestions are given for further developments which would provide a methodology to rank poses.

Contents

List of Figures

List of Tables

Acknowledgements

In scientific research, as in any other area of life, it is impossible to achieve success without help and contribution from those around you. I will thank here some of those who were helpful companions during the development of my PhD.

To my supervisor, Prof. Essex, for the opportunity. I have learned so much, developed abilities that will be key in whatever my professional future holds for me. This was possible because I felt my opinions valued, and the pressure was less than the trust I knew was placed in me. To all collaborators in this project, for your work and your inspiration. From some of you I have learned so much. This project would not have been possible without you.

To the ProtoMS developers. I had the most amazing of times coding with you, discussing issues, sorting problems, deciding we were going to revolutionize the world with the one and only software, and then deciding it was anyway all for nothing because no one was ever going to use it. It was not one nor the other, but I am incredibly proud of what we have done (I think you know this). My thank you extends here to all new additions to the ProtoMS team of developers and contributors, I could not have hopped for anything better.

Thank you to the whole of the Essex group. You made me feel welcome from day one, and every day. I have learned from each and every one of you.

Thank you to those who have shared with me these three years. To those who have been there always and will never leave. To those who have become my day to day, my home to come back to, “the promise of a new day”. Thank you for the music.

List of Abbreviations

- **ADMET:** Absorption, Distribution, Metabolism, Excretion, Toxicity
- **BAR:** Bennett Acceptance Ratio
- **CDK2:** Cyclin Dependent Kinase 2
- **DD:** Drug Development
- **FBDD:** Fragment Based Drug Development
- **FEP:** Free Energy Perturbation
- **GCMC:** Grand Canonical Monte Carlo
- **HCS:** High Concentration Screening
- **HSP90:** Heat Shock Protein 90
- **HTS:** High Throughput Screening
- **JAFS:** Just Add Fragment moleculeS
- **JAWS:** Just Add Water moleculeS
- **MBAR:** Multistep Bennett Acceptance Ratio
- **MC:** Monte Carlo
- **MD:** Molecular Dynamics
- **MM:** Molecular Mechanics
- **MUP:** Mouse Major Urinary Protein I
- **SBDD:** Structure Based Drug Development
- **T4Lys:** T4 Lyzozyme
- **TI:** Thermodynamic Integration

Chapter 1

Introduction

Through history, all societies have had an interest in the process of healing and maintenance of health.[1] From the use of plant extracts and fungus based on observation and experience, thousands of years of research have brought us to the use of chemistry and other related disciplines in the development of new synthetic medicinal drugs for individuals to experience longer, healthier and more comfortable lives.

The aim of each drug development (DD) project is to create a molecule capable of generating a desired measurable biological effect. However, owing to the complexity of biological systems, measuring biological effects at every step of the drug development process is hardly feasible. Simplified systems are used, commonly studying the binding affinity of the compound under development to a desired target, as a proxy for its biological effect. The most accurate measure of the binding affinity is the free energy of binding. It is hence easy to understand that the aim of the drug development process, up until its last stages, is to find a drug with high binding affinity and few side effects.[2,3]

There are many strategies to drug development. Differences can come from the reasoning behind selecting a particular target, to the process of hit discovery or its development into the final drug.[4] However drug discovery strategies generally differ on the global understanding of whole stages of the drug development process, such as lead generation. Structure based drug discovery is a clear example.[5,6]

Structure based drug discovery (SBDD) aims at including structural reasoning behind every step of drug generation and optimization. While other drug develop-

ment strategies, such as high throughput screening, base their success on the large number of chemical compounds examined as potential candidates (be it potential hits in the screening process, or modifications to a hit during optimization); SBDD works on the idea of *fewer compounds, carefully selected*.[6] To proceed with this *careful selection*, structural information on the ligand and its binding cavity is required.[6] Fragment Based Drug Discovery (FBDD) is one of the purest forms of SBDD.

FBDD is based on the understanding of drugs as the sum of their chemical moieties, where each moiety is called a fragment.[7] These fragments will be screened, virtually or experimentally, in the search for hits (those that bind to the target).[8,9] While this is a common process, the main difference relates to the small size of fragments. Fragment hits are only one chemical moiety of the final drug. Informed decisions are hence taken to include each chemical modification, from hit to lead and final drug, improving binding affinity while controlling any undesired effects.[9–12]

Computational chemistry is appealing to the pharmaceutical industry for its potentially fast and low cost calculations.[13–15] It is often considered as a tool to increase the efficiency of the drug discovery process; an approach to focus the process on the most promising candidates. Given the investment required for a drug generation cycle, both in time and economic terms, efficiency boosts are in demand.[14]

Different areas within computational chemistry offer different opportunities and the trade off between accuracy of results and speed of the computational process generally applies.[16] From identifying possible targets of a given drug in the search for potential off-target effects to detecting new drugs for a given protein target, or selecting the best binder among a set of possible ligands, the field of computational chemistry in drug discovery offers a wide range of applications.[14] In terms of selecting the best binder among a range of chemical compounds, concurrently detecting the correct binding geometry, docking and scoring is possibly the best known approach.[17] However, the performance of docking and scoring studies varies depending on the exact conditions under study and the systems included in the evaluation (see section 2.4.1). Scoring functions, in particular, are often found to generate unreliable results.[18] More rigorous (and computationally expensive) alternatives to docking and scoring can be found in free energy calculations.[19,20]

These can be the optimal choice if the set of potential binding molecules is small, or the accuracy requirements are high.

Owing to their small size, fragments are expected to have a low binding affinity, even when each of their atoms is involved in efficient binding interactions. It is in a challenging context like this, that methodologies like out-of-the-box docking and scoring are expected to be outperformed by those more sophisticated, such as free energy based methods.[21,22]

Alchemical transformations are computational methods available to accurately calculate the relative binding affinity between two ligands. However, they require previous knowledge of binding geometries, while generally assuming low variability of these geometries with time.[2,23]

Throughout this thesis we will present the JAFS methodology, which provides a flexible twist to classical alchemical transformations. Developed in the context of FBDD, JAFS focuses on increasing sampling of fragments in protein binding cavities, providing information on where, and how tightly, these fragments bind. This method offers the possibility of locating the binding pose of a given fragment molecule, automatically taking into account any potential water mediated interactions. It should also ideally estimate the relative binding affinity of different fragments without previous knowledge of their exact binding geometry. However this possibilities come at a cost, and our binding affinities must be considered estimates, rather than true relative binding free energies between potential binders.

The rest of the thesis may be divided in the initial chapters, providing background (chapters 2 and 3), methodology chapters providing all information required to follow the subsequent studies (chapters 4 and 5), the results chapters (chapters 6 and 7), future work (chapter 8) and the conclusion chapters (chapters 9 and 10), plus the Appendix (chapter 11). During the background chapters, the FBDD context in which JAFS is developed will be presented, and the required theoretical framework to understand the method will be explained, together with a few methods which might be considered *the competition* of the JAFS methodology. The methodology chapters will detail the JAFS methodology, both of its protocols, and its theoretical explanation, as well as the range of systems used in the subsequent studies, with their peculiarities and reasons for being selected. The results chapters will include results obtained during the development of the JAFS method, justi-

fying the choices and changes performed to the implementation of its predecessor, the JAWS method, as well as the results of the simulations performed to test the performance of the finalized JAFS method on a range of systems. The future work chapter will describe the most relevant path in which further development could benefit the outcome of JAFS simulations. The conclusion chapters will summarize the main lines of the analysis presented during the analysis of the JAFS simulations on the test systems as well as conclude on the goals achieved and current state of the JAFS method in context. The appendix will contain information which may be relevant but did not flow naturally within the presentation of the thesis, nor was required to its understanding.

Chapter 2

FBDD

The fragment approach to the development of new drug candidates, or fragment based drug discovery (FBDD), made its first appearances in the pharmaceutical industry during the 1990s.[11,24] Since then, a number of pharmaceutical companies decided to invest in this promising approach, with some even being exclusively dedicated to FBDD.[9,11,24] As a result, many success stories can be found in publications over recent years (where citations are only a few examples).[25–31]

Figure 2.1: Schematic representation of a drug size molecule (top) and its fragment components (bottom)

2.1 Concept

The idea of FBDD originates from early works on additivity of binding energies and the affinities of individual substituent groups.[7] If we can understand a drug-sized molecule as the sum of its smaller chemical components, we could study each of these components separately in our search for the optimal global binder. Building a drug from its components offers greater control on the chemical composition of the final molecule (see section 2.1.1 for further advantages of this approach).

While we cannot assume additivity of binding free energy between the chemical components of a larger molecule,[7] the binding of drug-like molecules to protein binding sites does generally revolve around certain key atoms, where a small chemical component forms strong interactions with the receptor.[9,32] FBDD is based on the idea of studying these small components rather than the final drug-like molecule, finding the most promising binders, and then adding further chemical moieties until the desired properties are achieved. Figure 2.1 illustrates the concept of small components forming a complex drug.

The previous paragraphs provide an overview of the ideas behind FBDD, but concepts need to be defined in practice. Next we will consider what we understand by each of these "chemical components" and the routes followed to transform them into drugs.

The best practical definition of these chemical components ("fragments" from now on) can be obtained by limiting the chemical properties that these molecules should obey. Following this viewpoint, the "rule of three" was proposed.[33] It sets a series of values for relevant properties which are considered as maximum limits to which fragment molecules must be constrained in order to fulfil their role of a minimal starting structure. These include a basic core with limits on the molecular weight, number of hydrogen bond donors and acceptors as well as a measure of hydrophobicity. Further useful criteria are suggested on number of rotatable bonds and surface area.[33] The limits assessed within the rule of three can be seen in table 2.1. The "rule of three" is built on the same basis as the previously defined and widely known "rule of five" for drug-like molecules.[34] It is important to note that while the "rule of three" is the most commonly accepted set of boundaries to the definition of a fragment, different pharmaceutical companies working

Property	**Limit**
Molecular weight	< 300
Hydrogen bond donors	≤ 3
Hydrogen bond acceptors	≤ 3
ClogP	≤ 3
Rotatable bonds	≤ 3
Polar surface area	≤ 60

Table 2.1: Limits set to properties in an effort to define fragment molecules within the Rule of Three[33]

in the field of FBDD apply slightly different limits, hence the "rule of three" must the taken as guidance rather than a strict set of rules that cannot be broken.

Within a FBDD procedure, a set (library) of fragment compounds are screened (see sections 2.2 and 2.4) generally searching for a compound binding to the macromolecular target of interest. The fragments found to bind to the target during screening are commonly referred to as hits. Different properties of these initial hits must then be improved, namely affinity towards target, cell activity and properties related to administration, toxicity and metabolism of the compound when administrated as a drug (ADMET properties[35]). This is done in the process generally called optimization, as well as "fragment to lead" or "fragment to drug". It can be understood that during the development of the initial hit into a lead ("fragment to lead") the increase in affinity towards macromolecular target and / or cell activity mainly prevails. During the "lead to drug" stage, other properties related to the drug administration would then be improved. Examples of FBDD processes can be found in the literature which fall within this two-stage definition.[25,36] This definition of stages, while useful in the understanding of the drug development process, is however arbitrary and does not need to be fulfilled. The processes followed to develop the fragment into the lead or drug molecule are described in section 2.2.2.

While different processes to develop drugs from fragments may be followed, all of them benefit from structural information of the macromolecule-fragment complex. This is linked to the concept of FBDD itself, as well as its advantages (see section 2.1.1). While starting from smaller, less tightly bound molecules, and hence often requiring a longer process of drug development from the initial hit, FBDD bases its successes on the level of control available to the researchers on the final structure and properties of the final drug.

It had often been observed that many conventional drug development projects failed in their latter stages, due to the drug's poor ADMET properties, or simply that it was not possible to improve affinity of the initial compound found to bind during screening (hit), while maintaining reasonable ADMET properties.[35] These observations led to the inclusion in the libraries of compounds selected for "lead-like" properties, rather than "drug-like" properties. That is, the required room for optimization, the process which more often than not requires an increase in molecular size and hydrophobicity, was taken into account in the compounds screened —

they were made smaller and less lipophylic. FBDD takes this idea a step further; the compounds screened are smaller and less hydrophobic than "lead-like", leaving far more room for subsequent modifications, and control of the resulting structure and properties, in the development of drugs from fragments.

The change in perspective regarding what we are looking for during the screening process needs to be accompanied by a change in properties used for selection of hits. If we keep compounds with higher affinity towards their target, it is more likely that we will keep the biggest compounds, where more chemical groups are available to interact with the target macromolecule. What should then be the measure that corresponds to the properties in the screened molecules that will provide a better drug once optimized? One of the options most commonly used is the concept of ligand efficiency. Ligand efficiency can be simply defined as binding affinity divided by the number of ligand heavy atoms.[37,38] This definition will be used throughout the thesis, while other measures are available.[39,40]

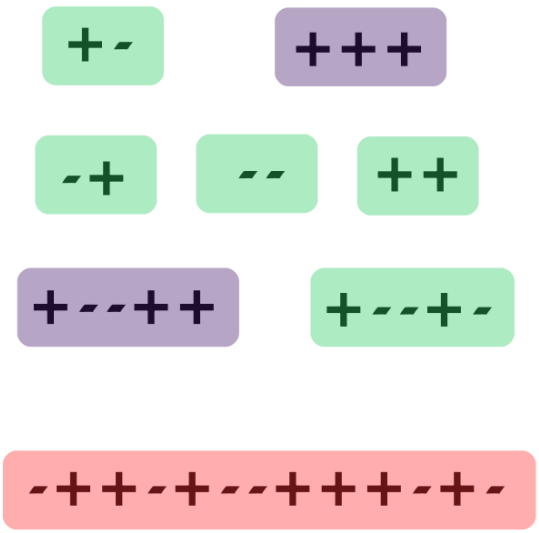


Figure 2.2: Representation of the complexity model according to its description on its initial publication. In the image, the orange string of pluses and minuses represents the protein binding cavity. The ligands are shown as shorter strings of pluses and minuses. The ligands shown in green are those which would be defined as "binders" according to the model and in purple are the "non-binders". Each plus or minus corresponds to one chemical functionality in the ligand or protein cavity and correct complementarity is that of a "+" with a "-".

So far the appearance of FBDD strategies has been analysed from the perspective of the further evolution of the hits into the final drugs. However, the "hit

finding" (screening) stage also benefits from the FBDD approach. The unachievable perfect screening procedure would be that which screened all possible binding compounds, finding only the tightest binders. If this could be done, no further optimization would be required, since the perfect binder would be detected from the beginning. Without aiming to such extremes, we would like to aim for the highest efficiency in sampling the chemical space, being able to detect binders by screening the lowest number of compounds possible.

A very simple model representing the likelihood of binding for ligands in terms of their size shows that, as long as no condition is applied on the number of possible binding configurations, the smaller the size of the ligand, the higher the probability of the ligand *binding* (where *binding* is understood as fulfilling all possible complementary interactions between ligand and protein simultaneously).[41] It is important to understand that this simple model assumes all binding can be detected, without taking into account the binding affinity, which is expected to decrease with decreasing ligand size. Equally, it does not consider the possibility of mismatches (that is, all possible interactions of the ligand must be satisfied), or ligands not necessarily employing all their functionalities on binding. A representation of this model is shown in figure 2.2. This simple model is called the complexity model, and a review on it was written ten years after its first publication.[42] Validations of the model have produced a range of different results, and it can be understood that the outcomes of such validations will depend on the properties of the compounds used in the study. Hydrophobic ligands, for example, are expected to bind tighter and to a wider variety of targets with increasing ligand size.[41–44]

Despite all their limitations, complexity studies highlight the basic reasons for the better screening of chemical space with fragments than drug-sized molecules. Simplistically, if we imagine a receptor with three binding functionalities, and an example where fragments would consist of only one functionality and drug-sized molecules are formed by three functionalities bound together, to find binding, fragments would only have to contain one of the complementary functionalities to the receptor, while the drug-sized molecules would have to present all complementary functionalities as well as presenting them in the correct bound order.

2.1.1 Advantages

The main advantages of FBDD have been implicitly described through the introduction of the FBDD concept and its reason for emerging. Here they will be summarized and clarified:[11]

- High quality interactions. Given the same affinity for a drug sized molecule and a fragment, the affinity of the drug may be achieved through a large number of weak interactions, while the fragment, due to its small size, can only reach that affinity through a small number of high quality interactions. This correlates with the high ligand efficiency of fragment binders (see above).

- Better screening of chemical space. As explained previously, the lower the number of functionalities within a ligand, the more effective the screening is expected to be, in terms of the number of ligands required to find a binder to target. In a study of the chemical space of organic molecules of up to eleven atoms of carbon, nitrogen, oxygen and fluorine, 26.4 million compounds where found, while only 1850 compound where found with up to 6 atoms of the same species.[45]

- Greater control of the structure of the final drug. Since the process of drug development in FBDD starts with a small sized hit, and the addition of each of the extra functionalities is made as a conscious decision, the majority of the structure and associated physicochemical properties of the final drug are controlled through the development process. This is expected to yield better ADMET (such as lower hydrophobicity, where hydrophobicity is linked to unspecific binding and side-effects) and binding properties (ligand efficiency) of the final drug.

2.1.2 Drawbacks

In the previous section, it has been explained how the small size of fragments is associated with advantages of FBDD. The small size of the fragments is also the reason for its major drawback: the low binding affinity of fragments and associated difficulty of detection during screening.[46–48] Even with fragments efficiently exploring high quality interactions, the affinity arising is generally lower than that which

can be reached by a much larger number of medium quality interactions formed by a drug-sized (or lead-size) ligand. While the advantages in section 2.1.1 account for a high affinity of the final drugs developed from these initial weak binders, the detection of the hits is necessarily done at the fragment stage, giving a requirement to detect low affinity binders during the fragment screening process. This drawback conditions the methods used when studying the binding of fragments experimentally, as well as the application of computational methods (sections 2.2 and 2.4).

2.2 Experimental Methods in FBDD

2.2.1 Finding Fragment Hits

Owing to the low affinity of fragments towards their targets, detection of fragment hits through standard screening procedures such as HTS (high throughput screening)[49] in their standard conditions is unlikely. As an example, biological assays would require higher ligand concentration to detect binding of lower affinity ligands.[50] Consequently, the level of affinity towards which the screening for FBDD is calibrated must differ. FBDD follows a highly rational perspective where the number of chemical groups which need to be added to the initial hit is often high (compared to other drug development strategies) but the resulting molecular structures are consequently under tight control.[11] Standard drug discovery protocols tend to start the development at the "lead" stage, closer in size to the fully functional drug, given their search for high affinity hits during screening. Consequently, much of the structure of the final drug is provided by the initial hit. From these two perspectives follows that, usually, the activity of the hit compound in the traditional approach needs to be relatively high (since a lower number of functionalities are to be added), while this is not a concern in FBDD where further cycles of compound modifications are expected.[11]

Consequently, two different perspectives may be adopted to achieve detection of fragment hits: a higher concentration of potential binders must be applied over traditional methods, or different, specific methods need to be applied to fragments. We will first present the former idea below, and then proceed to explain several of the specific techniques that may be applied within FBDD, together with some of

the associated advantages for each.[9,51]

High Concentration Screening

While, owing to the lower affinity of fragment hits with respect to lead-sized hits, fragments are not likely to be detected in a biochemical assay given the same conditions, a higher concentration of the weaker binder should compensate and allow detection. This is commonly known as High Concentration Screening (HCS). A variety of assay types have been applied to FBDD. Depending on the character of the assay, either sensitivity of the detection technique must be improved to be able to account for the weaker binding of fragments (see the use of X-ray crystallography as a sensitive detection technique within HTS for fragments[52]), or the concentration of ligand must be increased in biochemical assays[53] (note that this limits its applicability to fragments with solubility higher than the screening concentration[54]).

Structure-based Screening

One of the approaches to tackle the required increase in sensitivity due to the low affinity of fragments is to change the screening method from biological assays (which are typical in HTS) to biophysical techniques, which offer greater sensitivity.[11] One of these biophysical techniques (X-ray crystallography) has already been mentioned in the previous subsection, which highlights the overlapping of these two concepts, as biophysical methods can be incorporated within HTS.[52] In this subsection we will focus on discussing the biophysical methods which, besides providing high sensitivity, offer structural information on the ligand-protein binding geometry.

As explained in section 2.1, FBDD may be defined as a subclass of Structure Based Drug Discovery (SBDD). FBDD inherits from SBDD the rationale behind the development process. In principle, the full sized final drug could be obtained from the initial fragment by non-structural approaches such as dynamic combinatorial chemistry.[55,56] However, this limits the control of the structure and chemical composition of the final drug. To take full advantage of the FBDD approach, structural information on the binding mode of the fragment is often needed. In the cases where a non structure-based screening method is followed, a structure determina-

tion technique may be applied *a posteriori*. However, structure-based screening avoids the intermediate stage and provides structural information through the same screening process. Different variants of structure-based screening of fragments will be briefly presented below.[9,51]

- X-ray crystallography. X-ray crystallography is a common technique to study conformations of macromolecules on their own or complexed with bound ligands. In its more traditional approach, each of the potential binders would be put in solution with the target, the complex would be formed in the cases where favourable interactions may be expected, and a crystallization process would then take place. However, taking into account that we are aiming to use this method for screening purposes, a faster process is desired. This is obtained with fragment soaking into crystallized macromolecular targets. Solutions of 3 to 10 fragments are left in contact with the target and the electron density of the resulting complexes analysed.[57] Note that the structure of the protein is studied in a rigid (crystallographic) conformation.

- SAR by NMR. The other typical method to obtain structural information from macromolecules is Nuclear Magnetic Resonance (NMR). This method studies the structure of proteins in solution, where they often present flexibility. Information on the structure of complexes between protein and bound ligands may be obtained with this technique. Specific *2D* NMR methodologies are required for the study of macromolecules where the behaviour of more than one atom species (i.e. carbon and hydrogen) in the system needs to be studied. In 2D NMR, peaks are generated which may be associated with each of the residues in the protein. Upon ligand (fragment) binding, changes can be observed in some of the peaks, representing a change in their chemical environment. Following this procedure, the binding region and binding affinity of the ligand can be determined. The application of 2D NMR techniques to the study of fragment binding in this manner is referred to as Structure-Activity relationship by NMR, and its development is considered to be one of the key steps at the beginning of FBDD.[9,48]

Other Biophysical Methods

The structure-based screening methods are not the only biophysical methods available with sensitivity specifically applicable to affinities of the order of those expected for fragment hits. It may be considered that information on the binding configuration is only relevant for those binders which will be developed further into leads. Obtaining structural information during the screening process may hence not be considered a requirement, and structural techniques will then be applied *a posteriori* exclusively to fragments of particular interest. In that case, a wider range of techniques is available for the detection of fragment binding.

One of these techniques is 1D NMR — NMR measurements of the ligand, rather than the protein. As mentioned in the previous subsection, the analysis of the protein structure (or changes in its chemical environment) through NMR require complex 2D methods where more than one atom species (such as carbon and hydrogen) is studied at once. Much simpler is the NMR study of small molecules such as fragments. Ligand binding to a macromolecule produces changes in ligand properties which can be captured by NMR: upon binding, a change (loss of signal intensity) can be observed in the ligand NMR spectra.[9]

Other available methods include surface plasmon resonance,[58] mass spectrometry[59] and Isothermal Titration Calorimetry (ITC),[60] all of which require high fragment concentrations, which may make them more prone to artefacts.[9]

Often, several of the described techniques are combined to confirm fragment hits as well as to characterize their binding.[9,59]

2.2.2 From Fragment to Drug-sized Molecule

This thesis focuses on the study of fragment binding to their target, rather than the evolution of fragments into full-grown drugs. However, understanding the basic ideas behind this process is key to extracting conclusions on the best methods for studying fragment binding, as well as the requirements for the development of new methods.

There are different strategies for developing drug-sized molecules from the initial fragment molecules, and each will be briefly presented below.

- Linking fragments. In some cases, the screening of fragment binders to a par-

ticular target protein, may provide results regarding several fragment binding at once. Examples of such screening methods, which also provide structural information are X-ray crystallography with fragment soaking (as described in section 2.2.1) or computational screening techniques (described in section 2.4). In other cases, the structural information for several fragment binders may be obtained independently. When different fragments are known to bind to different subpockets within the same binding cavity, a logical approach to increase affinity is to link them together through a chemical linker to produce a bigger compound expected to bind tighter. While this approach has generated successful examples of hit to lead development in FBDD,[61–63] it is important to understand that the affinity of the linked compound cannot be predicted exclusively from that of the individual fragments.[7] Linking both fragments may increase affinity by decreasing the entropy of the (unbound) ligand in solution.[64] Besides this, the linker itself may have favourable or unfavourable interactions with the target.

- Growing fragments. Starting from the fragment hit, chemical functionalities may be added to the initial fragment structure to increase its affinity. The decision on which functionalities are added relies on the researcher, who is expected to avoid groups with associated toxicity or undesired side effects, as well as favour those which provide desired absorption / administration qualities. When structural information on binding is available, the decision on the added chemical functionalities is guided by the functional groups in the binding cavity which are available to interact, and the fragments may be grown keeping a high binding efficiency towards the target. This approach has proven successful in lead generation.[25,65–67] This manner of generating leads from initial fragment hits is sometimes divided into fragment evolution (as the process of modifying the fragment to improve affinity) and fragment optimization (referring to the modifications which improve other properties of interest, such as toxicity or absorption).[10]

- Alternative approaches can be found, for example, in fragment self-assembly, where fragments are allowed to react in the presence of the target, hopefully forming a bigger, tighter binder molecule within the macromolecular binding cavity.[10]

As can be seen from the above, the availability of structural information makes either of the chosen paths easier. This is because of the need to take decisions, such as which fragments to link, which linker to apply and at which particular points those should be linked (within fragment linking) or which functionality should be applied at which atom of the fragment (within fragment growth). Often, the bound structures of fragments different from those which are chosen to proceed into the lead optimization stage are used to guide the decisions of the functionalities or linkers to be added. This fact favours the use of techniques such as soaking of fragments into crystallized proteins where the binding configurations of many binders may be obtained.

This process of developing drugs or leads from initial fragment hits is presented here in the section on experimental work in FBDD, but would apply in the same manner to fragments which have been discovered as binders through computational methods. The rest of this report will focus on the study of the small fragments binding to the target, without considering in any depth the further development of these fragments.

2.3 Success Stories

When talking about success stories for FBDD, Plexxikon deserves the main mention for the first drug approved by the FDA (U.S. Food And Drug Administration), Vemurafenib (PLX4032).[11,68,69] While this is currently (to our knowledge) the only drug developed through FBDD which has been approved, more than 30 are (or have been) in the clinical trials stage.[70] As well as the number of drugs in development, research in the field of FBDD has also been increasing over the last ten years, with the number of publications in 2015 with the phrase "fragment based drug discovery" being more than five times those in 2005.[71] With all this, FBDD is still a young field, with the first paper acknowledging this approach appearing less than twenty years ago, in 1997.[72,73]

A few of the stories of fragment based drug development that have generated successful results will be presented below.

2.3.1 Plexxikon working on Vemurafenib

The studies that led to the discovery of PLX4032 started with a high concentration screening (HCS) with enzymatic (kinase) assays, of 20000 fragment-sized compounds. Of these, 238 were found to inhibit the activity of the three kinases used in the enzymatic assays, and were hence subjected to co-crystallography. One of the co-crystal structures revealed 7-azaindole binding to the ATP binding site of the kinase, with weak affinity and multiple different binding modes. A group of monosubtituted 7-azaindoles were then synthesised, finding a 3-aminophenil analogue which bound in a single configuration and with increased affinity.

This scaffold was studied on a range of different kinases, finding a consistent binding mode within the protein family as well as promising optimization sites on the ligand. Consequently, libraries of mono and disubstitued analogues of 7-azaindole where built and screened, finding a set of compounds containing a difluoro-phenylsulfonamide motif which presented excellent potency against oncogenic B-Raf and selectivity with respect to wild-type B-Raf and other kinases. These compounds were co-crystallized with $B - Raf^{V600E}$. Based on the structure of the complex, subsequent rounds of optimization lead to PLX4720, which inhibits $B - Raf^{V600E}$ at a 10 times lower concentration than required to inhibit wild-type B-Raf.[11,74] Later on during animal testing, a modified version of the compound with an added phenyl group, PLX4032, was chosen for its improved pharmacokinetic properties on animals.[68,69]

2.3.2 Astex working on HSP90 inhibitors

Astex studies on inhibitors of Heat Shock Protein 90 (HSP90) started with a 1D NMR screen (see section 2.2.1) against the N-terminal domain of HSP90[75] of approximately 1600 compounds from fragment libraries. The most successful compounds from this screen where taken to a further NMR study of competition with the native (and weak) binder ADP to select those fragments binding to the ATP/ADP binding pocket. Only the binders with a certain level of affinity towards this pocket were then selected by adding Mg^{2+}, which increases the affinity of ADP displacing the weaker fragment binders.

Based on the NMR data and considering chemical diversity, 125 fragments were taken into the crystallographic stage, and 26 fragment crystal structures

were obtained, which spanned a variety of chemotypes. Among these, four fragments were found of particular interest, three of which bound within the same subpocket, through conserved bridging water molecules, and one which was found to co-crystallize with the smaller of the other three, in a different pocket region.

Two of the fragments which bound through conserved waters where chosen for optimization. These fragments were also used at different stages (development, see section 6, and production, see section 7) in this project. Here we will follow the optimization of fragment one (which can be seen in section 6) and briefly summarize that of fragment three (see section 7), which ended up in the generation of a AT13387, which reached the stage of clinical trials.[36]

Optimization of Fragment One

The fragment chosen in this optimization strategy is formed by a pyridine ring bound to a pyrimidine, where the bond between these two rings is twisted (see section 6.2.2 figure 6.2 (a)). Torsion profiles and crystal structures of small molecules suggest that the optimal geometry of these rings is close to planarity, hence it should be possible to improve affinity through stabilization of the bound conformation.

With this in mind, the optimization started with the virtual screening of close analogues, finding a chloro analogue to the initial fragment hit which provided a 100 fold increase in affinity. The substitution of the phenyl ring with small groups, led to analogues which stabilized the bound structure, as well as filling a lipophilic side pocket. In particular, a 2-methoxy analogue led to a 5-fold improvement in affinity. While this compound did bind in the desired conformation, further studies, including that of the torsion energy profile led to the addition of another methoxy group which stabilized further the desired rotational geometry, providing a 5-fold gain in affinity. A 2-chloro analogue was then found to increase potency by reducing the energy difference between local and global minima of the torsion profile. This compound was the most ligand efficient from the initial synthesis. In the next iteration, a chlorine substitution was found to significantly increase activity by means of additional lipophilic interactions with HSP90. Solubilizing groups were then introduced to improve cell activity, leading to a compound with low micromolar activity in cell assays.[25]

Optimization of Fragment Three

From the fragment three hit (2dl in table 11.1), addition of isopropil groups to fill the proximal lipophilic pocket lead to a 100 fold increase in potency. Focus was then on replacing the diethylamide, with the aim of displacing a Lys sidechain with larger tertiary amides which might then form good hydrophilic interactions with the protein. This gave compounds with several hundred times improvement in affinity, with the Lys indeed displaced and forming a salt bridge. An extra hydroxyl group was then added to form a hydrogen bond with an adjacent protein residue, an idea driven by the similarity to the natural compound radicicol. The final compound from this optimization shows subnanomolar affinity, excellent ligand efficiency and good cell activity, and was chosen as the lead compound.[25]

Further optimization of this lead compound was focused on the physiochemical and pK_a properties, and lead to the generation of a compound that went into clinical trials: AT13387.[36]

2.4 Computation in FBDD

The binding affinity of a ligand to a macromolecular target is of interest in a number of stages of the fragment based drug discovery process. From the initial screening of a large set of fragments to the analysis of the effects of modifications applied on the initial hit, calculations of binding affinity are often performed and the time required to obtain these data and the accuracy of the results vary in their importance as the process of drug development progresses. At the beginning of the drug development process, when a large number of compounds are considered, it may be practically impossible to spend a large amount of time in obtaining information on each of them. However, in later stages, when only a few ligands are being developed, small affinity differences may be key to select the most optimal binder.

The required binding affinity results could exclusively be obtained experimentally, but computational methods offer a generally faster and less expensive alternative. More often than not, however, computational calculations are used, not as an alternative *per se*, but as a preparatory step, which allows the experimental studies to focus on the most promising candidates at every stage of the develop-

ment processes. They can also provide extra pieces of information (such as binding geometries) and allow for the flexibility associated to the lower economical expense of attempting new developmental pathways within computational methods. While diverse in their resource and time requirements as well as the theoretical background on which they are based, a key generality of computational methods to study binding affinity is their generation of structural information for the expected binding mode.

It is important to highlight, however, that, the uses of computational tools within FBDD are not limited to the study of binding affinities; computational tools can be involved in the generation of fragment libraries, as well as the exploration of the chemical space offered by replacement of substituents to a given scaffold.[76,77]

As previously stated, the focus of this project will be on methods to study binding configurations and relative affinity of small ligands (fragments) within the drug development process.

We may define a scale regarding computational methods to study the binding of ligands (particularly fragments) to macromolecular binding sites — of which, from now on, we will focus exclusively on proteins. The scale measures the time and resources required to proceed with the calculation of affinity and / or binding geometry of the fragment, where a lower cost (in both time and resources) is generally associated with a less theoretically accurate representation of the system. Within this scale, docking and scoring methods and free energy techniques would lean towards different ends of the spectrum, particularly the low cost (fast) end and more accurate representation end, respectively. While the theoretical background on which these methods are based will be described in chapter 3, here we will present their applications to fragment based drug discovery, what advantages and disadvantages they can provide to the field and any developments or techniques of particular interest to FBDD.

2.4.1 Docking and Scoring

Docking is a method to obtain possible binding configurations of a ligand to a target binding cavity, while scoring is the manner in which different binding configurations or different ligands are assigned an estimate of their affinity towards the target, as well as ordered in terms of their relative affinity estimates. While

docking and scoring can be defined independently, they are often understood as two parts of a common method, since the scoring function is applied directly on the poses obtained by docking, deciding which poses are presented in which order as an output of the docking run. While further information on docking and scoring can be obtained from section 3.2, suffice to say here that the estimations used to generate possible configurations (i.e. study of the available interactions between fragment and ligand) as well as those employed to calculate the affinity with the scoring function are approximations rather than strict definitions of affinity based on physically robust derivations. Their focus is on the speed at which the binding configurations and relative affinity estimates can be produced, with the key aim in mind that among the set of fragments predicted to bind with higher affinity, an enrichment of actual binders is found, so that subsequent experimental (or more theoretically robust) study of that set of compounds is more efficient that direct screening of the initial fragment library.

Docking and scoring was applied to the drug discovery process before the rise of the field of FBDD.[78–82] A critical study of its general performance on ligand binding (within the context of drug discovery) has been attempted often,[18,83–87] generating a variety of different outcomes. Arguably, the only general conclusion which can be extracted from these studies is that no one docking protocol is optimal for all systems, with different software packages performing better in different systems, and where the choice of parameters for each software significantly affects the quality of the outcome. Generally, it is less problematic to locate the correct binding configuration than it is to successfully select it as the configuration of higher binding affinity, or selecting the best binder among a set of ligands. Consequently, it can be argued that problems mainly relate to the scoring part of the protocol.

Regarding the application of docking and scoring methodologies to fragments, it may be logical to argue that their performance is expected to be worse than that of bigger ligands (drug-sized molecules). This is because of the weak binding affinity of fragments due to their small size, and the small number of interactions. To select the best binders among a set of fragments, the sensitivity required is expected to be higher than that to perform the same task on drug-like molecules, since the difference in affinity between the best binders and the non-binders is expected to be much lower in the case of fragments. However, some studies addressing this issue

found no significant difference in the performance of docking methods on drug-like and fragment compounds. In particular, for ensemble docking, the combined docking performance of the different scoring functions tested produced a percentage of top-ranked (best) docking solutions within a threshold of the crystal pose of 42% for both their fragment and drug-like sets.[88] These same studies found, however, that the docking and scoring behaviour of fragments and drug-sized molecules differed. For fragment molecules, the behaviour described above, where locating the correct binding configuration is less problematic than selecting it as the tightest binder, is even more extreme than was found for drug-like molecules. While the sampling of different configurations still needs to be improved for drug-sized molecules, the major problem related to fragments is the lack of accuracy in the estimation of the affinity.[88] The treatment of the binding affinity with more theoretically robust methods than those applied within scoring functions could then be envisaged as a logical step forward for the study of fragment binding configurations and affinity. Application of these more robust methods could, arguably, be considered more crucial for fragments than drug-sized molecules, where an increase in the sampling algorithms for docking may offer improvements to the current docking and scoring results.

Whether on drug-sized of fragment-sized molecules, the assessment of docking and scoring methods is often made difficult by the existence of different types of docking (explained in section 3.2.2) depending on the relation between the ligand and protein structure used in the docking study. Equally, when assessing the performance of a scoring function, its performance on ranking by affinity different binding configurations of the same ligands, or different ligands to a same protein target, can be studied (see section 3.2.1). Generally speaking, much better results are obtained for docking calculation when the protein structure used is obtained from a complex structure with the same ligand docked (self-docking or native docking) than when a different protein structure (apo or obtained from complex with a different ligand) is used (non-native docking).[89] Scoring functions, in turn, tend to perform better at selecting the binding geometries of higher affinity among all possible binding geometries of a single ligand, than selecting the tightest binding ligand among a set of ligands.[90] It is hence clear that the assessment of the performance of docking and scoring is not one-sided, and different results are likely to

be obtained depending on the exact functionalities under study.

2.4.2 Free Energy Techniques

In a similar manner to above with docking and scoring methods, the impact of the free energy techniques on FBDD will be presented in this section, while their theoretical background and a deeper understanding of their function shown within the appropriate section of chapter 3 (to which we will refer as appropriate). We will be considering here a series of methods which are, directly or indirectly, applied to fragments and which will hopefully set the context in which the development of the JAFS methodology can be understood.

GCMC Applied to Small Molecules

A method to calculate the binding affinity of small molecules that fall within the spectrum of those generally considered fragments, is Grand Canonical Monte Carlo (GCMC, see section 3.5.4) as shown in the work of Clark et.al.[91,92] A series of approximations are applied, such as the lack of protein flexibility and only indirect account of ligand flexibility, or the simplistic model employed in the calculation of the solvation free energies. Some further (and not theoretically sound) approximations are included inadvertently in the calculation of the binding free energies (see section 3.6.4 and citations[93]).

Their method offers the possibility of locating binding cavities within a protein surface for fragment-like molecules and estimating the binding affinity of each pose. Within their latest publication on this topic, their method was applied to the simple system of the hydrophobic cavity of (mutated) T4 Lysozyme (also used in this project, see chapter 5). The binding free energy of rigid ligands was calculated with their method and the results, with a mean signed error to experimental free energies of 0.5 kcal/mol and a standard deviation on the error of 1.5 kcal/mol, do not allow for a correct ranking of the different ligands by their affinity to target. No error is reported for the computed binding affinities. The rough calculation of their hydration free energies has a big influence on the deviation of the computed affinities to the experimental values. While a more robust method for calculation of free energies could be applied, this would negatively affect one of the main strengths of the method: its speed compared to other free energy based methods. The poor

performance of the method on a simple, hydrophobic and rigid test case suggests that further work is required in the development or improvement of the method. Related to its approximations, it is important to note that this protocol requires the availability of the protein conformation co-crystallized with the biggest of binders to minimize issues related to rigidity of the receptor. Equally, it does not offer the opportunity to account for potential water mediated interactions between ligand and receptor.[92]

The location of crystal binding geometries is not a feature offered with their latest implementation, since their calculation of binding affinities is obtained from the average concentration of ligand in the binding cavity.[92]

Fragment Mapping

While a series of fragment mapping methods have been developed to find binding cavities on protein surfaces,[94,95] their application to FBDD is best represented by the SILCS approach.[96,97] The method, which is further explained in section 3.6.4, provides probability maps of fragment types binding to different target regions (FragMaps) as well as an affinity estimate (LGFE) to rank-order ligands by binding affinity to target.

The SILCS methodology may be divided in two parts, the generation of FragMaps and the calculation of the affinity estimate. The assessment of FragMaps is generally done qualitatively by comparison of the location of known ligand functionalities with the regions of higher probability for each fragment type predicted with SILCS, and they have been found to suggest functionality which is present in known binders.[98] FragMaps *per se* provide no information on the binding geometry or affinity of the ligands of interest. While they might be used to study where further functionalities might be added to a ligand whose binding geometry is known, arguably, this same study can be done directly by observation of the complementary protein functionality. Furthermore, other factors (such as entropy, section 3.6) may play a role in affinity that FragMaps (and SILCS in general) are not taking into account. FragMaps are a step required for the posterior calculation of binding affinities of the ligands of interest.[96,98]

To calculate an estimate of the binding affinity (LGFE), previously, an ensemble of configurations must be obtained (unless the binding geometries of the ligands

are known previous to their study with SILCS). It is important to note that, to obtain the (meaningful) ensemble of configurations, one or several relatively short molecular dynamics (see section 3.5.1) or Monte Carlo (see section 3.5.2) simulations must be run. The need to run these simulations considerably increases the computational cost of the method.[96,98]

Once an ensemble of configurations and the FragMap are known, the LGFEs may be calculated. In terms of the binding affinity estimate (LGFE), the results vary strongly depending on the target protein, and application of LGFE calculation on an ensemble of configurations (see section 3.6.4) appears to be a requirement to obtain satisfactory correlation between LGFE and experimental affinities. The test cases for which better results were obtained within their presentation of the method, $R^2 = 0.79$ was obtained while in the worst case, anti-correlation was observed. This highlights the difficulty to predict the performance of SILCS for particular desired test case and makes the method unreliable.[98]

In summary, SILCS offers an easy to visualize representation of protein affinity towards different functionalities (FragMaps), while the information provided with these representations may not exceed by much what a chemist may obtain by observation of the protein surface. Their study of ligand affinity requires either previous knowledge of the binding geometry, or this binding geometry to be studied with (computationally expensive) sets of short molecular dynamics or Monte Carlo simulations. Once the binding geometry (or ensemble of these) is obtained, the estimation of affinity is found to perform well in some systems while failing in others, making the methodology unreliable when used in any new system.

Other methods derived from SILCS have been developed, such as the oscillating-μ_{ex} GCMC-MD, which aims to increase the sampling of the previously mentioned representative fragments to improve the quality of the FragMaps, as well as allowing study of occluded protein cavities. This method also provides an alternative for the calculation of hydration free energies and an estimation of the binding free energies of the chosen representative fragments. The only binding affinity shown during their method presentation does not appear to be of satisfactory quality for a simple and well studied protein-ligand system (benzene to T4 Lysozyme, see section 5, with a binding affinity difference to experiment of 1.94 kcal/mol). Conceptual problems with this approach are shown in section 3.6.4. A successful correlation

between LGFE and binding free energies seems to be obtained for this simple system (see above for the variability of these results between different systems).[99]

Other examples of fragment mapping techniques can be found in the literature.[100,101]

λ Dynamics

λ-dynamics has had many stages to its development.[102–107] First, its main applications will be summarized, highlighting those that allow for the consideration of ligands with completely diverse structures. Next, the results obtained from the latest developments are presented.

λ-dynamics increases the sampling of ligand configurations by scaling the interaction energies of the ligands with a parameter (λ) which is variable during the simulation. To sample all relevant binding configurations, and to obtain a meaningful measure of binding affinity, the scaling of the interaction energy must be sampled in all its ranges (from fully interacting to non-interacting). Biasing potentials need to be applied to achieve this sampling. The choice of biasing potential will influence the outcome of the sampled configurations, making the result of the configurational sampling dependent on the simulation input or settings.[104,106] Further information on the implementation of λ-dynamics can be found in section 3.6.4.

With respect to the calculation of relative binding free energies, again the application of bias is found to be required to avoid trapping in the sampling of the λ variables.[102] To find the correct biasing potential, often an iterative procedure is required, increasing considerably the required computational time to obtain a reliable relative binding affinity within λ-dynamics. This aspect is particularly relevant since one of the main advantages of λ dynamics is that it reduces the required computational time for the calculations of relative affinities between sets of several ligands to a common target, with respect to more traditional relative free energy calculations (see section 3.6.2).[103]

λ-dynamics has had many stages to its development.[102–107] The results obtained on estimation of binding affinities for the tests cases studied with its latest version (see section 3.6.4) are summarized here.[107] Note that this latest development is within what has been called multisite λ-dynamics, linked to the study of

the different substituents attached to a common ligand core (for more information see section 3.6.4). For both test cases (benzoquinone derivatives in solvent and geldanamycin derivatives bound to HSP90) a positive correlation between the calculated and experimental affinity is observed, where, for its more realistic test case (geldanamycin derivatives bound to HSP90) — where additional chemical space is sampled between the different substituents to a common ligand core — the correlation is found to be R = 0.56. A discussion of the validity of some of the experimental data is undertaken and a consequent filtering of the experimental values available was performed. For the experimental values deemed more reliable, the correlation between calculated and experimental affinities increases to an astonishing R = 0.98. It must be noted, however, that only four experimental and calculated affinities are included in this correlation.[107]

While this latest development seems to produce very appealing results, its applicability is restricted to ligands with a common core. Note that, within FBDD, this approach is probably more relevant at the fragment optimization stage than it is during calculation of affinities for initial fragment hits (see below). While the substituents are kept flexible, and the particular conformation of each substituent may be presumed to be correctly sampled, the initial binding configuration of the ligand core with all its associated substituents must be previously decided and no sampling of the configuration of the whole molecule may be expected given that the core always remains fully interacting with the environment during the simulation. Hence this latest version of λ-dynamics does not provide binding geometries for new ligand structures.

Note that the nature of fragments (small and generally formed of only one functional moiety) makes the study of their affinity much more amenable to methods which can consider ligands with completely different structures, rather than those requiring some common chemical core.

Free Energy Perturbation

Recently, traditional alchemical transformation (see section 3.6.2), methods have been used to calculate free energy differences in the context of FBDD.[108] Note that the calculation of free energy differences through alchemical transformations is often referred to as Free Energy Perturbation (FEP). Often standard FEP

approaches have been discarded from routine application to the drug discovery process, particularly in its earlier stages — those that concern the initial fragment hits in FBDD — due to their high computational cost. However, the increase in computational power — with better availability of GPUs and access to Cloud computing — could enhance the interest of the FEP methods on fragment molecules. In this section we will summarize results and relevance of FEP calculations on fragment molecules as performed by Steinbrecher et.al.[108]

In this study, their particular implementation and protocol of FEP is referred to as FEP+. Details of their protocol are summarized in section 3.6.4. The use of physically accurate free energy methods has advantages, providing information which is not available to other, less physically realistic, methodologies. Mostly, this extra information relates to the possibility of comparing binding affinities between different methodologies. This comparison is particularly useful in studies of ligand binding selectivity between different protein targets, or to those interested in the effect of protein mutations on ligand binding affinities. Besides, the physical accuracy relates to not needing to parametrize the protocol for different systems. While parametrization goes into the generation of the molecular mechanics force fields (see section 3.4) this is performed to capture all physical interactions of interest and no modification should be required in a system specific manner.

They present a study on eight different protein systems, with several ligands of conserved charge binding to each protein. Their results can be considered satisfactory, with an average correlation between calculated and experimental binding affinities for all studied systems of $R^2 = 0.65$, and two thirds of the ligands presenting absolute errors lower than 1 kcal/mol. Given their "fragment" classification, ligands typically present lower affinity than expected for typical drug development studies when starting from more drug-like hits. The affinities range from those exceeding 1 mM to those in the high nanomolar range. No significant difference in performance is observed for this method at the different ranges of affinity tested.

A difference in performance is however observed between different test systems, with two out of eight systems performing exceedingly well (DNA Ligase and MUP-I), which suggests low errors on both experimental and computational results for these systems, providing a high correlation.

Perhaps not surprisingly, FEP+ was found to outperform less physically ac-

curate (and faster) methods — particularly docking and MM-GBSA — for the majority of systems. The correlation between calculated affinities and experimental ones is found to be significantly higher for the FEP+ results. Interestingly, the correlation between molecular weight and experimentally measured affinity outperforms those of the less physically accurate methods, getting closer to the results obtained with FEP+. This is, however, likely due to a publication bias, where only active ligands are published, while the inactive ones will not be available in the literature to test new methodologies. Cases are analysed in their publication[108] where molecular weight does not correlate well with experimental binding affinity, while FEP+ calculated binding free energy does.

When comparing the calculated relative affinity between pairs of compounds in each particular system with the equivalent experimental results, FEP+ is found to predict correctly the effect in binding free energy for most structural changes, miss-predicting a low proportion of these. In particular, out of the 116 direct transformations described, FEP+ is successful in predicting the most potent binder in over 85% of the cases. Running FEP+ prior to compound synthesis in a drug development protocol would hence enrich the proportion of higher active compounds synthesised.

The summarized results show successful studies of fragment binding predictions by alchemical transformation methods. Despite this success, it is the prior knowledge and computational cost to obtain them that may be problematic. In terms of the required structural information, the binding pose of each of the ligands must be known in advance, or else its prediction must be performed with some extra methodology, where the accuracy of the FEP calculation will depend on the quality of the predicted bound configuration. The other major problem associated with the usability of the FEP calculations is the computational cost. As pointed out by Steinbrecher et.al.,[108] the accuracy of the FEP calculations, combined with their high requirements in terms of computational expense may make them the optimal tool for cases where ligands present important challenges from the synthetic perspective. In these cases, the time spent performing a FEP calculation may be worth the time and cost which may be required for the equivalent experimental step. However, in most standard cases, where synthesis may be relatively straight forward, the time scales required for FEP calculations may struggle to compete with

experimental measurements, particularly at the initial stages of drug development, which are associated with fragment-sized molecules in FBDD.

2.5 Summary

In this section, the basic concepts that lay the foundation of Fragment Based Drug Discovery (FBDD) have been presented, together with its major advantages and its drawbacks. The experimental methods of particular interest to FBDD have been described, including biochemical and physicochemical methods for fragment screening, as well as the different approaches to develop the initial fragment into a lead or drug-like molecule. Consequently two examples of successful projects in FBDD have been summarized, from its initial screenings to the entry of compounds in clinical trials and, in one of the cases, the release of an approved drug to the market. At last, the computational methods within FBDD have been described, focusing on docking and scoring as well as the free energy methods available, which are the context in which the development of JAFS should be understood.

Next, the theoretical concepts and computational developments required to understand the rationale of JAFS will be presented. The understanding of computational chemistry, and molecular mechanics will be described, fields of which the JAFS method is a small part. Techniques which are key in the context of computational chemistry for drug development (and which will be briefly used in this thesis) will then be introduced. Tools which are essential to the correct functioning of the JAFS methodology, such as the Monte Carlo sampling algorithm and the molecular mechanics force fields, will be explained in detail. Then, free energy calculations, which are key to the understanding of the implementation, as well as the objectives, of JAFS simulations will be studied in depth. Enhanced sampling methods will be shown, as they will be used in the thesis in one of its variants and could be considered alternatives as some of the objectives that JAFS tries to achieve. At last, the JAWS methodology, of which JAFS is a further development and modification, will be presented, with its advantages and caveats, and the reasons why it was chosen as the basis for the development of JAFS, presented.

Chapter 3

Theoretical Background

In the previous chapter, the necessary background on the concepts of FBDD, methods used within the field and latest developments have been summarized. Understanding the context of the application of the method presented in this thesis is essential to determine its requirements and analyse its successes, hence generating the most optimal method, and allowing the reader to understand the strengths and weaknesses of JAFS. Such context is provided in the previous section with the explanations on FBDD.

A knowledge of atomistic modelling, the Monte Carlo method and free energy techniques in particular is required to understand the JAFS method. All these concepts, and others of relevance to the project will be presented below, and hopefully described in enough depth for the reader to be able to comfortably follow the following sections of the thesis.

3.1 Computational Chemistry

A traditional understanding of chemistry would envisage a particular chemical reaction or the equilibrium state of a particular system as a process or situation which can be observed within a real environment (i.e. a laboratory) or described on paper, with a representation of its components via chemical formulae. Computational chemistry expands that traditional idea, providing an alternative representation of chemical processes and states which, capturing a range of traits and phenomena hard (or impossible) to obtain from a static two dimensional represen-

tation, does not require the study of the real process with its associated potential dangers, expense or difficulty on extracting from a complex environment, as well as offering an associated visualization of small scale — in size and time — events, which may be hard or impossible to detect from a traditional experiment.

3.1.1 Visions of Computational Chemistry

The use and applicability of the additional tools provided by computational chemistry can be understood within the context of experimental work (*real-life* experiments). From this perspective, computational chemistry can provide the means to speed up and optimize practical experiments by providing relevant knowledge on the system under study, as well as a guide, highlighting the possible research paths more likely to reach a particular goal.

Alternatively, computational chemistry can be addressed as an alternative vision of chemistry all together. A future of chemistry can be imagined where experimental work is obsolete or exclusively useful in the context of validating computational work. It can be considered as the equivalent of the futuristic vision of robots fabricating robots. The accuracy of results increases with the development of the different fields within computational chemistry, and the development of technology, and particularly computational power, continuously increases how far these computational methods can be applied. At the moment, computational methods are often not regarded as reliable and used simply as a tool to aid experiments. However, there is no theoretical limitation on the development of computational methods which may ultimately be as reliable as experiments are nowadays regarded.

While such futuristic perspective may well be regarded as exaggeration, it does illustrate the tendency towards developing computational chemistry as a field of its own, and the interest applied to automation. The idea of such a future seems necessarily interlinked with easy to use and robust computational methodology, where human intervention and expert knowledge applied to each of the calculations is minimal. This idea is generally used as a descriptor to methods with the term "black-box". While not applied entirely or in its more pure forms, a tendency towards the development of black-box methodologies will be apparent through this thesis.

The reasons for the development of JAFS as a black-box type method are related

to the increase in throughput this approach provides, together with an emphasis on the reproducibility of the results and a keen interest on developing a method robust enough so that it requires as little human intervention as possible in its application. The increased throughput is obtained because of the little intervention required from the user, and this same reason is related to its reproducibility. Given identical input, a piece of software must always produce the same output; a human being, on the contrary, may provide different answers depending on non-controllable external factors. For these reasons, an emphasis on the black-box style development was applied to our method development and consequently will be mentioned throughout the thesis as the justification for some of the decisions taken on the development of the JAFS methodology.

3.1.2 Applications within Computational Chemistry

Computational chemistry studies are present in a wide range of chemical disciplines, from the structural study of the air-water interface,[109,110] through the study of the bond formation and patterns in defects of crystalline structures,[111,112] to research on the size of the chemical space of all available small organic molecules and how to better explore it to obtain compounds of interest to the pharmaceutical industry.[113–116] This thesis will focus on techniques which are applicable to biomolecular systems, with an emphasis on proteins and particularly protein-ligand interactions.

As well as dividing research within computational chemistry by scientific fields, the level of computational resources required to perform each particular task is a common differentiating trait to further partition each area of expertise. While the computational expense of a particular group of methodologies might seem a minor consideration in their description, in computational chemistry, the trade-off between accuracy of results and computational expense of the methodology generally applies. Besides, the computational and time resources required for a particular study very much limit its applicability, defining the context of its usability. Within the study of protein-ligand interactions, we will be presenting two different levels of complexity, which lie within opposite extremes of the accuracy-expense range, namely the fast (low computational requirements) docking (and scoring) and the expensive free energy calculations based on molecular mechanics

simulations. While the method developed throughout this thesis (JAFS) will be presented from section 4.2 onwards, all the theoretical requirements for its understanding will be explained in this section. JAFS lies somewhere in between the two extremes in terms of computational expense, being closer to the requirements of traditional free energy calculations than those of docking. Both docking and traditional free energy calculations will also be applied through the development of the methodology as means of comparison of results and / or to perform required calculation to the correct functioning of the JAFS methodology.

3.2 Docking and Scoring

The fundamentals of the docking and scoring techniques, as well as a brief classification of the different approaches to scoring will be presented in this section. There are two main reasons why these techniques must be explained within the context of this thesis. First, through the explanation and analysis of the results obtained with the methodology developed throughout this project (JAFS), docking and scoring (simply referred to as "docking" through our thesis) will be used as a measure of reference. While the developed technique will necessarily require a higher computational cost than docking, we hope to provide functionalities which are either not found in the common docking protocols, or where docking is deficient. In particular, docking calculations on the same systems to which the fully developed JAFS functionality has been applied, will be presented towards the end of the thesis. This will provide some direct comparisons.

Second, as will be explained shortly, the limits on what is called docking and scoring are not necessarily as neat as one could imagine. While the most traditional docking and scoring methodologies can be clearly encompassed under this nomenclature, we can find techniques which are on the limits of what can be considered docking and scoring. In fact, depending on how we choose to interpret the definition of these terms, the method developed in this thesis can be considered as a docking method. It is hence important that we understand the basic idea underlining docking and scoring.

3.2.1 Concept

In their most basic form, docking and scoring techniques could be described as computational methodologies which aim to obtain the most favourable binding configuration of small molecules to a target macromolecule, as well as ordering these small molecules by their affinity towards this target. While this is the basic idea, docking and scoring are terms generally applied to those methodologies which can be used in the context of studying a large number of small molecules within limited computational time. In its most typical expression, this would refer to ligand screening within a drug discovery context.[17,117–119]

Docking

Docking and scoring are commonly used in the pharmaceutical industry and related research. Docking techniques are those employed to obtain the preferred binding configuration of small molecules (ligands), to a (generally protein) target. Scoring functions are intrinsically related to docking protocols as judges of the binding affinity of each configuration (score). Most common docking protocols may be divided in to two different stages. During the first stage, docking as such is taking place, and a set of potential binding configurations are produced as output. During the second stage, the scoring function of choice is applied on the generated poses. As a result the most favourable binding configuration is selected. Notice that this two-stage definition implies different scoring functions may be combined with the same docking approach (with one scoring function used for pose generation and a different one for pose scoring). For the first stage, in principle, a thorough scan of the whole configurational space of the ligand within the binding cavity would be applied. This would however, be far from optimal, exploring non-relevant regions and providing the scoring function with a large set of binding modes, most of which will be highly unfavourable. To select sensible binding configurations and pass them to the scoring function, different docking algorithms are applied. Some examples are highlighted below.[120]

- Filtering strategy, where an extensive set of configurations of the ligand are filtered initially by their shape complementarity to target, subsequently applying further filters based on pharmacophores or scoring functions.[121]

- Monte Carlo based strategy, where random perturbations are applied to the ligand, and then the ligand configurations are subject to minimization methods.[122]

- Simulated annealing, where the ligand starts in an initial random configuration and at "high temperature". Changes in the configuration of the ligand are then attempted, where the probability of keeping the new configuration is related to the difference in their affinities (estimated by scoring functions) and the current temperature. The temperature is decreased throughout the process, making the acceptance criteria more restrictive, hence forcing to keep only configurations with high scores.[123]

- Genetic algorithms, where an initial set of candidates are generated randomly, which then evolve through selection, recombination and point mutation. Genetic algorithms can be used directly or combined with minimization techniques.[123,124]

- Incremental ligand construction, where a ligand *core* is selected and initially docked in the binding cavity. Successive ligand *moieties* are docked, bonded to the correct atoms of the ligand core (or previously placed moieties), sampling the different conformations of each moiety upon placement.[125]

Scoring

While, as seen above, scoring is intrinsically linked to the docking process, it can also be conceived as an independent tool to rank either different binding configurations of a particular ligand, or different ligands, to a particular target protein. Scores assigned to each ligand or binding configuration are measures of their binding affinity towards the target.

The concept of free energy and how to calculate it in an accurate manner computationally will be presented in section 3.6. Suffice here to say that, taking all relevant factors into account correctly, the ligand, or configuration, with lowest free energy of binding towards the target is, by definition, that which must be found most frequently bound to target. However, the accurate calculation of free energy of binding is computationally expensive. The application of this accurate calculation of free energy to a large number of ligands or bound configurations would take

the whole process out of the scope of feasible time which can be invested within the early stages of drug development. For this reason, scoring functions are used. These are approximations of the accurate calculation of free energy of binding for a particular ligand or binding geometry, usually applied to rigid binding geometries (hence missing the entropic term, see section 3.6). The different types of scoring functions will be summarized in section 3.2.2.

3.2.2 Different Approaches of Docking and Scoring

Types of Docking

There are different classification strategies for docking methodologies. Here, we will present a classification of docking based on the relationship between the ligand and protein structure used. Note that another common option is to classify docking techniques depending on the flexibility (sampled degrees of freedom) applied to both ligand and target.

Based on the relationship between the ligand and the target protein conformation, docking can be divided into the following categories:[89,126]

- Native docking. A docking experiment is considered native docking when the ligand is docked against the conformation of the protein to which that particular ligand is bound. This is the type of docking performed when the conformation of a protein is obtained from a co-crystal of the protein and a particular ligand, and that same ligand is the re-docked into that protein conformation. In this case, the ligand bound configuration is known (the crystal structure is considered to be equivalent to that adopted *in vitro* and *in vivo*). This kind of docking is often performed to validate docking protocols, comparing the results of docking against the known crystal binding modes.

- Non-native docking. Non-native docking is performed when the target protein is found in a conformation which may be different (to some unknown degree) to that which is adopted upon binding of the studied ligand. This is the case when the target protein is taken from the structure obtained when bound to a different ligand, or to no ligand at all (apo structure). This is the most common case for realistic docking situations, where the binding geometry of the ligand is unknown (no structure of the ligand bound to the protein is

available). In most cases, several ligands will be docked to a single protein structure which may not have been obtained from a structure bound to any of them. Most real docking experiments are performed under these conditions.

- Ensemble docking. Ensemble docking is a docking approach developed to account for receptor flexibility and / or increase the success rate from non-native docking to a single structure. It consists on docking the ligand of interest to a set of structures of the target protein, rather than a single one. While commonly, none of the target structures will have been obtained from the complex with the ligand being studied, the results obtained are still generally more favourable than those obtained with non-native docking.[89,126]

As should be clear from the explanation above, a problem may arise in the assessment of the docking methods when analysis and assessment of the results obtained is not careful.[89]

Types of Scoring Functions

It is important to note that different classes of scoring functions perform better for different target proteins. In some cases, this is related to the functions being parametrized for proteins similar to the target. The selection of the correct scoring function for the particular target in hand is often problematic, and one of the important challenges present in docking and scoring.[118]

The different types of scoring functions differ in the approximation used to estimate the binding affinity of ligands (or binding geometries) to the target protein. Typically, three types have been used within the literature (namely force-field based, empirical and knowledge-based) which may be combined in the consensus scoring approach.[17,127] The classification presented is in an updated form[128] reviewed to fit to widely accepted categories. In the cases where the new classification is used, the equivalent traditional naming will be provided as well. The different categories of scoring functions are shown below:

- Force field based scoring functions. These scoring functions rely on force fields (see section 3.4) to calculate the interaction between the the ligand configuration and the target protein. While force fields are used in molecular

mechanics simulations to calculate free energies (see below) when being applied to one unique binding configuration, the resulting value is the potential energy of binding the ligand in that particular pose from the gas state. Implicit solvent models can be used to obtain the potential energy of binding from solvent rather than gas (more representative of the real binding process). However, free energy (see section 3.2.1 and section 3.6) can never be obtained from a single configuration, and the potential energy of binding to that exact configuration is obtained instead. These scoring functions tend to be computationally more demanding than other types, while their ability to capture previously unseen (or uncommon) and non-intuitive interactions is higher than for alternatives.

- Empirical or regression-based scoring functions. These scoring functions are based on the choice of intuitive energy terms, both favourable interactions (such as hydrogen bonds and lipophilic interactions) and penalties (such as steric clashes) which define the basic energy function. This energy function is then "trained" using a training set of known affinities, and coefficients are fitted for each of these energy terms. The value of each of these energy terms for each ligand-protein configuration is calculated when applied to any new case, multiplied by their coefficients and added to obtain the global score. These are generally faster than force field based methods, and easy to modify by adding new relevant terms and re-fitting against a training set.

- Knowledge based scoring functions. The scoring functions that fall within this category study the distance between different atoms types within a large training set of binders to protein targets. The atom types here are degenerate, depend on the chemical environment of the atom, as well as its chemical type. When the scoring function is to be applied to a new case, the atoms of the system are classified within these degenerate atom types. The distances between atom types are then calculated. When atom types in the system are found close together which correspond to atom types which have been equally found close among the training set, a favourable factor will be included in the score. A negative factor will be applied to the score when atoms found close in the system where not so within the training set.

- Machine-learning based scoring functions. In traditional classifications this category may fall within knowledge-based scoring functions. In this new category defined using modern classifications,[128] a large number of descriptors are chosen to account for the protein-ligand interactions. Then, machine-learning techniques are applied on a training set and generate statistical models to derive the final models for ligand binding affinities, which will generally involve a selected subset from the previous large pool of descriptors.

As described above, within all but the most pure of the force field based scoring functions, a training set is used to generate the final functional form of the scoring function. For all of these, the choice of training set may have a big influence on the results obtained for the estimate of binding affinity. In some cases, specific scoring functions trained on limited sets of proteins work better for studies on those particular families of proteins. On the contrary, diverse training sets are ideal to generate scoring functions which can be applied to any system of interest.

3.2.3 Docking and Scoring in the Context of FBDD

The docking and scoring methodology explained within this section is commonly applied in the drug discovery process. While often results are not optimal, an enrichment of binders is obtained among ligands with highest scores (estimates of their binding affinities), and non-binders are most common among those with low scores.

As explained in section 2, fragments are a specific type of ligand, and due to their small size, their behaviour compared to bigger (drug-sized) ligands may differ. It is logical to presume that, due to the smaller affinity linked to the smaller size of fragment molecules, the capability of differentiating between the affinity of a binding fragment and that of a non-binder using scoring functions may be lower than that on drug-sized molecules. Studies have been performed on the difference in docking and scoring performance between fragments and bigger ligands, whose outcomes seems to contradict the common thought that molecules of lower affinity should perform poorly in docking studies (see section 2.4.1).[88]

3.3 Molecular Mechanics

We can understand molecular modelling as the capacity to represent a real system at the molecular level (chemically) *in silico* (computationally). Given this definition, molecular mechanics (MM) refers one of the levels of accuracy available to molecular modelling representations.

Working in the realm of molecular mechanics implies accepting the Born-Oppenheimer approximation. It is known that the displacement of electrons is much faster than that of the nucleae. Electrons are then considered to adapt instantaneously to any change in configuration of the nucleae. When studying interactions in molecular mechanics, hence, the only positions considered are those of the nucleae (atomic positions), with the distribution of electrons around each nucleus considered constant.[129]

In particular, molecular mechanics encompasses representations of the system where molecules may sample different configurations, both internally and with respect to one another. However, they are not allowed to change their composition: bonds cannot form or break. In molecular mechanics simulations, energies are calculated between each component atom (or bead) of the system based on a force field (see section 3.4), where the energy is determined by a given functional form and a set of associated parameters. Neither the functional form nor the parameters are typically based on theoretical first principles. A functional form is provided which determines the behaviour of the components of the system, and the associated parameters are fitted to either experimental data, or calculations from first principles (quantum chemistry).

As for any of the molecular modelling techniques, the aim of molecular mechanics is to represent the real system as accurately and efficiently as possible (where the trade-off between efficiency and accuracy applies). The difficulty of representing a real system at the molecular level with the (fairly) accurate energy estimation provided by force fields is two fold: it relates to size of the system and time scales available.

When events are studied experimentally, a large number of molecules are studied at the same time. Even for an uncomplicated system, such as a solution of glucose in water, a big number of glucose and water molecules will be present within any measurable amount of the system. In a similar fashion, without the

use of special techniques, it could be said that the fastest changes observable to the systems studied experimentally are those occurring in the order of seconds. In the subsections below, the difficulty of representing these volume and time scales typically used in experimental measures will be explained.

3.3.1 Sampling the Correct Configurational Ensemble

In molecular mechanics, every single particle in the system is represented at the molecular level. This involves, if we were studying simply water (assuming a molarity of water of 56 M), for every millilitre of experimental sample, within molecular mechanics, roughly $5.6x10^{22}$ water molecules would have to be simulated. Now let us assume we wanted to calculate the total interaction energy generated according to our force field for such a system. Assuming that we need to calculate the interaction between all water molecules present, we would face the calculation of roughly $(5.6x10^{22})^2$ interactions — actually $(5.6x10^{22})^2 - 5.6x10^{22}$ or $((5.6x10^{22})^2 - 5.6x10^{22})/2$, assuming only one interaction point per water molecule. While the most common functional forms involved in force fields will be presented in section 3.4, these numbers probably serve as an indication of why the size of systems as they are simulated with molecular dynamics (or indeed most molecular modelling techniques) will be much smaller than that of experimental systems.

For the reasons exemplified above, systems modelled using molecular mechanics represent a small section of those seen experimentally. When a solvated ligand is simulated with molecular mechanics, typically, only one copy of the ligand is present, surrounded by water, in the simulation. However, the properties of this solute, as studied experimentally, will correspond to the ensemble average of the properties presented by all different copies of the ligand, with their corresponding different configurations (given a flexible ligand). How can one single copy of the ligand accurately represent the properties of the ensemble of configurations present in solution at one given time?

In statistical mechanics, the ensemble of configurations adopted through (simulation) time by the single copy (or few copies) of the system simulated is replaced by the ensemble comprising the different copies of the system present in an experimental setting at any given moment. For these two ensembles to be equivalent, we must make sure that, throughout the simulation, the most favourable states of

the system are sampled more frequently (and in the correct proportion), just as experimentally there will be more copies of the system in its most favourable states (Ergodic hypothesis). In summary, we must sample the free energy landscape (see section 3.6) of the system in an accurate and efficient manner.

In particular, both the experimental and computational ensemble must correspond to the Boltzmann ensemble, where the probability (p) for each configuration fulfils the Boltzmann law in equation 3.1.[130]

$$p \propto e^{-E/k_B T} \tag{3.1}$$

where E is the energy of the configuration, T the temperature of the system and k_B is the Boltzmann constant.

Both sampling methods explained in section 3.5 follow different procedures to ensure that this rule of probabilities is fulfilled. Each of the procedures will be explained in the corresponding section below.

3.3.2 Time Scales Available

In a similar fashion to the limitations in the size of the simulated system, in molecular mechanics there are limitations to the time scales available to the simulation. Before explaining the time limitations it is important to clarify that not all sampling methods directly account for a representation of real-life time within the simulation. While the reasons for this will be further described in section 3.5, molecular dynamics does produce an output which represents a particular time length of the experimental system, while Monte Carlo methods do not. In principle, sampling with both methodologies, using equally optimized software packages should be identical for infinite simulations. The reasoning will be applied here to lengths simulated by molecular dynamics simulations, but they should be conceptually applicable to both sampling techniques.

Let us start by clarifying the concept of limitations in time scales available. It is important to recognize that no conceptual limit exists, in principle. If we could leave one particular simulation running for an indefinitely long period of time within a particular computational engine, there is no theoretical reason why that system could not be simulated for any time length required. When we talk about the limits

to available time scales, often the phrase *within reasonable simulation time*, or in this case, *within reasonable computational time*, appears. The idea transmitted would be that there are limits to the time sampled by the system with molecular mechanics, if we are to run a simulation within reasonable computational time. This *within reasonable computational time* is a subjective measure, but clearly a simulation must not run take longer to run than a researcher takes to finish a project!

But what are the limits on the simulation time available? These limits will depend on the level of molecular representation (further explained below). Focussing on atomistic simulations of proteins, the longest simulations currently reach the order of milliseconds of sampled simulation time.[131,132]

Let us address now the issue of why are simulations so "slow" compared to real life events. This explanation is similar to that of the previous section, related to the limits in the size of the simulated section of the system. By reducing the size of the simulated system, we have decreased the number of interactions (and internal) energy components to calculate. However, within a simulation, every time a change is generated in the system, these energies (at least some of them) will have to be re-calculated. The energy calculation takes computational time, which may be diminished by increased computational power. While the calculation of the energies often takes up a good part of the computational time of the simulation, we should not forget that every single requirement is a computational task that takes up time. In Monte Carlo simulations (see section 3.5.2), random numbers must be generated, sets of conditionals must be checked to decide what will exactly happen in the next step, to proceed, statements need to be evaluated to be true or false. In molecular dynamics (see section 3.5.1), velocities and forces must be calculated as well as energies. In both cases, coordinates of the system will have to stored, files printed out, some may need to be read. Every single one of those processes will take computational time, making it difficult to imagine for computational simulations to happen as fast as their real life counterparts.

3.3.3 Exploring the Free Energy Landscape

As seen in sections 3.3.1 and 3.3.2, simulations cannot sample an unlimited time scale and are limited in the size of the system they can represent. Nevertheless, we

still aim to represent accurately the real system, where we are capturing an average over a much larger system size and the difficulty in time scales often relates to capturing fast events, rather than being faced with too short lengths. Simulations and experiments are connected by the Boltzmann distribution, which simulations in the canonical ensemble (NVT, see section 3.5.3) must follow provided that sufficient system configurations are sampled. The key resides in equation 3.1 and the Boltzmann law. Both experiments and computational simulations must follow it. Accordingly, both sampling methods explained in section 3.5 will make sure that it is fulfilled — as far as enough simulation time is provided to sample all relevant system configurations.

Note that equation 3.1 provides the probability of a particular energy state. The Boltzmann law will not provide the configurations of the system, but given a configuration, its associated energy can be calculated, and the Boltzmann law will indicate how likely that energy state is. These possible configurations must be found in some other way, and the sampling methods that will be presented in section 3.5 take care of that. Any sampling method could theoretically be used to generate hypothetical configurations for the simulated systems. All atoms within the system could just be placed in random positions within a range of simulation dimensions. However, we can see that, in most cases, this will generate atoms which should be bonded to appear far away within the simulation box. While the manner in which energies are specified will be explained in section 3.4, it is intuitive to assume that such a configuration will be associated with a high energy (will be unfavourable) and hence your system will be unlikely to adopt it.

Intuitively, we have started to draw a free energy landscape. The free energy landscape is nothing other than an imaginary multidimensional surface which represents the energy associated with each configuration of the system. The lowest free energy states would be those sampled more frequently.

Notice that we talk about the free energy surface, but we have described an analysis of the energy (enthalpy) of the system at each particular state. In molecular simulations, the entropic term (see section 3.6) is taken into account implicitly, where wide wells within the landscape represent states of the system with higher entropy, that is, which can adopt a larger number of similar configurations without big changes to their energy. For an equivalent enthalpy (depth in the landscape)

the wider wells (higher entropy) will be sampled more frequently (will be more favourable, lower free energy), simply due to more conformations being available.

An important factor to note in the description of free energy landscapes is that the height in the energy corresponds (for equal entropy) to a lower likelihood of the system present in that state. While the manner in which new configurations are generated differs between sampling algorithms (section 3.5), to avoid generating completely unlikely configurations as in the previous example of bonded atoms ending up far apart, new configurations are generated based on previous ones, and hence the free energy landscape is sampled starting from one initial configuration, moving in a particular direction, but never going too far with every step. This way, the system will be unlikely to proceed out of its local minimum, while a deeper and / or wider well may be available elsewhere in the landscape. This is the problem of trapping the system in local minima, while the properties of the system will be determined by all low energy configurations of the system.

This is a known problem within molecular simulations, where the system can get trapped and it is conceptually impossible to know whether all the required energy landscape has been sampled. Approximations to this assertion can of course be made, and we tend to consider a simulation has sampled enough when we succeed in matching experimental results and / or when several simulations, started from different points within the energy landscape provide the same results.

3.3.4 Levels of Molecular Representation

Until now we have made general references to the representation of the systems, and implicitly referred to the classical mechanics representation of "balls and springs" as representation of atoms and bonds, respectively. Throughout this project we will remain within the realm of classical mechanics, where covalent bonds are treated as ideal springs that cannot be broken. However, even within the realm of molecular (classical) mechanics, different representations of the systems are common, which diverge in what exactly the balls and springs represent in the system.

It must be noted that the expression "balls and springs" is an over simplification of the representation of the system. In practice, every "ball" is associated with a set of parameters and mathematical formulae that define its interactions with

other non-bonded "balls" and each "spring" corresponds to sets of parameters and mathematical formulae that control the interactions between specific bonded "balls". These parameters and mathematical representations will be explained in section 3.4.

Possibly the most intuitive representation of systems within classical mechanics is what is commonly known as atomistic representation. In (strict) atomistic simulations, every atom of the section of the experimental system that we are including in our computation will be represented by a *ball*, and all bonds between these atoms will be associated to *springs*. This is also the representation that will be used throughout this thesis.

Some other representations can be found which do not fulfil this definition but are still often referred to as atomistic, however, one might argue they are not *strictly* atomistic.

Among those above we may find what is commonly known as united atom representations.[133] In these, non-polar hydrogens (generally those bonded to carbon) are not included directly in the simulations but implicitly taken into account when describing the *ball* (parameters) associated to the carbon atom. This is generally considered a fairly safe approximation, since the behaviour of non-polar hydrogens is not expected to change much throughout a molecular mechanics simulation.

We have just described how two or more atoms (particularly carbon and one or more bounded hydrogens) can be described as one unique *ball* and hence be encompassed within one unique set of parameters, which will be behave as a unit throughout the simulation. When this is applied beyond non-polar hydrogens, this is commonly called Coarse Grained representation and *balls* are here called beads. The level at which a system is coarse grained can vary from a representation of an aromatic 6 atom ring with three beads, to representing a full protein residue with one unique bead, or even greater.[134–136] The difficulty with coarse grained representation is often in determining the correct parameters that must be associated with each bead and their respective bonds, and these might have to be determined specifically for each simulated system. General parametrizations are available, which require associating each of the beads within a system to a particular bead type, based on properties such as hydrophobicity and charge,[134] but the accuracy of the system representation using these is not necessarily expected to be

as good as in the cases where specific parameters for the beads in each particular molecule are generated.

Specific concepts of representation within classical mechanics refer to the solvent. While solvent (commonly water) molecules are most often represented just as any other molecule in the system (either atomistic or coarse grained) this does not need to be the case. Alternatively, the solvent can be represented as a continuum, with an independent mathematical formulation. This representation is commonly referred to as implicit solvent, as opposed to an explicit solvent representation (where the solvent molecules are treated as any other molecule in the system).[137]

All these alternatives to the fully atomistic representation of the system reduce the degrees of freedom to sample (there are fewer configurations available to the system when this is represented by a fewer number of *balls*), reducing as well the computational expense associated with energy calculations (fewer *balls* correlates with lower number of computations for the interactions between them). These two factors will increase the sampling speed, making it less likely that the system gets trapped within an energy well for a long simulation time. The free energy landscape effectively becomes smoother.[138]

3.4 Force Fields

Force fields can be defined as the tool employed in molecular mechanics simulation to calculate the energy of the system at any given configuration. Force fields take the positions of every atom (or bead) in the system, and generate an estimate of its potential energy, taking into account both intra and inter molecular interactions. The energetics are usually calculated pair-wise, so that to calculate the potential energy of each atom (or bead), its interaction with each other atom (bead) of the system is considered individually. As far as our current explanation is concerned, the interactions of each two atoms will then be added to result in the total potential energy (pair-wise additive force fields, see section 3.4.3 for exceptions).

Force fields encompass a functional form and a set of parameters, which are used to calculate this potential energy. In some circumstances, however, the term "force field" can refer to the set of parameters exclusively.

The functional form tends to be applicable in the same manner to all molecular

and atom types. The interactions between different atom types differ in their energy values thanks to the application of different parameters to the same functional form. These functional forms may differ between different force fields (see below). Functional forms contain different terms, each representing an energy component. Each of these terms is a mathematical approximation to the shape of the graph generated when plotting that energy term for any given pair of atoms (or beads) as a function of geometry (a particular geometry value, such as distance or angle, depending on the component of the energy measured).

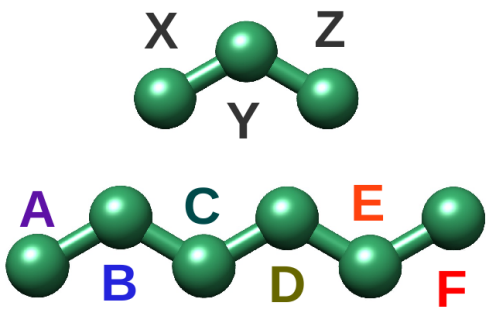


Figure 3.1: Simple model example of a system to run in molecular mechanics, where each sphere represents an atom and every cylinder a covalent bond.

Once the functional form is decided, the parameter sets for each atom and molecular type are calculated by fitting the functional form to experimental data or rigorous simulations (atomistic force fields will often be fitted to quantum mechanical calculations, while coarse grain ones may be fitted to atomistic simulations). While functional forms may vary between different force fields, some common traits can be extracted (see section 3.4.1), and generally the functional forms remain unchanged between different versions of the same force field family. Parameter sets, however, often change, and parameters for new interactions not previously considered may be included. The functional forms are included within the source code of the software packages used to run the simulations, while parameter sets are generally included as external files, so that new versions of the sets of parameters are easily applied to new simulations within the same software.

As mentioned above, the interaction energy of a particular atom with its environment, as calculated by force fields, will be dependent on its atom type and those of the atoms around it. Here, "atom type" refers to a degenerate typing, where the type of every atom is defined by its chemical type (carbon, hydrogen...) and its chemical environment (a carbon within an aromatic cycle will commonly hold a

different atom type than that of a methyl group). For the application of any force field to a new system, a process of atom assignment must take place. In most force fields, the atom type assignment of macromolecules will be done automatically (often through some provided tools) from the structural file of the macromolecule (generally in the pdb format). Owing to the huge diversity and variability of small molecules, the process of atom assignment is often more complex. The decision may have to be taken by the user, given a description of all possible atom types, where a comparison with the chemical structure of the small molecule is expected. In some cases, however, the process has been automated, and a tool or script is provided with the force field (or is available on-line). Further information on this point will be given, for our particular choice of force field, in section 3.4.2.

Below, a comment on the most common functional forms and frequently used parameter sets will be presented. Besides explaining the concepts, reference will be made to some of the most common and early classical force fields used in biomolecular simulations, namely AMBER,[139,140] CHARMM,[141,142] OPLS,[143,144] and GROMOS.[133,145] These force fields have been modified (improved) from their original development,[146–148] and references to these modifications may be found as well.

3.4.1 Functional Forms

Functional forms of force fields are generally divided into the different energy contributions, with their associated parameters. Energy contributions are typically divided in two groups, depending on whether the two interacting atoms are connected by up to three covalent bonds. When the two atoms of interest are within three bonds of each other the "bonded interactions" are taken into account, and the "non-bonded interactions" may be scaled or not present. If two atoms are not linked by three or fewer bonds, no "bonded interactions" are accounted for, nor is there any scaling applied for the "non-bonded interactions" (generally, for further details see section 3.4.4). By looking at figure 3.1, when calculating energies between atoms A and either B, C or D, the bonded interactions would have to be considered, but not for those of A with either E, F, X, Y or Z.

For the non-bonded interactions there is a consensus of including two major terms, representing the van der Waals and electrostatic contributions. However, less consensus is present within the bonded interactions. While it is commonly accepted

that bonded interactions include terms representing bonds, angles and dihedral angles, further terms are often added to these three basics. In this section the common elements within most well-known force fields will be presented, with their mathematical representations. Some of the alternatives implemented in common force fields will be mentioned as appropriate.

Bonded Interactions

Common terms within the bonded interactions include bond, angle and dihedral terms, which, in figure 3.1, would be calculated between atom A and B (bond), C (angle) or D (dihedral angle).

Bonds

The energy of two atoms connected by one bond (i.e. A and B in figure 3.1) is commonly calculated following equation 3.2.

$$U_{bond} = \sum_{bonds} k_{bond}(r - r_0)^2 \tag{3.2}$$

which represents a harmonic spring. The parameters to be fitted are to k_{bond}, the force constant, and r_0, the distance of null potential energy for the spring. r is then the current distance between the two bonded atoms and U_{bond} the contribution to the total potential energy of all bond interactions.

Representations of this energy term with a different use of exponents on the interatomic distances can be found.[149] Representations where k_{bond} is divided by a constant also exist,[149] however note that the division (or multiplication) by a constant may well be contained within the value of k_{bond}, as k_{bond} is simply a parameter.

Angles

The most common representation of angle (also called bond-angle) interactions, between atoms separated by two covalent bonds (i.e. atoms A and C in figure 3.1) is shown in equation 3.3.

$$U_{angle} = \sum_{angles} k_{angle}(\beta - \beta_0)^2 \tag{3.3}$$

which, as can be seen, represents a harmonic spring, where k_{angle} is the force

constant and β_0 the equilibrium angle for the spring at null potential energy. β is the current value of the angle and U_{angle} is the contribution to the total potential energy of all angle interactions. This representation is common in AMBER,[150] CHARMM,[151] and OPLS.[144] However, a completely different representation can be found within GROMOS force field.[149] Just as previously mentioned for bonds, note that the division (or multiplication) by a constant may well be contained within the value of k_{angle}, as k_{angle} is simply a parameter.

Dihedral Angles

Dihedral angles, also called torsion angles, are most commonly represented by equation 3.4. These interactions would be calculated between atoms A and D in figure 3.1.

$$U_{dihedral} = \sum_{dihedrals} \sum_{n} \frac{V_n}{2}(1 + cos(n\omega - \gamma)) \tag{3.4}$$

where the parameters are V_n, often called the barrier height, n, the multiplicity, and γ, the phase angle. V_n gives an indication of the height of the energy barrier between minima as the torsion angle rotates. n is a measure of the number of minima through the 360° rotation of the dihedral. γ controls where each of the minima are located.[129] In the equation, ω is the current value of the dihedral angle and $U_{dihedral}$ is the contribution of all dihedral angles to the total potential energy of the system. Again, this is a common representation in AMBER[150] and CHARMM.[151] In OPLS, a similar form to that of equation 3.4 is used, but signs and coefficients diverge from those in the equation above as well as between different values of n (multiplicity).[144] A different representation is found in GROMOS.[149]

Extra Terms

While the terms for bond, angle and dihedral angle are similar and often the only ones included within bonded interactions in force fields, some force fields add extra terms to account for specific interactions. A particularly common extra term is the improper dihedral term, also called out-of plane. It can be found in CHARMM[151] and GROMOS[149] force fields (in force fields where this term is not defined, out of plane motions are specified using the dihedral angle term). Improper terms are used to keep four atoms in a specific configuration. An example would be maintaining the planarity of hydrogens bound to aromatic carbons.

Some force fields add a number of extra terms. This is particularly characteristic

in force fields such as MM3.[152–154] As an example, in this force field, terms linking the sampling of angles with bonds (stretch-bend interactions), bonds with dihedral angles (torsion-stretch interactions) and angles with dihedral angles (torsion-bend interactions) are present.[152]

Non-bonded Interactions

Non bonded interactions are typically represented as van der Waals and electrostatic terms. Regarding their mathematical implementation, some variations can be found (such as in the van der Waals representation of the AMOEBA force field[155]), but again the representations shown below are more common.

Van der Waals

The most common representation of the van der Waals interactions between a pair of atoms is the Lennard-Jones potential. Its formula is shown in equation 3.5.

$$U_{vdW} = \sum_{ijpairs} 4\epsilon \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^{6} \right] \tag{3.5}$$

In the equation, σ and ϵ are the parameters to be fitted. ϵ is called the well depth and measures the how low is the energetic minimum of the van der Waals interaction between two particular atom types. σ is the distance at which the energy is zero for a particular pair of atoms, or "contact distance". In certain implementations, the equation can be expressed in terms of r_0 rather than σ, where r_0 is the distance between two atoms at which their van der Waals interaction energy is minimum. r_{ij} is then the current distance between both atoms (i and j) and U_{vdW} the contribution to the total potential energy by the van der Waals interactions of all atom pairs.

Electrostatics

In pairwise additive force fields (see 3.4.3), the representation of the electrostatic interactions is made using the coulombic expression, as shown in equation 3.6

$$U_{ij\ ele} = \sum_{ij\ pairs} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \tag{3.6}$$

where $U_{ij\ ele}$ is the contribution to the total potential energy of the electrostatic interaction of all atom pairs, q_i represents the charge of one of the atoms in the

pair and q_j that of the other atom and ϵ_0 is the permittivity of free space.

It must be understood that not only molecules with net charge will present electrostatic terms in their interactions. Most atoms, as defined by non-polarizable force fields (see section 3.4.3), will present partial charges, which must add up to the net charge of the molecule of which they are part (often zero). This is a computationally simple representation of molecular multipoles.

3.4.2 Parameter Sets

In section 3.4.1, the most general functional form of common force fields in biomolecular simulations has been presented. However, force fields are a combination of their functional form and the associated parameters for each atom type, or combination of atom types. In this sections some generalities of the parameter sets will be described and some of the most common available parameter sets for biomolecular simulations will be mentioned highlighting their main focus or some specific characteristics.

Within the parameter sets of a force field the data provided is generally divided in two sections:

- The list of atom types (and atom type combinations) with their associated parameters.
- The assignment of each atom in the simulated system to the most suitable available atom type.

The list of atom types may be identical for all molecules in the system but each simulated molecule must then present associated information regarding which atom type is associated to each of its atoms.

Assignment of Atom Types

The second part of the required information, the association of the molecule's atoms with the atom types differs between molecules which are common to biological systems and hence expected to appear very frequently in simulations, and those which are specific to a particular simulation (such as system specific biological ligands or drugs).

Some molecule types are expected to appear very frequently in our simulations, such as waters and proteins, common ions like sodium or chloride, as well as, in some cases, other biomolecules like lipids or nucleic acids. It is to be expected that every time a protein is run in any simulation with one particular force field, the same atoms within a particular protein residue are associated to the same atom type within the force field, independent of the particular protein in which the residue is located. A similar reasoning can be applied to water. Identical assignment of atom types is then expected every time for either identical molecules run very frequently (i.e. waters) or polymeric molecules, whose atom type assignment must be based on their common monomers (proteins). To avoid the performance of repetitive tasks and provide a consensus on atom type assignment, the information on the assignment of atom types for commonly simulated molecules is generally provided and packaged together with the list of available atom types and their related parameters.

A different case is that of molecules which are not present in most simulations, nor are they polymers of common monomers, such as system specific ligands (small molecules). The atom assignment in these cases must be done on a *per case* basis, and it is often a non-trivial task on which the accuracy of the simulation rests. There is thus a level of expertise required which slows down the process of simulating new systems. To avoid this problem, automatic protocols are available which computationally generate a guess on the optimal atom assignment for any molecule.[156–158] While it is always advisable to inspect and verify the generated results, these automated methods provide the consistency that human decisions cannot and tend towards the black-box approach (see section 3.1) that we favour in this project.

Just as the atom assignment differs between biomolecules and small molecules, so does the list of atom types and their associated parameters. While there is no particular reason why all atom types could not be included within the same file, it is common practice to separate force fields into their standard (often protein based) and general (associated to small molecules) force fields.[159,160] In the same manner, solvents are often provided as a separate "force field" (commonly referred as models). While the general force fields are commonly generated to be compatible with a specific standard (macromolecular) force field, solvent models are often

shared between different standard force fields. In particular, for the most common case of water, often macromolecular force fields are developed using a few of the available water models, and it is most optimal to perform simulations with the same combinations used during development.

Atom Types and Their Associated Parameters

The parameters *per se* provided in the parameter sets are associated to specific atom types or their combinations. It is important to remember here, as it has been previously noted (at the beginning of section 3.4 and equivalent to that described in section 3.2.2) that the atom types we are referring to here are degenerate, they describe both the chemical identity of a particular atom and its chemical environment, which influences its chemical properties.

It is important to remember that it is the set of parameters that has been fitted to either theoretical or experimental data, and hence combination of different parameter sets within a simulation is strongly discouraged.

Regarding the parameters required for the non-bonded interactions, it must be noted that, while for the electrostatic interactions (see equation 3.6) individual terms are included for each atom (namely q_i and q_j), for the van der Waals interactions (see equation 3.5), unique terms are included for the interaction between two atoms. However, in the parameter sets, individual terms are provided for each atom type. Consequently, a combination of the parameters must be undertaken. This combination is performed following certain *combining rules*. Note that two different parameters must be obtained to include in the van der Waals equation, σ and ϵ (see equation 3.5). Two different sets of combining rules are most commonly found. The first set of rules is known as the Lorentz-Berthelot combining (or mixing) rules, where the σ for a pair of atoms A and B is obtained as the arithmetic mean (see equation 3.7), while the *epsilon* for the same two atoms is obtained as the geometric mean (see equation 3.8). Alternatively, both parameters can be obtained with the geometric mean of the individual atom parameters. Note, in equations 3.7 and 3.8, the $_{AA}$ and $_{BB}$ indices indicate the parameters of atoms A and B respectively, which would be used directly for an interaction between two

atoms of the same species.[129]

$$\sigma_{AB} = \frac{1}{2}(\sigma_{AA} + \sigma_{BB}) \tag{3.7}$$

$$\epsilon_{AB} = \sqrt{\epsilon_{AA}\epsilon_{BB}} \tag{3.8}$$

The parameters associated with the bonded functional forms (see section 3.4.1) are generally assigned to the groups of atoms involved in a particular interaction, rather than defined for individual atoms and then combined, as is the case for van der Waals interactions. Notice that this limits the number of possible bonded interactions between atom types to those defined in the force field, and while these definitions are generally wide, it can be the case of a particular bonded interaction which lacks the required parameters in the force field. The parameters assigned to each non-bonded interaction will vary depending on the functional form of choice for each particular force field.

Different Force Fields

As mentioned previously, every force field must provide a functional form or mathematical expression of the interactions, and a set of parameters to specify and distinguish interactions between different atoms. While the functional form is rarely changed, due to required changes into all software using the force field as well as a complete re-parametrization, the parameter sets are quite often changed, or simply an additional small set of parameters is provided as a complement.

Commonly, at any one given moment, several versions of any particular force field are being used in different projects, due to preference or simply for legacy reasons. Common versions of the Amber force field — the default force field in ProtoMS and that used throughout this project — are Amber99,[147] Amber99SB[148] and Amber14SB,[161] where all share the same functional form but differ in their sets of parameters, and where Amber99SB improves a small set of the parameters (in particular protein backbone parameters) provided in Amber99 (rather than being a change of parameters in full). Amber14SB is an evolution of Amber99SB, including slight adjustments to protein backbone parameters and improvement on side chain terms.[161]

While originally certain force fields where developed for particular macromolecular types, nowadays, most of the common force fields provide support for proteins, lipids and nucleic acids within the same or compatible parameter sets. Similarly, force fields for small molecules are generally provided as separate force fields compatible with one particular macromolecular force field (see above for differences in atom assignment for small molecules).

3.4.3 Alternative Force Fields

The force fields described so far have been developed for atomistic representations of the system (see section 3.3.4) and a partial charge based representation of electrostatics. Different force fields are required for the simulation of coarse grained systems, of which ELBA[162] and Martini[134] are two commonly used examples.

Within the same (i.e. atomistic) representation of the system, the manner in which electrostatics are estimated may still differ. The simplest and most common approach within biomolecular simulations is to assign partial charges to each atoms of a molecule, accounting in this manner for molecular dipoles and the different electronegativity of different atoms. Adding all charges within a molecule must give the net charge of the molecule. While convenient, this representation does not reflect accurately the behaviour of the experimental system, where the proportion of the electron cloud placed over a particular atom will change with time, depending on the position of other atoms around it. While the partial charge approximation often provides satisfactory results, there are some particular cases where an accurate representation of the variation in charge on each atom with time may be essential, and it could provide a general improvement of results. Examples of these particular cases are the study of ionizable residues in proteins (particularly those in binding sites) or the study of biological electron transport processes (such as photosynthesis).[163] This alternative representation of electrostatics is achieved via polarizable force fields of which AMOEBA[155] and the CHARMM-Drude[164] force field are two common examples.

3.4.4 Extra Elements

The basic components of the energy calculation of a system through the use of force fields have been described above. However, extra terms can be added

to improve the speed of the calculation or to make sure the system fulfils some required configurational restrictions.

Restraints and Constraints

The application of restraints and constraints is common within biomolecular simulations when the position of a particular component of the simulated system determines the correct representation of the desired state of the experimental system. Restraints and constraints may also be used as a requirement of specific simulation set-ups. The most typical types of restraints and constraints are listed below.

- Harmonic restraints: A harmonic restraint is sometimes applied to the position of an atom or centre of a molecule, or a particular (real or artificially defined) internal coordinate of a set of atoms. A harmonic force is applied on the restrained variable during the simulation (see section 3.4.1).

- Half-harmonic restraints: A half-harmonic restraint works in a similar fashion to harmonic restraints, but in the case of half-harmonics, a buffer zone (distance from optimal) is defined which can be sampled by the restrained degree of freedom without experiencing any extra force. Once the buffer is exceeded, the forces are applied following the functional shape of a harmonic restraint.

- Hard-wall constraints: While restraints allow sampling of the affected degree of freedom by simply adding a force affecting the final energy calculation (hence modifying the preferred sampled regions), constraints forbid the sampling outside a given configurational region.

In ProtoMS simulations, when droplets of water are used (generally in protein simulations), half-harmonic restraints are used to maintain the shape of the water droplet. Besides, within GCMC (see section 3.5.4) and JAFS simulations, a cubic hard-wall constraint is used to limit the region where molecules must be inserted or deleted (GCMC) or to limit the sampled area of the JAFS particles.

Cut-offs

In the cases where simulations are performed with periodic boundary conditions (that is, the system simulated is assumed to be repeated in all directions to infinity), given that, in principle, the interaction energy between all atoms in the system must be calculated, the energy calculations would be performed to infinity, and no calculation would ever end. Since, obviously, energy calculations must eventually end so that the simulation can proceed, a cut-off is applied. The cut-off indicates the maximum distance between two atoms whose interaction energy is calculated. When calculating the energy associated with one particular atom in the system, all atoms which fall beyond the cut-off will be ignored.

Within ProtoMS cut-offs are evaluated on a molecular basis (except for proteins, where they are applied on a residue basis). This implies that it will be the distance between molecules that is measured, rather than between atoms, and all atoms within molecules which fall beyond the cut-off (as measured from the current molecule) will be ignored. The ProtoMS default cut-off of 10 Å with a 0.5 Å *feather* is used in all our simulations, where the feather acts over the last part of the cut-off scaling the energy quadratically to zero, to avoid an abrupt change to zero which might cause energetic inconsistencies.

It must be noted that calculation of a correction for electrostatics, which may remain relevant at long distances (beyond the cut-off) are often calculated, while corrections for van der Waals interactions are also available. However, none of these terms in applied typically in ProtoMS (nor in any of our calculations). These corrections will be briefly described here, and the reasons why they have not been applied mentioned.

The radial distribution function is a measure of the change in number density of atoms in a system with distance from any given atom in the system.[165] As the distance from any atom in the system increases, the number density of atoms tends to one for Lennard-Jones fluids.[129] If we assume the value of the radial distribution function to be one in the simulated system beyond the applied cut-off, the corrections for the electrostatics and Lennard-Jones terms can be calculated as in equations 3.9 and 3.10 respectively.[129] These corrections would be applied as a

post-processing step after the simulations has finished.

$$E_{correction} = 2\pi\rho N \int_{r_C}^{\infty} r^2 v(r) dr \tag{3.9}$$

$$E_{correction} = 8\pi\rho N\epsilon[\frac{\sigma^{12}}{9r^9} - \frac{\sigma^6}{3r^3}] \tag{3.10}$$

In equations 3.10 and 3.9, ρ is the number density, r_C is the cut-off distance, r is a measure of interatomic distance, $v(r)$ is the pair potential at a distance r, and the ϵ and σ parameters are those found in equation 3.5.

Note that, while cut-offs may be applied to any simulation set-up, they are particularly relevant when the system is placed within periodic boundary conditions. In our simulations, periodic boundary conditions are applied to the calculation of hydration free energies (see section 4.2). Simulating in periodic boundary conditions basically implies implementing the simulations so that the system feels the interactions provided by copies of itself repeated infinitely in all directions. However, interaction energies may not be calculated for an infinite number of pair-wise interactions, hence the relevance of cut-offs.

In this context, long range electrostatics may be treated with the Ewald method. The Ewald summation is a sum over all electrostatic interactions between the atoms in the central representation of the system and those of the periodic replicas of the system in all directions. However, such a sum converges slowly, and hence the Ewald method describes a path to transform that slow converging sum in a series of sums that converge much more rapidly. This is done by considering each charge in the system to be surrounded by a neutralizing charge distribution, commonly in the shape of a Gaussian. The summation is then performed on such a system, and it is commonly called the "real space" summation. Subsequently, a second charge distribution is considered on the system, which exactly neutralizes the previous one. A summation is then performed to account for the contribution from this second charge distribution. This is commonly called the "reciprocal space" summation. The interaction of each Gaussian with itself must also be subtracted and consequently another term is subtracted to account for this.[129]

However, the implementation of the Ewald summation is computationally expensive. Particularly for the case of (our implementation of) Monte Carlo, where

in each move only one molecule is sampled. Despite this small change in the system, the interactions between all particles would have to be calculated each step to implement the Ewald method. Simplified versions of this implementation, where the Ewald term is only considered every few number of steps are also a possibility. Given the high computational cost of the Ewald implementation together with reasons presented below, it was considered that the benefits provided by the application of the Ewald method were not enough to justify the excess in computational cost of the simulations for the JAFS method.[129]

While ignoring the long-range electrostatics beyond a cut-off may seem a crude approximation, studies have been done in the group[166] which show the ranking of small neutral compounds in terms of their hydration free energies is not expected to be affected by ignoring long-range electrostatics. Since the interest in the JAFS protocol when estimating affinity is that of the relative affinity between different compounds and their ranking, it was considered safe to ignore long-range effects in our calculations.

3.5 Sampling Methods

As described in section 3.3.3, the free energy landscape available to the system of interest must be sampled so that all relevant free energy minima can be found. Only by sampling all low free energy regions of the landscape can we make sure that the properties studied are representative of the experimental system, at equilibrium. Besides, when only unique configurations are considered, free energy cannot be taken into account within the context of molecular mechanics calculations (since entropy is accounted for implicitly via the number of configurations available to each particular state).

There are two main approaches to sampling the free energy landscape: either the configurational evolution of the system, starting from an initial configuration and given a set of velocities, may be studied (molecular dynamics), or a set of configurations may be generated by randomly repositioning elements of the system, with limitations to the random motion aimed at avoiding high free energy states (Monte Carlo). Each approach has different advantages and provides some different information. They will be described in more detail in sections 3.5.1 and 3.5.2.

3.5.1 Molecular Dynamics

Molecular dynamics (MD) is arguably the most intuitive of both sampling methodologies. Molecular dynamics represents what we would see if we could take our experimental system — in the lab — and visualize the behaviour of only one copy of protein in solution, while slowing down our observable time scale to visualize the events as they occur in femtosecond steps. From a molecular mechanics perspective, we may choose to visualize the energy surface available to our system as a sandy landscape with dunes and hollows, and picture the position of our system in this landscape as that of a ping-pong ball on the sand. Within this analogy — if we discard friction, but not gravity — starting from a particular configuration of the system, a set of velocities is applied to the system (the ball appears on the landscape at one particular position and moving in any one particular direction). From here on, it will be the shape of the energy landscape itself (dunes and hollows of the landscape) that decide the trajectory of the system, and the configurations (positions of the ball on the sand) that will be adopted.

In practice, in molecular dynamics, once the initial configuration and velocities are provided to each atom, the system is let free to evolve with those velocities for a very short period of time (on the range of femtoseconds), after which the forces that atoms apply on each other are calculated, generally the current configuration and energy of the system is saved, and the new velocities of each of the atoms is calculated from the forces. The recently calculated velocities are then applied to the current configuration of the system, which is left to sample for another very short period of time.

To accurately represent the experimental system, is not enough that we sample its configurations, but we must also capture certain conditions of the experimental environment, such as temperature or pressure. In molecular dynamics these are maintained via the use of tools such as thermostats, which will periodically correct the velocities of the system, to make sure they remain within the desired values. It is important to remember that the temperature correlates with the vibrational energy of any system. The higher the temperature, the higher the (average) velocities, the easier to overcome free energy barriers. This is the reason why the energies of the system must be corrected to match the desired temperature.

Molecular Dynamics is the Molecular Mechanics simulation technique which

directly provides information on how the system evolves with time. Two traits of molecular dynamics are directly associated with obtaining this information: first, we are concerned with kinetic as well as potential energy; and second, we need to calculate the forces that the elements of the system exerts on each other to be able to determine their changes in velocity and configuration throughout the simulation. Given the configuration of the system at one given time, using force fields (see section 3.4), we can determine the potential energy of the system. The forces suffered by each atom in the system can be obtained by differentiating the potential energy. The acceleration can then be obtain from the forces using Newton's second law $F = ma$ (were F is force, m is mass and a is acceleration). Hence, given an initial set of configurations and velocities for all particles in the system, their time evolution may be studied.[165]

If we take Newton's second law and express the acceleration as a derivative of the velocity, and the velocity in turn as a derivative of the positions, we get equations 3.11 and 3.12 respectively, where F is force, v is velocity, t is time and r is position.[165]

$$F = m\frac{dv}{dt} \tag{3.11}$$

$$v = \frac{dr}{dt} \tag{3.12}$$

MD deals with the time evolution by studying the changes on the system every short period of time (time step). Consequently, some common algorithms are available to study the energies and forces of the system at one given moment in the simulation and determine the velocities and positions of the atoms in the system after that time step. One of the simplest of such algorithms, the *Verlet algorithm* will be presented here.[165]

To obtain the Verlet algorithm we need to take the Taylor expansion of r(t), such that:

$$r(t + \Delta t) = r(t) + \left(\frac{dr}{dt}\right)_t \Delta t + \frac{1}{2}\left(\frac{d^2r}{dt^2}\right)_t (\Delta t)^2 + ... \tag{3.13}$$

$$r(t - \Delta t) = r(t) - \left(\frac{dr}{dt}\right)_t \Delta t + \frac{1}{2}\left(\frac{d^2r}{dt^2}\right)_t (\Delta t)^2 + ... \tag{3.14}$$

In equations 3.13 and 3.14, Δt is the time step, t is time and r is the position. Combining these two equations and assuming third and higher order terms are negligible we get equation 3.15.

$$r(t + \Delta t) = 2r(t) - r(t - \Delta t) + \left(\frac{d^2r}{dt^2}\right)_t (\Delta t)^2 \tag{3.15}$$

And the velocity is then obtained from equation 3.16.

$$v(t) = \frac{r(t + \Delta t) - r(t - \Delta t)}{2\Delta t} \tag{3.16}$$

It is important to note that Verlet algorithm is not the most efficient in terms of information storage. It has been chosen here for its simplicity.

An initial temperature is set at the beginning of MD simulations as linked to the kinetic energy of the system, and hence to the initial set of velocities of the system atoms. During the simulation this temperature is likely to change due to exchange between the kinetic and potential energies through collisions. To keep the system at an average constant temperature, a thermostat may be applied, of which the simplest form is a rescaling of the velocities of the system to adapt to the desired temperature.[129]

In a similar manner, when running in the NPT ensemble, the pressure must be kept constant, and consequently, a barostat is applied. Equivalently to the thermostat described above, the simplest barostat would consist of a rescaling of the volume of the system to keep the average pressure constant.[129]

In the next subsection we will present a whole different approach to obtain the relevant thermodynamic properties of the simulated system, while the information on the time dependent evolution of the system will be lost. Molecular dynamics is hence the method of choice when the time evolution is the desired information.

3.5.2 Monte Carlo

The Monte Carlo (MC) algorithm is not intrinsically related to molecular mechanics, but rather a statistical sampling method applicable to any problem which can be expressed in terms of rules applied to probabilities of events. It is based on the idea of expressing computationally a set of rules which define the real probability of a set of events, and then select one of these events from the computational

algorithm previously defined. Doing this a large number of times, the resulting distribution of events must be the same as the underlying probability distribution.[167] This idea is fairly simple and can be further understood with a very simple example. We can imagine some marbles of different colours in a bag. We are then asked which proportion of marbles display which colour. To provide an answer we could simply take all the marbles out of the bag, count how many display which colour and divide by the total number of marbles. But let us assume we are not allowed to take more than one marble of the bag at any given time. The Monte Carlo approach would be equivalent to randomly picking one marble, writing down its colour, and throwing it back into the bag. Then repeating this same process a number of times (say hundred), every time adding one to the total number of marbles of that colour. At the end, we could obtain the proportion by dividing the number of times we took a (say) green marble out of the bag by the total number of trials (100).

While proceeding with this experiment may seem (and probably is) fairly useless for a simple case, Monte Carlo has proven highly useful in situations where the set of rules are known but obtaining the consequent mathematical expression on the proportion of events expected of each type is very costly or impossible.

The initial development of the Monte Carlo sampling algorithm for molecular systems diverges slightly from this simple initial idea, being linked to the development of Markov-Chain Monte Carlo algorithms.[168] This initial development was later generalized, originating the Metropolis-Hastings algorithm now commonly used in molecular mechanics Monte Carlo simulations.[169]

Markov-chains fulfil two requirements: each new element depends on the preceding element, but not on any previous ones; and each element belongs to a limited set of possible elements.[129] These features directly link to how Monte Carlo is implemented in molecular mechanics simulations: we start from a particular initial configuration somewhere within the free energy landscape, which we may call the initial state or *state 0* (we can imagine the sandy landscape and ping-pong ball of section 3.5.1). What we call a *move* is then attempted, where a small displacement of one (or several) of the system components is performed. The probability of the new state in our set of possible finite states is calculated. Within molecular mechanics, as our probability rules must be related to the fulfilment of the Boltzmann

distribution (see section 3.3.1), the probability of the new state is related to the difference in potential energy between the initial and the new states. If the new state is deemed "acceptable", we say the move has been accepted, and the process will again be repeated originating from this new state. Otherwise, the move is not accepted, the old state will be recounted and the process will be repeated again from the initial state, selecting a different random displacement. Note that following this algorithm, the result of every move will depend on its *state 0*, but never on the configuration from which this *state 0* originated in turn, hence fulfilling the Markov-chain requirements.

The exact algorithm followed to generate new configurations within a current Monte Carlo molecular mechanics simulation is show below.[165170168]

1. Select an element of the system at random and calculate its energy
2. Give the element a random displacement, and calculate its new energy
3. If the energy at the new position is lower than the energy at the previous position, accept the move
4. If the energy at the new position is higher, we allow the move with a probability of
$$p = \exp\left(\frac{-\Delta U}{k_B T}\right) \tag{3.17}$$
where ΔU is the difference in energy between the new position and the old position ($\Delta U = U_{new} - U_{old}$).

   To assess that probability, we generate a random number γ between 0 and 1. If $\gamma < p$, the move is accepted.
5. If the move is not accepted, the old position becomes the new position, and the process is repeated.

where equation shown in point 4 is equivalent to equation 3.1.

Note that the moves which generate a new configuration lower in energy than the original will always be accepted. This is logically intuitive, as our aim is to sample more often the lowest free energy states. However, there is also a certain probability of accepting moves when the new configuration is higher in energy than the initial one. This fact will serve to overcome free energy barriers between several

minima. Within this context, the acceptance ratio associated with a particular Monte Carlo simulation will simply be the proportion of accepted moves with respect to the total number of attempted moves. The acceptance ratio for moves of a certain type or which fulfil a particular characteristic may also be calculated.

It is important to note that the acceptance test presented above is that used for simulations in the canonical (NVT) ensemble. To obtain the acceptance test for other ensembles, the fact that the condition of *detail balance* must be followed is used. The detail balance condition states that the probability of going from state A to state B must be the same as that of going from state B to state A. The probability of going from A to B can be taken as the product of the probabilities of being in state A, generating configuration B from state A, and accepting the move from A to B.[170] The acceptance test for MC simulations in the NPT ensemble results in equation 3.18, which applies to a change in volume and gets reduced to the same as equation 3.17 when the volume in a move remains constant and the coordinates of the system change.

$$p = \exp\left(\frac{-\Delta U}{k_B T} - \frac{P(V_{new} - V_{old})}{k_B T} + N \ln\left(\frac{V_{new}}{V_{old}}\right)\right) \tag{3.18}$$

where the difference in potential energy ΔU may be caused by a change in volume or in the system coordinates. The process followed in the generation of moves is the same as the one described above but with the volume added as a new degree of freedom to sample in the simulation.

In Monte Carlo simulations, keeping constant parameters such as temperature or volume is much easier than in MD runs. A value is provided to the simulation (in the acceptance test), and this is simply applied accordingly. As can be seen in equation 3.1 and the Metropolis-Hastings algorithm above, temperature is simply a factor to include on the calculation of the probability of a move to be accepted. Just as a higher temperature in MD will correspond with higher velocities (on average), hence making it easier to overcome free energy barriers, in Monte Carlo, a higher temperature will correspond with a higher probability to accept a move which generates a new state of higher energy than the original one.

Monte Carlo Implementation in ProtoMS

The algorithm presented above is used in most software packages that run molecular mechanics using the Monte Carlo sampling method. However, the sampling efficiency will vary between different packages depending on the implementation of this algorithm. In this section, the key factors of the implementation of the Monte Carlo sampling algorithm in ProtoMS will be presented. ProtoMS is the MC software used (and improved, see section 4.3) throughout this project.

Within ProtoMS, all molecules in the system are divided into molecular types, entering either of the categories; protein, solvent, solute or gcsolute; where solutes generally refers to ligand molecules, gcsolutes refers to particles which are allowed to sample the "insertion" and "deletion" GCMC moves, and protein and solvent are self-explanatory. The reason for this distribution is the different sampling capabilities of each molecular type. As such, solvent sampling can be optimized, where no internal degrees of freedom need to be sampled. Another example, proteins, sample individually each of their residues since moves applied to the whole protein are unlikely to be accepted due to the expected big energy changes suffered by the system. Each time a new move is to be attempted, the steps shown below are followed.

1. The type of particle to be moved is chosen randomly. The probability of choosing each molecule type is specified by the user as simulation input.

2. If more than one molecule of that type is present in the simulation, the exact molecule is chosen at random. All molecules of one given type are weighted equally.

3. A new configuration of the molecule is generated. Changes on translation and rotation as well as the internal conformation of the molecule are attempted in the same move.

4. The energy of the system with the new configuration of the molecule is calculated.

5. The difference in energy between the old and new state of the system is passed to the metropolis test, and the move is consequently accepted or rejected.

There are a few key features of the implementation in ProtoMS that need to be detailed. Particularly relevant is the use in ProtoMS of internal coordinates. This implies that rather than treating each atom based on its Cartesian coordinates, and each new move attempting a change on this Cartesian coordinates, changes in bonds, angles and dihedrals of the molecule are attempted. Within the ProtoMS input, flexibilities for each of the internal degrees of freedom (bonds, angles, dihedrals), as well as for translation and rotation, are provided. Equally, a z-matrix is given, which defines the relation of each atom in the molecule with the rest in terms of internal coordinates. When a new move is attempted, translation and rotation are attempted on the molecule (except for proteins, which only sample internally), with a maximum displacement limited by the flexibility given. All internal coordinates are then displaced, again with a limitation provided by the flexibility assigned to each. The new values of each internal degree of freedom is stored, and the molecule is re-built from the first atom defined in the z-matrix, and the newly calculated values for the internal coordinates. It is important to note that the flexibility assigned may be zero, hence not all degrees of freedom have to be sampled.

The input provided to the ProtoMS simulation package can be automatically generated with easy-to-use tools (shipped as part of the software package). Within these tools, default values for all input parameters (except for protein and / or ligand coordinate files) are provided. For further information of the software, please see section 4.3 (or visit the website *protoms.org*).

3.5.3 The Different Ensembles

So far we have talked about the free energy landscape (or phase space) as well as mentioning the importance of keeping parameters, such as the temperature, constant and correct during our simulations. We have not discussed, however, which parameters define the phase space sampled at a particular simulation. This actually will depend on the ensemble sampled. There are four main ensembles available to molecular mechanics simulations:[129,165]

- Canonical or NVT: where the total number of particles in the system, the volume and the temperature are kept constant.

- Isothermal-isobaric or NPT: where the total number of particles, the pressure and the temperature are kept constant.

- Microcanonical or NVE: where the total number of particles, the volume and the total energy of the system are kept constant.

- Grand Canonical or μVT: where the chemical potential, the volume and the temperature are kept constant.

Throughout this project, no simulations will be run in the microcanonical ensemble, hydration free energy simulations will be run in the isothermal-isobaric ensemble, while relative binding free energies with dual topology will be run in the canonical ensemble (see sections 3.6, 4.2 and 7). The JAFS simulations are run in a modified version of the grand canonical ensemble.

The Grand Canonical ensemble is the only one which allows the number of particles to vary through the simulation, keeping the chemical potential constant instead. The chemical potential can be understood as the change in energy on the system as the number of particles changes ($(\partial E/\partial N)$) — at constant entropy and volume.[171] However, it is more commonly defined in terms of free energy as $\mu_J = (\partial G/\partial n_J)_{T,P,n_B}$ where μ_J is the chemical potential for substance J, G is the Gibbs free energy, n_J is the number of particles of substance J, and $_{T,P,n_B}$ represents constant temperature, pressure and number of particles of the rest of the substances in the system.[172] Given that the variability in the number of particles throughout the simulation is key in the development of this project, the Grand Canonical ensemble and related methods are further discussed in section 3.5.4.

3.5.4 Variable Number of Particles

As seen above, the grand canonical ensemble provides unique properties to the sampling environment — the associated phase space — by allowing the number of particles to change. It is not one of the most common simulated ensembles, being generally used only for specific aims. Through this project we will see its use (and that of its modified version) related to an increase in sampling efficiency of the particles allowed to vary in number. Equally, its implementation is generally not straight forward within common simulation packages and we will see how

implementation pathways which are not necessarily intuitive may be the most effective, having the same final effect on the simulation as an intuitive (but harder) implementation. The concept of a simulation with a variable number of particles and its relation to sampling Cartesian coordinates will be key to understanding the development of the JAFS methodology.

Grand Canonical Monte Carlo

Grand Canonical Monte Carlo (GCMC) is the name given to the simulations sampling the Grand Canonical (GC) ensemble, with the Monte Carlo sampling method. Note that the statistical nature of the Monte Carlo sampling method grants a greater flexibility on its implementation and sampling possibilities than those offered by molecular dynamics. The requirements to the implementation of the grand canonical ensemble in a Monte Carlo software are, conceptually, limited to:

- The existence of the appropriate moves to handle the appearance and disappearance of molecules
- The existence of a Metropolis test associated to these moves / changes in the system

Within MD simulations, the dynamics of the process of inserting and deleting particles would have to be taken into account, which complicates the conceptual understanding of the GC ensemble. This difficulty is caused by the discontinuity at the points of insertion and deletion of particles.[173] The implementation of the grand canonical ensemble in molecular dynamics simulations generally implies a hybrid MD-MC simulation, where the inclusion of the MC steps allows for the particle insertion and deletion.[174]

Implementation in ProtoMS

The problems associated with the implementation of GCMC are related to the variable number of particles in the system. They could be formulated through questions such as *Where do the inserted particles come from?* or *How can the software handle a system of variable size (number of particles) and where no limits to this variation are intrinsically defined?*. In ProtoMS these problems are avoided by choosing an alternative implementation, where no particles are effectively added

or deleted from the system, as far as the software implementation is concerned, but which produce an equivalent effect to the energetics and configuration of the system, effectively following the GC ensemble.

This implementation relies on an idea which could be related to the expression *what you don't know won't hurt you*. Within the context of GCMC, if the particles *are* in the system as far as the software is concerned but the rest of the system does not *know* about them (it cannot interact with them), effectively, the effect is the same as if the particles *were not* there.

In practice, a GCMC cubic region (box) is defined, which will be the section of the system where particles are allowed to be inserted and deleted. Within this section, a number of particles are added which are in excess of the number of those required (in the case of water molecules in binding cavities, the concentration of bulk water to the whole gcmc box is commonly applied). These particles (gcsolutes) are provided with two additional moves: insertion and deletion. At the beginning of the simulation, all gcsolutes have their interaction energy set to zero (non-interacting). These non-interacting particles are free to sample the Cartesian space limited by the GCMC box, and since their interaction energy is zero, all (translation and rotation) moves must be accepted. When an insertion move is accepted on one of this particles, the (zero) scaling factor applied on its interaction energies is removed, and the particle is now fully interacting — the rest of the system now *knows* about it. When a deletion move is applied on an interacting particle, its interactions are zeroed, hence returning to the pool of non-interacting particles, freely sampling the GCMC box, awaiting an insertion move.

It is important to mentioned that, since GCMC was developed in ProtoMS to be used on solvent (water molecules), gcsolutes do not present internal degrees of freedom, hence their internal energies are constant, and the only variable to their energies being their interactions.

Applications

The most common application of GCMC within biomolecular simulations is the study of the solvation of cavities and the affinity of water molecules within those. While there is no theoretical obstacle to the use of GCMC on bigger molecules (see section 3.6.4), the sampling of insertion and deletion moves provides a limitation. Acceptance probability of insertion and deletion moves is generally low, and the

bigger the molecule, the lower the insertion and deletion probability is expected to be. This is simply related to the available space in the cavity.

While the GCMC methodology has been known for a number of years, its ability to predict the optimal number of water molecules in a particular cavity, together with their location and affinity has been shown in a recent study.[93] The particular relevance of this aspect relates to the difficulty of sampling different configurations of particles bound to a cavity throughout the course of a molecular mechanics simulation. Given a particular initial configuration of solvation of a protein cavity (be it empty or with a number of waters in a particular disposition) we cannot necessarily expect the solvation level and location of the waters to equilibrate with solvent water through a typical length Monte Carlo or Molecular Dynamics simulation. It is this capability of non-interacting particles to freely sample the cavity and hence the possibility of inserting a water in any position and orientation of the cavity, associated with the possibility of simply deleting it from the simulation (transfering it to the reservoir) without having to access bulk solvent to be released, that offers the possibility of sampling different solvation configurations in a simulation time that would otherwise be impossible.

3.6 Free Energy

3.6.1 Definition

The free energy can be understood as a measure of how likely a particular state of a system is. The measure of this "likelihood" always refers to the comparison between two different states, describing which of the two will appear more frequently for a particular set of equilibrium conditions. The relative ratio of expected observations for each of the states of the systems in that set of conditions can be extracted from their difference in free energy.

Any free energy difference is formed of two components, enthalpy and entropy, as shown in equation 3.19.

$$\Delta G = \Delta H - T\Delta S \tag{3.19}$$

where ΔH represents enthalpy differences, ΔS differences in entropy, T is the

temperature of the system, and ΔG the relative Gibbs free energy (obtained in the NPT ensemble, see section 3.5.3). In molecular terms, differences in enthalpy between two states are correlated with the differences in chemical interactions they present, while differences in entropy can be correlated to the available degrees of freedom to sample in the different states of the system. The process of going from state A, with lower number of favourable interactions between its molecules and lower number of available degrees of freedom available to sample, to state B, with higher number of favourable interactions and higher number of free degrees of freedom, will have an associated negative free energy, negative enthalpy and positive entropy — $\Delta G < 0$, $\Delta H < 0$, $\Delta S > 0$ —. The process of going from state A to state B will be favourable.

This is an important concept to understand. It is the reason why, for example, the configuration of a ligand binding to the protein with tighter (enthalpic) interactions is not **necessarily** the most favourable, the one that is found more often when the real system is in equilibrium. Another binding configuration with less tight interactions, but more available degrees of freedom to sample (e.g. where some dihedral angles are free to rotate) may present a higher affinity towards the protein.

It can be observed that we are using a different mathematical formula here to that used to calculate free energies with MC or MD. While information on the enthalpic terms can be obtained from these simulations (potential energy in Monte Carlo and potential and kinetic energy in molecular dynamics), no direct calculation of entropy is performed. This is accounted for implicitly, since states with a higher number of conformations associated to them will be sampled more frequently (for probabilistic reasons).

The most common mathematical expression of the free energy in the context of statistical mechanics is shown in equation 3.20, which defines the Helmholtz free energy (A) — free energy associated with the canonical ensemble, see section 3.5.3 — can be defined as in equation 3.20.[130]

$$A = -k_B T ln Q \tag{3.20}$$

In this equation, the concept of partition function (Q) is introduced, and will be defined below, together with the associated concept of the density of states.

We must say, that to define free energy in equation 3.20, an arbitrary zero value has been assumed, which cancels for the calculation of relative free energies (see above, beginning of section 3.6.1). For the calculation of relative free energies, this same expression is transformed into equation 3.21, where the initial and final states are represented as 0 and 1, respectively.[175] Note that in this equation, the notation $\beta = k_B T$ is used.

$$\Delta A = -\beta^{-1} ln \frac{Q_1}{Q_0} \tag{3.21}$$

Density of States

The density of states (W) is a manner of describing the energy levels available to the system as a continuum (opposite to what they would be in quantum chemistry). $W(E, N, V)dE$ is then the number of states the system can present whose energy lies within the infinitesimal range from E to $E + dE$, where the number (of each species) of particles in the system (N) and the volume (V), define the system.

Partition Function

The Boltzmann equation has been discussed previously (see section 3.3.1), and in equation 3.1, the probability of finding our system in one particular state was defined. However, probabilities are generally defined in a normalized manner, such that the sum of all probabilities for one particular choice (the sum of the probabilities of finding our system in any possible state), adds up to 1 (or 100%). To comply with this normalization, we must define the probability (p_i) of finding our system in any one state as in equation 3.22.[130]

$$p_i = \frac{exp(-E_i/k_B T)}{\sum_i exp(-E_i/k_B T)} \tag{3.22}$$

where the normalization factor, $\sum_i exp(E_i/k_B T)$, defines the partition function.

It turns out that the partition function is dependent on the temperature T (obvious from its expression), as well as the number of particles N (the number of particles of each species, in fact), and the volume of the system V, since the number of particles and volume of the system delimit all the available energy levels of the system E_i. We can then provide the expression in equation 3.23 for the partition

function.

$$Q(T, N, V) = \sum_i exp(-E_i/k_B T) \tag{3.23}$$

Given a system complex enough (i.e. a macroscopic system) so that the energy levels are not distinguishable, we can use the density of states (W) to define the partition function, such as in equation 3.24.[130]

$$Q(T, N, V) = \int_{E_0}^{\infty} e^{-E/k_B T} W(E, N, V) dE \tag{3.24}$$

3.6.2 Alchemical Transformations

Now that we have defined the free energy and its associated concepts, we can explore its applications.

As has been mentioned in section 3.6.1, only calculation of the relative free energy between two different systems, or states within a system, is applicable. While the term "absolute free energy" is sometimes used (as in "absolute binding free energy"), it still refers to the free energy between two different states (in the binding case, between the ligand free, and the ligand bound to target).

There are different ways of calculating the relative free energy between two different systems, and some will be explained in section 3.6.3. However, the set-up of the calculations required to run all the methods presented below are basically identical. In this section, we will present the common set-up and simulation mind-set.

First, it is important to understand that to accurately calculate the free energy difference between two systems, we must sample the phase space associated with each. If the energy difference between two systems was calculated using the initial configuration instead, a few problems would arise. First, we would not know whether we are at a high free energy position within the phase space (unlikely configuration), and even a minimization procedure, generally only grants access to a local minima, disregarding the possibility of deeper global minimum. Second, even if the single configurations could be placed at the global minima of the free energy landscape, often a number of minima are low enough in energy so that their contribution to the properties of the system is considerable. Furthermore, even if one unique configuration was placed in every low energy minima, the entropic con-

tribution would not be properly taken into account if no sampling of the degrees of freedom is captured. Hence, to properly capture all components of the degrees of freedom for each end state and obtain a valid free energy difference, both systems need to sample their free energy landscape, and this will be done following MD (section 3.5.1) or MC (section 3.5.2). All the free energy calculations (and indeed all sampling, except for the docking runs) performed throughout this project follow the Monte Carlo sampling method. From now on, concepts and applications will be described exclusively in this context.

Another important requirement is that the free energy landscapes of the systems whose relative free energy is to be calculated must overlap, particularly at the low free energy areas, that will dominate the simulation. The calculation of free energies generally requires that, during sampling, the energy of each configuration is calculated with the parameters associated with the opposite system as well as with its own parameters. The idea is that, by sampling only one of the systems, relative free energy information is obtained for both. From the difference in energy for each of the configurations, the free energy difference is obtained following equation 3.25, where the fact that the phase space has been sampled is expressed with the ensemble brackets ($\langle\rangle$).[129] This equation represents the Zwanzig approach to calculating free energies.[176] The reason for the need for overlap between both phase spaces is that the free energy of the second system (that we are not sampling) is based on configurations reached by the first system. If those configurations are not a good estimate of the ones sampled by the second system, we will not be capturing accurately the free energy properties of the second system. It is important to understand that the Zwanzig equation (equation 3.25) is not an approximation, it will always hold, given enough sampling. However, its convergence properties are poor, hence it may not be possible to reach "enough sampling" within feasible simulation time, if the overlap between the free energy landscape of both states is poor. To check that there is a good overlap between the free energy landscapes, the Zwanzig free energy calculation may be done in both directions (also sampling with the second system, and calculating the energies with parameters of the first). If not enough common phase space is present, both answers will diverge. This situation is called hysteresis. While the methods to calculate free energies in this project are more elaborate than the Zwanzig equation, the need for the phase space overlap

still applies.

$$\Delta A = -k_B T ln\langle e^{-\Delta E/k_B T}\rangle \tag{3.25}$$

The requirement for the two systems to present overlapping phase space limits the range of systems on which the free energy difference can be studied. Since the interest in free energy differences will generally span a range of more diverse systems than those similar enough to have overlapping free energy landscapes, a method to bypass this issue is needed. To understand the method, we need to introduce the idea that the free energy is a state function. This means that alternative paths can be chosen for the calculation of the relative free energy between two states. As long as the initial and final state remain the same, the final free energy difference should be the same. Simplistically, to go from A to B, in the calculation of any property that happens to be a state function, the results should be the same whether obtained following the direct route A to B or an alternative path A to C to D to B. While conceptually the same, sampling can be an issue for any given path (such as is the case with the Zwanzig equation between different systems).

In summary, we require overlap of the phase space between the two systems for which we are to calculate the free energy difference, but we are allowed to imagine any path between the two desired systems, as long as we start at the initial system and end at the final one. The logical solution is to build intermediate states, generating a net of overlapping states, so that we can go from A to an overlapping state Ab, then to a neighbouring overlapping aB, which finally should overlap with B. This is common practice; all states, from the initial to the final and all intermediates are commonly assigned values of a parameter widely referred to as λ, in such a way so that the initial state is called $\lambda = 0$, the final state is $\lambda = 1$, and all intermediates fall in between these values depending on how similar they are to the initial or the final state.

The key factor to remember here, is that the path does not have an influence on the end result (as long as good sampling is achieved), and we are not in the experimental but the computational realm. Hence, the intermediate states do not have to be "realistic": they do not need to correspond with any system that could be generated experimentally. Here is where the concept of **alchemistry**, and alchemical transformations originates.

It is important to define the concept of Potential of Mean Force or PMF which will be referred latter on this thesis. The PMF can be defined as a profile of the free energy changes experienced by the system as some variable is changed. Within the context of alchemical transformations, a PMF is often depicted to represent the free energy changes associated to the changes in the system between its different λ states (see section 3.6.3 for further information on free energy changes).

There are two common methods to generate intermediate states, particularly within the context of relative free energies of ligands within a common environment (be it the protein cavity or solvent): single topology and dual topology. The manner in which each of them constructs intermediate states will be presented below.

Single Topology

Single topology follows the strategy of directly transforming one ligand into the other, within its environment. The intermediate states (at intermediate λ values) in single topology will actually be hybrids of both end states.[177]

We can envision the transformation between two molecules which are identical in all their atoms but one. Let us say we are transforming A-B-C (molecule 1, initial state) into X-B-C (molecule 2, final state). The assignment of values of $\lambda = 0$ and $\lambda = 1$ will correspond to molecule 1 and molecule 2, respectively. We could then think of generating one intermediate state, at the middle point $\lambda = 0.5$. In single topology, this intermediate state (molecule 12) should present parameters which are half of those of the initial state plus half of those of the final state. Note that, for all the common atoms, parameters at molecule 1 and 2 should be identical, and hence so will be those at the intermediate state. If we decide to generate an intermediate state at $\lambda = 0.25$ this should display the parameters of molecule one scaled by 0.75 plus those of molecule two scaled by 0.25.

In the simplest example of linear scaling, given parameter k, at any given λ, and where k_0 is the value of the parameter k for molecule A-B-C (above) and k_1 is the value of the parameter k for molecule X-B-C (above):

$$k_\lambda = \lambda k_1 + (1 - \lambda)k_0 \tag{3.26}$$

In summary, in single topology, the force field parameters are scaled with λ.

Single topology works really well for molecules that are similar enough in their

parameters and topologies, where it can be easily decided which atom maps on to which. While the size of molecules can be increased by growing atoms with single topology, and equally atoms can be faded to nothing, the similarity between compounds is always a factor to take into account when planning single topology calculations. In principle, any transformation could be successfully computed with single topology, no matter how different the two end states are, given enough intermediate states and clever planning of which atom should be converted to which, and at which time. But out-of-the-box single topology is only encouraged for similar ligands. The level of difference at which single topology is not recommended can generally be quite well estimated when the mapping of atoms between the initial and the final state (which atom of the initial state should be transformed into which of the final) becomes an over complicated and non-intuitive decision.[178]

In the cases where two ligands share a common structure that makes them suitable for a single topology calculation, but their mapping requires the creation or annihilation of atoms, the singularity problem, or end point problem, may originate. This problem is associated with the shape of the Lennard-Jones equation (see equation 3.5), with the repulsion increasing very rapidly as two atoms come closer than their contact distance (σ). While the interaction energy for the disappearing atom (which will be fully decoupled from the environment) would be zero by definition no matter how close other particles may be, a partially interacting atom with its interaction energies scaled close to zero, would already present very high repulsion to particles much closer than its contact distance, as described by equation 3.5. If an atom decoupled from the environment found a neighbouring particle in its close proximity, its energy gradient with respect to λ would be a lot bigger than that calculated from its neighbouring λ window. This situations would create numerical instabilities in the calculation of the relative free energies. The approach commonly taken in single topology to avoid atoms decoupled from the environment to encounter particles within their contact distance is to decrease the length of the covalent bond of the disappearing atom at the same time as its interactions are being faded. It is important to note, however, that other approaches, such as the non-linear scaling with λ, which are explained in the following section (for energy scaling), may also be applied here for the scaling of the force field parameters.[179]

Dual Topology

Dual topology is the recommended approach when the two molecules to interconvert are not particularly similar. In dual topology simulations both ligands are present in the simulation at all λ values, but their interaction energies are scaled differently at each λ window.[177] Also, for any of the values of λ, their mutual interaction energies are zero (they do not interact with one another). This last feature is essential since they are expected to occupy the same simulation volume.

We can imagine a hypothetical system where we have ligand A as our initial state ($\lambda = 0$) and ligand B as our final state ($\lambda = 1$). In its dual topology representation, at $\lambda = 0$ the interaction energy of ligand A will be unscaled, while the interaction energy of ligand B will be zero (scaled by zero). At $\lambda = 1$ the opposite will be true, with the interaction energy of ligand B being unscaled and that of ligand A set to zero. At intermediate λ values, the interaction energy of the ligand unscaled at $\lambda = 1$ (ligand B) will be scaled by the value of λ, while the ligand unscaled at $\lambda = 0$ (ligand A) will be scaled by the value of $1 - \lambda$. Following this principle, we will find that, for $\lambda = 0.5$ both ligands present their interactions halved, while at $\lambda = 0.25$, ligand A will have its interaction energies scaled by 0.75 and ligand B scaled by 0.25. Note that, in dual topology, a non-interacting ligand is taken as a proxy for a ligand not being present in the simulation.[178] This idea has already been mentioned in section 3.5.4.

In the simplest example of linear scaling, the interaction energy from the ligand, at any given λ, and where E_0 is the interaction energy of ligand A (above) without scaling and E_1 is the interaction energy of ligand B (above) without scaling is:

$$E_\lambda = \lambda E_1 + (1 - \lambda)E_0 \tag{3.27}$$

In summary, in dual topology, the interaction energies are scaled with λ.

One of the key factors mentioned in the reasons to develop these alchemical transformations is the generation of transformation pathways between both end states where the free energy landscapes overlap between contiguous states. When plotting the gradient in free energy as λ changes with respect to λ, this will result in smooth plots, which are both essential *per se* in some of the forms of calculations of free energy (see section 3.6.3) and a valuable tool to indicate the validity of the

chosen path.

In dual topology calculations there are a few sources of potential lack of overlap between free energy landscapes in neighbouring λ, one of which will be studied in below, and it is sometimes referred to as the end point problem.

One of the possible sources of non-smooth gradients relates to the presence of one of the ligands being non-interacting at each of the extreme λ. Given that the interaction energies of this particular ligand will be zero, it will experience a flat energy landscape in terms of its position (and orientation) within the simulation space. It may very well drift into an area of the simulation filled by protein or solvent atoms. This is not that much of a problem during the simulation, but it will be a problem during the free energy calculation, since the free energy landscape available for this ligand to sample will change drastically from the value of λ where some interaction is experienced to that where the interaction energy is zero. Some measure of restraint may be applied to avoid this drifting, but these restraints would then have to be accounted for during the calculation of the free energies.[180] A simpler solution is to link the translation and rotation of both ligands.[178] This is the implementation which is applied by default in ProtoMS and what has been used in the relative free energy calculations performed in this project (see section 7).[181]

Softcores

The reason behind the appearance of the intermediate λ values and their implementation in both types of alchemical transformation is to generate a path between both end states through overlapping energy landscapes. As mentioned previously (see *dual topology* above), the good overlap between the phase space of all neighbouring values of λ will result in a smooth plot of gradients of energy with respect to λ changes for all λ values. However, the changes towards the end states ($\lambda = 0$ or $\lambda = 1$) where one molecule or part of it must make its interaction energies zero is often problematic when the scaling of the energies follows a linear correlation with λ. In the case of single topology (section 3.6.2) where it may be one atom or a small part of a molecule that disappears, this problem is solved by modifying bond lengths as well as van der Walls and electrostatic parameters of the fading atom, to make sure that, at the time when the energies turn to zero, the atom

falls within the van der Waals radius of an atom to which it is bonded. In dual topology this approach cannot be taken, since one full molecule must appear and one disappear at each of the extreme λ values. The approach hence taken in dual topology is to adopt a modified scaling of the energies with respect to λ. Note that, as always, the intermediate λ are simply a tool to achieve a convenient free energy pathway, hence modifications to the energetics of intermediate λ do not distort the correct calculation of the relative free energy between both states, as long as the end states $\lambda = 0$ and $\lambda = 1$ are represented accurately. One of the possibilities of alternative scaling of energies with respect to λ is referred to as softcores. This alternative allows for the smooth annihilation and creation of atoms and molecules at the end states ($\lambda = 1$ and $\lambda = 0$) without creating numerical instabilities during the calculation of free energy differences.[179]

Note that another possible alternative is the generation of non-linear scaling of the interaction energies with λ, such that equation 3.27 would become equation 3.28. This approach would change the rate of energy coupling with λ, which offers only a partial solution to the problem, since the gradient may become smoother at non-extreme λ values, but the issue would remain for the end states.[179,182]

$$E_\lambda = \lambda^k E_1 + (1 - \lambda^k) E_0 \tag{3.28}$$

where k may take any desired value.

The two different softcores used throughout this project are shown below, where that of equation 3.29 has been used for the dual topology calculations (both relative binding and absolute hydration free energy calculations) and that in equation 3.30 has been used for the JAFS calculations.[181] The choice of these softcores is simply related to trial of a few different examples and choice of the one providing the smoothest gradients of energy with respect to λ.

$$\begin{aligned} U_{VdW} &= \sum_{ijpairs} 4\epsilon(1-\lambda) \left[\frac{\sigma^{12}}{\left(r_{ij}^6 + 0.2\lambda\sigma^6\right)^2} - \frac{\sigma^6}{r_{ij}^6 + 0.2\lambda\sigma^6} \right] \\ U_{ij\ ele} &= \sum_{ijpairs} \frac{(1-\lambda) q_i q_j}{4\pi\epsilon_0 \left(2\lambda + r_{ij}^6\right)^{\frac{1}{6}}} \end{aligned} \tag{3.29}$$

$$
\begin{aligned}
U_{VdW} &= \sum_{ijpairs} 4\epsilon(1-\lambda)\left[\frac{\sigma^{12}}{(r_{ij}^2+1.5\lambda\sigma)^6} - \frac{\sigma^6}{(r_{ij}^2+1.5\lambda\sigma)^3}\right] \\
U_{ij\ ele} &= \sum_{ijpairs} \frac{(1-\lambda)q_i q_j}{4\pi\epsilon_0\left(\lambda + r_{ij}^2\right)^{\frac{1}{2}}}
\end{aligned}
\tag{3.30}
$$

Note that equations 3.5 and 3.6 will be replaced by either 3.29 or 3.30 (depending on which softcore is used), which will affect the energies at intermediate values of λ. But both softcores will be reduced to equations 3.5 and 3.6 for extreme (0 and 1) values of λ. All variables in the softcore equations (3.29 or 3.30) are equivalent to those in the original force field equations (3.5 and 3.6) except for the λ parameter.

Understanding the reason for the need of softcores is relatively easy when an example is considered. Let us think of the calculation of the relative free energy of solvation between two different ligands, with ligand A (non-interacting at $\lambda = 1$) being bigger than ligand B (interacting at $\lambda = 1$). For the simulation run at $\lambda = 1$, molecules of water may very well overlap with the van der Waals radii of atoms of ligand A, as this is fully non-interacting and hence there is no associated energetic penalty. However, for values of λ which may be very close to 1, but not exactly 1 (say $\lambda = 0.99$), these atoms will present van der Waals parameters being 0.01 times their full van der Waals parameters. However, taking into account that the energy associated with the van der Waals interactions, as measured by the Lennard-Jones potential (in equation 3.5), increases very rapidly as two atoms get closer than their optimal distance, the associated repulsion will still be high. Consequently the gradient at this point will be very high. However, in a simulation run at that value of $\lambda = 0.99$, given than some (even if small) van der Waals parameters are associated with the atoms of ligand A, it is unlikely that any solvent atom will overlap with them, hence making the energy gradient with change in λ a lot smaller. These will give rise to non-smooth gradients, which are a sign of non-overlapping free energy landscapes. The softcores modify the Lennard-Jones potential (as seen above) for intermediate values of λ, hence having an effect on the energy gradients. In particular, softcores will make the Lennard-Jones repulsion suffered by particles with low interaction energies (their interaction energies scaled down close to zero) much lower, for short interatomic distances, that they would be with the original

Lennard-Jones formulation. This will, in turn, reduce the energy gradient at the end states (at $\lambda = 1$ in the example above).

3.6.3 Calculating Free Energy Differences

To calculate the free energy difference between two systems, or two different states of the same system, each of these states must be simulated to ensure that the most relevant states of the system, contributing to their free energy, are sampled. As described previously (in section 3.6.2), the phase space of the two end states must overlap or, alternatively, a pathway of intermediate states with overlapping free energy landscapes must be designed. Alchemical transformations (see section 3.6.2) are one of the most relevant (but certainly not the only) methods to construct such intermediates. While a basic methodology to calculate free energies (namely the Zwanzig equation) has already been presented in section 3.6.2, in this section we will show more advanced methodologies to calculate free energy differences. All of the techniques shown here are available within the ProtoMS analysis tools, and all three of them were performed on the dual topology simulations shown on this thesis. The results displayed here and those used to incorporate as input data in the JAFS calculations shown in section 7 are those generated with the MBAR[183] technique.

Thermodynamic Integration

The calculation of free energies through the Thermodynamic Integration (TI) method requires the generation of the gradients of energy with λ changes for all λ windows. These gradients (particularly in their graphical form) have been previously mentioned as a tool to visualize the phase space overlap between different λ windows, where a smooth variation on the gradient (with no sudden peaks) corresponds to a good overlap and associated reliable free energy estimates. The calculation of the free energy with TI involves the integration over these gradients for all the range of λ (from 0 to 1). By looking at the plot of gradients used as an indication of phase space overlap, the free energy between both end states would then be the area under the curve. The equation which represents this integration corresponds to equation 3.31.[129,170] Effectively, the integration over all values of λ of the ensemble average of the gradient of the potential energy with respect to λ,

as sampled at each λ window.

$$\Delta A = \int_{\lambda=0}^{\lambda=1} \left\langle \frac{\partial U(\lambda)}{\partial \lambda} \right\rangle_{\lambda} \mathrm{d}\lambda \tag{3.31}$$

In equation 3.31, the free energy is ΔA, the gradient of (potential) energies with changes in λ for each λ value is shown as $\frac{\partial U(\lambda)}{\partial \lambda}$, and the angular brackets represent the ensemble average over all sampled configurations at each λ window.

Within ProtoMS, two different methods to calculate free energies with TI in dual topology simulations are available, the numerical and the analytical route. In the analytical route, to calculate the gradient, rather than the whole of the potential energy, each of the energy components is studied independently, and their functional form (see section 3.4) is then differentiated with respect to λ. The ensemble average is then applied by taking the average over all stored values of energy throughout the simulation. Finally, the integration over λ is fulfilled from our independent λ windows using the trapezium rule.

$$\Delta A = \int_{\lambda=0}^{\lambda=1} \frac{\partial A}{\partial \lambda} \mathrm{d}\lambda \tag{3.32}$$

To follow the numerical route, the equivalence of equation 3.31 with equation 3.32 is used. The finite difference approximation is applied to assume that $\partial A/\partial \lambda$ can be estimated from the difference between the free energy that our system presents at the current λ and that it would present at a slightly different value of λ (i.e. ± 0.001 in λ values), divided by this small variation in λ. Given that now we are trying to calculate the difference in free energy between two (very similar) systems, by sampling only at one of them (the current λ) the Zwanzig equation, as presented in section 3.6.2 can be applied here. Just as in the analytical route, the trapezium rule is then applied for the integration from individual values of λ. The numerical route has been defined as finite difference thermodynamic integration (FDTI) in the literature.[184]

BAR

The BAR methodology (Bennett Acceptance Ratio) was designed as a more efficient manner of estimating the free energy than the Zwanzig equation, and it is considered more efficient than TI.[185,186]

For every free energy difference, an iterative process to determine a variable C is followed so that $N_j\langle f(U_i - U_j + C)\rangle_j = N_i\langle f(U_j - U_i - C)\rangle_i$ where f is the Fermi function in equation 3.33, U is the (potential) energy, N is the number of configurations (snapshots) and i and j are the two states (neighbouring λ values) for which we are calculating the free energy differences. Notice how at each side of the previous equation, the ensemble average is obtained by sampling at different states. It is important to understand that the Fermi function is chosen as an optimal observable, in terms of providing the highest precision on the free energy calculation for a given simulation length. However, different observables would be equally valid in the limit of infinite sampling.

$$f(x) = \frac{1}{1 + exp(x)} \tag{3.33}$$

Note that, in equation 3.33, x will be replaced by the term inside the fermi function f seen above in the expresion to determine the variable C.

The difference in free energy between the two states (neighbouring λ) can then be obtained with the calculated C following equation 3.34 where N_j and N_i are the number of stored energy (and coordinate) frames at λ_j and λ_i respectively.

$$\Delta A_{ji}^{BAR} = -k_B T ln\frac{N_j}{N_i} + C \tag{3.34}$$

The process described above will be followed for every two neighbouring λ values, and the free energy differences of every λ step finally added to obtain the total free energy difference between the initial ($\lambda = 0$) and final ($\lambda = 1$) state.[186]

MBAR

The MBAR method (Multistep Bennett Acceptance Ratio) is an optimized version of the BAR methodology presenting a lower variance on the free energy estimate. It is a generalization of BAR to any number of states (rather than only two).[183] The MBAR method presents an estimation of the free energy for

a particular λ_i as shown in equation 3.35, which must be solved self-consistently for $\hat{A}_i$ (which can be found in the equation also as $\hat{A}_k$, since k is the index of a sum over all λ). Note that "absolute" free energies are obtained, but these are calculated only up to an additive value, hence only relative free energies as taken by the difference in free energies calculated following this equation are meaningful.

$$\hat{A}_i = -\beta ln \sum_{j=1}^{K} \sum_{n=1}^{N_j} \frac{exp(-\beta U_i)}{\sum_{k=1}^{K} N_k exp(\hat{A}_k - \beta U_k)} \tag{3.35}$$

In equation 3.35, $\hat{A}_i$ is the free energy (up to a constant value) associated with λ_i, K is the total number of λ windows, N_j is the number of energy (and configuration) frames stored through the simulation and U is the potential energy at a particular value of λ.

Equation 3.35 can be understood as the equivalent to the combination of equations 3.34 and 3.33 in BAR, but applied to a number of states rather than only two, and with a different choice of observable (rather than the Fermi function in equation 3.33), as the optimal choice changes when applied to an indefinite number of states.

3.6.4 Free Energy Methods for FBDD

In this section we will describe the theoretical background of some specific free energy methods, not described previously for not being part of the basic core of free energy techniques, but which have a relevant impact on FBDD. More directly, their impact on FBDD has been described in section 2.4.2.

GCMC Applied to Small Molecules

So far, the theoretical explanation of GCMC (see section 3.5.4) has focused on its main application to water molecules. While all the principles are equivalent between different applications, a few key points affecting the application to small molecules will be highlighted here. Please, refer back to section 3.5.4 for an explanation on the basic principles on the method, as well as its implementation within the ProtoMS software.

An implementation of the GCMC protocol applied to ligands, where the particular example systems involve fragment-sized molecules, is presented by Clark et

al.[91,92] In their representation, GCMC on each ligand binding to a target protein is studied in vacuum, with the ligand hydration free energy estimated with implicit solvent (see section 3.3.4) calculations, and the contributions of solvating the protein-ligand complex assumed to be equivalent for different ligands in a series (hence cancelling for relative binding calculations). Both protein and ligands are treated as rigid bodies, and the flexibility of ligands that is expected to be present is captured through the use of several ligand configurations. The study of the binding over the whole surface of the protein is performed with this method. The chemical potential (μ) of the system — in fact a related variable, Adams value, B — is annealed through the simulation, observing a high concentration of particles for high values of B, and decreasing the value of this variable through the simulation, to observe a decrease in the number of particles, where only the tightest binding pockets will present ligands at the lowest values of B. The binding affinity at each pocket is obtained from the study of the effect of annealing B at that position, in a similar manner as it would be performed with a titration curve. As argued by Ross et al.,[93] assumptions are implied in this calculations which do not apply to small protein cavities.[91,92] As seen, a number of approximations are applied to this method, related to flexibility, solvation and calculation of affinities. Most of these approximations are, however, the trade off required to be able to explore the whole protein surface and estimating ligand binding affinities for a number of different poses. The results obtained and relevance of the work in FBDD can be seen in section 2.4.2

Fragment Mapping

While several related methods have been developed based on similar principles to map protein binding sites,[94,95] we will focus here on the SILCS methods given its specific focus on FBDD.[96,97] SILCS bases its study of fragment binding to protein target in molecular dynamics (see section 3.5.1) all-atom explicit solvent (see section 3.3.4) simulations. It studies the preferred regions of binding within a protein surface for different fragment types by performing an MD simulation of a protein soaked in a solution of different fragment species at 0.25 M (or 1 M) concentration. These fragment species are meant to be representatives of different fragments (and interaction) types, where small sized fragments are chosen as the

representatives to increase sampling. Sampling is increased for small fragments both by means of their higher diffusion and the lower affinity (hence higher turn-over of bound fragment), due to fewer number of atoms (see the concept of *ligand efficiency* in 2.1). A repulsive potential is implemented between these fragments to avoid aggregation and guarantee effective equal concentration of all fragment species. Equally, weak positional restraints are applied to the protein to allow sampling while avoiding denaturation (which can appear in the SILCS simulation conditions).

With the resulting snapshots from the molecular dynamics simulation of the target soaked in solution of representative fragments, the fragment atom locations are binned, creating a 3D histogram called FragMap, which defines the probability of each fragment type at each location of the target surface.

When quantitative affinity information is desired, an equivalent simulation to the previous MD run is performed, but without the presence of the macromolecular target. The occupancy with the target is then normalized with respect to that of bulk and inverse-Boltzmann weighted. With this, a "Grid Free Energy" (GFE) is computed per FragMap atom type. To obtain an estimate of affinity for any ligand, its atoms are transformed into FragMap atom types and the GFE of each of these atoms in the bins defining a particular ligand binding configuration are summed. This process may also be applied to ensemble of thermodynamically weighted configurations.[96]

Note that, even though SILCS relies on a MD simulation to obtain the most likely binding regions of representative fragments, this method does not provide binding geometries for the ligands or fragments of interest. SILCS provides an estimate of the binding affinities of ligand configurations and information which may be used for possible ligand modifications (FragMap). While obtaining the initial FragMap for a particular target may be costly, given a particular ligand binding configuration (obtained elsewhere) the process of getting an estimate for its binding affinity is fast. It is also important to note that this binding affinity estimate is based on a principle that, while intuitive and easy to understand, is not theoretically accurate. The binding free energy of a particular ligand is not the result of adding the binding free energies of its component functionalities (or atoms).[7] The binding affinity (free energy) will be influenced by the entropic

change of the ligand binding from solvent to target. This entropic change is mostly unrelated to the presence of a particular chemical functionality within the ligand.

Some further methods have been developed from the initial SILCS approach, with the oscillating-μ_{ex} GCMC-MD being an example. In this method, each GCMC (see section 3.5.4) step is followed by a very short MD simulation to improve configurational sampling. This process is applied repeatedly on the representative fragments of the previously defined SILCS protocol, to allow for a better sampling of their preferred configurations, including the reach of occluded cavities, hence improving the convergence in the generation of FragMaps. During the simulation, the excess chemical potential (μ_{ex}) as applied in the GCMC steps is varied following an algorithm with the aim of reaching a desired target "concentration" of fragment, and then maintaining an osculation of μ_{ex} around this equilibrium value. While this guarantees a better sampling within the GCMC protocol (where sampling can be problematic, see section 3.5.4 and section 3.6.4) the concept of a target concentration in the binding region (cavity) of a ligand is a problematic concept, since this will actually vary depending on the affinity of the fragment whose estimation is attempted.[99]

λ Dynamics

The λ-dynamics[102] approach shares features with the GCMC approach while diverging in many aspects. λ-dynamics presents a variable numbers of particles in the simulation also as a way of improving sampling, as GCMC does. However, the definition of this variability is different between both techniques. They also diverge in their applicability — λ dynamics focuses of relative affinities of ligands, while GCMC (although it has been applied to small molecules) focuses on water location and affinity — and their method of sampling — with λ dynamics using MD while GCMC uses Monte Carlo.

Let us start with the definition of variability in the number of particles. The Grand Canonical ensemble (see section 3.5.3) sampled with GCMC defines a system in which the number of particles is variable. However, we must remember how the ProtoMS implementation of this technique took an indirect interpretation of this idea where the deletion of particles is equated to zeroing their interaction energies and the insertion to resetting their full interactions. λ-dynamics can be understood

as an expansion of this idea where how much a molecule is present in the system is measured by the scaling of their interactions. The key difference is that, in λ-dynamics, intermediate scaling is allowed, rather than the switch-like behaviour presented in the GCMC implementation in ProtoMS (where particles where either fully interacting, or their interaction energy was zero). The variable λ measures this scaling (see the reason for this choice of variable and its relation to dual topology simulations in section 3.6.2). The variable λ is here sampled as a continuum, with limitations, that will be explained below.

The sampling method of choice is another one of the big differences between GCMC and λ-dynamics. As mentioned in section 3.5.4, MC offers a greater flexibility which makes the implementation of a new kind of move rather simple. In MD, the dynamics associated with every move must be defined, however the dynamics of inserting and deleting a particle are non-intuitive. λ-dynamics benefits from the sampling of the scaling variable as a continuum, which facilitates its treatment within MD. A set of fictitious particles are created within MD simulations, one associated with each of the particles which will be allowed to appear or disappear from the system. The "position" of these fictitious particles is the scaling of their associated particle in the system, and a mass is associated with them so that a kinetic energy is provided to sample the scaling ("position").

One of the applications of λ-dynamics is the calculation of relative free energies (of binding or hydration). These can be obtained from the proportion of the simulation time that each of the solutes spends "deleted" or "inserted". The scaling of the different solutes are generally coupled (i.e. the sum of the scaling for all solutes is one) and the application of biasing forces to the simulation is often required to obtain the correct sampling for all the "inserted" and "deleted" states.[102–105]

When the JAFS method, developed through this project is presented, it will be noted that both JAFS (which deals with small solute molecules) and its predecessor, JAWS (which studies water molecules) are based on similar concepts to λ-dynamics. They also present particles with variable scaling, sampled as a continuum. They are, however, implemented as Monte Carlo methodologies, their application does not involve coupling between the scaling for the different particles, and, in JAFS, no application of bias is required.

In their latest studies, λ-dynamics developed into biasing potential replica ex-

change multisite λ-dynamics.[107] In this latest version, the common core for a series of ligands is simulated with a set of different substituents in one or more of its potential modification sites, with one λ variable assigned to each substituent. For each site, typical λ-dynamics conditions apply (namely the sum of λ values for all substituents at that site must add up to one, where each λ is sampled as a continuum between zero and one). The relative free energy estimate between two particular ligands is then obtained from the ratio of probabilities of all substituents being at $\lambda = 1$ for each of the ligands. Given the sampling of λ as a continuous variable, an approximation is taken in defining $\lambda = 1 \approx \lambda > 0.8$.

Common in λ-dynamics is the application of biases to allow for adequate sampling of all ligands at $\lambda = 1$, where the choice of this bias has typically not been straight forward and often required iterative procedures. In this latest development, both typical λ-dynamics bias, fixed bias and variable bias are applied in a slightly different manner, which is expected to provide satisfactory sampling within the applicability limits of the method.[107] The application of the fixed bias is related to the use of replica exchange, where each bias is assigned to a different replica. These bias are optimized by running short simulations and searching for those providing the desired level of sampling in λ coordinate, where the initial bias value for the middle replica is taken as the free energy of hydration of each substituent as calculated with implicit solvent (see section 3.3.4). Their variable biases are meant to increase sampling and are only applied when $\lambda < 0.8$, hence not having an influence on the states which are later used in the calculation of the relative free energies. Results of this latest developments of λ-dynamics and their relevance to fragment based drug discovery are presented in section 2.4.2.[107]

Free Energy Perturbation

Free Energy Perturbation (FEP) is the name commonly applied to the calculation of free energies through alchemical transformations. While the process of alchemical transformations has been described previously (see section 3.6.2) as has the calculation of free energies (see section 3.6), some of the specific details for the protocol followed by Steinbrecher et al. to apply FEP to fragment molecules[108] will be summarized here.

Their FEP calculations are based on molecular dynamics sampling (see section

3.5.1) explicit solvent simulations (see section 3.3.4) using the Schrödinger molecular modelling suite. They have used solute tempering (see section 3.7.1) to improve their sampling. The structures of their ligands were generated by hand, energy-minimized and either docked into the protein receptor or manually placed on top of the binding configuration of a ligand co-crystallized with the protein, when this provided a clear binding mode. Standard protocols within the Schrödinger suite where used throughout the study, without an optimization of protocol applied to each individual system (what might have been expected to enhance accuracy of results). Please, see their study for further details.[108]

3.7 Enhanced Sampling

We have described so far the concept of phase space or free energy landscape, as well as mentioning the possibility of trapping the system in certain wells of the landscape, hindering visits to other regions which may present deeper or wider wells (see section 3.3.3). We have also mentioned that, for systems where this is a known problem, often when none of the region of the landscape one wishes to be studied is reached within any reasonable simulation time, a new representation of the system may be advisable, provided by coarse-grained models (see section 3.3.4). A representative example of the latter is the study of protein folding thanks to coarse-graining.[187,188] We might, however, be faced with systems where the sampling and trapping challenge is not as large, where the minima sampled are relevant, but still, an increase in sampling is required, for example, because the desired convergence between repeats starting at different initial configurations is not satisfactory. We may also be presented with a system where the application of a coarse-grained representation would defeat the purpose of the study itself, such as when the particular interest lies in the exact conformation of the internal degrees of freedom of a small molecule, some of which would be missed by the application of bigger beads including several atoms. In these cases, the enhanced sampling techniques may be applied to avoid trapping the system in particular free energy wells.

Enhanced sampling techniques are diverse. They can be based on the increase in kinetic energy (or probability of accepting a move) provided at higher temperatures (e.g. parallel tempering[189,190]) or in the application of a bias to the previously

visited states — free energy wells, naturally visited more often, should suffer higher bias — (metadynamics[191]), where these are only two of the most common basic ideas behind a wide range of different enhanced sampling methods.[192]

In this project we will focus on a particular group of enhanced sampling techniques, which rely on the exchange of configurations of the system between several replicas run simultaneously under different conditions. These are called replica exchange methodologies, and parallel tempering mentioned earlier is one example of these.

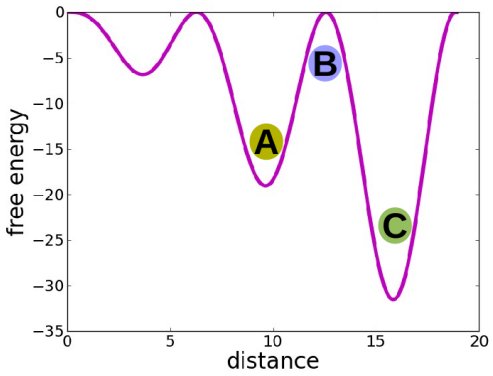


Figure 3.2: Simplified free energy landscape for an arbitrary system, with arbitrary units of energy (y) and distance (x), with distance measured from an arbitrary origin.

A simplified representation of a free energy landscape where enhanced sampling methods may be required can be seen in figure 3.2. In this hypothetical free energy landscape, a situation could be imagined with the system being in energy well A. Potentially, the change in energy between A and B may be too large, hence the majority of moves attempting to go in the direction of B from A would be rejected. Moves to change the system directly from A to C may not be available. Hence it would seem logical that C is not sampled within the available simulation time, while being the global minima within our landscape.

3.7.1 Replica Exchange

While replica exchange methodologies have been applied to molecular dynamics,[190] they originated within the realms of Monte Carlo sampling and they are easier to define within those terms. Since throughout this project, only the Monte Carlo sampling method has been used, the explanations within this chapter will focus on this sampling method.

In principle, the only requirement for a system successfully sampling the Boltzmann distribution within Monte Carlo is the generation of new configurations of

the system, which are then subject to the Metropolis test (see section 3.5.2), to determine whether they will be included within the ensemble of configurations sampled. While, in the standard Metropolis-Hastings implementation of the method, the new configuration is originated from the previous one, this is done with the purpose of generating realistic configurations of the system and any other method of generating possible configurations of the system is not forbidden.

Replica exchange makes use of this flexibility provided by Monte Carlo and gives the opportunity of generating possible configurations of the system in an alternative manner. In particular, several replicas of the system are run simultaneously at different conditions and, at given intervals, a Monte Carlo move to exchange the configurations of the system at different replicas is attempted. The nature of these "different conditions" between replicas is what defines the different kinds of replica exchange methodologies.

Please note that, since parallel tempering (presented in section 3.7.1) was the original example for which replica exchange was developed, the term replica exchange is often used to refer, specifically, to the parallel tempering methodology. Throughout this thesis, however, we will use them as distinctive terms, where replica exchange reflects a more general class of enhanced sampling methods characterized by the exchange of configurations between replicas.

λ Replica exchange

As mentioned, the different applications of replica exchanges will be defined by which conditions vary between different replicas of the simulation. If we think about it carefully, we have already talked about specific simulation types which require running several copies of almost the same system with one differing variable. These are the alchemical transformations defined in section 3.6.2. Whether single or dual topology (see sections 3.6.2 and 3.6.2 respectively), running alchemical transformations requires the simulation of the system at different values of the λ parameter, that accounts for the scaling of certain energies or parameters in the system. While, to perform an alchemical transformation it is perfectly possible to run all the different replicas of the system at different λ values (often called λ windows) sequentially, we can imagine as well that these are run in parallel (commonly in different cores of the same computational engine). Given this setting of

alchemical transformation simulations it is easy to conceive of a replica exchange methodology which will attempt to swap the configurations of the system between neighbouring λ windows at certain intervals. These might provide certain λ windows with the sampling of system configurations which might be easier to reach at other λ values. Notice, however, that, thanks to the application of the Metropolis test, these configurations will be accepted with a probability related to their likelihood (free energy) in the λ window for which they are destined.[193] The test applied to this configurational swaps necessarily diverges from the original Metropolis test presented in section 3.5.2, since the difference in energy of two different replicas of the simulation needs to be taken into account for the same move. For a λ value of a and b the move will be accepted if the p in equation 3.36 is higher or equal than a randomly generated number between 0 and 1.[193]

$$p = e^{(-\Delta E_a - \Delta E_b)/(k_B T)} \tag{3.36}$$

where k_B is the Boltzmann constant, T is the temperature and ΔE is the difference in energy for each particular lambda window (a or b) between the system after and before the configuration swap ($\Delta E_a = E_{aafter} - E_{abefore}$ and $\Delta E_b = E_{bafter} - E_{bbefore}$ with $after$ and $before$ defined with respect to the configuration swap).

Since the setup for this particular replica exchange procedure is given by the alchemical transformation definition, this is a practical and intuitive manner of aiding the convergence of the simulation at each of the λ windows and increasing the sampling of the common regions of the free energy landscape between neighbouring λ windows, as seen in section 3.6, a requirement to the correct calculation of the free energy.

Parallel Tempering

As mentioned previously, parallel tempering was the first application of replica exchanged envisioned. In fact, it may be fair to say that originally parallel tempering was developed,[189] and then the idea of swaps between replicas was generalized to other types of replica. In parallel tempering, the difference between replicas is the temperature at which the simulation is run. Note that, the higher the temperature, the more likely Monte Carlo moves are to be accepted, following the Metropolis test (see section 3.5.2). So we could, in principle, run any simulation at a higher

temperature to increase sampling. However, because the temperature is affecting the acceptance test, it will have an influence on the proportion of time spent in the wells of the free energy landscape with respect to the peaks. This will have an effect on the behaviour and properties of our system. Consequently, most frequently, we are interested in running our simulations at temperatures which replicate the experimental conditions (commonly, room temperature, 25° C). Hence we would like to have the increase in sampling and lower probability of trapping provided by higher temperatures, but with the free energy landscape and associated proportion of sampled configurations of a lower temperature (room temperature). To obtain this, we apply the replica exchange idea to a set of replicas of the simulation run at different temperatures. While the resulting data will only be studied from the temperature that replicates experimental conditions, the other replicas at higher temperatures will sample faster different regions of the phase space (notice that the phase space is not modified *per se*). The swaps between configurations at different temperatures should drive the configurations sampled at high temperatures, but which correspond to a well in phase space, to lower temperatures, through acceptance of Metropolis tests associated with configuration swaps.

The Metropolis test associated with parallel tempering in an NVT ensemble (as used in this project) corresponds to the acceptance of a swap move between temperature replicas a and b if p in equation 3.37 is bigger than a random number between 0 and 1.[193]

$$p = e^{\Delta E/k_B T_b - \Delta E/k_B T_a} \tag{3.37}$$

where, as previously, k_B is the Boltzmann constant, T is the temperature and ΔE is the difference in energy between the configurations swapped, such that $\Delta E = E(j) - E(i)$ if i is the configuration that was at temperature a before the swap, and j the configuration at temperature b before the swap. Notice that the energy of each configuration will be the same at both replicas, since the energy landscape remains unchanged.

Notice that, because we are generally only interested in the configurations sampled at the end of the simulation at one of the temperatures (generally the lowest temperature), it is essential that configurations sampled at the highest temperatures reach the ensemble of the lowest temperature replica. This must be done

through individual configurational swaps between neighbouring replicas. A good sampling in replica space is hence required. This associates with the idea of the random walk. When plotting the path between different temperatures of a particular trajectory of configurations (or the computational task associated with one starting configuration), the path must be varied in the temperature replicas reached and not appear to follow any particular direction, hence the term *random walk*. This concept of a desirable random walk, while possibly more obvious for parallel tempering, applies equally to λ replica exchange (see above) and solute tempering (see below).

Throughout this thesis the terms "parallel tempering" and "temperature replica exchange" may be used interchangeably.

Solute Tempering

So far we have seen two different examples of replica exchange, one in which a change in the free energy landscape varied between different replicas (λ replica exchange) and another where a different variable was changed (temperature), but the free energy landscape was left unaltered (parallel tempering). The replica exchange applications where the free energy landscape differs between different replicas are commonly known as Hamiltonian replica exchange. Solute tempering is another specific example of Hamiltonian replica exchange.

We explained above how higher temperatures can be used through temperature replica exchange as a resource to increase the sampling of our lower temperature replica of interest. The temperature will, however, have an effect of increased sampling on the whole system. Often, only a particular part of the system is of interest, and it is the sampling of that region that requires focus. This can be done by modifying the energies associated with this particular region of the system at different replicas, while keeping the original energetic representation of the system at our "lowest" replica. These energy modifications are commonly a scaling of the energies, and the particular region of the system is often a particular solute, hence the given name "solute tempering".

While the general idea of solute tempering is that described above, the implementation of the scaling of the energies, and which energies are to be scaled exactly, will depend on the particular implementation of the technique.

In ProtoMS, the scaling of the energies is associated with a "temperature" term (called "solute temperature" throughout this thesis), which does not really correlate with any temperature of the system, but does measure how the solute energies are scaled, given the real temperature at which the simulations are run. The equation that represents this implementation of the energy scaling can be seen in equation 3.38, where T_0 (or the equivalent β_0) refers to the real temperature of the simulation, also being the solute temperature of the lowest replica. Equivalently T_i refers to the solute temperature of the current replica.

$$E_i(X) = \frac{\beta_i}{\beta_0} E_1(X) + \frac{\beta_0 + \beta_i}{2\beta_0} E_2(X) + E_3(X) \tag{3.38}$$

where $\beta = 1/k_b T$, where T is temperature and k_b is the Boltzmann constant. X refers to a set of coordinates. E_i is the total energy of each replica, E_1, E_2 and E_3 its energy components with different associated scaling factors. Subindex i refers to each particular replica and 0 the reference, generally associated with the lowest replica (but not necessarily). Please notice that the scaling of the energies described above is purely empirical. Any scaling may be chosen for those replicas that are not the reference replica (used on the final collection of information), but the scalings presented above are the ones available in ProtoMS. Any energy term related to the solute may be included within E_1, E_2 or E_3 depending on the scaling desired for that energy component. Again, this choice will be purely empirical. A similar implementation has been published by Wang et al.[194] In this project, the internal energies of the solutes are left unchanged (being part of E_3) while all interaction energies of the solute are altered with E_1.

In this thesis, the terms solute tempering and solute replica exchange will be used interchangeably.

3.7.2 Other Enhanced Sampling Methods

Up to this point we have discussed enhanced sampling methods based on the exchange of configurations between replicas of a simulation which differ either in their temperature, or the scaling of (some of) the energies of the system. In this section we will summarize other relevant and widely used enhanced sampling methods, in which reaching configurations which may require a very long simulation time within standard MC or MD simulations is achieved by some other means.

Metadynamics

Metadynamics[191] belongs to the class of enhanced sampling methods which reach configurations which would be otherwise difficult to sample by the application of an energy bias (or forces) to the simulation. Metadynamics can be considered similar to replica exchange methods which require a scaling of the energies (see section 3.7.1), in the sense that the energy of the system is modified. However, in the case of metadynamics, an external energy contribution is added to the system, rather than a scaling applied to already present energetic terms. As a further difference, in particular for the case of metadynamics, simulation of several replicas is not required as the biases are applied within one single simulation.

Metadynamics is based on the application of energy biases, in the form of Gaussians, which are applied to any visited configuration. In this manner, already visited configurations (generally energy minima) are increasingly less favourable to visit (due to the applied bias) and new configurations eventually become preferred (lower in free energy). If we take the simplified free energy landscape shown in figure 3.2 as a guide, we may start our simulation in the local minima A, where the jump to a different energy minima is unlikely due to high energy barriers (B). If metadynamics is applied in the simulation, successive Gaussians will be applied on A as we visit and re-visit its configurations, effectively elevating the minima until, eventually, it will reach the same height as B. At this point, there is no energy barrier for the system to move towards C (or equally the smaller minima shown on the left of the figure). Once the system reaches C, since this minima has not been sampled yet, initially, no Gaussian biases will be present. However, as C is sampled, the same behaviour as previously seen in A will be observed. As this happens to all available minima, eventually the simulation will sample a flat energy landscape. At this point the metadynamics simulation is considered to be converged, and may be stopped. The original free energy landscape may then be regained from the information of all Gaussian bias being applied in the simulation.[195]

While this conceptual explanation is valid, some difficulties appear in practice. To proceed with a metadynamics simulation, the free energy landscape of the system is effectively re-defined in terms of a few selected degrees of freedom (generally called collective variables, CVs). The selected degrees of freedom must be all those presenting the high energy barriers that are difficult to sample within a standard

MD/MC simulation. These degrees of freedom should be able to distinguish between the different relevant states. The wrong choice of collective variables may lead to a metadynamics simulation which cannot converge, or not lead to the exploration of the full energy landscape of the system. While some guidance on the choice of collective variables is available, their correct selection is not trivial and one of the main drawbacks of metadynamics.[195]

While arguably the most crucial, the choice of collective variables is not the only decision required to run a metadynamics calculation. Parameters associated with the size of the applied Gaussian (energy bias) are also required, and again, their selection is not necessarily intuitive without previous information on the shape of the energy landscape to be sampled.[196]

Markov State Models

Markov State Models (MSMs) can be understood as a kinetic interpretation of the information provided by a number of simulations on a given system (i.e. molecular dynamics, see section 3.5.1). Typically, these simulations will cover part but not all of the desired section of phase space. As well as interpreting the information available from a kinetic perspective which may not be obvious from any of the individual simulations, MSMs serve as a tool to indicate the most optimal starting point for subsequent simulations to optimize sampling. The enhanced sampling provided by MSMs can hence be understood as two fold; first their ability to extract information from a number of short simulations which would normally only be available from much longer ones, second, on their suggested starting points to optimize any further sampling (adaptive sampling).[197]

A number of short molecular dynamics simulations are input to a MSM to which an initial configurational clustering is applied, where the number of resulting clusters is a parameter. Each of the resulting clusters is then called a microstate. The trajectory of each of the available simulations is then "translated" into a trajectory of microstates, by assigning each of the configurations to the closest microstate. The frequency of transition between each microstate with respect to all possible transitions is then calculated, generating a microstate transition matrix. On their own, these transition matrices can provide information on experimental observables of the system, however their human interpretation is not necessarily intuitive. To

facilitate human understanding, MSMs can be "coarse grained", that is bigger macrostates can be generated from the original microstates and the kinetic evolution of those may be studied. To generate the macrostates from the microstates, a kinetic clustering is performed based on the information provided by the microstate transition matrix. To proceed with adaptive sampling (see above) the uncertainty associated with each of the elements of the transition matrix is employed. Those elements of the transition matrix which are most imprecise are chosen as starting point for additional runs.[197]

MSMs have been used for the study of protein-ligand binding,[198] from their processes of association and dissociation to relevant intermediates.[199] All these pieces of information are often useful in the discovery of new possible binding sites and structural design of alternative binders.

3.8 JAWS: Just Add Water moleculeS

In this section, the JAWS[200] (Just Add Water moleculeS) methodology will be described. Its close relationship with GCMC and λ-dynamics points towards the inclusion of this technique within sections 3.5.4 or 3.6.4. However, given that the method presented in this thesis (JAFS) is a further development from the initial JAWS methodology, its explanation was deemed relevant enough to be assigned its own separate section.

JAWS is a method developed to study binding configuration and affinity of water molecules to protein cavities. JAWS is implemented as a Monte Carlo simulation and follows a similar idea to that adopted in GCMC — a method which has been widely used to determine the location of water molecules binding to protein structures — but addresses its sampling problem associated with the insertion (or deletion) of new water molecules in the system. The low acceptance ratio of the insertion and deletion moves in GCMC (see section 3.5.4) is related to the large change in energy most often arising by the appearance or removal of a new particle. For the case of insertions, this low acceptance ratio can be intuitively associated to the low likelihood of enough space being available within the (solvated) protein binding cavity, at any given point in the simulation, for a new water molecule to be inserted.

JAWS addresses this sampling issue in a similar fashion to that presented with

λ-dynamics. All water molecules available to move between the reservoir and the simulation box are associated with a parameter (θ). The θ in these simulations can be considered in some sense equivalent to the λ value in alchemical transformations (see section 3.6.2). Like λ in alchemical transformations, θ accounts for an scaling of the interaction energies of the particles with which it is associated. In fact, we could understand $\theta = 1 - \lambda$. However, some differences remain. Arguably the main difference is that θ are sampled as continuous variables throughout the simulation, rather than the set of predefined values (windows) being assigned to different replicas in the alchemical λ. Another key distinction is the assignment of an individual θ to each of the waters, which sample independently of one another. In summary, we have a protein binding cavity delimited by a cubic box, to which a reservoir of waters is associated, with a θ variable associated to each water which corresponds to a scaling of their interaction energies. Together with moves sampling Cartesian space, moves in θ space are attempted throughout the simulation. When a θ move is chosen, a change in the scaling of the interaction energies for one of the waters is attempted. Please keep in mind that, while we are describing the scaling of the interaction energies, water molecules are treated here in the same manner as solvent water molecules, which do not sample any internal degree of freedom (the only variable energy component is the interaction energy).

Note that, while the theoretical explanation is consistent, the specifics of the setup described in this section are different to the setup reported in the literature.[200] Here the setup associated with the implementation of the method in ProtoMS is described. This is particularly convenient since many of these details are shared between JAWS and JAFS simulations in ProtoMS and may be referred to later.

3.8.1 Water Binding Affinity with JAWS

In this section we will explain the calculation of the binding affinity of water at one particularly defined site (pose, binding geometry) — which may be obtained with the protocol described below (section 3.8.2). To study the affinity of water for a particular site, first that site must be defined. Within the ProtoMS implementation of JAWS, the site is defined as a small cubic region, typically 3 Å x 3 Å x 3 Å (JAWS box). Within this region one unique water molecule is included. This water molecule will have the capability of sampling the space between the reservoir and

the binding site by sampling its associated θ variable (with $\theta = 0$ being gas-like or in the reservoir, and $\theta = 1$ being fully interacting with the surrounding binding site). This water will also be allowed to sample Cartesian space, but will be constrained to the previously defined JAWS box, and no other water (solvent) molecule will be allowed inside. After sampling both Cartesian and θ space, the free energy of transferring this water molecule between gas (reservoir, $\theta = 0$) and the binding site ($\theta = 1$), can be calculated from equation 3.39).

$$\Delta G_{trans}(water, sitei) = -k_B T ln\left(\frac{P(\theta_i \rightarrow 1)}{P(\theta_i \rightarrow 0)}\right) \tag{3.39}$$

where k_B is the Boltzmann constant, T is temperature, $P(\theta_i \rightarrow 1)$ is the probability of finding the water at θ one and $P(\theta_i \rightarrow 0)$ the equivalent for θ zero.

By looking at equation 3.39, and analysing the simulation setup just described, two relevant issues may be noted:

- Since θ is being sampled as a continuum, it needs to be decided what exactly is considered $\theta_i \rightarrow 1$ and $\theta_i \rightarrow 0$, since the values of 0 and 1 will rarely (or never) be exactly sampled.

- For equation 3.39 to provide realistic results in practice, a good sampling of both states, $\theta_i \rightarrow 1$ and $\theta_i \rightarrow 0$, must be achieved.

To sort the first issue, a threshold is defined, where $P(\theta_i \rightarrow 1) \approx P(\theta > 0.95)$ and $P(\theta_i \rightarrow 0) \approx P(\theta < 0.05)$.

The second problem is solved by applying a biasing potential, which is then accounted for when calculating the free energy of binding from bulk water (see equation 3.40).[201] Typically, a set of different biases are attempted until a relation $P(\theta_i \rightarrow 1)/P(\theta_i \rightarrow 0) \simeq 1$ is obtained, since this correlates to the most optimal sampling of both states. The data from simulation at this biasing potential is then used to calculate the free energy of binding from bulk water following 3.40.

$$\Delta G_{bind}(\theta_i) = -k_B T ln\left(\frac{P(\theta_i \rightarrow 1)}{P(\theta_i \rightarrow 0)}\right) - \Delta G_{hyd} - \Delta G_{bias} \tag{3.40}$$

where ΔG_{hyd} is the free energy of solvation of a water molecule in bulk water and it is taken as a constant value $-6.4 kcal/mol$, ΔG_{bias} is the biasing potential and the rest of terms are as defined for equation 3.39.

3.8.2 Water Location with JAWS

To detect the location of water molecules in a protein binding site, the volume comprising the protein cavity is first delimited as a cubic region (JAWS box). A number of water molecules with the possibility of sampling the θ space between the reservoir and the cavity (JAWS waters) are included within the limits of the JAWS box, and their Cartesian sampling is constrained, so that they may not leave this region during the simulation. The number of water molecules is user-determined, but commonly, the bulk concentration within the volume of the JAWS box is used. All the JAWS waters typically start the simulation in the reservoir (i.e. with their interaction energies scaled to zero, non-interacting, at $\theta = 0$). The simulation starts including the protein, solvent water, JAWS waters, and possibly a bound ligand.

As the simulation progresses, Cartesian space is typically sampled for all simulation species (as in any other MC simulation, see section 3.5.2), with the water molecules and any possible ligand starting inside the JAWS box constrained to that space. Equally, no solvent waters are allowed inside. Together with Cartesian sampling, a sampling in θ space is available to the JAWS waters, where θ space is sampled as a continuous variables within its limits of 0 and 1.

We know that the free energy of transferring a water molecule from the gas phase (reservoir) to the specific site (pose, binding geometry) i is defined by equation 3.39. Hence, the most favourable binding sites should be those which most often present JAWS waters with $\theta_i \to 1$. For a simulation without any bias (see section 3.8.1), waters at $\theta_i \to 1$ are only expected at certain sites, where a bias would be required to observe them at other positions within the binding cavity. While the densities of waters throughout the θ distribution may be visualized, commonly, to determine the particular water binding sites, the waters presenting θ above a particular threshold are clustered. The most commonly used threshold is $\theta > 0.95$, just as before.

3.8.3 Application of JAFS to fragment molecules

At the onset of this project, it was considered what may be the most optimal approach to the development of a new technique to study the binding geometries and affinities of fragment molecules to binding cavities of proteins, in a manner

which improved on the accuracy of standard docking and scoring methodologies, but with results, for a small number of fragments, that could be obtained within feasible time scales. The advantages or reasons for choosing to develop JAWS into a methodology that could account for fragments will be presented in this section. These could be divided into two categories: JAWS as a methodology to improve sampling within a binding cavity and JAWS as a method to estimate binding geometries and affinities. Obviously, these two perspectives are interlinked, but they will allow the comparison with different sets of methodologies in this section.

It is only fair that the drawbacks of JAWS in its application to fragments may be understood. These correlate to two characteristics of the JAWS methodologies that have been discussed above: the sampling of the scaling of the interaction energies (θ) as a continuous variable, and the clustering process during the analysis of the results of the water location JAWS simulations.

While the sampling of θ as a continuous variable provides sampling advantages to JAWS with respect to GCMC, it is associated with the presence of particles (waters) in the simulation at intermediate values of θ (i.e. particles that are not present nor absent from the simulation). These intermediate particles diminish the realism of the representation of the experimental system, since, in experiments, no water molecule will be "partially on" (i.e. only half interacting with its surroundings).

Clustering is a powerful and very useful tool for the visualization of simulation results that would otherwise be tedious to analyse and very hard to compare to experimental data (i.e. crystallographic data). However, on clustering, information is simplified, and, potentially, simulations that have sampled in a manner consistent with experiment could generate clustering results which do not match experimental data. As an example, two neighbouring water binding locations may have been observed experimentally, and sampled in the simulation, but clustered incorrectly as a unique cluster in the middle of both binding geometries rather than two distinct clusters. While clustering is considered to be the best tool available for the analysis of results such as those generated by JAWS, it cannot be denied that complications may arise from its use.

JAWS to Improve Sampling

JAWS can be understood as method to improve the sampling of water molecules within protein binding cavities with respect to standard Monte Carlo and molecular dynamics simulations. The improvement in sampling can be related to that obtained with GCMC. We can take the ProtoMS implementation of GCMC (see section 3.5.4) as well as JAWS, where the particles in the reservoir are actually present in the simulation, constrained to the GCMC or JAWS box, with their interaction energies scaled to zero, to understand the benefits provided by these sampling techniques. While both the GCMC and JAWS simulations typically start with all the cavity waters in the reservoir (interactions scaled to zero), to better understand sampling we can imagine a protein cavity with a particular network of water molecules and possibly a ligand bound (this may represent a point during a GCMC calculation or a standard MD or MC calculation). Now we can assume that there are two different low free energy distributions of the water network within that particular cavity. Within a standard MD or MC simulation, to sample both configurations of the water network, the concerted motion of a number of particles would have to occur for the disposition of the water network to change. It can be assumed that a number of the water molecules would have to exit the cavity, for the rest to have enough space to re-position. To achieve this, in turn, solvent molecules would have to make space for the water molecules in the cavity to exit. Waters within the cavity would have to re-position and then the correct number of water molecules find their way back into the cavity. This combination of events is rare. While the re-organization of the water network in the cavity would still require a number of concerted moves within the GCMC and JAWS approaches, the number of required moves is reduced by the extra phase space (reservoir) available to the cavity waters in GCMC and JAWS. With their interaction energies scaled down, the cavity waters are free to sample the whole GCMC or JAWS box, since no energy difference will be found between any of the different configurations for the reservoir waters. This sampling while in the "gas" phase allows the waters to be inserted exactly as the point of interest, without the need to sample their way from the exterior of the cavity (solvent) to the exact required bound configuration.

In terms of comparing GCMC and JAWS, as explained previously, JAWS provides an improvement on the sampling of the particles changing between the reser-

voir state to the (fully interacting) binding cavity. The rationale behind this sampling improvement is simply the performance of smaller steps associated with the scaling of the interactions. Smaller changes in the interaction energy are linked to a smaller difference in energy, which increase the probability of accepting the move (see the beginning of this section for a further explanation).

However, if we are talking about techniques to improve sampling, naturally, JAWS is not the only alternative, and previously (see section 6.3.2) other enhanced sampling methods have been described. However, JAWS offers advantages for the specific case of the configurational sampling of molecules within a protein binding cavity. In comparison to parallel tempering (see section3.7.1), JAWS improves on the focus of the improvement in sampling to the specific area of interest within the simulation (the configuration of the molecules within the protein binding cavity). The main advantage over metadynamics (see section 3.7.2) is related to the simplicity of running the simulations, avoiding difficulties related to choosing the correct collective variables and additional simulation parameters. While the Markov State Models (see section 3.7.2) allow for the study of processes of binding to protein cavities, they have a kinetic focus, and their enhancement of sampling is centred on an iterative process, where information on how to better run further simulations is provided by MSMs. JAWS provides an enhancement in sampling intrinsic to every single run. Of the previously analysed enhanced sampling methods, solute tempering seems to be the more applicable to our particular case of interest, with its focus being specific to the molecules which require further sampling and no complex decisions being, in principle, required for its application. While solute tempering is still considered of interest for our protocols (as will be further explained in section 6.3.2) JAWS was chosen for its previous application to obtaining both affinity and binding geometry results.

JAWS to Study Binding

As shown in the previous section, JAWS can be understood as a way of improving the sampling from that which would be obtained in an standard molecular dynamics or Monte Carlo simulation of the protein cavity. However, JAWS is truly a method to study the binding geometry and affinity of water molecules to a protein cavity, and it needs to be expanded to account for fragments. Hence, arguably,

the most logical comparison is that of the capabilities offered by JAWS to those of other available methods to study fragment binding. These have been described in section 2.4 and can be separated into docking (docking and scoring) and free energy based methods. It is important to note, first that, were JAWS to be included in this classification, it would logically fall within the free energy based methods.

While docking algorithms are referred to here in the context of fragment (or ligand) binding, there are docking methods available for water molecules as well.[202] Standard docking is expected to perform more efficiently that any of the free energy based methods (including JAWS). While an study on water location methods is out of the scope of this work, an example in the comparison between JAWS and WaterDock[202] is the superior ability provided by JAWS of locating water networks within the cavity (WaterDock only provides a first water layer on top of the molecular surface per simulation run).

Some comparative examples on how the development of JAWS applied to fragments aims to improve the available free energy methods will be presented below. For further details on the free energy methods applied to fragments, please see sections 2.4.2 and 3.6.4.

- JAWS applied to fragments will present the advantage over the GCMC method applied to small ligands (Clark et al[91,92]) of providing the ligands with sampling of their internal degrees of freedom, besides avoiding the inherent sampling problem associated with GCMC and its insertion and deletion moves.

- On the SILCS[96,97] and related fragment mapping methodologies, JAWS improves by studying both the binding configuration and affinity of any desired fragment. SILCS is limited in this respect since binding geometries are only studied for a low number of simple "representative" fragments, and the binding geometry of any other ligand which affinity is to be calculated with SILCS needs to be known in advance or studied with supplementary methods (e.g. MD or docking).

- JAWS applied to fragments will improve on λ-dynamics[102] by avoiding the use of any tunable biasing potential. An improvement on the configurational sampling may also be expected from JAWS with respect to λ-dynamics, due to the independent θ parameters associated to each particle in JAWS.

- With respect to the FEP+ method,[108] JAWS applied for fragments provides the advantage of studying binding geometry as well as affinity. Besides, the calculating of relative affinities with JAWS applied to fragments is not expected to be limited to pairs of compounds, but rather a number of them might be included in the same simulation. A considerable gain in computational expense could be obtained from this implementation. Another gain in computational expense, even when considering pairs of ligands with both methods is required to the number of replicas required for a single simulation. When no enhanced sampling method (i.e. solute tempering) is applied to any of the methods, FEP+ will still require a number of replicas to run at different λ windows while JAWS for fragments would require a unique simulation replica.

In conclusion, we believe that the development of JAWS for its application to fragment binding to protein cavities, both in terms of its affinity and binding configuration and automatically taking into account the solvation of the cavity as well as possible water mediated interactions fills a void within the field of computational studies on fragment based drug discovery.

Chapter 4

The JAFS Methodology

4.1 Aims

The objective in the development of JAFS methodology is to provide a computational tool for FBDD which can assess binding of fragments to protein cavities providing all the accuracy of free energy methodologies, sampling within feasible time scales with no previous knowledge of the binding mode, while exploring unique features that no other available technique can offer. Through the development of the method this initial objective was further specified in the aims seen below:

1. Estimate the binding affinity of fragments to protein cavities
2. Estimate the binding geometry of fragments to protein cavities
3. Allow for competition between different fragment species
4. Capture water mediated interactions between fragments and proteins
5. Develop methodology applicable to fragments of realistic complexity

Each of these goals crystallized during the development of the JAFS methodology. These goals will be outlined here, but further expanded in the next sections. Briefly, an estimate of binding affinity (score) is generated with the 'JAFS score' routine, where competition between different fragment species is expected. The binding geometry is estimated in both 'JAFS score' and 'JAFS pose' procedures,

but emphasised within 'JAFS pose', which captures potential water mediated interactions. Both protocols have been developed within the framework of test cases provided by pharmaceutical companies (see chapter 5).

4.2 Method

The aim of this work was to generate a method that could be applied generally to any system of potential fragment binders and their target protein cavity, with a high level of automation. Two different approaches were developed by the end of this work, and they will be highlighted here. The results obtained with each will be later presented in chapter 7. These two distinctive protocols share a common outline, described below (section 4.2).

Theory

The JAFS (Just Add Fragment moleculeS) method is based on JAWS,[200] sharing its theoretical background. It was developed from the application of the JAWS method to small ligand (fragment) molecules. The theory behind JAWS has been described previously (see section 3.8). Briefly, it can be understood from the perspective of GCMC, and in particular the GCMC implementation in ProtoMS. Just like in GCMC, in JAWS particles are allowed to vary in number during the simulation. However, the sampling problem associated with the "insertion" and "deletion" of particles in GCMC is tackled in JAWS (and JAFS) by sampling the appearance and disappearance of the particles from the system with a continuous variable.

The JAWS and JAFS theory applies this idea by defining a continuous variable rather than a switch between the "on" and "off" states. This variable is called θ, which is allowed to vary freely between 0 and 1, where $\theta = 0$ corresponds to the "off" state and $\theta = 1$ represents "on". A flexibility (move size) is associated with this degree of freedom, just as any other, to control sampling.

The Monte Carlo algorithm associated with moves in θ space is as follows:

1. Assuming a move on θ space has been chosen among all possible simulation moves

2. Choose randomly which of the JAFS particles will be affected by the θ move

3. Choose a random displacement of θ (within its assigned range)

4. Calculate the energies of the system for the new and old θ values

5. If the energy at the new θ is lower than at the old θ, accept the move

6. If the energy at the new θ is higher, allow the move with a probability

$$p = exp(-\beta(\Delta U_{MM} - \Delta U_{PMF})) \tag{4.1}$$

   where ΔU_{MM} corresponds to the potential energy change of the system, and ΔU_{PMF} corresponds to a bias applied to take into account the affinity of the particle to water (hydration free energy — the need to apply the hydration free energy is explained in section 4.4). ΔU_{PMF} is defined as

$$U_{PMF} = m_0 + m_1(1-\theta) + m_2(1-\theta)^2 + ... + m_n(1-\theta)^n \tag{4.2}$$

7. If the move is accepted, the new θ becomes the old θ

The U_{PMF} term above will be referred as the solvation penalty or hydration penalty. The different use of the hydration penalty will be explained later in this thesis (see section 6.3.1). However, it is worth pointing out here that, during production runs, all m_x terms in U_{PMF} were set to 0, except m_1 which was set to the free energy of decoupling one molecule of that JAFS particle species from water; resulting in:

$$U_{PMF} = -\Delta G_{hyd}(1-\theta) \tag{4.3}$$

where ΔG_{hyd} is the hydration free energy of the JAFS particle.

Outline

The main steps to be followed when running a JAFS calculation in any of its forms are highlighted below. These will serve as a framework for the subsequent explanations of the different JAFS protocols in the next subsections.

Following in order as presented, to set-up, run and analyse a JAFS simulation the steps required are:

- Set up of fragment (ligand) molecules

- Running the fragments' absolute hydration free energy simulation
- Calculation of hydration free energies with MBAR (see section 3.6.3)
- Generation of the combined template file for all JAFS particles
- Set up of the protein file
- Definition of the dimensions of the JAFS box surrounding the binding cavity
- Random distribution of the JAFS particles in the cavity
- Set up of appropriate protocol of JAFS (see sections 4.2.1 and 4.2.2)
- Running JAFS
- Clustering the positions and / or scoring with θ for JAFS particles during the simulation

4.2.1 JAFS score

The JAFS score protocol is ideal when the binding region within the protein cavity is known (or for small protein cavities). The aim is to rank a number of fragments in the order of their binding affinities to the target, simultaneously locating their correct binding geometry within a small binding region. The ranking order of the fragments is defined by their score (where the highest score is associated with the top ranked fragment), an estimate of their relative binding affinity.

Theory

One of the standard, free energy based approaches to calculate the relative affinity of different fragments to a target is relative free energy simulations by (single or) dual topology alchemical transformations (as have been described in section 3.6.2). The free energy of hydration of our ligands (section 4.4) is obtained with dual topology alchemical transformations.

These methods are theoretically sound and have been widely used for many years,[23,177,203] as well as recently applied to fragments.[108] However, they have some major drawbacks, particularly when applied to the drug discovery process.

One of their main problems is the computational time and resources required. This has been frequently tackled from the perspective of directly trying to reduce the computer time required by improving sampling of the free energy landscape.[203,204]

Another hindrance to the application of these techniques in the context of drug discovery is the requirement to know the exact binding geometry for each of the ligands. Within common simulation lengths, the interconversion between different potential binding poses is often unlikely.

With JAFS score we tackle both these issues from an alternative perspective. In JAFS score, the relative affinity of a number of ligands can be estimated at once, eliminating the limit present in dual or single topology alchemical transformations, where the relative affinity of only two ligands is calculated at once. Besides, previous knowledge of the exact binding geometry is not required, only the binding region.

This is achieved with the JAFS approach presented above. All ligands, whose relative affinity in which we are interested, will have their Cartesian sampling constrained to lie within a small binding region. Each will be assigned a scaling parameter to their interaction energies (θ).

All ligands are constrained to the same, small, binding region, so that they are on top of one another. This is a highly unfavourable conformation for fully interacting particles (with many Lennard-Jones clashes). But ligands in JAFS do not need to be fully interacting, they can sample the scaling of their interaction potential. Hence, for the system to avoid unfavourable clashes, only one of the ligands is expected to present a high θ value (be mostly "on") at any one given time in the simulation. All ligands will hence be competing in θ space. In system terms, a lowest free energy minima will be sampled when the most favourable ligand (that with lowest free energy at high θ) is "on", while the rest remain with their interaction energy close to zero. Since the system is expected to spend more time at the lowest free energy, the relative affinity of these ligands can be estimated by the proportion of simulation time each of them spends with high θ (being close to fully interacting). In JAFS score simulations, relative affinity of fragments is assessed based on their simulation time spent with high values of θ.

In addition, as all ligands are expected to spend part of the simulation time

at low θ values, with their interactions close to zero, they will have the chance to freely change orientation and position within their constrained binding region. This is why the exact knowledge of their binding geometry is not required, and their most favourable binding pose should be located during the simulation time.

Protocol

Typically, two to ten potential fragment binders will be selected and one copy of each of the fragment species included as a JAFS particle in the simulation (a free sampling θ variable assigned to each fragment copy). A box of 125 $Å^3$ will be defined around the binding region. These measures were estimated as a compromise to allow enough configurational sampling while forcing overlap between fragments (desired in JAFS score calculations). The box will limit the Cartesian space available for the JAFS particles to sample, with their centre of geometry always remaining within the box limits. The initial position of each JAFS particle is set randomly within the box limits, with their θ variable set to zero.

Solute tempering (see section 3.7.1) is included in the set up of the simulation, with specifics as presented in section 4.4.

The setup of the JAFS score simulation is outlined in the list below:

1. Select the fragments and protein target

2. Follow the simulation setup outlined in section 4.2 and further described in section 4.4

3. Include, as JAFS particles, one copy of each fragment species

4. Set the JAFS box to 125 $Å^3$ around the binding region

Once the simulation is finished, analysis is required to obtain both a score to proceed with fragment ranking, and clustering to analyse the binding geometry. For the test systems and our production runs, a comparison between the clusters and the known true binding geometry is performed by calculation of RMSD. The tools used for analysis are described in section 4.5.

Analysis

Since the main objective of a JAFS score simulation is to obtain information on the relative affinity of binding of a set of ligands to a target protein cavity, its analysis focuses on obtaining the score associated with each of the fragments, consequently providing an ordered list of ligands by binding affinity. Plots for the study of the sampling of the θ variable throughout the simulation can also be generated.

The study of the binding geometry of each fragment within the small binding region can be done through clustering methods. The generated clusters are understood as binding poses, and their distance to the crystallographic binding pose can be studied by the calculation of RMSD.

One possible analysis protocol is presented below:

1. Run *plot_thetas.py* (section 4.5.2) for each fragment to:

   1.1 Obtain a score, based on the proportion of the simulation each fragment spends with $\theta > 0.5$ (see section 6.4)

   1.2 Generate plots of θ sampling

   1.3 Extract all configuration of that fragment found in snapshots where its θ was above 0.9 (see section 6.4)

2. Run *calc_clusters.py* (section 4.5.2) on each fragment to obtain their binding geometries

3. (with test systems) Calculate the RMSD of each of the poses to the crystallographic binding pose with *frag_RMSD.py* (section 4.5.2)

4. (with test systems) Study the sampling of fragments in Cartesian space with *plot_RMSD.py* (section 4.5.2)

   4.1 Of particular interest to the study of the correct functioning of JAFS pose simulations is to compare plots showing the evolution of the RMSD to crystal pose with respect to simulation snapshot (generated with *plot_RMSD.py*) with those of fragment θ with respect to snapshot (generated by *plot_thetas.py*). Abundant Cartesian sampling (variable RMSD) is expected when the

fragment presents low θ, while, ideally, the RMSD to crystal pose would be low (and stable) at times where θ is high.

4.2.2 JAFS pose

The JAFS pose protocol is ideal when the binding region of the fragment is unknown, for systems with big binding cavities, as well as systems where waters might play an important role in fragment binding. The aim is to locate the correct binding geometry of the fragment, automatically taking into account the possibility of any water mediated interactions in fragment binding. No previous knowledge is required on how well hydrated the binding cavity is. In fact, the results presented in section 7.2 include examples of cavities expected to be dry, as well as others which are known to contain large numbers of waters (see chapter 5 for a description of the protein systems used in this study).

Theory

The idea behind the JAFS pose protocol relies on the sampling potential of JAFS particles at low θ values and the realistic representation of the system presented for high values of θ.

Sampling different binding configurations with standard Monte Carlo or Molecular Dynamics simulations is challenging due to the high number of degrees of freedom whose sampling needs to be coordinated for the change between different binding configurations to happen. A particularly clear example involves highly hydrated binding cavities, where two different binding configurations involve moving the ligand between two different regions of the binding cavity. In this scenario, besides any internal degrees of freedom which need to be sampled in the ligand to change from one configuration to the other, a number of water molecules will have to change their positions and / or orientation, together with the ligand molecule. Any intermediate configuration is expected to be a lot less favourable than either of the binding geometries, hence making the interchange between binding geometries a rare event.

JAFS pose tackles this issue by adding alternative free energy paths between binding geometries. Both ligand and water molecules are defined as JAFS particles. A set of particles with high θ values represents one particular binding disposition.

While their θ values remain high, their ability to sample Cartesian space is reduced. However, any of these particles may reduce their θ, hence scaling down their interaction potential. When the θ value of one of any of these particles is low, its interaction energy will be close to zero. Consequently, the particle will be able to freely explore the whole space of the JAFS box. The energy landscape for this particle in Cartesian space is expected to be nearly flat, and the acceptance ratio consequently high for any change in position and orientation. When a favourable configuration is reached, it will be energetically favourable for this particle to increase its θ and a new binding geometry will be then formed by all JAFS particles which present high θ values at that moment in the simulation.

Protocol

In each simulation, our estimate of binding geometry for one single fragment will be obtained. While it is possible to set up the JAFS pose simulation for a number of fragment species at once, it was considered advisable to develop and test the method in its simpler version. Typically, a box of 12 x 12 x 12 $Å^3$ (measures estimated to cover the whole of the binding site for the studied systems) is defined around the binding region. Several copies of one single fragment species and several copies of water are randomly distributed within the box limits. The number of copies of fragment and water to be included in the simulation is automatically calculated as explained in section 4.4.1. There will always be the same number of fragment copies as water molecules within the box. Furthermore, the number is chosen so that particles are always in excess (i.e. there should be more JAFS particles than those required to fill the space of the binding cavity - within the limits of the JAFS box). The solute tempering set-up will be identical as that of the JAFS score protocol (see section 4.2.1).

Solute tempering (see section 3.7.1) is included in the set up of the simulation, with specifics as described in section 4.4.

The setup of the JAFS pose simulation is outlined below:

1. Select one fragment and protein target

2. Follow the simulation setup outlined in section 4.2 and further described in section 4.4

3. Include, as JAFS particles, the same number of copies of fragment and water. The specific number is simulation specific and calculated as specified in 4.4.1

4. Set the JAFS box 12 x 12 x 12 $Å^3$ around your binding region

Once the simulation is finished, a clustering analysis is performed on both fragment and water molecules. The result of this clustering should include the binding geometry of the fragment, as well as all waters relevant in the binding event (i.e. waters bridging interactions between the fragment and target protein). The RMSD between the clustering results and the crystallographic binding mode is calculated in this study to assess the success of the protocol in predicting the binding geometry of the fragment.

Analysis

As explained previously, the main objective of the JAFS pose protocol is to find the optimal binding pose of a given fragment against a target protein cavity, concurrently locating any relevant water molecules. Consequently, the analysis focuses on obtaining a set of likely binding poses from the simulation snapshots. To this end, a clustering protocol is applied to both fragment and waters.

The sampling of the θ variable during the simulation can also be studied. This can serve to corroborate frequent exchange in theta space between different copies of the fragment, where this exchange is to be expected in a well sampled simulation.

One potential analysis protocol is outlined below:

1. Run *plot_thetas.py* (section 4.5.2) on the fragment and water to:

   1.1 Extract all configurations of each particle above clustering threshold ($\theta >$ 0.9, see section 6.4)

   1.2 Generate plots of θ sampling for fragment and water

2. Run *calc_clusters.py* (section 4.5.2) on the configurations extracted in the previous step, to obtain the binding geometries

3. (with test systems) Calculate the RMSD distance between the poses and the crystal binding mode with *frag_RMSD.py* (section 4.5.2)

4.3 Software: ProtoMS

All JAFS and other free energy calculations presented in this thesis have been run in the ProtoMS software, developed in house, and freely available on its website (*protoms.org*).

ProtoMS is a biomolecular Monte Carlo free energy simulation package, whose capabilities include absolute and relative free energy calculations with single and dual topology methodologies, as well as Grand Canonical Monte Carlo.

ProtoMS started life in 2002, however tools to deal with both input generation and output analysis were limited at the time when work on this project started. The lack of appropriate tools increased the time required for system preparation and simulation set-up and analysis. It was thus considered pertinent to invest effort on the improvement of the software. The development of a new version of ProtoMS was undertaken as a group project by several members of the research group (including me).

The latest developments in ProtoMS include improvements and additions to the main software, together with a wide range of consistent and user-friendly set-up and analysis tools, which expand the usability and applicability of the code. The first version of ProtoMS which provides these features is ProtoMS3.

All calculations whose results are presented in this thesis have been run with the version of the software ProtoMS3, with all production runs using subversion ProtoMS3.1.2 plus some additional tools.

While version 3 included most of the tools that consistently saved time and effort, as well as reducing the probability of errors, during preparation and analysis of simulation data, subversion 3.1 added the parallel tempering and solute tempering capabilities that have been key to the generation of the observed results in section 7.

Within these additions to the code, the parallel tempering (temperature replica exchange) functionality to the code was mainly my individual effort (notwithstanding the relevance of sound advice from colleagues).

4.3.1 Simulations in ProtoMS

Setup

To run simulations with the ProtoMS software, command files (called *cmd* files, after their extension) are used. Each *cmd* file includes the parameters and flags to run a particular simulation type, together with paths to the required structure and parameter files and other required information (such as dimensions of the JAFS box for JAFS simulations). The *cmd* file can also be understood as the coordination input file for ProtoMS, which provides all information expected by the main source code. While *cmd* files can, in principle, be generated by hand, most commonly, they are created with the main setup coordination tool, *protoms.py* (see section 4.4.1), or with sets of individual setup tools provided as part of the ProtoMS package. The automatic setup within ProtoMS also uses external tools such as antechamber from AmberTools (from Amber12 in the productions runs of this project).[150] In some specific simulation types, as is the case for JAFS, extra setup steps may be required, such as modifications of the initial *cmd* file, or the use of particular tools after the general setup with *protoms.py*

Force Fields

Simulations run with the ProtoMS software are fully atomistic and use explicit solvent. They evaluate energies based on classical force fields, with the default force field being AMBER[139,140,147,148] for proteins and GAFF (general amber force field) for ligands. In particular, in all simulations presented in this report, the AMBER99SB[148] was used for proteins and the TIP4P[205] was the chosen water model. The TIP3P and TIP4P water models are commonly used in combination with AMBER protein force fields. In our case, TIP4P was chosen over TIP3P as no extra cost is associated with the extra interaction point within the Monte Carlo simulation with rigid solvent, while greater accuracy in the description of the energetics is expected. In terms of force field versions for ligands, for the production runs, the GAFF14 force field is used in all absolute hydration free energy simulations. In JAFS calculations, the GAFF force field is used in all cases except when extra parameters where required which are not present in this version. When that was the case, for all examples, simply the substitution of GAFF by GAFF14

provided all required parameters. Information on force field parameters for ligands are stored in *.tem* files. Partial charges stored in *.tem* files were automatically calculated during setup with the semi-empiric AM1-BCC level of theory.[206,207]

Proteins in ProtoMS

Other specifics of the default setup for ProtoMS used in the production runs of this report must be introduced. One of these peculiarities is the protein scooping. To reduce the computational requirements of protein simulations in ProtoMS, some of the external residues of proteins, which are located far from the main point of interest of the simulation (i.e. the centre of geometry of the ligand) are removed from the simulation. All residues which are next to the deleted ones are then fixed (not sampled), to avoid unrealistic changes in protein conformation. Default sizes of protein scoops are used throughout this report (i.e. 20 Å radius from the centre of geometry of the ligand, with a region of residues with fixed backbone from 16 Å to 20 Å radius).

Solvation

With regards to the solvation of the protein in water, again to gain computational speed (lower computational requirements), a sphere of water centred around the protein scoop was used as solvent, rather than a box of waters. A water box, however is the solvation method used for simulations of ligands in solvent, when no protein is present (i.e. absolute hydration free energy simulations). Both hydration protocols are performed automatically by the ProtoMS setup tools (specifically *solvate.py* which may be automatically called from within *protoms.py*). For the generation of water boxes, a pre-equilibrated water box is duplicated around the ligand, which is not translated during the simulation. In the case of spheres of water, waters are placed randomly within a radius of 30 Å all around the protein scoop, from its centre, deleting any water molecules which fall within the Atomic radius of any previously present atom of the simulated system.

Sampling

All components of the simulated system in ProtoMS can sample Cartesian space. Cartesian sampling of each component can be divided to the sampling

of external and internal components.

External degrees of freedom include translation and rotation. Solvent molecules will, by default, always (and only) sample these degrees of freedom. Ligands sample external degrees of freedom generally only when a protein is present in the system, while they are placed as the constant centre of the simulation box in simulations of ligand and solvent, or ligand in gas phase. When a protein is present, it is set as the reference frame of the simulation; the protein is always located in the centre and not allowed to sample translation or rotation.

While external degrees of freedom control the position of each element of the system with respect to one another, internal degrees of freedom control their internal conformation. Internal degrees of freedom which can be sampled in ProtoMS include bonds, angles and dihedral angles, however bonds are not sampled by default. Proteins and ligands sample their internal coordinates, while solvent molecules generally keep these internal coordinates fixed throughout the simulation.

In certain simulations in ProtoMS, such as JAFS, other degrees of freedom which are not Cartesian, can be sampled for certain particles. JAFS particles — particles included within the limits of the JAFS box — sample in θ space, besides Cartesian space, in JAFS simulations. The θ variable controls the scaling of the interaction energies of JAFS particles. Further information can be found in sections 3.5.4, 3.8 and 4.2.

In one Monte Carlo move, one or a group of these degrees of freedom may be sampled (grouping of degrees of freedom for sampling will be explained below). As explained in section 3.5.2, for each Monte Carlo move, a new configuration of the system may be generated, the energies of this configuration are calculated and compared with that of the previous configuration. The difference in energy will be entered in the Monte Carlo test whose output will determine whether the new configuration is accepted or the old one is kept. When the energy of the new configuration is higher, the bigger the difference in energy, the lower the likelihood of accepting the configurational change. As seen in this brief summary, the first step of the move is the generation of the new state of the system. While this new state could in principle be chosen at random from all possible states of the system, this method would decrease the sampling efficiency by often generating states with high associated energy, hence increasing the possibility for rejection of

the configurational change. Instead, some limits are applied to how much each of the sampled degrees of freedom can be changed in each attempted Monte Carlo move. Within ProtoMS, these limits are called the flexibilities of each degree of freedom. While the flexibilities associated with the ligand degrees of freedom may be defined by the user, those associated with protein and solvent moves are provided within the force field files in ProtoMS.

Every time a new move is to be performed, the type of move is first randomly chosen according to a user-determined probability per move type. "Move types" in this context include each of the molecule types ("protein", "solute", "gcsolute", "solvent") related to the Cartesian space sampling of molecules within this type. It also includes move types only relevant in specific simulation classes, such as "insertion" and "deletion" for GCMC or "theta" for JAWS and JAFS. Extra move types exist, such as "volume", for simulations run within the NPT ensemble. When the move related to one of the molecular types (i.e. "solute") is chosen, firstly, one of the molecules of that class present in the simulation is chosen, all weighted equally (for the case of protein, one of the residues is chosen). Within that molecule, all internal degrees of freedom, as well as translation and rotation may be altered (according to the previously defined "flexibility" term). In the case of proteins some differences are observed, where only backbone, only side chain or both might be moved. When the backbone is moved (either with or without moving side chain), neighbouring residues are also altered in their conformation to allow for an adjustment of the protein conformation. For the case of the simulation specific moves, these are related to a specific molecule type (that which presents the associated flexibility, i.e. gcsolutes for GCMC). One of the molecules present of such type is chosen and the change associated with the move type is applied. In the case of "theta" moves, for JAWS and JAFS sampling of θ space, a flexibility is also defined as ProtoMS input for each of the ligands. Moves which are not associated with a particular molecule type (i.e. "volume") are performed on the system when chosen.

After the change to the system (or particular element of the system) has been performed, the new energies are measured and, as described above, the move is accepted or rejected based on its energy difference with the previous configuration of the system.

4.4 Simulation Setup

Protein Setup

The protein file for each system to run JAFS was chosen among the structures present in the Protein Data Bank (*www.pdb.org*) in complex with one of the fragments, with no particular criteria.

1. Choose one conformation for residues which might present several alternatives in the crystallographic structure
2. Choose the optimal protonation state for titrable residues and tautomers (histidine, acidic and basic residues). The protonation state is chosen based on observation of optimal hydrogen bond pattern
3. Change of residue names corresponding with the protonation state chosen
4. Protonation, following the patterns of the residue names assigned, performed with tleap (from the AmberTools[150])
5. Leaving the crystallographic waters, proceed with the scooping and solvation of the protein. This will be done automatically with *protoms.py* when setting up any simulation where the protein is included. The most simple example, with a ligand present could be run with
   `python $PROTOMSHOME/protoms.py -p protein.pdb -l lig1.pdb -s none`

Ligand Setup

When talking about ligands in our calculations, we refer to the fragment molecules whose binding configuration and / or score we are interested in studying.

Ligand structures were obtained from two different sources. For systems and ligands associated with publications, where their crystal structure complexed with the target protein was publicly available, the initial ligand crystal structure was taken from the complexed protein pdb (from the Protein Data Bank *www.pdb.org*). Several sets of ligands were provided directly by collaborators within pharmaceutical companies. Among those, binders followed the same set-up as outlined below, with the initial protein structure being that given to us by our collaborators. In the case of decoys, no minimization step was followed.

Starting from the initial ligand pdb, the ligand set-up followed certain steps:

1. Protonate ligand using the Chimera 1.8 software[208]

2. Generate the required *antechamber* configuration and parameter files by running

   ```
   $PROTOMSHOME/protoms.py -l ligand.pdb -s none
   ```

   where *ligand* must be replaced by the corresponding ligand name.

3. Minimize the configuration of the ligand (see below)

The minimization of the ligand is run in sander (within AMBER 12[150]). It comprises a hundred steps of steepest descent minimization in the gas phase with a non-bonded cut-off of 20 Å. The script used for the minimisation in sander is provided in the Appendix, in section 11.1. Notice that, by running this form of setup, the flexibilities assigned to internal degrees of freedom of the ligand, as well as its translation and rotation moves, are assigned automatically by *protoms.py*, based on sets of empirical rules.

Hydration Free Energy Runs

The hydration free energies are obtained commonly in ProtoMS by calculating the free energy of decoupling the ligand from water. 16 equally spaced λ values are distributed between $\lambda = 0$ and $\lambda = 1$. A replica-exchange dual topology (see section 3.6.2) simulation is performed in which the ligand is transformed into a dummy (non-interacting) atom. At $\lambda = 0$, the fully coupled ligand is simulated while at $\lambda = 1$ only a dummy particle in a box of water is present. In this project, the MBAR methodology (see section 3.6.3) was used to calculate the decoupling free energy from the simulation results. This free energy is needed as input for the `jpmf` parameter in the input file of each JAFS particle.

Three different repeats of hydration (decoupling) free energy calculations were performed. The setup for these runs is automatic with *protoms.py* by using the following command:

`$PROTOMSHOME/protoms.py -l ligand.pdb --absolute -s dualtopology -r 3`
where *ligand* must be replaced by the corresponding ligand name. In all hydration free energy simulations presented in the report, 5 million equilibration and 40 million production moves are run per λ window (*protoms.py* default).

The average decoupling free energy for the three repeats was calculated. Again, the analysis is run automatically using the tool provided with ProtoMS *calc_dg.py*. The calculation of MBAR requires the installation of some dependencies. The required analysis command is as follows:

```
$PROTOMSHOME/tools/calc_dg.py -d out1_free/ out2_free/ out3_free
```

where `out?_free` (with `?` in `1 2 3`)corresponds to the default naming of out folders provided in the *cmd* file by *protoms.py* (the correct names of the out folders must be provided if these have been changed).

JAFS Runs

JAFS ligand templates

Not all parameters available to ligand template files are consistently used; some are specific requirements for certain simulation types. This is the case for JAFS.

All JAFS particles will be treated by ProtoMS as ligands, hence requiring an associated template file. For JAFS particles to present their full correct behaviour, extra parameters have to be included in their template files. These include `jtheta`, `jcorr` and `jpmf`.

The main distinctive feature of JAFS particles is their ability to sample the scaling of their interaction energy (sample θ space). Just as for any other degree of freedom, the flexibility for the θ moves needs to be set. This will be assigned with the `jtheta` keyword. See section 3.5.2 for further information on the flexibility of MC moves.

A concentration correction is available on JAFS simulations, however the applicable scaling factor was set to 1 (no scaling) for the simulations presented in the report. This parameter can be used to account for the different concentration in solution of diverse JAFS particles. The original idea links to the expected higher concentration of water than fragment molecules in solution. Consequently applying the concentration correction in this case would favour the interactions of water with respect to fragments. It was found, in studies previous to my intervention in

the project, that such effect resulted in seldom sampling fragment configurations at high θ values and was hence deleterious for obtaining useful data from the simulations. The scaling associated with the concentration correction is applied with the `jcorr` keyword.

The last of the JAFS specific parameters in the template files is particularly relevant for JAFS score calculations, where the relative binding affinity of several ligands is estimated. `jpmf` is the keyword used to specify the hydration penalty of the fragment. The penalty is applied to each of the JAFS particles to account for their affinity for the solvent water. Throughout a JAFS calculation, the JAFS particles repeatedly sample the exchange between their vacuum state (for $\theta = 0$, when their interaction energy is 0) and their bound state (when $\theta = 1$, fully interacting, since their Cartesian space is limited and they are forced to remain within the protein binding cavity). If no hydration penalty is applied, the score obtained with JAFS will be an estimation of the binding free energy from vacuum. But both *in vivo* and *in vitro*, the process of ligand binding occurs between the ligand in solvent and the ligand in the binding cavity. The hydration penalty will then account for the free energy of taking the ligand from solvent to vacuum, completing the free energy cycle.

The need for the hydration penalty (`jpmf`) is explained in figure 4.1. In the figure, when no hydration penalty is applied, the difference between the free energy of going from E to A and that of going from F to B is what the JAFS score estimates. The hydration penalty on the template for ligand *orange triangle* accounts for the free energy of going from C to E, just as the hydration penalty in template for ligand *green square* accounts for that of going from D to F. The score (proportion of simulation time with $\theta > 0.5$) obtained from a simulation with correctly applied hydration penalty will be an estimate of the difference in free energy between going from C to A and going from D to B.

JAFS box

As mentioned previously, the protein structure used is taken from the complex with one of the fragments in the system. Consistently, the JAFS box is centred on the centre of geometry of that same fragment. The idea behind this decision on the protocol is based on a hypothetical situation where a JAFS calculation is made on a set of fragments for a particular system, where only one of them has been

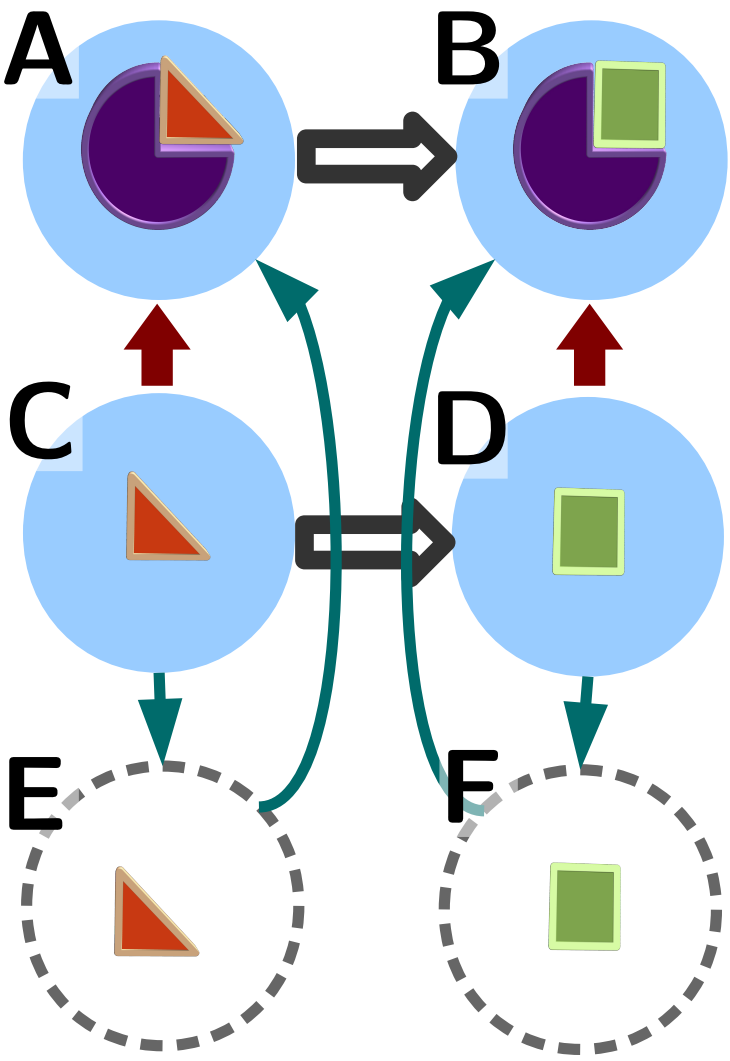


Figure 4.1: Schematic representation of several possible thermodynamic cycles to obtain the relative binding free energy between two different ligands to the same target protein. In the figure, the purple pie-shape (full circle minus a quadrant) represents the protein, the orange triangle and green square are ligands, the blue circle represents the solvent water and the dashed circumference means vacuum. The difference between both dark red arrows is the relative binding free energy between those two ligands. Two different free energy cycles are present in the figure, both being different routes to reach the same relative binding free energy. Arrows in turquoise represent the free energies estimated with the JAFS score protocol. Black arrows are calculated with standard binding free energy techniques (i.e. alchemical single/dual topology transformations).

previously studied and the crystallographic structure of that complex is known. In a different scenario, the exact placement of the JAFS box may be decided based on the observation of the dimensions of the protein cavity — directly, or based on cavity prediction software.[209–212]

Solute Tempering

The solute tempering setup chosen comprised of a maximum 'solute temperature' of 100°C, with 14 equally spaced intermediate solute replicas and a lowest (and main) replica at 25° C (16 replicas in total, with 5°difference between replicas). Swaps between replicas were attempted every two hundred thousand Monte Carlo moves.

The application of solute tempering to all the simulations presented throughout this report corresponds to a scaling of all interaction energies of the ligand (namely ligand-ligand, ligand-protein and ligand-solvent, both coulombic and Lennard-Jones terms) with $\frac{T_0}{T_i}$ while all other energies (internal energies of the ligand, and all energies not involving any ligand) remain unscaled (that is, they would be identical for any "solute temperature"). Here, T_0 corresponds to the reference "solute temperature" (here always 25°C) and T_i to that associated with each given replica.

The application of this setup in ProtoMS requires the two lines below to be included in each *cmd* file:

```
temperaturere 200000 25.0 30.0 35.0 40.0 45.0 50.0 55.0 60.0 65.0 70.0
75.0 80.0 85.0 90.0 95.0 100.0
solutetempering 25.0 bndang 3 dih 3 lj 3 coul 3 solu 1 prot 1 solv 1
```

Further information on solute tempering and its implementation in ProtoMS can be found in section 3.7.1. Correlating the aforementioned scaling with equation 3.38, interaction energies of the ligand are included in E_1 and all other interactions in E_3.

The solute tempering set-up was optimized for best results (described in section 6.3.2), but also taking into account the most convenient structure in our high performance computing facility (i.e. 16 replicas in 16 cores per node). A different choice in number of replicas or spacing might be better suited for other facilities.

4.4.1 Tools

Several tools provided with ProtoMS are used throughout the setup and analysis of (mainly JAFS) simulations presented in this report. Each of these will be presented below. Note that all these tools were developed as a group effort by several members of the research group (including me).

protoms.py

protoms.py is the initial script and coordinates the setup tools offered with ProtoMS. The whole setup to standard ProtoMS simulations can be completed with *protoms.py*. Defaults are provided for all but the essential parameters (i.e. ligand and/or protein files, type of simulation), while a wide range of customization is available through an extended set of flags. Help within the script is provided in its reduced version which highlights the most commonly customized parameters, as well as in its full version (`--fullhelp`) with all provided options. Simulations can equally be set up using individual tools provided, with certain levels of customization only available through this route. This modular structure allows for full flexibility as well as complete automation of the most common simulations in ProtoMS.

Even for non-standard simulations, where full, direct setup through *protoms.py* is not provided, this coordinator of tools is still frequently used to create files that may be subsequently manually modified, or complemented with additional tools. In this way, even if several steps are required in the desired setup of a non-standard simulation, these can easily merged together in user-created simple scripts.

An example of the use of *protoms.py* for standard simulations is the setup of the hydration free energies of fragments, where the required files, ready to run, are generated as direct output of *protoms.py* given the correct parameters as flags (see above).

An example of the set-up of non-standard simulations is that of JAFS simulations. As has been previously presented, the *protoms.py* script is used at several points in the setup process, such as ligand setup (see above), and generation of a, later modified, initial version of the JAFS *cmd* files.

distribute_waters.py

distribute_waters.py is the script used within ProtoMS to randomize the position of solvent and solute molecules within a provided cubic space. It is used within the set-up of the JAFS calculations to distribute the JAFS particles (fragments and water molecules) within the JAFS box around the protein binding cavity.

distribute_waters.py takes the dimensions of a box, the number of particles that should be included within that space and a pdb file with the coordinates of the particles to be distributed. It can also be run with the name of a water model ("t4p" = TIP4P or "t3p" = TIP3P) instead of a pdb coordinate file, in which case, waters will be automatically generated and randomly distributed. Particles are placed within the box in such a way so that their geometry centre is always within the box limits, while part of their atoms can lie outside the provided volume. This arrangement corresponds perfectly to that required by JAFS runs, as JAFS particles are constrained to keep their geometry centre within the limits of the JAFS box throughout the simulation.

The header of the file generated with *distribute_waters.py* must be changed to match that of ligand files to proceed with JAFS simulations.

cavity_volume.py

cavity_volume.py is a script provided within the ProtoMS tools which can be used to estimate the volume of a protein cavity as well as (an excess of) the number of copies of a molecule that would be required to completely fill the cavity.

The calculation of volumes within the script is grid based. For the calculation of the volume of the cavity, a grid of imaginary points is generated to cover a user-specified cubic volume. Within that volume, the grid points that lie within a given distance of one of the protein atoms are deleted, the remaining grid points are added up and multiplied by a unit volume associated with each grid point, the result being the estimated volume of the cavity. A similar technique is used to estimate the volume of each of the provided "ligand" molecules, where each of the grid points that lies within a given distance of one of the ligand atoms is added up and multiplied by their associated unit volume. To calculate the excess of copies of the molecule required to fill the cavity, the estimated volume of the cavity is divided by that of the ligand.

This script is used within JAFS calculations to determine the number of copies of fragment and water that must be included in the calculation to overfill the cavity. The script can estimate the number of copies of two different molecules, being the same number of each, required to fill the cavity. This result is the number of fragment and water molecules included in a JAFS pose simulation, where the cubic volume corresponds to the limits of the JAFS box.

merge_templates.py

While individual template files can be used for each ligand in ProtoMS, it is advised that the templates of all ligands are merged into one unique file which is then included in the ProtoMS *cmd* file. The reason for this recommendation is that individual template files share common numbering for certain parameters that need to be different for each ligand included in the simulation. While it is possible to access manually each of the ligands and make the numbering compatible, it is highly error prone. The *merge_templates.py* tool takes care of merging the templates of several ligands in one unique (separate) template file, automatically assigning them compatible numbering.

Within the JAFS calculations this script will be used to merge templates for all JAFS particles that must take part in a simulation in one combined template file.

4.5 Simulation Analysis

4.5.1 Clustering

Clustering can be defined as the process of dividing a bigger set of entities into smaller sets or clusters.[213] In particular throughout this thesis the term clustering will be used to refer to the grouping of sets of molecular configurations into smaller sets and choosing a representative configuration of each set (which we will call cluster representative or simply cluster). The aim of this process is to study the relevant configurations visited during a simulation in a more comfortable (and human readable) manner.

Throughout this thesis, the clustering method of choice will be hierarchical clustering. In all production runs and some of the development runs (chapters 7

and 6 respectively) the clustering will be performed with the *calc_clusters.py* tool provided with ProtoMS (section 4.5.2) using default parameters, unless otherwise specified. However, the runs presented in section 6.2 follow different clustering methods (the current version of *calc_clusters.py* was not yet available), which are specified within that section.

The clustering methods used within this thesis are hierarchical and deterministic. In most cases (all but one clustering method used in section 6.2), no information on the number of desired clusters is provided *a priory* to the algorithm. Instead, information is always given on the cut-off distance between posses to be part of the same cluster.

The cluster representatives shown in the graphical representations throughout the thesis are obtained as the copy of the molecule taking part in a cluster which is closest to the centroid of that cluster. Here, the centroid of a cluster is obtained as the mean position of each atom of the molecule in that cluster.

4.5.2 Tools

calc_replicapath.py

calc_replicapath.py is a tool designed to visualize the efficiency of swaps between replicas when running λ, temperature or solute replica exchange simulations (also known as reti, parallel tempering and solute tempering respectively). It follows the processor id for each replica exchange simulation through the simulation snapshots. The value of replica (λ, temperature or "solute temperature") that each processor presents at every simulation snapshot is then plotted (with replica value on the y axis and snapshot on the x). Examples of the results of these plots can be seen in figures 6.16 and 6.15. Note that each coloured line corresponds to one processor id.

frag_RMSD.py and plot_RMSD.py

frag_RMSD.py and *plot_RMSD.py* are in-house tools not included with the ProtoMS package version 3.1.2 (used for production runs).

frag_RMSD.py is an in-house tool used for the calculation of RMSD between the crystallographic binding pose and clusters obtained as result of JAFS simu-

lations (both JAFS pose and JAFS score). This is a simple tool and does not account for any symmetry during RMSD calculation. Symmetry was hence taken into account manually, by generating any required symmetrical equivalents of the crystallographic pose as required per ligand. Furthermore, this tools calculates RMSD by assuming atoms correlate based on their order of appearance in the pdb file structure of the molecule. Since ProtoMS generates output file with a custom ordering of atoms for each molecule, the pdb file of the crystallographic pose (and any required symmetrical equivalent) were converted to the same atom ordering as that of the ProtoMS output before the RMSD was calculated.

plot_RMSD.py is, as its name suggest, used to plot the RMSD of a copy of a fragment molecule to crystallographic binding mode with respect to simulation snapshot. The "time" evolution of the RMSD distance of a fragment to the (considered) "correct" binding mode is deployed. Calculations of RMSD follow the same procedure as those in *frag_RMSD.py*. Examples of these plots can be seen in figure 7.3 and figure 7.4.

plot_thetas.py

plot_thetas.py is a tool included with the ProtoMS software. It is commonly used to analyse the output of JAWS and JAFS simulations. It can, in principle, be used on the output of GCMC simulations as well, but other tools might be better suited for these data.

As indicated by its name, its main use is to produce plots of the evolution of the θ variable throughout the simulation. This information is represented by a set of plots, with a display of θ values for a given JAFS particle in the form of histograms as well as the evolution of θ with respect to simulation snapshot ("time" evolution of θ in the simulation). When several copies of the same fragment type are included in a simulation, the evolution of all copies are plotted in together, hence the number of copies with high or low value of θ can also be estimated from these plots. A modified version of this tool was used to analyse the production runs. This modified version also provides the proportion of simulation snapshots that one particular fragment spends above a user specified θ threshold. Equally, the mean value of θ for the given fragment is printed. Note that this modified version of the tool is not provided with the ProtoMS package version 3.1.2 (used

for production runs).

calc_clusters.py

calc_clusters.py is a tool provided within the ProtoMS suite. Together with *calc_densities.py*, they offer different options to visualize the space sampling of a given molecule through a ProtoMS simulation. The visualization offered by *calc_densities.py* can be of particular interest when dealing with water molecules, and the script has mainly been designed for this purpose. As suggested by its name, *calc_densities.py* does not provide individual configurations of the molecule of interest as a result. Instead, it generates a map of the positions of the most central atom of the relevant molecule using all provided simulation snapshots by means of a grid distribution.

calc_clusters.py on the other hand, gives unique configurations of the molecule of interest as output. This is useful for JAFS pose calculations, which aim to provide the most likely binding configurations of particular fragments (and water molecules). *calc_clusters.py* is hence the main analysis tool of the output of JAFS pose simulations.

The generated clusters are printed in order of occupancy (the number of configurations included in each cluster), with the exact value for each cluster included in the last column of the pdb file. While occupancy of clusters may be used as a mean of ranking, unsatisfactory results (see section 8.1) suggest this is not a recommended practice here.

The clustering method used in *calc_clusters.py* as of ProtoMS version 3.1.2 is hierarchical clustering, where the default clustering threshold is 2 Å but, as will be obvious in section 7, this can be customized.

4.6 Summary

In this section, JAFS, the method developed throughout this thesis has been presented. The theoretical understanding of this method has been described, which is common to the previous JAWS method and based on GCMC and inspired by the ProtoMS implementation of GCMC. The two protocols developed within the JAFS methodology have been presented: JAFS score and JAFS pose. The specific theory

behind each of this protocols as well as their specific set-up has been presented. Furthermore, the tools required for the set-up and analysis of JAFS have been described individually.

In summary, in this section we have presented the method developed to the study of the relative affinity of different fragment molecules to a common region of a protein cavity (JAFS score) as well as the study of the exact binding configuration of fragments to a protein cavity together with the solvation pattern of that same cavity (JAFS pose). In chapter 6 the development of the method until it reached its latest state of development (as presented in the current section) will be described. In chapter 7 the results obtained when applying the fully developed JAFS method to the systems seen in chapter 5 will be shown.

Chapter 5

Systems

A range of systems were chosen to test both protocols of the JAFS method (JAFS score and JAFS pose, see section 4.2). Some of these systems are expected to present no particular difficulty in order to test basic functionality of the methodology. Others present different challenges to assess the limits of what JAFS can achieve. In this section the systems will be introduced, the potential challenges presented and their interest within the field of drug development or computational chemistry explained. Please note that this section does not intend to explain each of the systems in depth from their biological perspective, nor to explain all studies these systems have undergone within computational chemistry. Within the brief introductory section to each of the systems, references to biological and computational developments for that system will be provided.

All systems, with their respective ligands, can be found in figure 5.9. 2D representations of the ligands can be seen in the Appendix, table 11.1.

5.1 Model Systems

Some of the test systems chosen to develop and study the performance of the JAFS methodology can be qualified as model systems. These are comparatively simple systems, or present a particular property which tests some specific aspect of the JAFS method. They may not be therapeutic targets and have not been chosen for their interest to the pharmaceutical industry. In these cases, the interest lies in their physical properties rather than their biological functionality.

The two model systems used throughout the development of the JAFS methodology will be presented in this section.

5.1.1 T4 Lysozyme

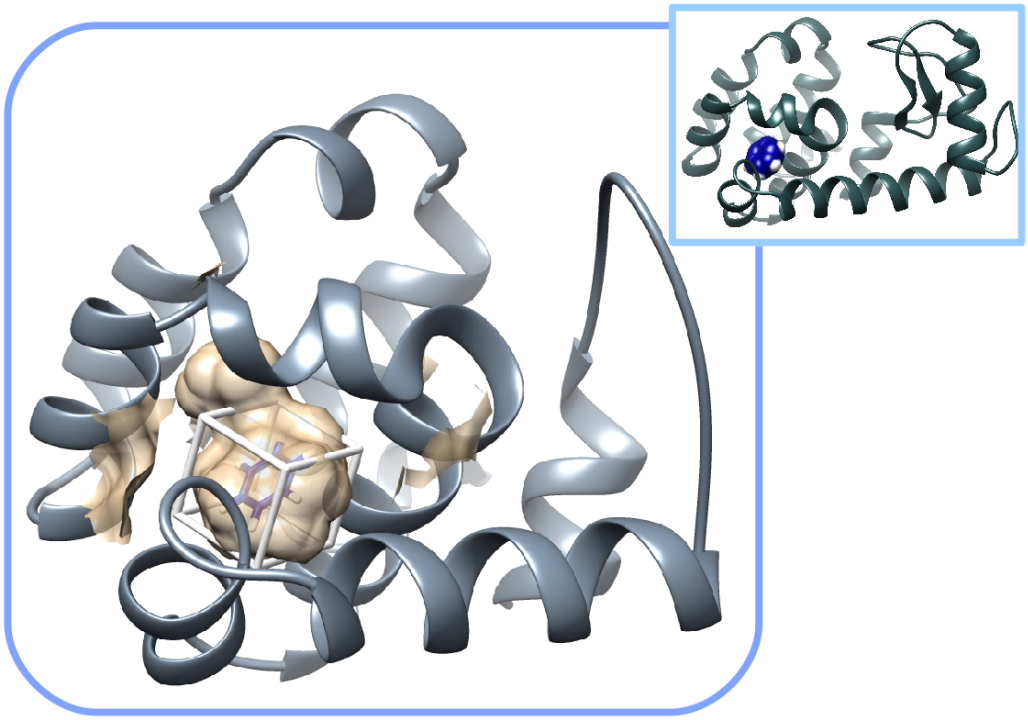


Figure 5.1: T4Lys scoop (left, bigger) and full protein structure (right, smaller). Note that the scoop mostly retains a subdomain of the protein. In both cases, the protein is displayed as ribbons in grey and the ligand (bnz) in dark blue. On the left, the surface of the binding cavity is shown in light brown. The JAFS box is shown in white. PDB code 181L.

T4 Lysozyme (T4Lys) was chosen as a test system due to its simplicity and well studied behaviour. It is a small globular protein for which crystal structures have been known for a number of years (since 1986).[214] This protein has been widely used on studies of protein structure, and, in particular, on the structural effect of mutations on protein residues.[214–220] A number of these mutations generate cavities within the structure of T4Lys, which have been commonly used in computational chemistry as simple test systems.[220–222]

In this project we have taken T4Lys with its L99A mutation, which generates

a small apolar cavity, as a simple test case on the performance of JAFS. Fragments of known experimental binding affinities have been chosen.[223]

The T4Lys mutant studied here presents an apolar, non solvent exposed cavity, of such small size that binders generally fill most, if not all, of the binding cavity. This can be observed in figure 5.1, where benzene (bnz) is shown bound to T4Lys L99A (referred simply as T4Lys, from now on). As can be seen, the cavity is formed by two linked (nearly) spherical subcavities. Most ligands bind to the main (bigger) subcavity. However, one ligand (nbb, see figure 5.9 or the Appendix, table 11.1, for the 2D structure) will be presented in this project which does bind its side chain into the smaller subcavity (to the side).

This system will be used exclusively on JAFS score calculations, since most of the cavity can already be covered by the small box that delimits these simulations, and one unique region of the cavity where ligands could bind can be comfortably defined. In this context, the JAFS box constraining the location of the centre of geometry of the fragments in the JAFS calculations is also displayed in figure 5.1. Given the steric restrictions provided by the protein and the limits to conformational sampling imposed by the JAFS box, we can be confident no two fragments could fit within the volume available without overlapping.

Fragments chosen to study performance on this system can be found as molecular models in figure 5.9, as well as 2D structures in the Appendix, table 11.1, under the names: bnz, 1mp, nbb, dcb and wa1. The structures of the co-crystallized protein-ligand complexes are those with PDB code 181L, 2OU0, 186L and 2OTY respectively (the water molecule, wa1, is not expected to bind as its structure corresponds to that of the Tip4P water model and binding cavity is hydrophobic). During most of the study the protein structure used was that of 181L. In some cases 186L was used, this will be indicated where appropriate.

5.1.2 Mouse Major Urinary Protein I

The mouse Major Urinary Protein I (referred as MUP or MUP-I throughout the thesis) was chosen as a test system for the JAFS method due to its particular property in terms of cavity hydration. MUP presents a hydrophobic cavity, but in contrast to that of T4Lys, this is considerably bigger. While the cavity is big enough to accommodate a number of water molecules at the same time as any of

the fragment binders within this thesis, the rest of the cavity is found empty, with the minimum energy configuration corresponding to no water molecules present in the cavity, on average (GCMC studies performed within he group — Dr. G. Ross, personal communication). Equally, when the binding cavity is empty of ligand, enough space is available for a water network to be present. On studies of solvation run with GCMC within the group,[93] the conformations associated with 0 to 6 water molecules bound were found to be thermally accessible (the number of water molecules is expected to fluctuate). The minimum energy configuration is associated with two water molecules where the two most occupied clusters of water configurations correspond to the conserved waters in the crystal structure of the apo protein (PDB code 1I04).[93] Note that there are other ligands, not used in our simulations, where bridging waters are found in the cavity, co-crystalized with the ligand, mediating interaction between ligand and protein (e.g. PDB code 1I05).

It is for cases like this, where the solvation state of the system (dry cavity) may prove hard to predict, that JAFS can be of particular use. In the case of JAFS pose, to correctly predict the solvation state of the binding cavity, and hence obtain the correct binding poses, JAFS will need to automatically detect the hydrophobicity of the pocket, to prevent the appearance of water molecules. In the case of JAFS score, the available sampling volume is higher than that of other systems (e.g. T4Lys, see above) due to the absence of waters and protein residues close to the limits of the JAFS box. In MUP, ligands displaced to the edges of the box (where it is only their centre of geometry that is constrained to the JAFS box limits) may not find any steric clashes disfavouring their conformation. In systems such as T4Lys, any ligands displaced to the edge of the box are likely to find steric clashes with protein residues. The same effect would be expected if a bigger cavity were heavily hydrated. It is hence both the big volume and dryness of the binding cavity that is effectively increasing the volume available to sample by the interacting fragments within the JAFS score calculations. It is important to remember that JAFS score relies on the fragments overlapping, so that only one of them presents high values of θ at any given time. If several fragments can present high θ at the same time, the assessment of relative affinity of binding through the scores (proportion of simulation time with $\theta > 0.5$) provided by JAFS might be compromised.

Fragments chosen to study performance on this system can be found in figure

5.9 and in table 11.1 under the names: ipz and prz. The structures of their protein complexes have PDB codes 1QY2 and 1QY1 respectively, where 1QY2 was the protein conformation chosen to perform the simulations.

In figure 5.2, the structure of MUP with ipz bound can be seen. Both boxes for JAFS score (smaller one) and JAFS pose (bigger one) are also shown surrounding the binding cavity.

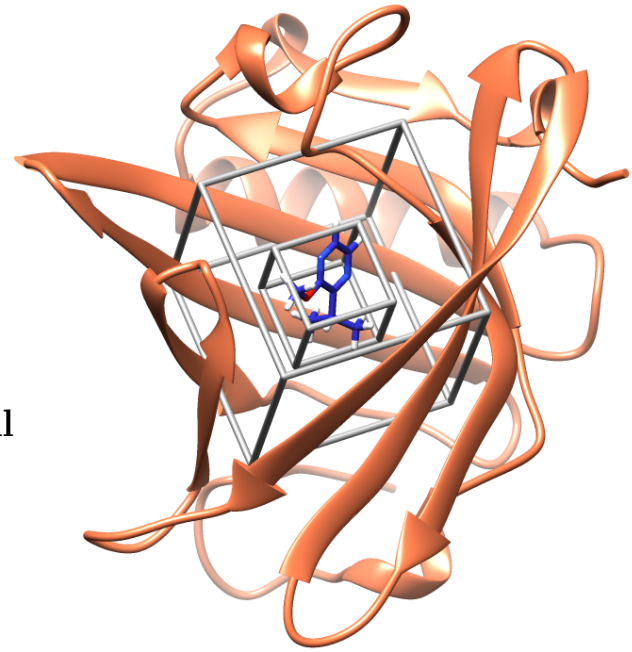


Figure 5.2: MUP-I structure (orange ribbons), PDB code 1QY2. The bound ligand (ipz) is shown in dark grey and the two JAFS boxes (JAFS score the smaller and JAFS pose the bigger one) are shown in light grey. Note that, by following the default setup as provided by *protoms.py*, no scoop is required in this case, due to the small size of the full protein structure in the PDB file.

5.2 Pharmaceutical Targets

Certain test systems were selected due to their relevance within the pharmaceutical industry. These systems are expected to present equivalent properties and level of difficulty to those on which JAFS will likely be applied in the future. In these systems, their biological relevance and reasons for their targeting by the pharmaceutical industry will be presented. While mainly chosen for their pharmaceutical relevance, together with the availability of binding affinity experimental data and X-ray crystal structures, they also present physical and structural properties affecting the binding process which will be explained

below.

5.2.1 Heat Shock Protein 90

The Heat Shock Protein 90 (HSP90) system was provided by our collaborators at Astex Pharmaceuticals. Their interest in this protein is related to its role in cancer cells and cancer development. This role is complex and wide spread throughout cancer types, being related to the control of cell apoptosis and the maintenance of the functional state of oncogenic mutated proteins.

HSP90 for Development

A system had to be chosen for the initial studies on the project. This system would be used throughout the development of the methodology. The system was chosen to fulfil certain criteria:

- Possibility to validate the results obtained by the application of the new methodology to the system. Reliable experimental data has to be available to compare to the new results.
- Representativity. The system of choice must be a model for those to which the new methodology will be applied.
- Simplicity. Initially, the basic performance of the methodology must be assessed. Once this step has been fulfilled, further traits can be added to the complexity of the test case.

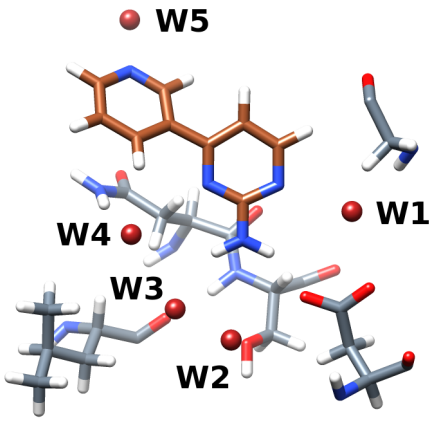


Figure 5.3: Water mediating interactions (dark red) between ligand f01 (brown) and HSP90 residues (grey). PDB code 2XDK. Names to water molecules assigned for reference W1 to W5.

With this criteria in mind, HSP90 and the co-crystallized fragment f01 were chosen (see figure 5.4).[25]

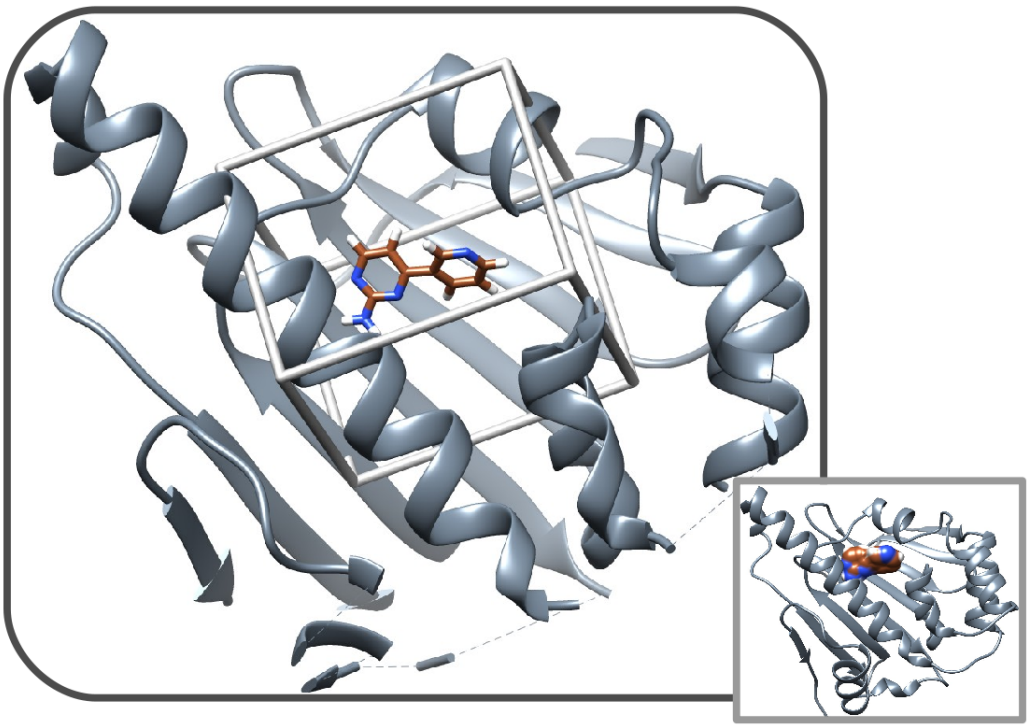


Figure 5.4: HSP90 scoop (left, bigger) and full protein structure (right, smaller). Note that the scoop includes most of the structure of the full protein. In both cases, the protein is displayed as ribbons in grey and the ligand (f01) in dark brown. On the left, the JAFS box is shown in white. PDB code 2XDK.

The fragment used through my work on the development of the methodology, f01, can be seen in figure 5.5. It establishes interactions with the protein through four conserved water molecules, which are "W1" to "W4" in figure 5.3. The key dihedral angle to be sampled is that between both aromatic rings. Notice that waters from "W1" to "W3" share their bound conformation with those found in figure 5.6 (from right to left, "W1" to "W3", respectively).

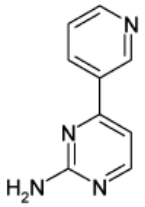


Figure 5.5: 2D structure of f01.

HSP90 for Production

As well as the relevance of its biological function and its role as a target for cancer therapy (see below), HSP90 also presents physical properties of particular interest to our analysis of JAFS performance. It

should help assess whether JAFS can provide all the automation expected in its treatment of waters in the binding cavity. The hydration of the binding cavity of HSP90 presents opposite challenges to those offered by MUP (see section 5.1.2). In contrast to MUP, HSP90 presents a highly hydrated binding site.[224] The JAFS pose protocol must automatically detect this hydrophilicity. During the simulations, the pocket must be filled with both the fragment and a number of water molecules so that the correct solvation states of the cavity are sampled and the binding geometry can be found. Following the black-box approach to JAFS, the setup for both, MUP and HSP90 systems, will be identical.

Not only is the cavity in HSP90 expected to be filled with waters, but the exact positions of these water molecules are essential to the binding of the fragments. In HSP90, three to four waters are conserved between several crystal structures of bound ligands (see for example structures with PDB codes 2XDK, 2XDU and 2XDL) mediating key interactions between the binder and protein. The three waters conserved throughout all binders studied in this project can be seen in figure 5.6. The right-most water in the image (called W1 in figure 5.3) mediates an essential hydrogen bond in the binding of all fragments presented in this thesis. In preliminary studies of the affinity of the conserved waters for HSP90 (data not shown) W1 has proven to be the most tightly bound of all these waters.

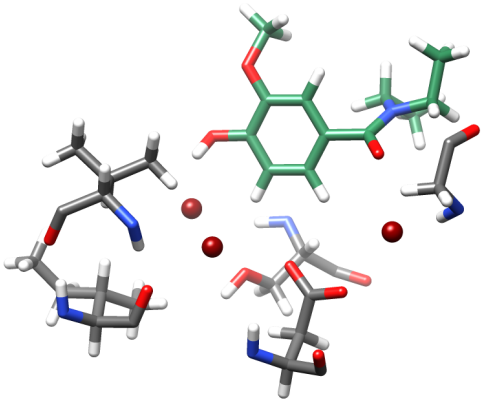


Figure 5.6: Water mediating interactions (dark red) between ligand 2dl (green) and HSP90 residues (grey). PDB code 2XDL

Fragments chosen to study the performance on this system can be found in figure 5.9 and in table 11.1 under the names: 2dl, ata, atb, atc atd ate and atf. The protein structure used during the simulations is that co-crystallized with 2dl (PDB code 2XDL). Note that the rest of ligands used do not have a complexed structure in the Protein Data Bank. Their RMSD to crystal binding geometry are calculated to crystal structures obtained by our collaborators at Astex Pharmaceuticals (Marcel

Verdonk, personal communication).

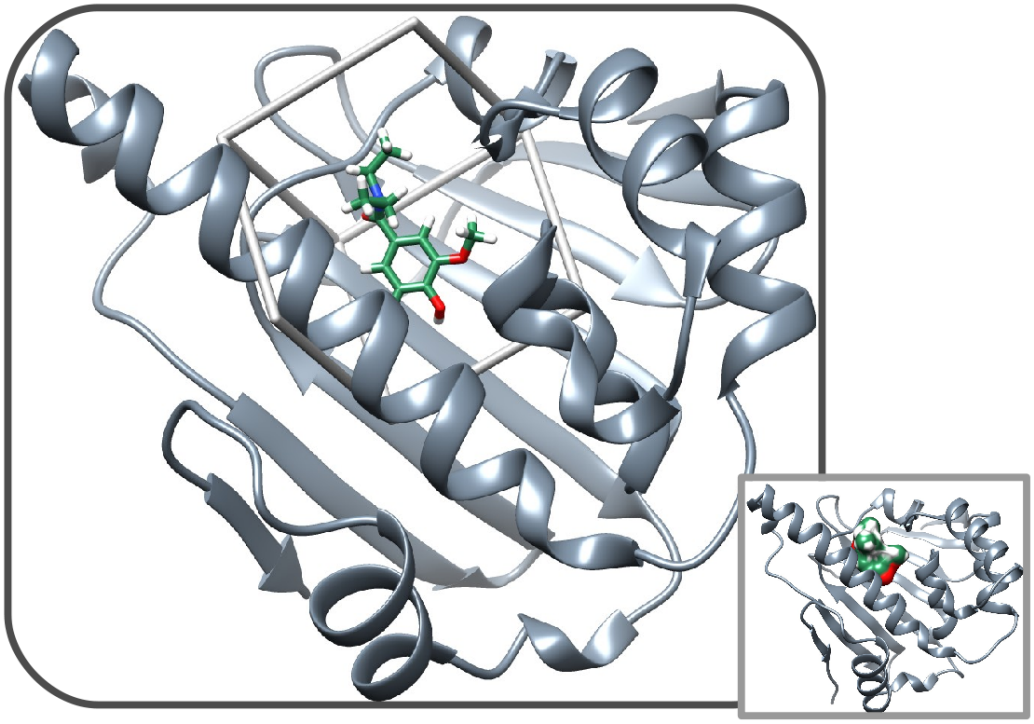


Figure 5.7: HSP90 scoop (left, bigger) and full protein structure (right, smaller). Note that the scoop includes most of the structure of the full protein. In both cases, the protein is displayed as ribbons in grey and the ligand (2dl) in green. On the left, the JAFS box is shown in white. PDB code 2XDL.

The Protein Structure

The section of the protein included in our system is the N-terminal domain — which is also the ATP binding section. It is the ATP binding site where the studied fragments bind (as can be seen in figures 5.4 and 5.7).

The N-terminal domain of HSP90 (simply HSP90 from now on) is formed by a $\alpha + \beta$ sandwich defined by nine helices and an anti-parallel β sheet (see figure 5.4).[225]

A cavity is formed in the centre of the $\alpha + \beta$ sandwich, accessible through the helical side of the sandwich. Hence, the pocket is limited by the β sheet at the front and a helical structure to the sides.

The Protein Function

HSP90 is part of the family of proteins known as molecular chaperones (or simply chaperones). The proteins in this family share common roles in the cell, related to protein folding and cell homoeostasis.[75,226]

Several subgroups form the protein family of chaperones, among them, HSPs are proteins of particular interest. HSP stands for Heat-Shock Protein. This name is due to the increase in expression found in these proteins when the cell suffers different forms of stress (heat among them).

Given their proteostatic role, chaperones are highly influential for the survival of cancer cells, which, due to their mutations and exacerbated growth, live in constant stress conditions. Moreover, inhibition of certain chaperones has been proven to specifically inhibit growth of some cancer cell lines.[75]

5.2.2 Cyclin Dependent Kinase 2

Cyclin Dependent Kinase 2 (CDK2) is one of the targets provided by our pharmaceutical collaborators at Astex Pharmaceuticals. It is, like HSP90, a target for cancer therapy.

The correct functioning of CDK2 in the cell is essential to the correct progression of the cell cycle, controlling the evolution of the cell between cycle states.[227] This cycle determines when (and if) cells should divide and proliferate, as well as leading the cell towards apoptosis if required. Consequently, the deregulation of this cycle can generate uncontrolled division and cellular growth, as well as inhibiting apoptosis, typical behaviour of cancerous cells. In cancer cells, the expression of the proteins of the CDK family, as well as the modulators of their activity, is often found to be altered.[227]

To fulfil their function as regulators of the cell cycle, Cyclin Dependent Kinases (CDKs) work in tandem with Cyclins, forming a heterodimer. In this dimer, CDKs perform kinase activity, while cyclins form the regulatory subunit.[227]

In the case of CDK2 binding to one of its possible partner Cyclins (Cyclin A), the binding of the cyclin activates the Kinase by releasing steric clashes which otherwise block the entrance to the active site. CDKs hence remain inactive when not attached to Cyclins, and it is the formation of these transitory complexes, and the associated conformational change, that initiates the kinase functionality and

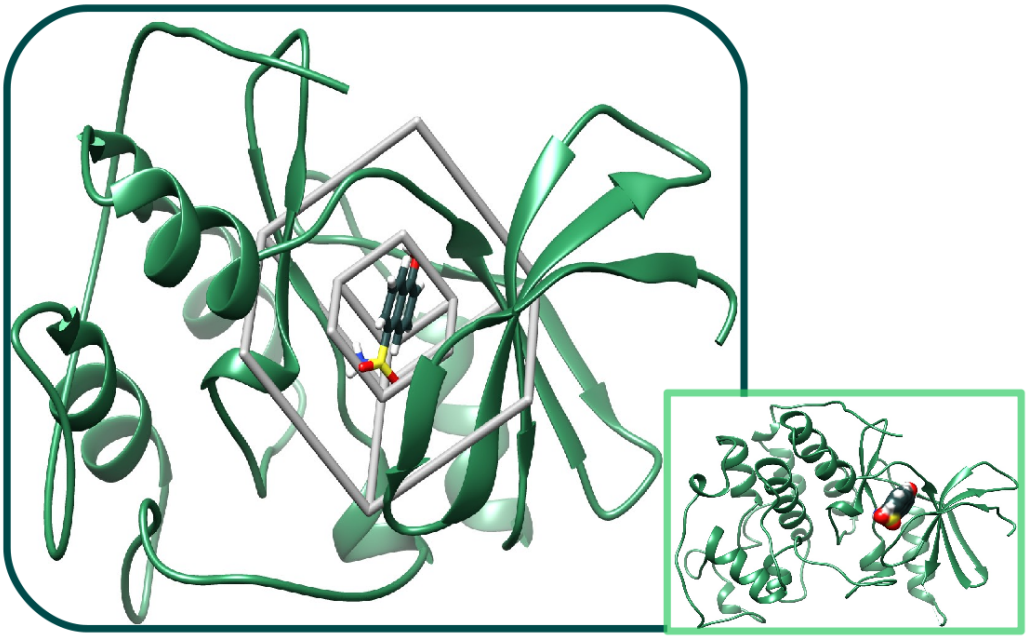


Figure 5.8: CDK2 scoop (left, bigger) and full protein structure (right, smaller). Note that the scoop includes mostly one of the subdomains of the full protein. In both cases, the protein is displayed as ribbons in green and the ligand (vth) in dark grey. On the left, both JAFS boxes (score, smaller and pose, bigger) are shown in light grey. PDB code 2VTH.

drives the progression of certain cell cycle events, such as the replication of DNA in the case of CDK2.[228]

CDKs share the same fold and tertiary structure with the rest of the eucaryotic protein kinases, while differences in their active site distinguish members of the protein kinase family. The structure of CDK2 is composed of two different lobes, a smaller one (N terminal) where β-sheets are dominant, and a larger one, mostly α-helical. As can be seen in figure 5.8, the region studied for the binding of our fragments is formed by a cleft between both lobes, with the β-sheet lobe to the right and down of the ligand and the α-helical lobe to its left in the image. This is the native binding site for ATP.[228]

In the dataset of ligands used with this system we find some which binding affinity has been determined experimentally (binders, actives) and those which experimental studies have found not to bind with detectable affinity (non-binders, decoys). Distinction between these two ligand classes is one of the main points

of interest in the initial stages of drug development processes. Optimization efforts are invested in initial active hits, since non-binders are not expected to evolve into effective drugs by optimization. The measure of enrichment factors (increased proportion of actives with respect to the total number of ligands) as measures of success of docking and scoring protocols are an example of the relevance associated to the distinction between binders and decoys.[229–231] Accordingly, it is highly relevant to assess the performance of JAFS within this context. The specific fragments chosen to study performance on this system can be found in figure 5.9 and in the Appendix, table 11.1, under the names: vta, vth, vtm and wcc (binders) and cd1, cd5, cd6, cd8, and cd9 (decoys). The PDB codes of the protein complexes with the binders are 2VTA, 2VTH, 3VTM and 1WCC. The protein structure used in all simulations unless otherwise specified was taken from the 2VTH complex.

5.3 Why These Systems

In the two previous sections we have briefly described the systems that have been chosen as our test cases for JAFS. However, besides their classification as model or pharmaceutical targets, little has been said about why these particular protein complexes have been selected. In this section the main reasons will be summarized.

5.3.1 Interesting Range of Challenges

Possibly the reason which has been mentioned the most in the previous sections is related to the range of difficulties they offer to test new JAFS method. From the different levels of solvation and cavity size, from the small and apolar T4Lys, to the big and dry MUP to the big and well hydrated HPS90, which also presents the challenge of conserved bridging waters. Equally, in the complexity associated with the number of fragments, the systems range from only two binders in MUP to 9 different ligands in CDK2, which is interesting as it includes both binders and decoys.

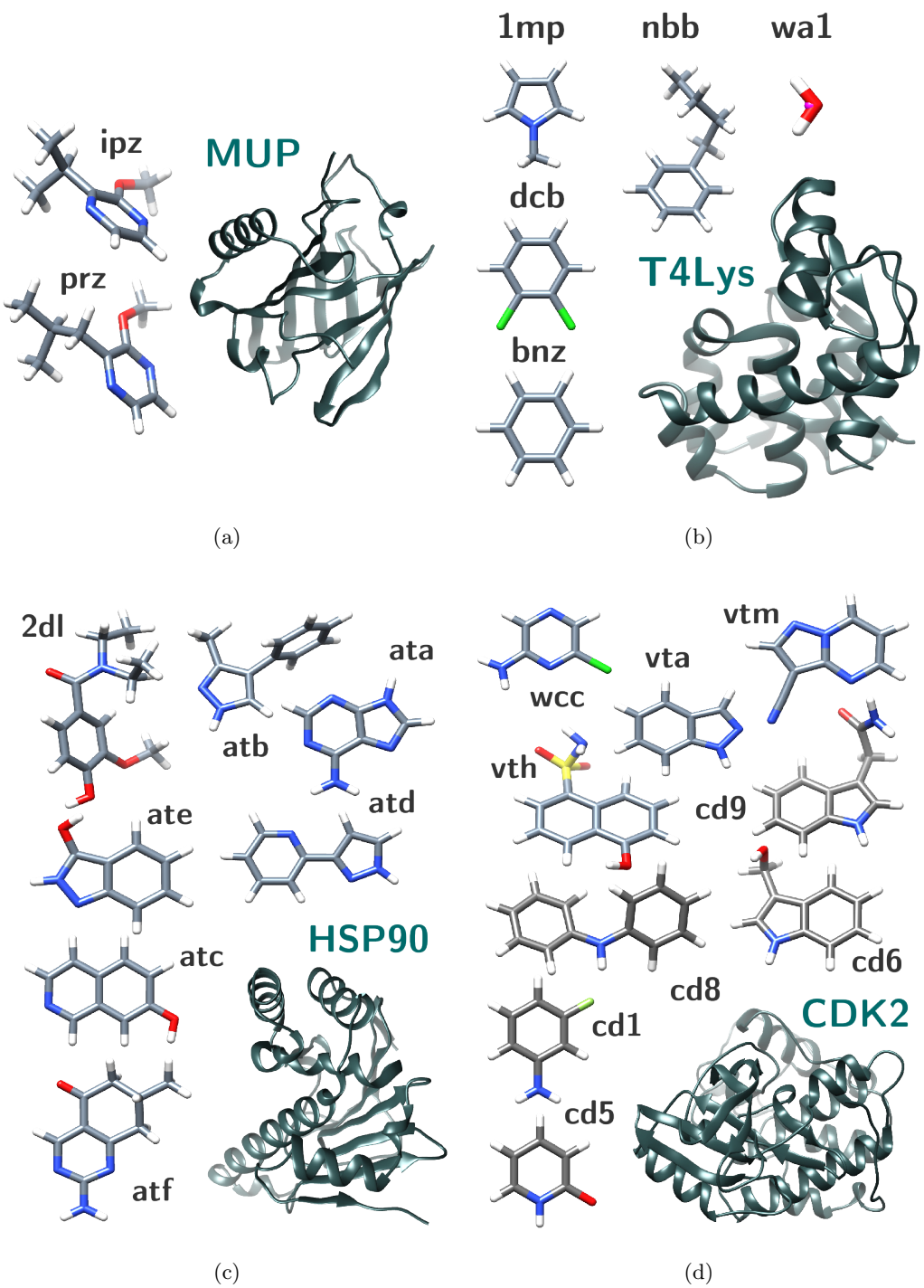


Figure 5.9: All systems used throughout the JAFS production stage. Fragments shown associated with their target. 2D representations of fragments can be found in the Appendix, table 11.1.

5.3.2 Availability of Structural Data

In all cases, a determining reason for the choice of these particular test cases was the availability of crystal structures of the protein-ligand complexes. These may be publicly available (in most cases) but in some cases (i.e. HSP90) part of the information was made available to us for this study through personal communications. The availability of good quality structural data is essential for the comparison of out calculated binding conformations with experiment.

5.3.3 Availability of Experimental Binding Affinities

The availability of binding affinity was of great importance, in particular for the systems where the JAFS score protocol (see section 4.2.1) was to be performed. This data allowed for a comparison of the ranking obtained by affinity of different ligands to a binding cavity as observed experimentally to that obtained computationally with JAFS.

Chapter 6

Development

6.1 Previous work: JAWS Applied to Fragment Molecules

The JAFS method is based on the JAWS methodology. For a description of the theory behind JAFS, sharing many principles with JAWS, see chapter 4.

6.1.1 JAWS

Just Add Water moleculeS (JAWS)[200] is a method developed to tackle the always-relevant aspect of water molecules binding to protein cavities. Water molecules may mediate interactions between bound ligands and their binding cavity and the solvation state of the cavity has an effect on binding affinity even when no direct mediation of binding interactions is present. Moreover, the displacement of bound water molecules is a method commonly used to increase affinity of ligands through the increase in the entropy of the system as the water is released to the solvent (bulk water). Despite all its importance and having been studied often, predicting the correct binding affinity and geometry of water molecules to protein cavities is still challenging.[202]

With the aim of sampling occluded cavities and increasing the sampling of crowded ones, a similar concept to that of GCMC (see section 3.5.4) is applied. However, rather than offering water molecules only two possible states (present in the simulation vs. the external bath), the JAWS waters are offered an extra degree of freedom to be sampled as a continuum, from absence ($\theta = 0$) to presence ($\theta = 1$) in the simulation. Both extreme states, as well as all intermediates, are

defined by the scaling factor (θ) applied to the interaction potential of the water molecules. The possibility of sampling this variable as a continuum allows for smaller Monte Carlo steps (attempted simulation moves) to be taken when sampling the presence or absence of JAWS waters. Consequently the energy difference arising from the moves can be small, and hence, the acceptance ratio (and sampling) can be improved (see section 3.5.2). JAWS has been proven to provide results on water binding affinity and location in agreement with experiment.[200]

For more information on the original JAWS method, see section 3.8.

6.1.2 Issues on ligand binding

While the study of water binding to protein cavities is still very much a *hot* topic,[93,202,232–234] similar issues can be found when looking to find the correct binding geometry of small organic molecules (ligands) binding to protein cavities.

Just like water molecules, small ligands can bind in occluded protein cavities,[223,235] where capturing the process of ligand binding to target would require protein conformational changes not expected in typically simulated time scales. Different methods to increase sampling within a given simulation time are described in section 6.3.2. Even when binding in fully solvent exposed cavities, sampling a range of different configurations in atomistic, molecular mechanics, explicit solvent simulations may require the concerted move of a set of water molecules, the ligand and potential re-arrangement of protein side-chains, all of which, again, cannot be guaranteed in a typical simulation time. Given these sampling difficulties, computational estimates of optimal binding geometries often rely on simplified system representations and energy calculations such as docking[17,83,124] (see sections 2.4.1 and 3.2). Other available methods with their contributions and drawbacks are summarized in sections 2.4.2 and 3.6.4.

6.1.3 Applying JAWS to Fragment Molecules

However, given ligands and waters face similar challenges in the study of their binding to protein cavities, why not try to apply the same solution to their problems?

When thinking of transferring water binding techniques to other protein ligands, fragment molecules can be seen as the most feasible step. Besides, fragment

molecules, due to their small size and related lower binding affinity are considered more difficult cases in binding prediction methodologies (see section 2.4.1). As detailed in chapter 2, fragments are small, comparatively rigid molecules, in the context of organic molecules binding to proteins. Both the size of the molecule and lack of degrees of freedom prove relevant measures of the transferability of water methodologies.

Water molecules are smaller in size compared to any other molecule present within a typical biomolecular system (excluding some inorganic ions). This small size is key to the efficiency of techniques such as GCMC, based on the sudden "appearance" of a molecule in the simulation. As the acceptance of a Monte Carlo move depends on the generated change in the energy of the system (see section 3.5.2), a move which, introducing a molecule in the system, generates Lennard-Jones clashes with those previously present, is very unlikely to be accepted. Consequently, the smaller the molecule to which GCMC is applied, the more efficient the technique is expected to be. The same reasoning can be applied to JAWS, while JAWS is expected to achieve better sampling than GCMC for any given molecule. As is generally the case, this increase in sampling does not come without associated drawbacks. To study the benefits and drawbacks of each method, the reader is advised to read sections 3.5.4 and 3.6.4 for information on GCMC, section 3.8 for further information on JAWS and sections 9 and 10 for a summary of the drawbacks and benefits found when applying JAWS to fragment molecules.

The number of degrees of freedom, on the other hand, simply increases the dimensionality of the problem. Given a particle where internal degrees of freedom are sampled, its insertion move in a particular point of simulation space may generate a different energy change (and hence be accepted or rejected) depending on the internal conformation of the inserted particle.

It is in this context that techniques applied to water molecules are more likely to be successful if extended to fragments, than to other bigger, more complex species.

It is important to note, however, that GCMC has been applied to small molecules, as described in section 3.6.4. The advantages expected from the application of JAWS to this same problem are noted in section 3.8.3.

The aim was then to expand JAWS — the technique developed with a similar approach to GCMC, but with increased sampling efficiency — to be used in the

fragment context. Some general ideas of this initial stage of the application of JAWS to fragment molecules are listed below:

- Definition of the binding cavity space within a cubic region surrounding the whole of the desired cavity

- Several fragment species and water molecules included within the binding cavity

- Fragment species and water molecules within the cavity are only allowed to sample the cubic region in terms of Cartesian space

- Fragment species and water molecules within the cavity allowed to sample the scaling of their interaction energy (θ)

- Fragments and waters within the cavity are expected to compete both in Cartesian and θ space

- As a result of this competition, each particle in the cavity must find its optimal binding pose, or remain at low interaction energy (low θ) if binding is disfavoured

The outlined protocol for the application of JAWS to fragments would estimate binding geometries of several fragments at the same time, with intrinsic competition in terms of affinity for the different binding areas within the cavity. It would provide useful information for the intuitive development of drugs from combinations of fragments binding to different areas of the cavity and / or potential modifications of the initial fragment in the process from hit to lead and drug (see chapter 2), given that potential binding geometries of fragments are partially overlapping.

While the initial ideas for the protocols would provide a number of advantages highlighted above, attempts to develop this complex protocol proved unsuccessful (Dr. S. Genheden, personal communication). The objectives to be fulfilled by a single simulation were deemed too ambitious and the dimensions of the free energy landscape to be sampled excessive, since convergence could not be achieved. The objectives where then simplified, by reducing the application of JAWS to one unique fragment species per simulation (plus water molecules). The new protocol would

allow for the binding geometry of this one fragment species to be studied, at the same time as the solvation of the cavity around it.

This protocol was used to study CDK2 with ligands *wcc* and *vta* (see chapter 5 and the Appendix, table 11.1). Finding the correct binding geometry for these fragments proved challenging, with solvent exposed configurations frequently found and the crystal geometry rarely appearing (Dr. S. Genheden and Dr. G. Ross personal communication). Key conditions of the simulations performed for these studies are shown below:

- 10 x 10 x 10 $Å^3$ box limiting the binding cavity
- 15 copies of fragment species (one species in each simulation) and 15 copies of water
- 5 million equilibration and 20 million production Monte Carlo moves

6.2 Initial Work: Sampling vs Mean Field Effect

6.2.1 Protocol

Key aspect of the protocol followed during section 6.2 will be highlighted here. Note the differences with the JAFS protocol as presented in chapter 4. This is to be expected as a reflection of the different stages of method development. A form of "naive" implementation is shown here, which was then further refined into the JAFS methodology as shown in chapter 4. Unless otherwise stated:

- four repeats (different random seed but common starting conditions) have been run for each simulation with a particular set of conditions

- the initial configuration of the JAWS fragments emulates that of the X-ray crystal structure. The centres of geometry of the JAWS waters are located on the same coordinates as that of the fragments. Note all JAWS particles overlap.

- the initial value of θ for all JAWS particles is zero

- before starting the main body of the simulation, ten thousand randomizing steps (Monte Carlo moves in Cartesian space) are applied to those initially centred JAWS particles, so that they are distributed within the simulation box. Note that, at this stage, the JAWS particles move freely within the JAWS box, given that for them all $\theta = 0$ (non-interacting with the environment)

- once the simulation is finished, only "on" fragments are clustered, where an "on" fragment is that for which $\theta > 0.95$ (threshold consistent with published work on JAWS[200])

Simulation Setup

JAFS box. The dimensions on the JAFS box were decided to make sure the whole binding site was covered. Looking at the reference where fragment f01 was studied and developed into more complex compounds,[25] three other fragments are also discussed (co-crystallized with HSP90 in entries 2XDL, 2XDU and 2XDS of

the pdb). The JAFS box was designed from the overlayed structures of these co-crystals to be at least 2 Å away from all atoms in the protonated structures of these fragments. The approximate size of the box resulted 16 Å x 14 Å x 13 Å(see figure 6.1).

JAFS particles. The number of JAFS particles in these simulations was chosen to be in excess of the number of molecules expected in the cavity (results of GCMC calculation on the cavity with and without the fragment and personal communication with Dr. G. Ross):

- Copies of f01: 1 or 5
- Copies of water molecule: 25 or 30

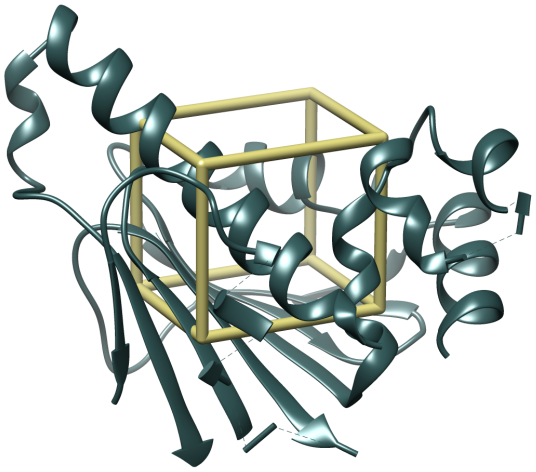


Figure 6.1: HSP90 scoop (from PDB code 2XDK) in dark grey, with the JAFS box used in section 6.2 in green

resulting in three different simulation conditions: 1 copy of f01 and 25 copies of water (1F25W), 5 copies of f01 and 25 copies of water (5F25W), and 1 copy of f01 and 30 copies of water (1F30W). Of these, in depth analysis will be presented for 1F25W and 5F25W simulations. Their comparison was found to provide the most relevant information to reach conclusions on the methodology.

This number of particles was taken as approximated by a GCMC (see section 3.5.4) simulation on the cavity, with the limit on the GCMC region of the same dimensions as the JAFS box. 25 water molecules were estimated to be the optimal number of waters in the cavity when fragment f01 was present, as 30 was the optimal number of waters generated for an empty cavity. While these calculations were performed before the GCMC methodology reached its most recent development,[93] our previous application of GCMC is likely to over estimate the number of water molecules (Dr. G. Ross personal communication). An over estimation of the number of waters is not expected to have any major effects in the development

of the JAFS methodology, since JAFS should take care of keeping any excess of JAFS particles at low θ values.

Simulation lengths and proportion of moves. Simulations of different lengths (number of moves) were run. Standard (20 million moves) length simulations were followed by approximately 70 million move simulations. Eight repeats of "extra-long" simulations (250 million moves) were run with 1 copy of f01 and 25 copies of waters. The idea behind these "extra-long" simulations was to check whether better convergence could be achieved for very long simulations. The proportion of moves attempted on each type of particle involved in the simulation can be seen in table 6.1. These were estimated to provide good sampling of the system based on experience of simulations with the same software in the group. A special focus is applied on the sampling of solutes, both in θ (*theta*) and the Cartesian (*solute*) space (note there is a high number of molecules of solvent, hence the sampling per molecule will be low for solvents). While the move size limit (flexibility) of Cartesian sampling is divided in different terms related to each of the internal degrees of freedom, as well as translation and rotation, the flexibility of the theta moves is determined by a single value, which is set to 0.15 for all work presented in section 6.2.

move type	moves
theta	50%
protein	4%
solvent	23%
solute	23%

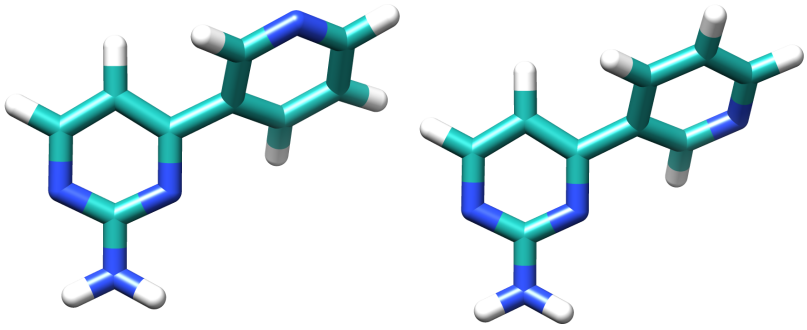
Table 6.1: Proportion of moves for each move type in the JAFS calculations throughout section 6.2. Solute refers to the move within the Cartesian coordinates of the fragments and waters treated as JAFS particles (internal degrees of freedom in the case of fragments, as well as translation and rotation for both fragment and waters). Theta refers to the moves in the θ space.

Clustering Analysis

Two different programs were used for clustering fragment and water structures from the snapshots produced through simulations: cpptraj, included in AmberTools, as part of Amber12,[150] and an in-house script (generated by Dr. Gregory

Ross, based on scipy hierarchical clustering, see docs.scipy.org/doc/scipy/reference/cluster.hierarchy.html). Two associated clustering methods were used: average linkage[213] with cpptraj and single linkage[213] with the in-house script (however, average linkage was used with the in-house script in specific examples specified later). Different thresholds were also used for each approach, explaining differences between figures which might be noted through the text (4 Å as maximum distance between clusters and 5 as the minimum number of clusters for cpptraj clustering, and 2 Å as maximum distance between clusters for the in-house script). Cases where clustering methods were problematic will be specified in the text. The different clustering approaches were used depending on visualization preference for different simulations and simulation conditions. Results regarding the success in locating the crystal binding mode were compared for some of the simulations, observing comparable results for the different clustering protocols. This diversity was considered acceptable at these early stages of method development where visualization of the results was key.

6.2.2 Results



(a) (b)

Figure 6.2: The f01 fragment in its crystal conformation (a) and the "flipped" conformation (b) where the dark blue atoms are nitrogens, light blue carbons and white hydrogens.

In this section the results of the simulations within the initial development of the JAFS methodology will be presented. Most importantly, the RMSDs to the crystal binding mode of the different resulting clusters will be shown. In some

cases, the RMSD to the "flipped" pose will be mentioned. Here, the "flipped" pose refers to an equivalent to the binding mode with the unique sampled dihedral angles in f01 rotated by 180°(see figure 6.2). This "flipped" pose was considered valid as its distinction from the binding mode displayed in the crystal structure would be hard from electron density maps (as a carbon with a hydrogen atom present the same number of electrons as an atom of nitrogen). The assignment of the ligand conformation in the crystal structure was done based on a water molecule found within hydrogen bonding distance in the crystal structure (Marcel Verdonk, personal communication). The non-flipped pose (image (a) in figure 6.2) may hence be considered the "correct" binding geometry. However, the ring might rotate in non-crystallographic conditions. When the (all-atom) RMSD of a given cluster to either of these poses was less than 2 Å, the cluster was considered to have successfully located the crystal binding mode.

The distribution of θ values throughout the simulations will also be shown for some particular cases, as a route to reaching conclusions on the reasons behind the differences in results for different simulation setups.

The cluster population was explored during this stage of the project as potential information which may be used to rank clusters by binding affinity and identify the crystal binding pose among those generated. Here cluster population refers to the number of simulation snapshots whose configurations compose one particular cluster.

Please note that the water model used in the simulations presented in this section (section 6.2.2) was later found to be a faulty version of the TIP4P model. The small differences between the parameters of the water model used in this section and that of TIP4P are not expected to produce significant differences in the simulations. Furthermore, the results of the simulations run at this initial phase were simply taken as experience and a guide to develop further the application of JAWS to fragment molecules. Nevertheless, this fact must be noted.

Finding the Crystal Pose

Tables 6.2 and 6.3 show the RMSD between the clusters (obtained by clustering 200 snapshots per simulation, with 100 000 configurations separating each snapshot) and the crystal binding mode. Clusters shown are the result of clustering all

Protocol: 1F25W

20M moves	population	RMSD (Å)
cluster 1	0.47	2.32
cluster 2	0.21	3.73
cluster 3	0.20	5.77
cluster 4	0.12	6.07

Protocol: 1F25W

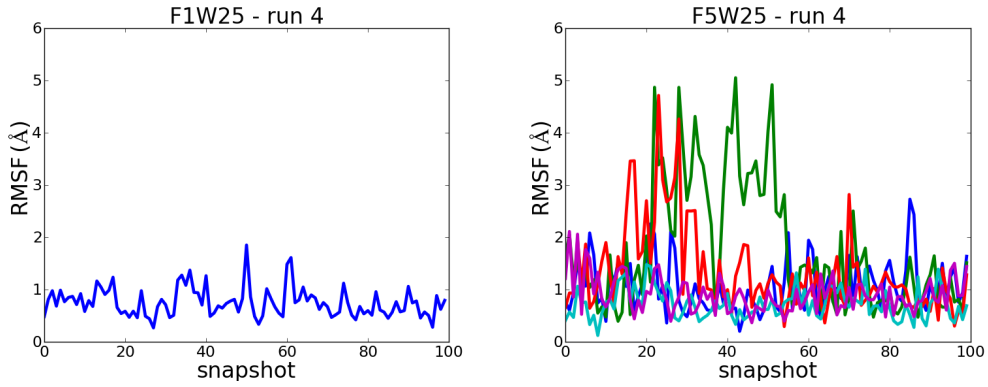
70M moves	population	RMSD (Å)
cluster 1	0.58	1.32
cluster 2	0.42	4.98
cluster 3	0.00	4.51

Table 6.2: Snapshot population of clusters in 20 M and 70M move runs of 1F25W. RMSD (in Å) from each of the clusters to the crystal pose.

snapshots for all four repeats run under those conditions, during that simulation length. Note how, for both simulation lengths (20M moves and 70M moves, table 6.3), none of the clusters found in the 5F25W simulations is within 3 Å RMSD of the crystal binding geometry. In the case of the 1F25W (table 6.2) for the 70M move simulations, the most populated cluster already presents an RMSD to crystal pose of less than 2 Å. The top populated cluster for the 20M 1F25M simulations presents an RMSD to crystal pose slightly higher (2.32 Å). However, when the RMSD to the "flipped" pose is calculated, this falls below the 2 Å limit (RMSD = 1.16 Å).

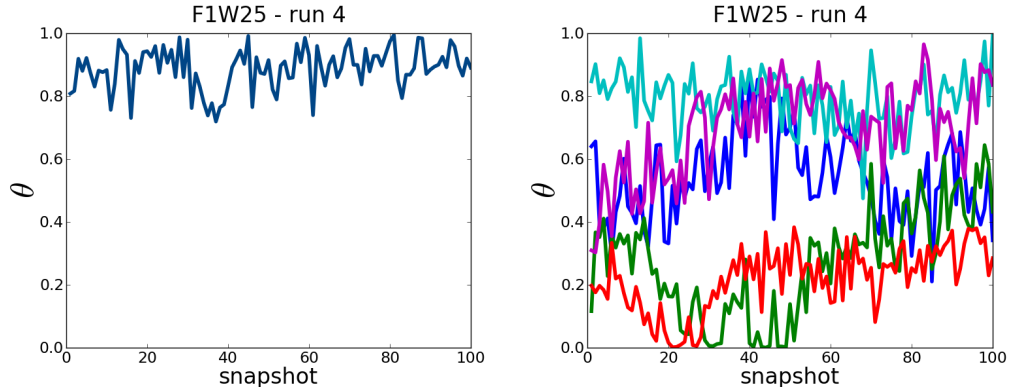
As previously indicated, these clusters were obtained from all snapshots in all different repeats within one set of conditions. The presence of snapshots from the different repeats in each of the clusters was studied.

For the 1F25W simulations, it was then found that, within each repeat, all snapshots (with fragment $\theta > 0.95$) were part of only one cluster (different clusters for different repeats). In all cases, four different simulation repeats were run, and all were clustered together. For the 20M runs, where four clusters were found, each cluster is populated by snapshots from only one repeat. For the 70M moves runs three clusters are found, with one of them being populated by snapshots from two different repeats and the remaining two clusters, each being populated by snapshots



(a) (b)

Figure 6.3: RMSF (RMSD with respect to previous snapshot) for f01 during the last 100 snapshots (10M moves) of the 20M move simulations for representative examples of 1F25W (a) and 5F25W (b) runs. Each different line colour corresponds to a different copy of the fragment.



(a) (b)

Figure 6.4: Sampling of θ for f01 during the last 100 snapshots (10M moves) of the 20M move simulations for representative examples of 1F25W (a) and 5F25W (b) runs. Each different line colour corresponds to a different copy of the fragment.

Protocol: 5F25W

20M moves	population	RMSD (Å)
cluster 1	0.38	5.66
cluster 2	0.31	3.70
cluster 3	0.13	7.94
cluster 4	0.12	4.68
cluster 5	0.04	7.65
cluster 6	0.01	3.38
cluster 7	0.01	3.41

Protocol: 5F25W

70M moves	population	RMSD (Å)
cluster 1	0.37	3.25
cluster 2	0.20	5.02
cluster 3	0.17	6.41
cluster 4	0.13	8.88
cluster 5	0.07	4.28
cluster 6	0.04	3.77
cluster 7	0.02	4.98
cluster 8	0.00	10.31

Table 6.3: Snapshot population of clusters in 20 M and 70 M move runs of 5F25W. RMSD (in Å) from each of the clusters to the crystal pose.

from only one repeat. In either case, each of the repeats does not sample more than one unique binding configuration (cluster) while the fragment presents $\theta > 0.95$.

In the case of the 5F25W simulations, some more variability in the distribution of snapshots was found, where one cluster was typically populated by snapshots from two to three different repeats, with only one cluster (the most populated cluster for the 70M moves runs) found to be populated from a single repeat.

The points below can serve as a summary of the results presented above:

- The 1F25W simulations seem to be more successful than 5F25W in finding the crystallographic binding mode

- Based on the snapshots clustered together from different simulation repeats, the sampling of the 5F25W simulations is understood to be better

To confirm the idea that the sampling of Cartesian space in the 5F25W is bet-

Protocol: 1F25W

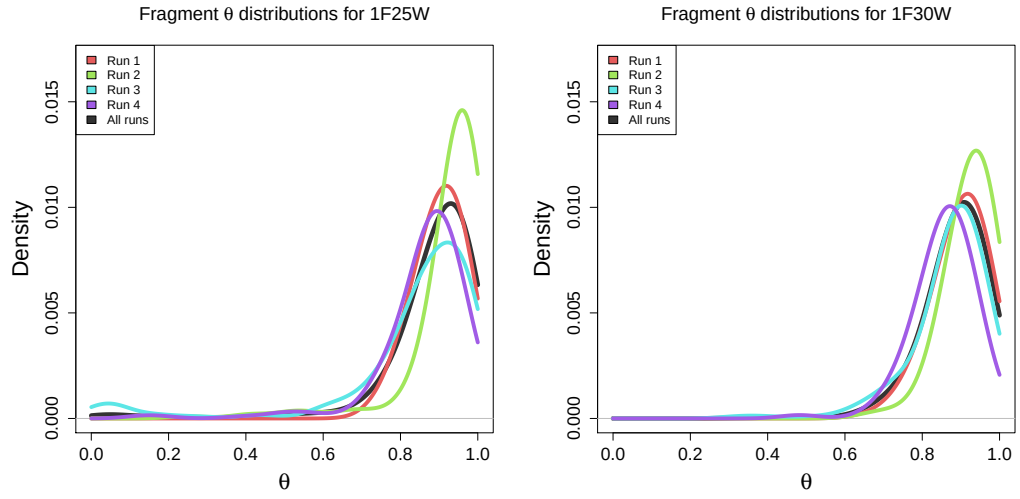
250M moves	population	RMSD (Å)
cluster 1	0.69	3.69
cluster 2	0.21	4.37
cluster 3	0.10	4.51
cluster 4	0.01	5.34

Table 6.4: Snapshot population of clusters in 250 M move runs of 1F25W. RMSD (in Å) from each of the clusters to the crystal pose.

ter than that of 1F25W, a study of the RMSF (RMSD with respect to previous snapshot) per simulation snapshot was performed. Plots of the last 100 snapshots (one snapshots per 100000 MC moves) of representative repeats of 20M move simulations for 1F25W and 5F25W can be found in figure 6.3. These plots support our previous understanding that increasing the number of copies of the fragments increases the Cartesian space sampled.

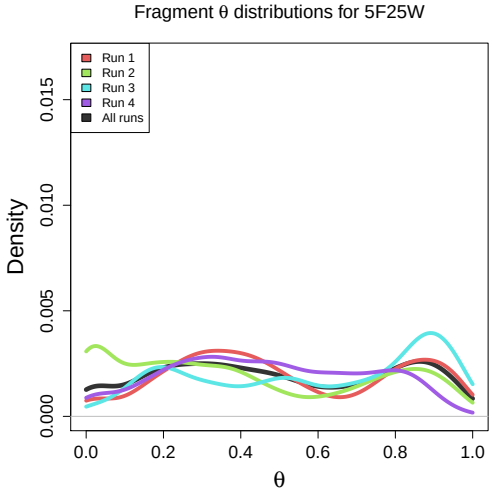
By comparing the plots in figure 6.3 with those in figure 6.4 the coupling of the θ sampling with sampling of Cartesian space can be observed. See how, for the case of 1F25W, the only fragment present remains at high values of θ (figure 6.4 plot (a)), correlating with a low sampling of Cartesian space (figure 6.3 plot (a)). As the number of copies of the fragment is increased, some of those copies sample low values of θ (see figure 6.4 plot (b), the green and red lines). When the lowest values of θ are sampled, an increase in the RMSD with respect to previous snapshot can be observed (see 6.3 plot (b), the red and green lines).

Given that the problem associated with the 1F25W simulations seemed one related to sampling, we attempted to run very long simulations, with the idea that these should then reach convergence and provide the crystal binding mode in all repeats. Results for the clusters obtained from all repeats of the extra long simulations (250M moves) are shown in table 6.4. As can be seen, no crystal pose is obtained for this simulation length. It was then concluded that increasing the simulation length was unlikely to provide the solution to the sampling problem observed, within any feasible simulation time.



(a)

(b)



(c)

Figure 6.5: Smooth histogram of the θ values of f01 in every snapshot in the 4 repeats of 20 M moves with simulation set-up: 1F25W (a), 1F30W (b) and 5F25W (c). A Gaussian kernel is used for the smoothing of the histograms (the density is displayed as a normalized sum of Gaussians). Note that, while this representation simplifies visualization, it necessarily associates with a non-accurate representation of the density at the extreme values of θ (0 and 1).

The Influence of θ Sampling

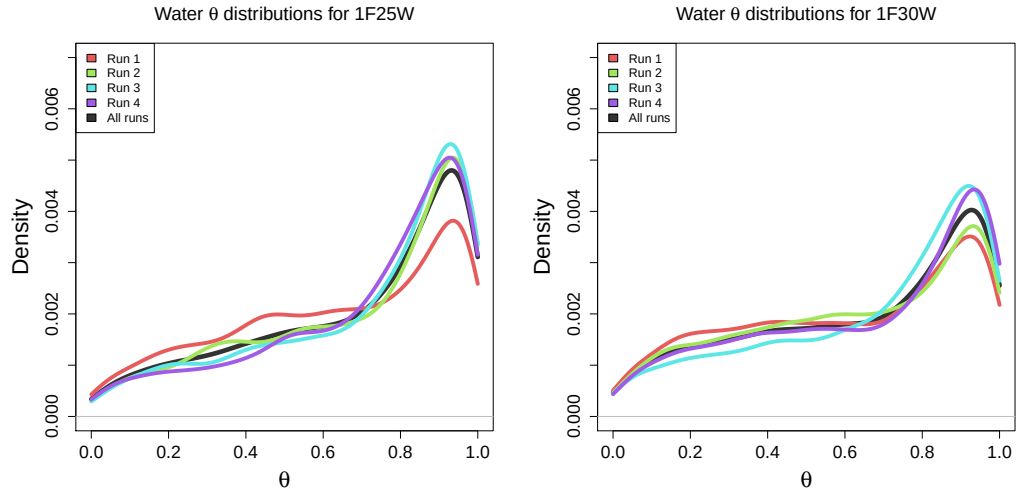
The distribution of θ throughout the simulations for a set of different simulation conditions is shown in figure 6.5. An obvious difference between the simulations where one copy of the fragment is present (1FW25 and 1F30W) and those with five copies of the fragment (5F25W) is presented. It can be seen how, for those with one copy of the fragment, f01 remains at high values of θ most of the simulation time, with the possibility of appearing at low value of θ at some stage in the simulation. However, for simulation where five copies of the fragment are present, there is a close to uniform distribution of θ values throughout the simulation for the f01 particles. It is important to understand here that particles at intermediate values of θ are not a true representation of any state in the real system. That is, in reality, a particle can be in the cavity (represented by high values of θ) or not be there (low values of θ), but no particle with half its interaction energy will exist in the real protein-ligand system. It can be hence concluded that simulations with less crowded cavities (those with one copy of the fragment) seem to represent a more realistic system that those with over-crowded cavities (five copies of the fragment).

While differences appear less prominent, a similar behaviour to that of the fragment molecule can be observed for the waters included as JAWS particles. The θ distributions for JAWS waters can be seen in figure 6.6, where the simulations with five fragment copies (5F25W) present a flatter distributions of θ than those seen for the simulations with only one copy of the fragment (1F25W and 1F30W).

6.2.3 Conclusions

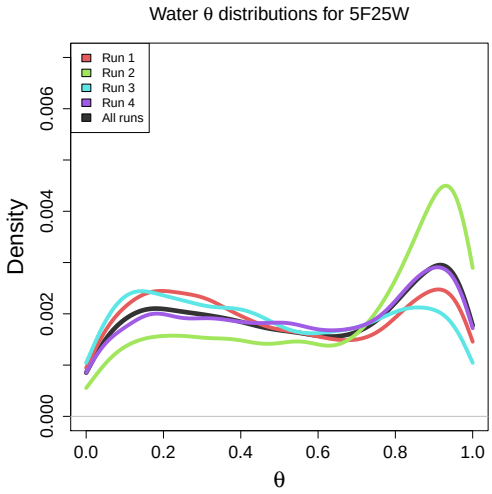
Based on the results presented above, it was concluded that:

- Over-crowding the JAFS box by including more particles of bigger volume (volume of f01 > volume of water molecule) seems to increase the sampling of the fragments, at the same time generating a less realistic environment (with a number of fragments with intermediate values of θ). Both facts are understood to be correlated, with the presence of fragments at intermediate and lower values of θ increasing the Cartesian sampling, due to their interaction energies being scaled down. Fragments can then increase their θ where favourable interactions are found.



(a)

(b)



(c)

Figure 6.6: Smooth histogram of the θ values of the waters in every snapshot in the 4 repeats of 20 M moves with simulation set-up: 1F25W (a), 1F30W (b) and 5F25W (c). A Gaussian kernel is used for the smoothing of the histograms (the density is displayed as a normalized sum of Gaussians). Note that, while this representation simplifies visualization, it necessarily associates with a non-accurate representation of the density at the extreme values of θ (0 and 1).

- A less crowded JAFS box generates a more realistic environment, where the crystal binding pose could be successfully found. However, Cartesian sampling was hindered by the single copy of f01 remaining at high values of θ during almost all of the simulation time. While the fragment copy remains at high values of θ, without the interaction energy being scaled, the increased sampling effect of JAFS simulations does not apply.

A dilemma arises, where increased sampling seemed to be associated with unrealistic simulations, while a realistic simulation can only be achieved with low levels of sampling. The rest of chapter 6 will aim to solve this dilemma, by increasing sampling in realistic simulations and / or making well sampled simulations better correlate with the real system.

6.3 Addressing the Sampling Problems

As described, the initial work on the development of JAFS seemed to end in an impasse, where the overall outcome of the simulations could not be improved. The sampling of the simulations could be improved at the cost of decreasing the realistic depiction of the system, while representing the system accurately seemed to be intrinsically linked to poor sampling. As explained in previous sections, an increased number of fragments in the simulation is correlated to both an increase in sampling and decrease in the accuracy in the representation of the physical system, through the appearance of fragments at intermediate values of θ — neither present nor absent from the simulation. Simulations with lower numbers of fragments, lacking this intermediate θ fragments, represent more accurately the physical system, since fragments can be assigned to either "present" (high θ) or absent (low θ) states. However, the low number of fragments is also associated with a lower sampling of the fragment binding configurations (with lower number of copies available to perform the search).

The logical question to answer was what was pushing the simulations towards this impasse. We would like to run simulations with conditions linked to better sampling (several copies of fragment) without getting, as a consequence, a sea of particles with intermediate values of θ (and hence a simulation which poorly reflects reality). Next, questions on how a realistic representation with improved sampling can be obtained, and why the intermediate values of θ seem to dominate the simulations, will be addressed.

6.3.1 Minimizing Intermediate θ: Changes to the Hydration Penalty

The hydration penalty has been defined before in this thesis (see sections 4.2 and 4.4, and refer to figure 4.1). However, an in-depth explanation of its application through the development of JAFS to the final state of the method has been missing.

To understand the effects and application of the hydration penalty, it is first required to understand the links between the θ variable in the JAFS simulations and the λ variable used in common simulations involving alchemical transformations for the calculation of relative free energies. (see sections 3.6.2 and 4.4).

Let us picture an absolute dual topology transformation (see section 3.6.2)

where a ligand is decoupled from its environment (transformed into "nothing" or a dummy atom, as implemented in ProtoMS). The different λ windows may be assigned so that $\lambda = 0$ corresponds to the ligand fully interacting with the environment, while $\lambda = 1$ corresponds to the ligand fully decoupled. A set of intermediate λ windows will be set in place equally. For simplicity, let us assume only two intermediate λ windows exist, at $\lambda = 0.33$ and $\lambda = 0.66$. At $\lambda = 0.33$, the ligand will be "0.33 off", that is "0.66 on", or more specifically, its interaction energy will be

$$E_{\lambda=0.33} = E_{\lambda=0} * 0.66 \tag{6.1}$$

where $E_{\lambda=0}$ is the unscaled (full) interaction energy of the ligand with its environment. Equivalently, $E_{\lambda=0.66} = E_{\lambda=0} * 0.33$. Notice that we are asuming linear scaling of the energies (and no softcore), for simplicity.

The assignment explained above is exactly that chosen for hydration free energies in ProtoMS (except for the use of softcores in ProtoMS simualtions) where, in fact, the free energy of decoupling the ligand from water, rather than the hydration free energy, is calculated (note that these two free energies are equivalent but with opposite sign, as they measure the same process in opposite directions). Simulations are then run for each of the λ windows (generally with more windows than the simplified example above), and the free energy is produced by calculating the PMF (see section 3.6.2) through the λ path.

As explained in sections 4.2 and 4.4, the free energy of hydration of each of the ligands in the JAFS simulation needs to be included so that their relative affinity of binding is correctly represented. This applies to both JAFS score and JAFS pose protocols (see section 4.2) since the binding geometry cannot be represented accurately if the relative affinity of waters and fragment is not accounted for properly. The hydration penalty is the means to include the hydration free energy into the JAFS calculations.

Let us look again at how the hydration free energy (or rather the free energy of decoupling the fragment from water) is calculated in ProtoMS. While λ is a static variable, their values correspond to particular scalings of the interaction energy of the fragment, in an equivalent fashion to the θ values. It is important to take into account that, for the decoupling free energy as defined in ProtoMS, $\lambda = 1 - \theta$, since a fragment with its interaction energy not scaled is represented by $\theta = 1$ in JAFS

and $\lambda = 0$ in the decoupling calculations.

It has been said that the hydration penalty is the way in which the hydration free energy is incorporated into the JAFS calculations, but what effect exactly does the hydration free energy have?

Within ProtoMS, the hydration penalty is implemented as an external factor, only calculated when a move in θ space is attempted. When this move is chosen randomly between all possible moves, the process explained in section 4.2 is followed, where the hydration penalty is expressed by U_{PMF} in equation 4.2. The idea is that, when a move in θ is attempted, the energy at the old and new θ are calculated both for the protein cavity system in which the particle is being simulated, and the associated change in free energy for the JAFS particle in solution. In this way, when two JAFS particles compete in θ space (when one of them may be "on" in a particular region of the simulation), that molecule with lower interaction energies with the JAFS system and higher free energy when solvated in water will be favoured. For equal affinity towards the binding region, the particle which is less favoured in water will adopt higher values of θ. This is equivalent to separately calculating the relative affinity of both particles in water and adding it to their "bound leg" when calculating relative binding affinities with typical alchemical transformation methods (see section 3.6.2 as well as section 4.2 and figure 4.1).

The variability in the application of the hydration penalty can be related directly to equations 4.2 and 4.3. During the initial work on JAFS, the hydration penalty is used directly as a polynomial (of 4th or 5th order) fit of the PMF generated from the simulation of decoupling the fragment from water. The PMF plotted with respect to λ can be seen in figure 6.7 for several of the ligands that will be seen in chapter 7.

By looking at figure 6.7, it can be seen that the PMF maximum is always between $0.6 < \lambda < 0.9$, with minima at both extreme λ values. As previously mentioned, initially the hydration penalty was expressed as a polynomial fit of these same PMFs. To understand the effects of the hydration penalty on the θ sampling in the calculation is important to remember that, for our particular scenario, $\theta = 1 - \lambda$, and also that, as seen in equation 4.1, it is actually the negative of the hydration penalty that will be applied to the energy of the system

for the calculation of the Metropolis test for θ. Taking the previous into account, and looking again at figure 6.7, we can tell that the hydration penalty will have the effect of, given a flat energy landscape for the ligand interaction energy as a function of λ (or θ), generate an energy minimum at $0.1 < \theta < 0.4$ and energy maxima at both extreme values of θ.

While these exact data may be only true for these examples, the general trend (parabolic shape of the PMF with respect to λ, with a maximum at $\lambda \neq 0$, $\lambda \neq 1$) has been observed for all fragments analysed. We can then tell that, for all examples studied, the application of hydration penalty as a high order polynomial of the PMF of decoupling the fragment from water, tends to favour intermediate values of θ over $\theta = 1$ and $\theta = 0$.

Consequently, the hydration penalty is promoting the presence of these intermediate values of θ that make our system less realistic. The next logical question is whether there is a manner of applying the hydration penalty which, still accounting for the relative affinity of each fragment for water, does not promote intermediate θ values.

Most obviously, it is the shape of the hydration penalty curve that is generating the preference towards intermediate values of θ. If we could apply a hydration penalty whose shape through θ-space diverged from that of the PMF of decoupling the particle from water, the bias towards intermediate θ could be avoided. But would a different shape of the hydration penalty accurately represent the real system? The first important fact to note, again, is that the intermediate θ states do not represent any real state of the system. Hence, how they are represented in our system should affect sampling but, conceptually, the representativity of our model should not vary. Furthermore, in practice, it is the abundance of these intermediate θ states that is pushing our model away from an accurate representation of reality, hence decreasing their presence certainly must take us closer to the real state, instead of further away.

In a more theoretically based form of reasoning, it has been explained previously (see section 3.6) that the free energy is a state function. In the calculation of the decoupling free energy of the particle from solvent, it is essential that a smooth path in terms of energy gradients between both end states is generated; the free energy landscape between neighbouring windows must overlap. However, the exact path

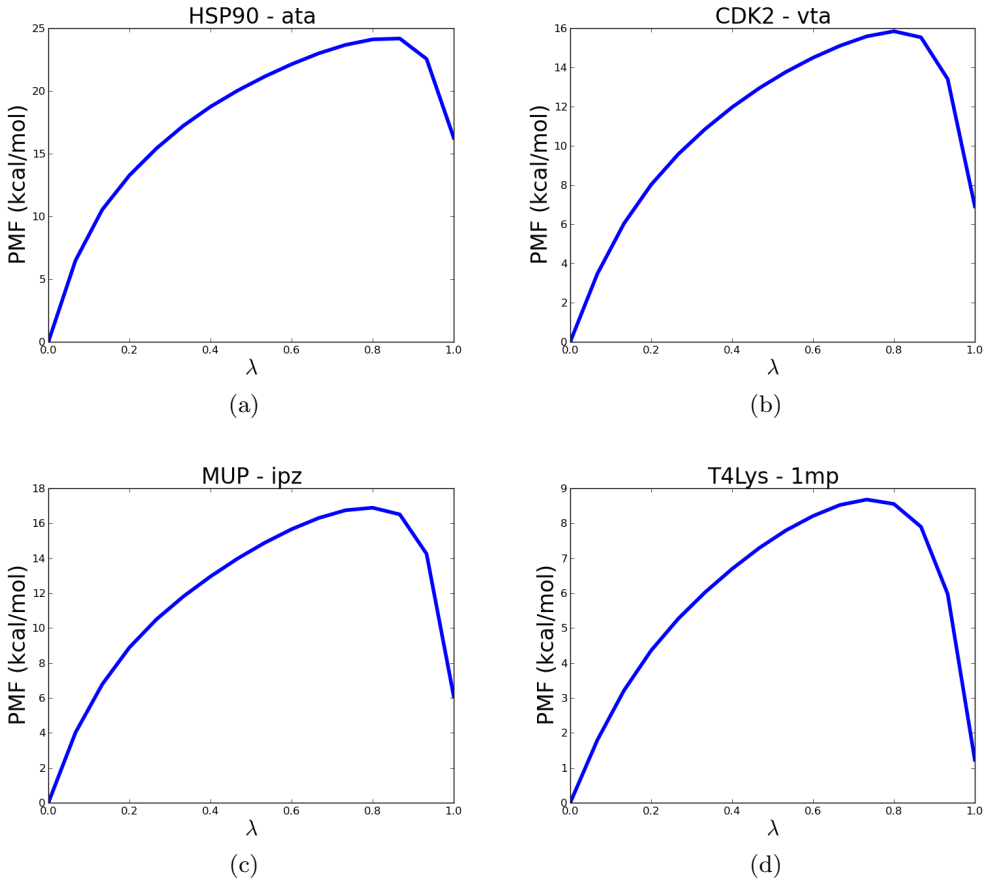
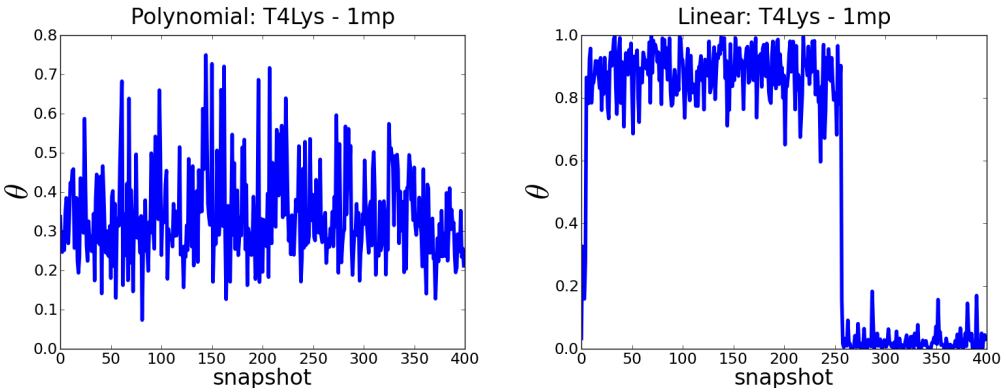


Figure 6.7: PMF (in kcal/mol) with respect to λ for the decoupling of fragment ata (a), vta (b), ipz (c) and 1mp (d) from water. See figure 5.9 and the Appendix, table 11.1, for a molecular and 2D representation of the ligands, respectively.

is simply chosen for convenience and a different implementation would be equally valid as far as it complies with the requirements stated. It was hence concluded that no theoretical or practical impediment was preventing the implementation of a hydration penalty with a different shape. The only requirement was that the PMF value for both end states ($\lambda = 0$ or $\theta = 1$ and $\lambda = 1$ or $\theta = 0$) remained equal to that of the PMF of decoupling the particle from water. To have the most neutral influence on the sampling of θ, without creating any minima or maxima, and still fulfilling the task of accounting for the hydration free energy of each particle, a linear hydration penalty was chosen, defined by a value of 0 for $\lambda = 0, \theta = 1$ and the value of the free energy of decoupling for $\lambda = 1, \theta = 0$.



(a) (b)

Figure 6.8: Sampling in θ space of ligand 1mp in a JAFS pose simulation with only two ligands (1mp and nbb) within system T4Lys (see section 5.1.1 for a system description and figure 5.9 and the Appendix, table 11.1, for a molecular and 2D representation of the ligands, respectively). Two different repeats are shown with the same simulation conditions except the shape of the hydration penalty is a polynomial in (a) and linear in (b).

The effect in θ sampling of the change in the shape of the hydration penalty for a JAFS pose simulation (see section 4.2.2) of ligands 1mp and nbb in T4Lys (see chapter 5) can be observed in figure 6.8. A very relevant change is obvious, between mostly sampling intermediate values of θ to remaining at extreme values of θ. While a clear example has been chosen to illustrate, this is a widespread effect, which has been observed for all examples where this comparative study was applied.

We can hence consider that, by changing the shape of the hydration penalty, the issue regarding the presence of abundant intermediate values of θ has been alleviated. However, by looking at plot (b) in figure 6.8, we can see that only one swap between low and high values of θ was observed through the 40 million production moves of the simulation. This repeat was chosen to display θ sampling more clearly due to most of the repeats within this particular simulation setup not presenting any swap between high and low values of θ (the fragment remains at either high or low θ during the whole length of the simulation). It hence became obvious that the other issue presented within our impasse after the initial work needed to be addressed. In the next subsection, the lack of θ sampling of the JAFS particles in our simulations will be examined.

6.3.2 Increasing Sampling: Parallel Tempering and Solute Tempering

From the two issues presented at the end of the initial work in this project, the abundance of intermediate θ states and the lack of sampling, only one has been addressed so far. The previously studied abundance of intermediate θ values is an issue intrinsically related to JAFS and JAWS, the methods within which θ is defined. The lack of sampling, on the other hand, while including an extra θ dimension, is a general problem in molecular mechanics simulations. A number of techniques have been developed to alleviate this issue.[192,236,237]

Among all possible techniques, several aspects were taken into account to decide to focus on parallel tempering and related enhanced sampling methods (based on replica-exchange):

- Simplicity. A well known and robust method was desired, which would not over-complicate either running or analysing JAFS simulations.

- Sampling the Boltzmann distribution. Since we are interested in binding affinities and optimal binding geometries of ligands, enhancing sampling must not come at the cost of the correct Boltzmann distribution of states. Besides, since no further complication is desired, this ensemble must be directly available, rather than requiring further calculations to obtain it.

- Implementation availability. While no enhanced sampling technique was

readily available within the release version of ProtoMS, the implementation of the new technique was restricted in time by the scope and demands of the project. Parallel tempering fits in well within the ProtoMS software, with available parallelization, meaning that the enhanced sampling technique will not reduce the speed of running the software.

Parallel Tempering

Parallel tempering is an enhanced sampling method based on the increase in sampling obtained at higher temperatures. A theoretical explanation of this technique can be read in section 3.7.1. It is based on running a set of replicas of the same simulation at different temperatures. It relies on the increased sampling provided by high temperatures, and on replica exchange Monte Carlo moves to swap configurations between replicas at high and low temperatures, hence providing the replicas at low temperatures with configurations that can only be reached within the simulation time-scale for high temperatures.

In the JAFS scenario, and by looking again at plot (b) in figure 6.8, we expect the simulations at high temperatures to have frequent swaps between high and low values of θ for each of our fragments, as well as directly and indirectly (through increased sampling of θ) increasing the Cartesian sampling of the JAFS particles within the cavity. Swaps between replicas should then hopefully provide both more varied configurations and θ values at the lowest temperature replica.

Implementation in ProtoMS

There were a number of reasons to choose parallel tempering as the enhanced sampling technique in this project, which have been highlighted before. However, in terms of the implementation of an enhanced sampling method in ProtoMS, parallel tempering was the obvious choice, since work had been published previously with ProtoMS and the use of solute tempering (a method related to parallel tempering both conceptually and in terms of implementation).[166] In previous versions of ProtoMS (see section 4.3), parallel tempering was available as a set of external scripts that could be applied to the simulation. With the latest optimizations of the ProtoMS source code structure, which involved simplicity of use and neatness of the source code, as well as an increase in speed, the use of external sets of tools throughout the simulation is not possible. Further development was hence

required.

The source code of ProtoMS is written in the Fortran programming language. Before the release of ProtoMS 3.0, for the performance of any other action besides out of the box Monte Carlo simulation moves, the source code had to stop the simulation, and then proceed with the next assigned task (i.e. printing energies to a file), proceeding then with the next action demanded. In other words, the execution of tasks by the source code was strictly linear, and so the instructions had to be provided by the user. To avoid the tedious task of writing a list of instructions which often would have to be repeated identically as many times as simulation snapshots where required, a *python wrapper* was available. This wrapper (written in the python programming language) would work by creating small input files which would periodically be produced and read by the source code, providing it with the information required until the next snapshot. The existence of this python wrapper would also increase the flexibility of the code. By arranging the simulation in this small sub-simulations controlled through the python script, different actions could be performed at intervals on the simulation by manipulating the simulation input fed to the source code. At this stage, the main source code of ProtoMS was also serial (non-parallel), and no way of simulating several simulation conditions (i.e. different temperatures or λ values) simultaneously had been implemented within the Fortran (source) code. Given this implementation of the (Fortran) source code, the previously mentioned python wrapper was required to perform any replica exchange related techniques (see section 3.7.1). If we take as an example λ replica exchange (see section 3.7.1), by using exclusively the (Fortran) source code, the simulation of different λ windows was possible, but not at the same time, making the exchange between different replicas (practically) impossible. With the use of the previously described python wrapper, extra python scripts had been generated and made available, allowing for the exchange of configurations between different replicas at different λ values at the points were new snapshots were printed, where the next sub-input file was generated by the python wrapper.

Please note that I was not involved in the generation of the original ProtoMS source code, nor its so far described *python wrapper*. What has been described so far is the state of the software at the beginning of this thesis.

As soon as ProtoMS 3.0 was released (which development was a group effort

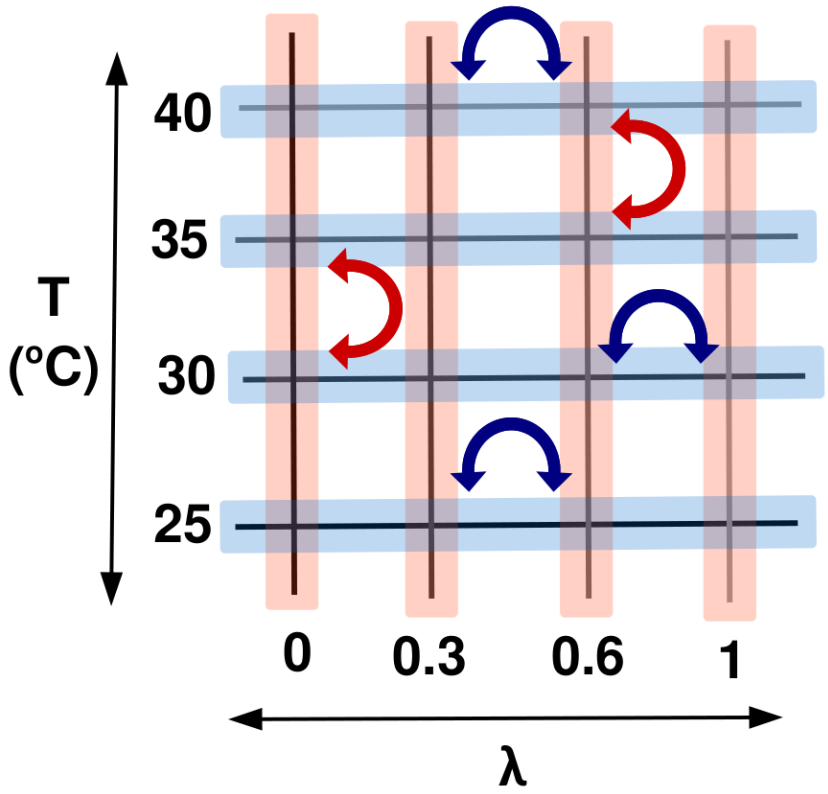


Figure 6.9: Schematic representation of the ProtoMS implementation of the combination of multiple replica exchange methodologies within a single simulation. In this example, the two replica types are temperatures (parallel tempering) and λ (alchemical transformation) represented by the two axes. Each processor is represented by the crossing between two lines. All processors within the same vertical line share the same value of λ and all those in the same horizontal line share the same temperature. The orange bars represent temperature ladders (all processors within a temperature ladder have different values of temperature and the same value of λ). The blue bars represent λ ladders (all processors within a λ ladder have different values of λ and the same temperature). The blue arrows represent swaps in λ and the orange arrows represent swaps in temperature. Note that, when a swap in λ is accepted (blue arrow), not only the values of λ must be swapped, but also the identifier that indicates the temperature ladder (orange bar) each processor corresponds to. Equivalent behaviour for λ ladders applies when temperature swaps are accepted. Using OpenMPI groups, the attempts of λ swaps in all λ ladders at the same time is easy to implement. Note that swaps between different values of λ may be accepted for different λ ladders at the same moment in the simulation. This concept applies also to temperature swaps and ladders.

by several members of the research group, including me), the python wrapper was removed, and each replica exchange methodology had to be implemented separately into the (Fortran) source code. The new implementation of the source code improved its speed by skipping all stopping and starting of the simulations, as it allowed for certain tasks to be provided in a non-linear manner (i.e. the printing of energy and configurational information to file). At the same time it made the handling of simulations simpler by avoiding the requirement of a wrapper, since all instructions could now be provided in a natural manner within one single input file. Specifically, tasks which must be performed regularly throughout the simulation (i.e. printing energy information) are indicated with a specific flag including the frequency of the action, and performed periodically, at the appropriate simulation times within the Fortran source code. However, this necessarily requires that the implementation of the replica exchange procedures, previously in python scripts, is now moved to the Fortran code, which is now allowed to run several replicas in parallel, all within the Fortran code.

Independent of parallel tempering being previously available for ProtoMS, a general structure of replica exchange was included within ProtoMS 3.0 for the simulation of λ replica exchange. This both simplified and made the inclusion of parallel tempering in the code more complex. It provided the general structure which the parallel tempering code needed to follow, but it also required both parallelizations to be fully compatible, allowing a customizable grid of λ values and temperatures.

While including either one of the replica exchange techniques within the source code required transforming the python instructions into Fortran code and adapting them to the slightly new structure of the source code, including several compatible replica exchange methodologies was more complicated. While previously all replica exchange related actions were performed by independent python scripts, now all had to be controlled and coordinated within the same source code. The coordination difficulty is related to the need to control different groups of replicas (i.e. those with different λ values and with different temperatures which may need to perform different actions at different moments of the simulation — for example if the swaps of different types of replica exchange are not happening at the same time in the simulation). The other issue is related to the storage of the informa-

tion by each of the processors taking each of the replicas. In the implementation of replica exchange in ProtoMS, each processor takes care of one configurational "trajectory". It is then the conditions of the simulation (i.e. λ value or temperature) that are swapped between different processors, providing effectively the same result as the swap of configurations between different simulation conditions. Each processor must then store its set of conditions associated with its replica, and this information must be swapped between the correct processors at the correct time in the simulation, without affecting the rest of the simulation conditions. Since the Monte Carlo acceptance tests have been developed for each of the replica exchange methodologies independently but not for a (conceptually strange) mixture of them, the implementation must make sure that exchanges between different replicas are only attempted between those with appropriate simulation conditions. For example, if a parallel tempering (variable = temperature) exchange is taking place in an alchemical transformation (variable = λ) simulation, swaps between different temperature replicas must only take place between replicas sharing the same λ variable. Equally, swaps between λ replicas must only be attempted between those sharing the same temperature.

A general structure was finally created with groups of tasks that controlled every "ladder" of temperatures or λ values. Temperature ladders would exchange processes with other ladders during a λ replica swaps and within themselves (between different λ ladders) during a temperature swap.

The implementation was made using the OpenMPI capabilities available in Fortran and, within those, the possibility of dividing the processors in different groups. For a simulation with replica exchange applied in the λ dimension (alchemical transformation) and the temperature dimension (parallel tempering), several temperature groups and several λ groups would be defined. The different temperature groups would correspond to the different temperature ladders, where within one temperature ladder all processors share a common λ value. This distribution would allow for temperature swaps to be performed at the same time in all temperature groups, with swaps restricted to replicas within the same group (sharing the same λ value). The same structure would apply to the λ dimension. It is key to note that the swap between replicas of the same group in the λ dimension will then correlate with a change of temperature group (as previously described, all

processors within the same temperature group must share the same λ value). A schematic representation of this implementation can be found in figure 6.9.

The new implementation of the software was tested in situations which required only λ replica exchange (which produced consistent results with the previous version of the software) as well as those requiring exclusively parallel tempering. Once these two basic test were successful, the implementation was tested on simulations which combined both λ replica exchange and parallel tempering. The path followed by the replicas as they swapped between different λ and temperature ladders was studied to search for any potential error in parsing information. Equally, the changes in the acceptance ratio of the swaps was studied as the number and proximity of replicas, both in temperature and λ dimensions, were modified.

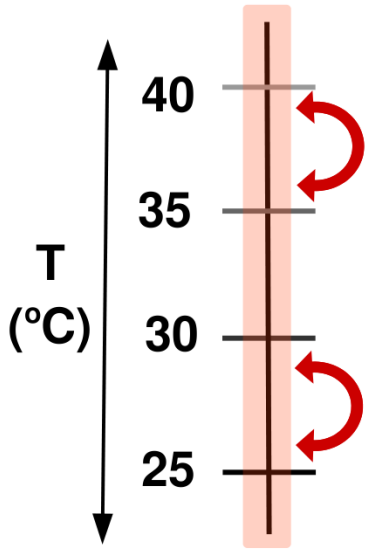


Figure 6.10: Schematic representation of the ProtoMS implementation of a replica exchange simulation with a single temperature ladder. This is the situation which would apply to the JAFS simulations. The orange bar represents the temperature ladder and the arrows the swaps between temperature replicas.

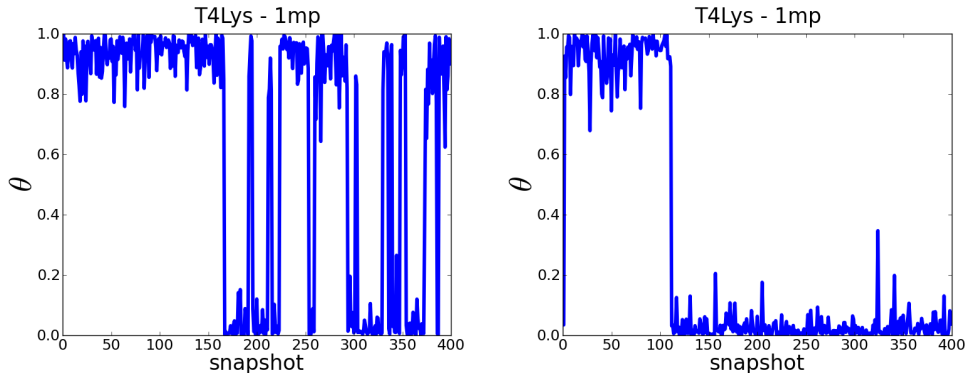
This implementation was then applied to the JAFS calculations in the mode of one unique temperature ladder (since no λ replicas were present, see figure 6.10).

Application to JAFS

To test the effects of parallel tempering on the JAFS calculations, simulations where setup with 25 different replicas per run, with equally distributed temperatures ranging from 25°C to 100°C. These runs are equivalent to those represented by plot (b) in figure 6.8. It must be noted, besides applying parallel tempering, the flexibility (move size) of the θ variable was increased between these two sets of results from 0.15 to 0.30 (the latter was observed to produce results closer to those observed experimentally in terms of ligand ranking by affinity).

The best (a) and worst (b), in terms of θ sampling, of the repeats of JAFS

score simulations of the T4Lys system with ligands 1mp and nbb (see chapter 5) are shown in figure 6.11. In this figure, the evolution of θ sampling with simulation snapshot at the lowest temperature replica for one of the fragments is shown. It can be appreciated how the worst sampled repeat appears very similar to the best sampled run obtained without parallel tempering (plot (b) of figure 6.8). While parallel tempering can be concluded to improve sampling (now all repeats present some swaps between low and high θ values), we would still hope to achieve further sampling in our JAFS simulations. How this is achieved will be presented in the solute tempering section below.



(a) (b)

Figure 6.11: Sampling in θ space of ligand 1mp in a JAFS pose parallel tempering simulation with only two ligands (1mp and nbb) within system T4Lys (see section 5.1.1 for a system description and figure 5.9 and the Appendix, table 11.1, for a molecular and 2D representation of the ligands, respectively). The lowest temperature replica (25°C) of two different repeats is shown, the best (a) and worst (b) in terms of θ sampling.

Solute Tempering

We have presented before the idea that the lack of sampling is a common problem in molecular mechanics simulations, and hence we could approach the issue by applying common methods to enhance sampling in biomolecular modelling. However, the best enhancement in sampling will be obtained when the method is chosen depending on the simulation conditions and the exact problem associated with the lack of sampling. In the case of JAFS, we are most interested in sampling the

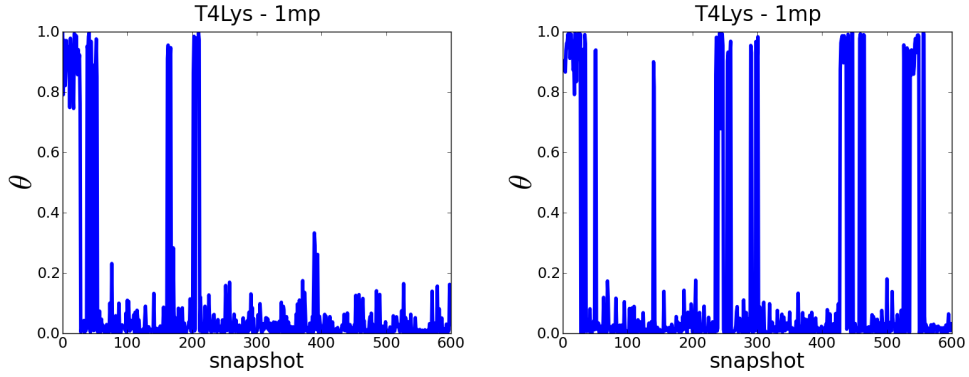
JAFS particles. It would be optimal, therefore, to apply some sort of technique to enhance the sampling of that particular region of the system, since the configurational sampling of the solvent or most of the protein residues can be expected to have little effect on our calculations (since their interaction with the ligands is expected to be minimal and hence their influence in the difference in affinity of several ligands or the ligand binding configurations, negligible).

A technique which is equally based on replica exchange, but which focuses on an increase of the sampling of the small molecules present in the system is available. It is called solute tempering, and it has been explained previously, section 3.7.1. In summary, the idea is that, rather than changing the temperature of the system between the different replicas, the energy of the system will be scaled. Modifications applied to any replicas which is not the base one (the one used to obtain the final results) can be any we like, since the acceptance test on the replica exchange moves will make sure the configurations that reach the base replica fulfil the Boltzmann ensemble. In our case, a scaling on the interaction energies of all our JAFS particles will be applied (a scaling per replica, on top of the θ scaling). Equations describing the energy scaling can be found in 3.7.1.

As explained in section 3.7.1, the difference between replicas in solute tempering is commonly described by referring to how "hot" the solute is, where further scaled (lower) interaction energies of the solute correspond to "hotter" solutes. While this is not an accurate description since the different replicas correspond to a change in the Hamiltonian (scaling of solute energies) rather than different temperatures, it simplifies the tags associated to each replica. We will be referring here to the different replicas by their associated temperature or "solute temperature". See section 3.7.1 and equation 3.38.

In these terms, in figure 6.12, plots presenting the θ sampling of solute tempering simulations with 16 replicas equally distributed, ranging from 25°C to 100°C are shown. It is worth noting at this stage that the ligand 1mp, when competing with nbb in this simulation setup is expected to spend most of its simulation time at low θ values (since experimentally, nbb is found to bind with higher affinity to the protein cavity). For all five repeats in these simulation conditions, the percentage of time spent with $\theta > 0.5$ (our definition of affinity *score*) is lower for 1mp than for nbb. For comparison, in the five initial repeats without any enhanced sampling,

only one out of five repeats showed a percentage of time spent with $\theta > 0.5$ is lower for 1mp than for nbb.



(a) (b)

Figure 6.12: Sampling in θ space of ligand 1mp in a JAFS pose solute tempering simulation with only two ligands (1mp and nbb) within the system T4Lys (see section 5.1.1 for a system description and figure 5.9 and the Appendix, table 11.1, for a molecular and 2D representation of the ligands, respectively). Two different repeats are shown. That with the worst θ sampling is shown in plot (a). Information from the lower "solute temperature" replica (with unscaled energies) is displayed.

In all repeats run with these simulation conditions, several swaps between high and low θ values are observed. The sampling of θ with simulation snapshot of the repeat with the lower number of swaps between high and low θ is shown in plot (a) figure 6.12.

Different Solute Temperature Settings

While the results of the application of consecutive enhanced sampling techniques is most definitely satisfactory, the choice of settings in terms of temperatures and "solute temperatures" ranges for the different replicas has been intuitive. At this point we wonder whether a more optimal distribution of replicas is available.

The logical option when optimizing a replica exchange-type protocol is looking at the acceptance ratio of the moves controlling the swaps between different replicas. When looking at these for different simulation setups with increasing temperature difference between replicas, it was found that these acceptance ratios were all higher than expected ($> 85\%$), and tending towards a plateau, without ever reaching what was considered optimal acceptance ratios ($40\% < \text{ratio} < 60\%$). This finding could have been expected and can be easily reasoned. We were expecting the acceptance

ratio to drop as the temperature difference between replicas increased, since the difference in the energy landscape must also increase, hence providing diverse configurations, with greater differences in energy, making the replica moves less likely. However, there are two factors we did not take into account. First, we were keeping constant the total number of replicas, hence, as the difference in temperature between replicas increased, so did the maximum temperature reached, and the acceptance ratio studied accounts for replica moves between all replicas. We were making the energy landscape (at least that concerning the solute) flatter for the top replicas as we were increasing the temperature difference between replicas, which could account for an increase in the acceptance ratio between the top replicas. At the replicas with high "solute temperature" the interaction energies of the ligand and scaled down close to zero. Hence, the difference in the energy of the system for different configurations of the ligand will be small. Consequently, the swap of configurations between different replicas (which are expected to diverge mainly in solute configuration, since the sampling of the rest of the system is not enhanced) at high solute temperatures will be associated with a small energy difference, accounting for a high probability of accepting the move. Second, the swaps between replicas can happen at any of the values of θ for the solutes. The values of θ for the JAFS particles can influence the acceptance ratio since they precisely account for scaling of the energies of the solutes.

The study of the acceptance ratio was hence considered not applicable to the optimization of the solute tempering, and focus was instead placed on the direct study of results and sampling for the different simulation settings.

JAFS score simulation for a pair of fragments, 1mp and dcb, on the T4Lys system (see chapter 5) were run with increasing "solute temperature" difference, in a search for the temperature difference which provided a better distinction between the tighter and the more weakly bound (greater difference between their scores, assigning a higher score to the ligand with higher affinity, nbb) as well as a better convergence between repeats (lowest standard error on each score). See section 4.2.1 for further information on the JAFS score simulations and how scores are obtained.

The study of the differentiation between binding affinity was performed by plotting the difference between the scores of both fragments for simulations with solute

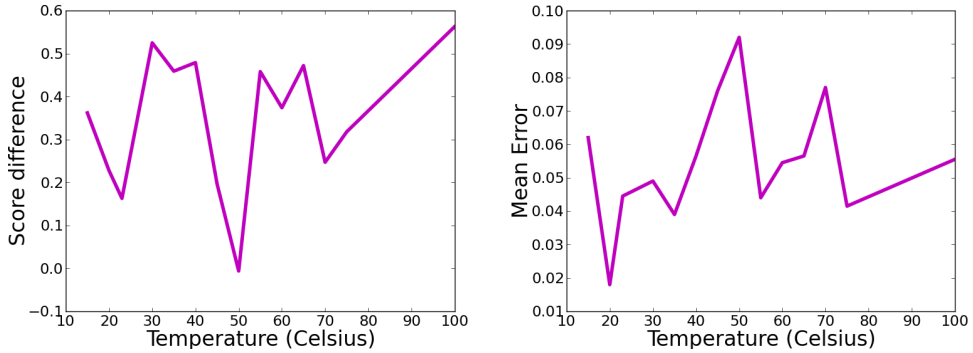
tempering with increasing difference in "solute temperature" between the replicas. In every case, the score was obtained as the proportion of simulation time each fragment spent at $\theta > 0.5$ in the lowest "solute temperature" replica (note the lowest "solute temperature" remained at 25°C — no scaling of the interaction energies — as the differences between replicas were increased for different simulation set-ups). The resulting plot can be seen in figure 6.13 (a).

The study of the convergence of the different repeats (measured as standard error) was performed by plotting the mean of the standard errors for the scores of both fragments for simulations with solute tempering of increasing difference in "solute temperature" between the replicas. Note again that, for every case, the score was measured at the lowest "solute temperature" replica (25°C, no scaling of the solute interaction energies). The resulting plot is shown at figure 6.13 (b).

These studies were performed with the hope of observing, for some particular difference in "solute temperature" between replicas, a maximum in the difference between scores (figure 6.13 (a)) and a minimum in the mean standard error of the scores (figure 6.13 (b)). If such scenario were found, the maximum in score difference and the minimum in mean standard error would define the optimal simulation set-up (clear differentiation between tightest and weakest binders with consistent results between repeats). However, as can be seen in figure 6.13, no clear trend could be found in either of the cases. The information was consequently discarded as noise.

As an alternative, measures of the sampling were used to select the optimal setup of the solute tempering replicas. Focusing on the Cartesian sampling (where, as explained previously, θ and Cartesian sampling are always expected to be coupled), the number of clusters identified was taken as a measure of sampling. It was hence decided that the initially attempted spacing of 5° C between replicas (16 replicas between 25° C and 100° C) was the optimal setup since a higher number of clusters, as well as high number of swaps between high and low θ was observed.

Plots showing representative examples of the sampling of θ throughout the simulation for two different setups (5° C and 100° C difference in "solute temperature" between replicas) are shown in figure 6.14. As can be seen, a higher number of swaps is obtained for the simulation with 5°C separation between replicas than for the 100° C separation setup. While a representative example was chosen, and this



(a) (b)

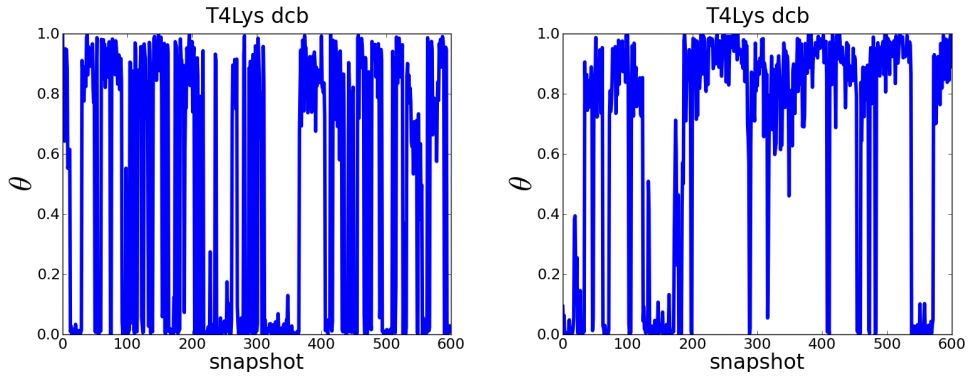
Figure 6.13: Study of the results of JAFS score calculations with fragments 1mp and nbb with different solute tempering simulation conditions. Study of the difference in score between both fragments (a) and mean error for the score of both fragments (b) for different separation between "solute temperatures" for the different replicas. Replicas are always equally spaced, and always 16 replicas are used in total. The lowest replica is always at 25°C (no scaling of the solute interaction energies).

trend is observed for all repeats run, it can also be observed that the difference is not as obvious as it might be expected from such difference in the simulation setup. The average number of clusters per repeat obtained from JAFS pose simulations with solute tempering and either 5°C or 25°C difference between replicas were studied for three different fragments (2dl, atd and ate) of the HSP90 system (see chapter 5). The results of this study can be seen in table 6.5. As it can be observed, for all three cases, a higher number of clusters per repeat is observed for the simulation with a 5° C difference in "solute temperature" between the replicas.

Fragment	Clusters 5°C step	Clusters 25°C step
2dl	10 ± 0.7	5.8 ± 0.6
ate	18.8 ± 2.6	10.6 ± 1.0
atd	17.6 ± 0.7	9.8 ± 1.0

Table 6.5: Average number of clusters per repeat and standard error of the mean for JAFS pose simulations on the HSP90 system (see section 5). Five repeats were run in each case.

In figures 6.15 and 6.16, examples of paths of the different replicas for solute



(a) (b)

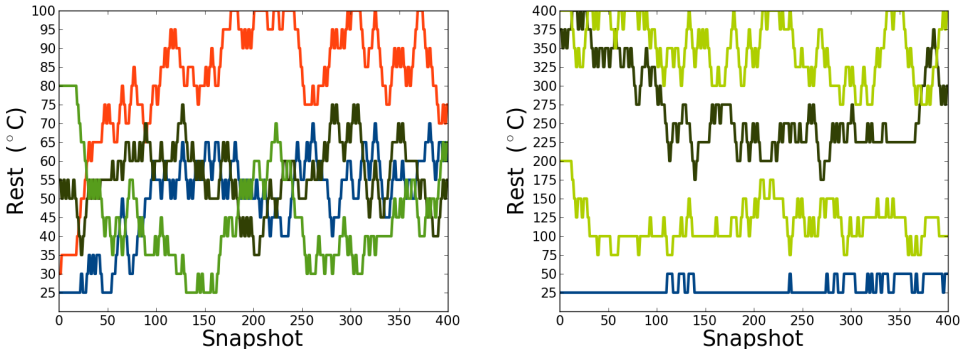
Figure 6.14: Plots of the evolution of the θ variable of fragment dcb in JAFS score solute tempering simulations of fragments 1mp and dcb in the T4Lys binding cavity (see chapter 5). Both plots have been obtained from the replica at lowest "solute temperature" (25°C, or no scaling of the ligand interaction energies) of 16 replicas with equally spaced "solute temperatures". In plot (a) the solute temperatures of the replicas were separated by 5°C while in plot (b) they were separated by 100°C. A representative repeat is shown.

tempering with different temperature ranges can be observed. Owing to our equal spacing distribution of replicas and consistent use of 16 replicas, a bigger range in "solute temperature" corresponds to a proportionally bigger difference in "solute temperature" (step) between replicas. As expected, a pattern more typical of a random walk is observed in replicas with a step of 5°C that when a 25°C step is chosen. While the choice of "solute temperature" ranges was based on the sampling as studied by the number of clusters per simulation, the path descriptions for replicas support the choice of lower step associated with optimal sampling.

Notice that a study of the "random walk" sampling of the different replicas is performed here rather than a comparison of the acceptance ratios for the replica swaps. The reasons why the acceptance ratios are not a good measure of the optimal sampling in this case can be found at the beginning of this subsection.

6.3.3 Conclusions

In this section we have addressed the issues found on applying the JAWS method to fragments during section 6.2. The two main improvements desired with respect to previous developments of the method were an increase in sampling



(a) (b)

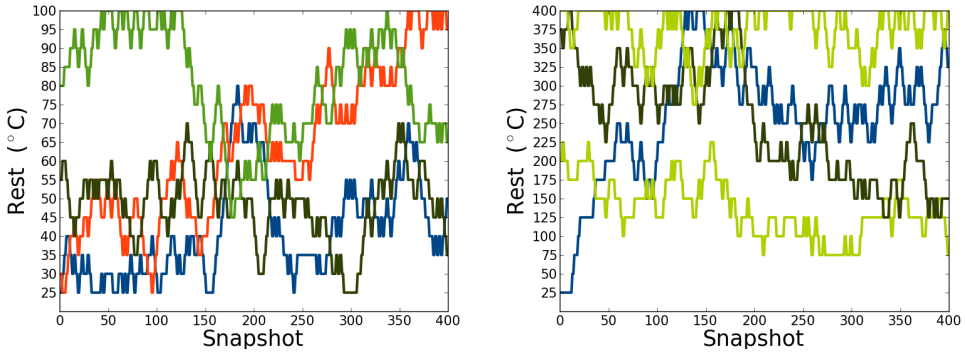
Figure 6.15: Plots produced with calc_replicapath.py (see section 4.5.2) from the JAFS pose simulation results on ate (system: HSP90). The x axes corresponds to simulation snapshot and y axes shows the "solute temperature" ("rest", replica exchange solute temperature) in degrees Celsius (see section 3.7.1 and 4.4) associated with each replica. Figure (a) corresponds to a solute tempering difference between replicas of 5°C while (b) corresponds to a difference of 25°C between replicas.

(both in θ and Cartesian space) and a reduction of the "unrealistic" states of the JAFS particles with intermediate values of θ. In this subsection we will summarize the main achievements in addressing these problems and conclude on the state of the JAFS method heading towards the production runs in chapter 7, where the finalized version of JAFS is used.

Applying a Linear Hydration Penalty

In this section we have seen how changing the shape of the hydration penalty could help reduce the presence of JAFS particles at intermediate values of θ. The previous shape of the hydration penalty, favoured intermediate θ values over extreme values of θ ($\theta = 1$ and $\theta = 0$).

The hydration penalty fulfils the function of accounting for the hydration free energy of each of the JAFS molecules in the simulation to correctly measure their relative affinity towards the protein. To fulfil this function in a theoretically sound manner, the only requirement is that the penalty applied at the end points ($\theta = 1$ and $\theta = 0$) correctly corresponds with the affinity of the fragment for bulk water (hydration free energy of the fragment). The behaviour of the hydration penalty



(a) (b)

Figure 6.16: Plots produced with calc_replicapath.py (see section 4.5.2) from the JAFS pose simulation results on atd (system: HSP90). The x axes corresponds to simulation snapshot and y axes shows the "solute temperature" ("rest", replica exchange solute temperature) in degrees Celsius (see section 3.7.1 and 4.4) associated with each replica. Figure (a) corresponds to a solute tempering difference between replicas of 5°C while (b) corresponds to a difference of 25°C between replicas.

at the intermediate θ states does not need to follow any particular shape, from a theoretical perspective, since the existence of these intermediate states themselves does not correspond to any real state of the experimental system whose binding affinities we are trying to reproduce. In practice, decreasing the abundance of these intermediate θ states should lead to a more accurate representation of the experimental system, avoiding interactions between particles present in the system ($\theta \approx 1$) and the unrealistic particles at intermediate θ.

Previously, the hydration penalty was taken as a polynomial approximation to the PMF of decoupling the particle from bulk water as obtained with alchemical transformations and understanding $\theta = 1 - \lambda$. Observing undesired effects with this shape, the hydration penalty was defined as a linear representation between the extreme points of the hydration free energy for $\theta = 1$ and zero for $\theta = 0$ (the hydration free energy of "nothing" or a dummy atom must be zero). The associated mathematical representations can be found in equations 4.2 and 4.3.

The results observed for this change are satisfactory as can be seen in figure 6.8, with sampling of θ clearly more focused on its extreme values for the linear implementation of the hydration penalty (plot (b)) than for the polynomial implementation (plot (a)) of the hydration penalty.

Applying Enhanced Sampling Methods

While the application of the linear hydration penalty helps in reducing the presence of the intermediate θ states, it does not help to improve sampling. In fact it can be considered to make this matter slightly worse, as the JAFS particles present now a certain tendency to remain at one of the extreme values of θ without sampling the opposite end of the range of θ values.

To address this problem, the application of enhanced sampling methods was undertaken. Parallel tempering (and subsequently solute tempering) where chosen since they fulfilled our requirements (i.e. not altering the Boltzmann distribution of states) as well as for implementation reasons. Parallel tempering was then implemented in ProtoMS since this was not available at the time in the release version of the code.

The effects of parallel tempering were studied on θ sampling, and are shown in figure 6.11. As can be seen when these results are compared with those in figure 6.8, plot (b) (where the best sampled repeat of linear hydration penalty simulations without enhanced sampling is shown) parallel tempering does considerably improve θ sampling. However, further sampling improvement was desired.

To this end, the application of solute tempering was attempted. Solute tempering was chosen for its ability to focus the sampling boost of the simulation species of interest, in our case, the JAFS particles, while leaving the rest of the system unaltered. Leaving the rest of the system sampling at the same temperature (and with the same energy description) for all replicas, improves the ratio of swaps between different replicas, allowing for a more efficient boost to the sampling of the configurations of JAFS particles.

The results of the application of solute tempering can be seen in figure 6.12 and, by comparing these with the plots in figure 6.11, a further improvement in sampling can be observed.

The specific simulation settings associated with solute tempering (such as the distribution of the different replicas and their associated "solute temperatures") was chosen based on the observed number of clusters generated and the study of the "random walk" for different sets of simulation parameters. A more thorough study of the optimization of parameters for solute tempering was attempted, but no conclusive result was obtained, hence these data were discarded.

Summary

Both issues observed in the application of JAWS to fragment molecules during section 6.2 were addressed in this section. The application of a linear implementation of the hydration penalty, together with the application of solute tempering generated a sampling of θ focused on its extreme values, while swaps between these both extremes ($\theta = 1$ and $\theta = 0$) were attempted frequently throughout the simulation, for the systems tested.

6.4 Choosing Thresholds

In the previous sections, the naive approach to the project has been presented, followed by the changes applied to the method to address the issues found during its initial implementation. While the major developments of the JAFS method have been shown, the thresholds applied to the θ variable, both in the JAFS score and JAFS pose methodology, have been presented as arbitrary so far. In this section, information on the study of different choices of threshold will be presented, concluding with the reasons why the used thresholds were chosen.

6.4.1 JAFS score Threshold

In the JAFS score protocol, a threshold is used to calculate the score (measure of binding affinity towards target) for each of the fragments studied. As explained previously, the measure of affinity in JAFS score is based on the sampling of higher values of θ during the simulation for the ligands with higher binding affinity (see section 4.2.2 for an in-depth explanation). The measure of sampling of higher values of θ is made, in the JAFS score protocol, by measuring the proportion of time each of the fragments in the simulation spends with a value of $\theta > 0.5$, where 0.5 is called the threshold. An equivalent measure could however be obtained with alternative values of threshold, or a different measure could be used, such as the average value of θ during the simulation sampled by each of the fragments. These alternative options were studied (following the JAFS score protocol as described in chapter 4) in the T4Lys system (see chapter 5) and the results are presented here.

The results of applying a threshold of $\theta > 0.9$ and a threshold of $\theta > 0.5$ to the calculation of scores for ligands 1mp and dcb binding to T4Lys are shown in table 6.6. Knowing that dcb has been calculated experimentally to be a tighter binder to T4Lys than 1mp (relative binding free energy of 2 kcal/mol),[223] we would expect a higher score to be obtained for ligand dcb. The relative binding affinities obtained with well established simulation methods, using the same force field and simulation package as the JAFS runs (as shown in section 7.2, table 7.5) predict the binding affinity of dcb to be higher than that of 1mp by 1.15 ± 0.28 kcal/mol. We might hence expect our simulations to slightly underestimate the difference in binding free energy between both fragments.

System: T4Lys – 1mp & dcb

Fragment	Score $\theta > 0.5$	Score $\theta > 0.9$
1mp	0.533 ± 0.072	0.385 ± 0.050
dcb	0.487 ± 0.065	0.206 ± 0.043

Table 6.6: Scores obtained for ligands 1mp and dcb on T4Lys (see chapter 5 and the Appendix, table 11.1) from five repeats of a JAFS score simulation including only these two fragments. Scores are shown for two different thresholds, $\theta > 0.5$ and $\theta > 0.9$.

As can be observed in table 6.6, using the $\theta > 0.5$ threshold, the scores for both fragments are within error, hence it could not be established, from that data, whether 1mp or dcb binds tighter to T4Lys. Meanwhile, the use of a higher threshold ($\theta > 0.9$) generated results which (wrongly) assessed 1mp as the tightest binder. It is hence considered that, by using a higher threshold, relevant information on the assessment of the best binder is discarded. A choice of $\theta > 0.5$ as threshold was hence considered more optimal. It is worth mentioning that this study was performed for a range of thresholds (0.5, 0.6, 0.7, 0.8 and 0.9) where the general trend of dcb being assessed a lower score for higher thresholds was followed (information not shown).

However, alternatives to the use of any threshold can also be presented, where, for example, the average value of θ sampled throughout the simulation for each of the fragments is taken as their score. This possibility was studied for different pairs (and one trio) of fragments binding to T4Lys (see chapter 5), as shown in table 6.7. Taking into account that the order in binding affinity of these fragments is, from tightest to weakest binder, nbb > dcb > 1mp > wa1 (as can be seen in section 7.1, tables 7.5 and 7.2) the use of threshold $\theta > 0.5$ was always associated with a higher score to the tightest binder and / or lower score to the weakest binder compared to those generated via the average θ sampled throughout the simulation.

In conclusion, the scores (estimated binding affinity) obtained as the proportion of simulation time spent by each fragment above a threshold of $\theta > 0.5$ allowed for a neater distinction of the tightest binder towards the T4Lys cavity, compared to the use of higher θ thresholds as well as the average θ throughout the simulation as a score. Consequently, throughout this thesis the scores have been calculated as

System: T4Lys – 1mp & nbb

Fragment	Score $\theta > 0.5$	Score mean θ
1mp	0.20 ± 0.04	0.21 ± 0.04
nbb	0.79 ± 0.04	0.67 ± 0.03

System: T4Lys – 1mp & dcb

Fragment	Score $\theta > 0.5$	Score mean θ
1mp	0.53 ± 0.07	0.51 ± 0.06
dcb	0.49 ± 0.07	0.43 ± 0.06

System: T4Lys – 1mp & dcb & wa1

Fragment	Score $\theta > 0.5$	Score mean θ
1mp	0.51 ± 0.08	0.48 ± 0.07
dcb	0.52 ± 0.07	0.47 ± 0.06
wa1	0.00 ± 0.00	0.08 ± 0.00

Table 6.7: Comparison of the scores obtained as the proportion of snapshots spent with $\theta > 0.5$ for each fragment, and the mean value of θ per fragment. Data shown for five repeats of three different JAFS score simulations on the T4Lys system, two of them with two ligands each (1mp and nbb or 1mp and dcb) and one of them with three ligands (1mp, dcb and wa1). See chapter 5 and the Appendix, table 11.1 for further information on the system and ligands.

stated in chapter 4.2, using a threshold of $\theta > 0.5$.

6.4.2 JAFS pose Threshold

Just as a choice of threshold is needed for the JAFS score simulations, a threshold is also required in JAFS pose. In this case the threshold limits the configurations selected to cluster (see section 4.5.2). The resulting cluster representatives will then be compared to the crystal binding mode and an RMSD obtained, with the correct crystal pose considered to be found whenever a cluster representative falls within a 2 Å RMSD. It is worth noting that the threshold used for clustering in JAFS pose will also be used for the selection of poses for clustering in JAFS score simulations, when the study of the binding configuration is desired.

While the same threshold as that of the scoring in JAFS score simulations could in principle be used for the selection of configurations for clustering, the require-

ments vary between both processes. In the case of the scoring threshold, a scoring value should be obtained for all ligands which present some affinity towards the protein (a null score, obtained when the ligand was not above the scoring threshold at any point during the simulation, must be reserved for non-binders, which will be discarded for any further development into tighter-binding ligands). In the case of the pose-selection threshold, a reduced number of poses is desired. Since JAFS pose does not offer a method to select the correct binding geometries among those generated, it is likely that a posteriori rescoring is desired, whose computational expense is expected to increase with the number of binding configurations (see section 4.2.2 and chapter 8). In particular, only configurations with the highest values of θ are desired. A binding configuration observed for a ligand with interactions scaled by 0.5 (or 0.7), may well contain atoms partially overlapping with protein residues. In general, any pose observed by ligands with their interactions partially "on", but not observed for fully interacting ligands could be an artefact, and is discarded.

Consequent with this line of thought, a threshold of $\theta > 0.9$ was chosen as the limit above which configurations sampled throughout the simulation are collected for clustering and pose finding.

Chapter 7

Production

In this section we will be presenting the results of the main production runs of this thesis. The JAFS methodology is considered to be fully developed at this stage of the project, and it is now tested on a variety of systems (see chapter 5) which should reveal the difficulties that might be encountered by the methodology.

As described in the methods chapter (chapter 4), the JAFS methodology has been divided in two different protocols – JAFS score and JAFS pose — with different aims. The results obtained from applying the two protocols to our systems will be described here in separate sections of this chapter. The two protocols provide different information and are appropriate for different situations.

JAFS score is suited for small cavities or situations where the region within a binding cavity where the ligands bind (or where the binding of ligands wants to be studied) is known. It requires several ligands (fragments) to be studied at once, and their relative binding affinities will be estimated, as well as their binding configuration within the user-limited binding region studied. It is important to keep in mind that this is not a method to calculate the relative affinity of different known binding configurations. The configurations of all ligands are expected to sample during JAFS score simulations.

JAFS pose is a protocol suited for bigger binding cavities, and in situations when the solvation state of the binding cavity might not be known and/or the presence of water mediated interactions between the ligands and the protein is to be studied. The output of these simulations is a study on the possible binding configurations of the ligand, with information on the solvation state of the cavity and automatic

placement of waters required for the ligand binding in each configuration. No competition between different ligands is currently implemented in this protocol.

Given that these two protocols are suited for different situations, not all systems have been run with both protocols. Only MUP and CDK2 systems have been run with both JAFS pose and JAFS score. T4Lys is an optimal system to run with JAFS score due to its small and occluded cavity. However, for this same reason, running T4Lys with the bigger user defined region of JAFS pose would not provide any extra information. In the same way, no space is expected to be available for any solvent molecule in the T4Lys cavity at the same time as any ligand is present making it unnecessary to run this system with the JAFS pose protocol. HSP90 has only been run with the JAFS pose protocol. While running JAFS score on HSP90 would be possible and might provide interesting information, one difficulty is added by this system, and it is the presence of known water mediated interactions between the ligands and the protein. While JAFS pose takes water into account automatically, JAFS score does not offer this possibility. A decision would have to be made on how to treat these bridging waters, and previous information of the system would need to be included in the JAFS runs, which goes against the black-box idea which is central to the development and study of the JAFS protocol. It was hence decided to leave HSP90 out of the system set used on the JAFS score protocols (however, we do admit this system might be of interest in future studies with JAFS).

From now on, results of the JAFS runs and their discussion will be presented in this section, as previously mentioned, divided by the two protocols, JAFS score and JAFS pose. We have decided to start this study with the JAFS score protocol, due to its simplicity compared to JAFS pose.

7.1 JAFS score

The theory, set-up and analysis of JAFs score have been explained previously in section 4.2.1, In this section, some specifics of the simulation set-up as it was performed in the production runs of JAFS score will be presented first. Next, the results of the production runs of JAFS pose will be presented, starting with a general discussion. The discussion of specific aspects that have been the focus of study throughout the development of the method will then be presented. Lastly, the results will be analysed on a system by system basis, with special focus on the most complex of all test cases.

7.1.1 Specific Setup

All production runs of JAFS score share the same simulation specifications:

- Simulations consisted of 5 million moves of equilibration and 40 million production moves
- Five repeats were run for each system
- The JAFS box was defined as a 5 Å x 5 Å x 5 Å simulation box around the centre of the ligand which was co-crystallized with the protein structure used
- One copy of all fragments within that system were included within the JAFS box for each simulation
- Solute tempering was applied to the simulations, with 16 equally spaced replicas with solute temperatures between 25°C and 100°C
- The proportion of attempted moves applied to all JAFS runs has been: solvent = 51%,protein = 9%, solute = 20% and θ moves = 20%

Protein	Fragment
T4Lys	bnz
CDK2	vth
MUP	ipz

Table 7.1: Co-crystallized fragment with each of the protein structures used in JAFS score production runs.

For each of the three systems presented here, the protein has been co-crystallized with one of the fragments taking part in the JAFS score simulations. The fragment co-crystallized with the protein structure used for each system can be found in table 7.1. This is relevant, since a certain advantage could be experienced by the co-crystallized fragment with respect to the other ones since the protein conformation is not expected to change much during the simulations. While the protein is allowed to sample freely, our Monte Carlo simulation setup does not promote efficient sampling of the protein conformational landscape.

The difference in the number of fragments can be taken as an *a priori* measure of the complexity of JAFS score simulations. All JAFS particles must sample their available configurational space and sufficiently sample the θ space to be assigned a meaningful score (measured as the proportion of simulation time spent at $\theta > 0.5$). Besides, the higher the number of binders (score significantly higher than zero), the higher the accuracy and precision required to satisfactorily distinguish their assigned scores (since scores can only range from zero to one).

7.1.2 General Discussion

The results of the production runs of JAFS score can be seen in table 7.2. The success of these simulations vary widely with the studied system. For the MUP-1 system, which presents the lowest complexity in terms of number of fragments to rank, as well as a dry cavity (little or no water is expected to have an influence on the binding process) the results are satisfactory. The two fragments are correctly ranked (ipz top, prz bottom), with their scores (proportion of simulation time spent at $\theta > 0.5$) being separated by several standard errors (where the score shown is the average between scores from all simulation repeats and the standard error is that associated with this average). We can hence be quite confident that choosing the correct ranking order is due to a higher affinity of binding of ipz with respect to prz under our simulation conditions. Hence, our conditions seem to represent the real system with sufficient accuracy to extract conclusions on binding affinity for this particular system and number of fragments.

The next system with increasing complexity is T4Lys, where five different ligands must be ranked. The cavity is occluded, with limited space, and again, no water molecule is expected to intervene in the binding process. As an initial guess,

System: T4Lysozyme

Fragment	Score	JAFS rank	Exp rank	Exp ΔG (kcal/mol)
nbb	0.59 ± 0.03	1	1	-6.7
dcb	0.11 ± 0.03	4	2	-6.4
bnz	0.15 ± 0.03	3	3	-5.2
1mp	0.19 ± 0.04	2	4	-4.4
wa1	0.00 ± 0.00	5	5	—

System: MUP-1

Fragment	Score	JAFS rank	Exp rank	Exp ΔG (kcal/mol)
ipz	0.91 ± 0.03	1	1	-9.2
prz	0.55 ± 0.06	2	2	-8.1

System: CDK2

Fragment	Score	JAFS rank	Exp rank	Exp ΔG (IC_{50})
vth	0.00 ± 0.00	9	1	120 μM
vta	0.29 ± 0.08	3	2	185 μM
wcc	0.22 ± 0.14	5	3	< 1 mM
vtm	0.22 ± 0.08	4	4	< 1 mM
cd1	0.57 ± 0.09	1	9	—
cd5	0.21 ± 0.07	6	9	—
cd6	0.39 ± 0.10	2	9	—
cd8	0.06 ± 0.03	7	9	—
cd9	0.02 ± 0.02	8	9	—

Table 7.2: Results of JAFS score compared to experimental binding affinities for T4Lysozyme[223] MUP[238] and CDK2.[239] The score obtained with JAFS, rankings for both JAFS score and with experiment, as well as the experimental binding affinities, are shown. The errors shown associated to the scores are standard errors of the mean. No error is shown on the experimental binding affinities for consistency, since some were found in the literature while others were not available. The experimental binding affinities were obtained with a range of different methods, including different methods applied to ligands of the same system. Note that, in the case of CDK2 ligands, the IC_{50} shown for wcc and vtm correspond to experimental 64% I at 1mM and 54% I at 1mM, respectively.

the increase in complexity from MUP-1 to T4Lys could mainly be associated with the increase in the number of fragments to rank. It can be observed in the results how the top binder (nbb) and the non-binder (wa1) are successfully ranked among all fragments. Errors associated with their scores remain low for all fragments and extremely low for the non-binder. The ranking associated with the intermediate binders does not match the ranking based on experimental results. Taking a closer look at the scores and associated errors of these intermediate binders (dcb, bnz and 1mp), we can see that their results are within error. We cannot tell them apart and hence, arguably, their ranking should be set to be the same. Comparing the errors with those in the MUP-1 system, we can see that those are not bigger. It can be speculated that the problems found in the T4Lys test case are intrinsically associated with increasing the number of fragments in a JAFS score simulation. However, further systems would be required to confirm this statement. In any case, errors typically associated with experimental binding affinities and inaccuracies of the force field are a limiting factor to the results that may be obtained when comparing simulations to experimental data. In this context the performance of JAFS score on T4Lys is considered satisfactory, particularly given that no fragments are ranked in an incorrect order (when errors associated with their scores are taken into account) but rather the binding affinities of a set of the fragments cannot be discriminated.

CDK2 is the most complex of the systems, presenting nine fragments to be ranked and a bigger cavity (same size JAFS box defined within a bigger protein cavity), when waters could potentially be involved in fragment binding. Results for CDK2, are distinctly different from the two previous examples. When comparing the ranking based on experimental results with that based on JAFS simulations, no exact matches are observed and no trend can be obviously spotted which pushes the best experimental binders to the top of the JAFS rank. Looking more carefully at the particular system, we can see that five (cd1, cd5, cd6, cd8 and cd9) out of nine fragments are classified as non-binders experimentally. Of these, one of them (cd1) is assigned as top binder following the results of JAFS. Equally, the weakest binder according to JAFS (vth) is the top experimental binder. The ranking trend is not followed inversely either, with three of the five decoys being ranked among the bottom five binders with JAFS (what could be considered "correct ranking").

Problems here can arise from sampling (further simulation time could be required for more complex systems), or from an incorrect representation of the system in our JAFS simulation (be it simulation setup, force field parameters, etc.). The increase in standard errors associated to the scores (proportion of simulation time with $\theta > 0.5$) for CDK2 compared to previous systems suggests sampling is more problematic here. However, the top binder for JAFS score and its bottom ranked fragment do not present scores which are within error of one another, suggesting sampling must not be the only issue.

In this subsection, a general overview of the results of the JAFS score simulations for all systems has been presented. In the following subsections, the main points highlighted here will be further discussed, key points in the analysis of the JAFS simulations will be studied in depth, and the particularities of each system will be presented in detail.

7.1.3 Correlation Between Sampling in Cartesian and θ space

At the core of JAFS is the idea of an increase in sampling. This idea flows from the development of JAWS (from which JAFS originated), as a method which could obtain equivalent results to GCMC but increasing sampling of water insertion and deletion, to JAFS as an approach to simulate a binding pocket where swaps between binding configurations happen within reasonable simulation time. Intrinsically, the sampling in θ space and that of Cartesian space are interlinked. The reason why JAFS can provide increased Cartesian sampling within a binding cavity is the presence of fragment copies at low values of θ, which, having their interaction energy scaled down, can explore the binding region without clashing with other molecules in the system. However, this increase in sampling at low θ will only be productive if particles sample effectively θ as well, being able to get back to their (at least nearly) full interaction energy when a region with favourable interactions is found. It is at this point when these interacting particles

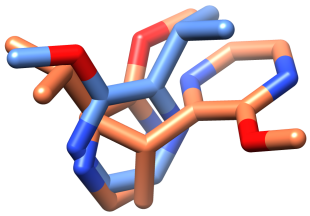


Figure 7.1: The crystal structure of ipz (blue) and cluster representatives obtained at the third repeat of JAFS score (orange). Hydrogens hidden for clarity.

will be captured and their poses clustered during analysis. It is only if particles are capable of finding their favourable regions of binding and increase their θ value accordingly that the measure of the proportion of time each fragment spends at high θ, as a measure of relative binding affinity makes sense.

box volume	MUP cavity volume	T4Lys cavity volume
125 Å^3	118 Å^3	114 Å^3
343 Å^3	297 Å^3	267 Å^3

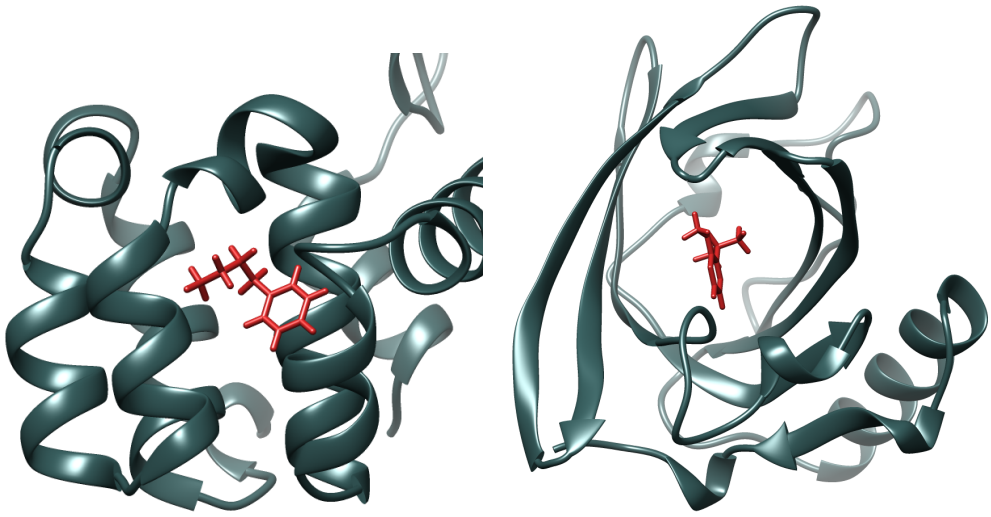
Table 7.3: Volume of the protein cavities of MUP and T4Lys systems (see chapter 5) enclosed by JAFS box of different sizes — centred around the same position as the JAFS box used for the JAFS score calculations — once the volume occupied by protein atoms has been discarded.

We will take here an example from JAFS score, where the JAFS box is smaller than that of JAFS pose simulations and only one copy of each fragment species is present, to analyse the expected correlation between the sampling in θ space and that in Cartesian, expressed in RMSD to the crystal pose. That is, we expect particles at low θ to present high variability in their RMSD, since they must be sampling most of the JAFS box. On the other hand, particles with high values of θ should present barely variable RMSD, since their interaction energies will be (almost) fully "on". Ideally, their RMSD must be both stable and low, meaning they have increased their θ after finding favourable interactions in the crystal binding geometry.

While this sampling strategy of θ and Cartesian space is perfectly true conceptually, the results observed in practice will depend on the environment of the JAFS box. The whole rationale of the concept explained above is based on the fragment at high values of θ not being able to sample Cartesian space, due to the interaction energy between the fragment and surrounding particles. However, given the extreme example of the JAFS cavity, simulating only one fragment in vacuum, the value of θ would have no influence on the sampling of the fragment. How close our system is to that extreme example of the fragment in vacuum will influence how closely our expectation is followed.

We will show here examples of the sampling of one of the fragments present in one particular repeat, for two different systems, T4Lys and MUP. While the

size of the JAFS box for both sets of simulations is identical (see section 7.1.1), the available space to sample for the fragments in not the same in both systems. This can be appreciated in figure 7.2. Their cavity volumes as estimated by our *cavity_volume.py* tool (see section 4.4.1) have been calculated. The volumes estimated to remain available once protein atoms have been taken into account within increased sizes of boxes — centred around the same position as the JAFS box used in our JAFS score simulations — is shown in table 7.3. In other words, the table shows the free volume for the ligands to explore. When looking at the table it is important to remember that, while the JAFS box used in the JAFS score simulations is 5 Å x 5 Å x 5 Å (125 $Å^3$) in size, only the centroids of the atoms are constrained by the box, hence the ligands will effectively be sampling bigger volume. We will see how results consequently differ between the T4Lys and MUP systems.



(a) (b)

Figure 7.2: Crystal structure of T4Lys with ligand nbb bound (a) and MUP with ligand ipz bound (b). Protein in dark grey, ligands in brown.

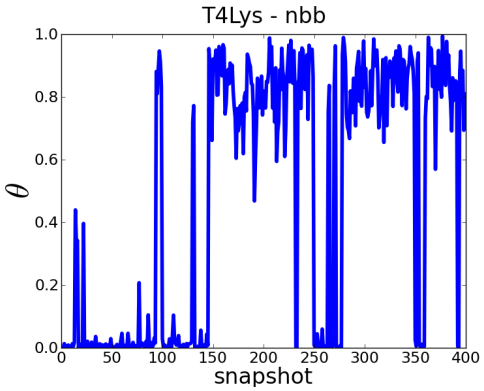
First, we will study the correlation between θ sampling and Cartesian sampling (as measured by RMSD to the crystal structure) for ligand nbb. These are shown in figure 7.3 for the second repeat of T4Lys JAFS score. We can see how, during the first 15 million production moves (150 snapshots), nbb mostly remains with $\theta < 0.2$, and consequently, there is a high variability in RMSD to crystal binding mode, of

between 2 Å and 8 Å. This is true for this period of the simulation except for a few snapshots around 10 million moves (snapshot 100), where $\theta > 0.8$ are sampled, and the RMSD seems to remain fairly stable and lower than 2 Å during those few snapshots. During the rest of the simulation, periods with $\theta > 0.6$ are associated in the neighbouring plot (b) with fairly stable RMSD values between slightly below 2 Å and 3 Å. Notice that it is the stability of the RMSD and not its particular values that we expect to correlate with high values of θ. Meanwhile, when $\theta < 0.2$ are sampled, an increase in RMSD sampling is observed. The variability in RMSD is harder to observe during the short periods spent at low θ simply due to lack of amplification in the plot. We can hence conclude that, for nbb in T4Lys, the correlation between sampling of θ and RMSD is entirely as expected based on our theoretical understanding of JAFS. See over for comments on these ideas.

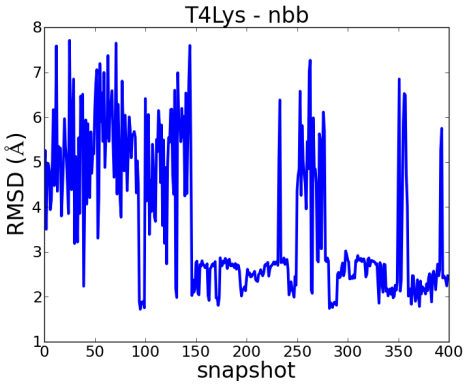
We will now examine the θ and RMSD sampling of ligand ipz during the third repeat (chosen as a representative example) of the MUP JAFS pose runs (see figure 7.4). No long periods with consistent low values of θ are observed. However, short-lasting decreases in θ to very low values are loosely correlated with sudden increases in RMSD sampling in an otherwise fairly stable plot. Periods (like that between snapshot 200 and nearly 350) which maintained $\theta > 0.5$ present more stable values of RMSD, while some RMSD variability is still observed. After the sampling of low values of θ around snapshot 350, the stability of RMSD is again decreased (higher RMSD variability). In conclusion, the expected general trend can be observed in these plots, where low θ is generally associated with high RMSD variability. However, results are more noisy here than in the previous example of nbb in T4Lys. It is also important to note that the stable RMSD here is not particularly low, hence the crystal binding pose has not been successfully found in this repeat, as it can be observed in figure 7.1.

Comparing both examples shown in this section, an increase in noise and greater differences to the ideal behaviour as explained at the beginning of this section is observed for systems with more volume available for the ligands to sample. This may be considered a limitation of the methodology. However, in the particular example of the JAFS score simulations on MUP as performed in this thesis, correct ranking of the ligands has been observed (see sections 7.1.2 and 7.1.4).

Rather than plotting the value of RMSD with respect to the crystal binding



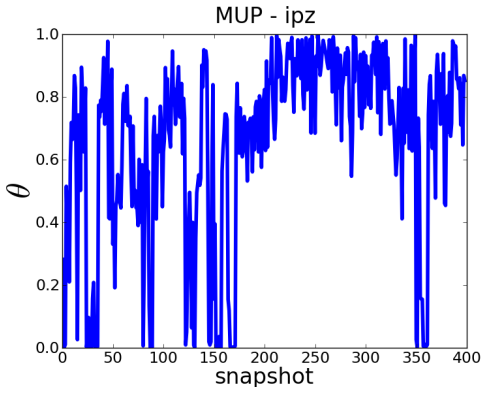
(a)



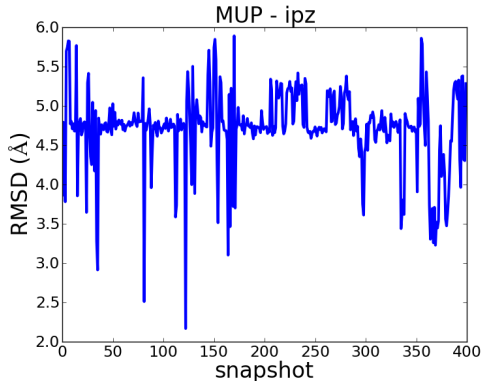
(b)

Figure 7.3: Evolution with simulation snapshot of θ (a) and RMSD with respect to crystal pose (b) of nbb in repeat 2 of its JAFS pose runs. Observe how higher θ corresponds to lower and more stable RMSD.

geometry, as has been shown in figures 7.3 and 7.4, the RMSD with respect to previous snapshot (RMSF) can be displayed. This representation of the same repeats and ligands as discussed previously is shown in figure 7.5. By studying this representation the situation between both systems appears more different. It must be taken into account that stable RMSDs with respect to crystal pose in plots (b) of figures 7.3 and 7.4 should correspond with low RMSDs with respect to previous snapshot (RMSF) in figure 7.5. While the correlation for nbb (figure 7.3 and plot (a) in figure 7.5) is satisfactory — high θ values correspond to stable RMSD values and low RMSF values — any form of correlation seems hard to find when looking at the ipz plots (figure 7.4 and plot (b) in figure 7.5). The reason for this lack of



(a)

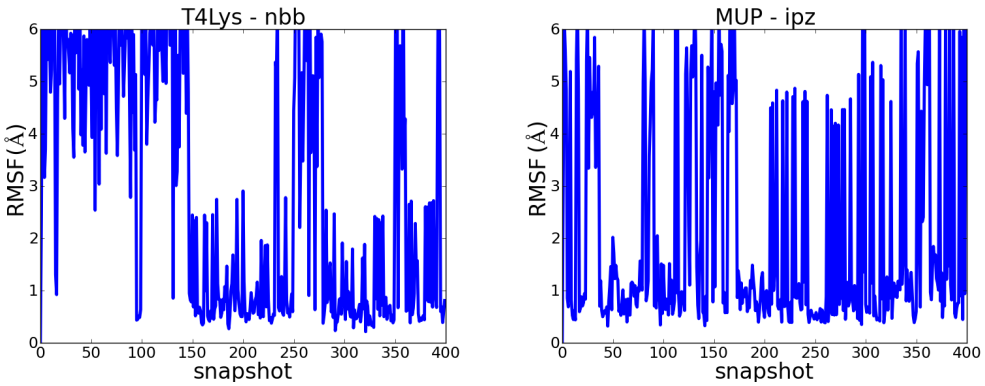


(b)

Figure 7.4: Evolution with simulation snapshot of θ (a) and RMSD with respect to crystal pose (b) of ipz in repeat 3 of its JAFS pose runs. Observe how the sampling of low θ corresponds with moments of high variability in RMSD.

correlation of the RMSF plot and the sampling of θ in the case of ipz can be related to further space available for the ligands to sample in the MUP cavity, a concept which is further explained in section 7.1.4. Indeed, there is enough space available so that both ligands in the JAFS score calculation on the MUP system can present high θ values at the same time (see the discussion of the MUP system in section 7.1.4), which is not the case in other systems such as T4Lys. In this situation, the MUP ligands (ipz and prz) are placed each at one extreme of the JAFS box, and they interact with each other. When studying the configurations of ipz during the period where high values of θ are observed (approximately between snapshots 170 and 350), swaps between different configurations of both ligands, which allow them

to remain at opposite edges of the JAFS box are observed. Hence, for this repeat of ipz, the stable RMSD values at high θ account for the sampling of the specific regions of space, those at the edges of the JAFS box, that allow for both ligands to present high values of θ at once.



(a) (b)

Figure 7.5: Evolution with simulation snapshot of RMSD with respect to previous snapshot (RMSF) of the nbb ligand (a) of the T4Lys system and the ipz ligand (b) of the MUP system. Repeat 2 and repeat 3 respectively of their JAFS pose runs.

In conclusion, a general trend correlating higher variability in RMSD with lower values of θ is observed. These observations support the theoretical understanding of the JAFS methodology and the sampling of θ and Cartesian space as explained at the beginning of this section. These are expected but encouraging findings since without this correlated sampling of θ and Cartesian space no increase in the configurational sampling of the ligands within the binding cavity would be expected with respect to well stablished simulation methods (standard MD or MC). Without this increase in sampling, the exact binding configuration of each of the ligands would be required prior to performing the JAFS score simulations, defeating one of its main objectives, namely the study of the relative binding affinity of ligands without previous knowledge of their exact binding configuration, taking into account all possible relevant binding configurations.

While this general trend is found we have seen that the traits of specific systems (namely the greater amount of sampling space in the MUP cavity) have placed the expected correlation between Cartesian and θ sampling in danger. While this will affect the scores obtained for the ipz and prz ligands, the simulation time where

this sampling correlation is not observed is characterized by high values of θ for both ligands at once. Hence the relative score is still expected to be higher for the ligand which dominates during the remaining period of the simulation. The problem associated with bigger sampling space for ligands in JAFS score must however be noted and it will be further discussed in section 7.1.4.

7.1.4 Discussion on individual systems

The Go-to Test System: T4Lys

As explained above (see chapter 5), T4-Lyzozyme has been used in a number of studies for its convenience and simplicity in the benchmarking and testing of computational methods, as well as the study of protein structure and structural changes. It is a logical choice as an initial test system for a new method since the system is well understood.

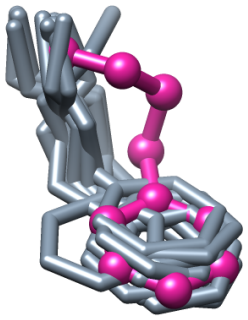
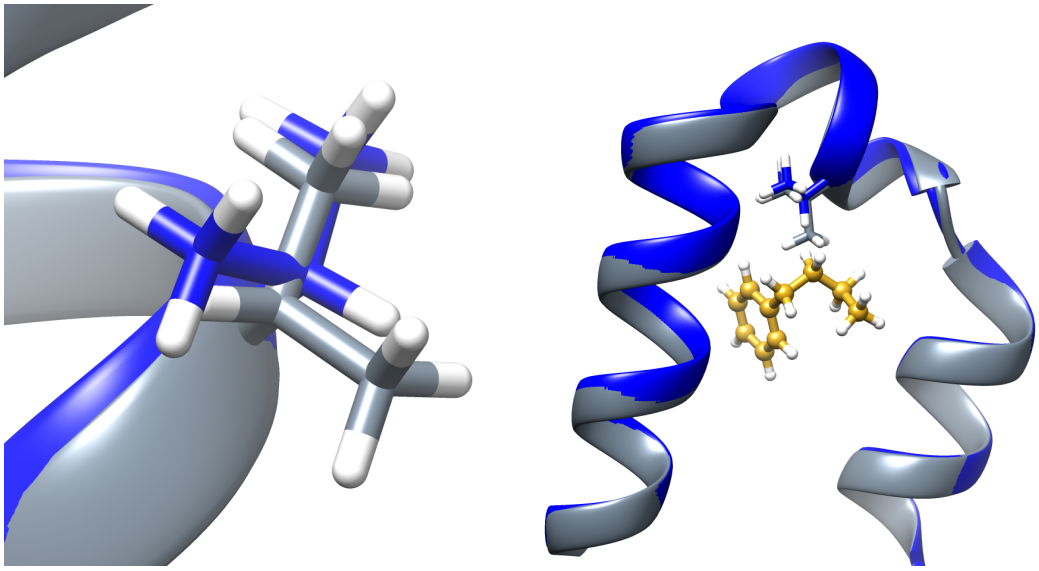


Figure 7.6: The crystal structure of nbb (magenta) and clusters obtained with JAFS score (grey). Hydrogens hidden for clarity.

In particular in the context of JAFS score, the small, occluded cavity avoids potential complications related to the different interactions of each of the ligands with solvent water molecules, whose disposition might otherwise have an effect. In sufficiently big and well hydrated cavities ligands might require different water network configurations to explore their correct binding geometry. This problem does not appear in small occluded cavities without water networks. Additionally, small cavities limit the available space for the ligands to sample when presenting high θ, with a "barrier" of potential Lennard-Jones clashes with the protein. The limited sampling and lack of solvent interactions should make this a simple system for JAFS score.

However, there were two potential issues which may still appear. The first is related to the sampling of the hydrocarbon chain of nbb — the tightest binder being also the one which might require more sampling to find the correct binding conformation. The second relates to the presence of Cl atoms in ligand dcb, whose interactions tend to be less accurately captured by standard molecular mechanics

fixed-charge force fields.[240] Of particular relevance is the anisotropic distribution of the charge around a halogen atom (such as Chlorine) when bound to an aromatic ring (as it is the case in dcb). The anisotropic distribution of charge allows for less electron-filled regions of the Cl atom bound to an aromatic ring to interact with electron-rich atoms (such as the oxygen in water). This trait cannot be captured by fixed-charge force fields since their charge distribution is homogeneous around each atom.[241]



(a) (b)

Figure 7.7: Key difference in the protein conformation within the binding site of T4Lys. The protein taken from the bound structure to bnz (PDB code 181L) is shown in grey and that bound to nbb (PBD code 186L) in dark blue. The crystallographic binding mode of nbb can be seen in yellow. In (b), the effect of the change in side chain conformation of the valine on the proximity to nbb can be appreciated — a Lennard-Jones clash is to be expected between the sidechain displayed in grey and nbb.

The ranking results obtained for the five presented ligands of T4Lys (table 7.1) are in general satisfactory, as the top binder could successfully be detected as could the decoy. Furthermore, rather than an incorrect ranking, the intermediate affinity ligands could just not be separated with JAFS score (their scores were within error of each other). Equally, for the location of the crystal binding pose (table 7.4), it was found for almost all repeats for all binders (note that wa1 is a decoy). However,

Clustering T4Lys

Fragment	nbb	dcb	bnz	1mp	wa1
Clusters per run	2.2	1.4	5.8	4.2	0
Runs with crystal pose	4/5	5/5	5/5	5/5	

Table 7.4: Average number of clusters per repeat and the proportion of runs presenting a cluster within 2 Å of their crystal pose for all T4lys ligands. Note that wa1 is a decoy.

a few issues do require further attention, including:

- Lack of correspondence between experimental binding affinity and scores obtained with JAFS
- Reasons for the slight lack of consistency in finding the crystal pose for nbb

When looking at the experimental binding affinities in table 7.1 a small difference in affinity between the top binder (nbb) and the second binder (dcb) can be observed, while the experimental ΔG for the next fragment (bnz), is considerably lower. This trend does not correspond to what is obtained with JAFS score, where the top scorer (nbb) is clearly picked among the rest, but the scores for all intermediate binders (dcb, bnz and 1mp) are identical to within error.

Alchemical free energy transformations in T4Lys

Transformation	Protein structure	ΔG_{calc} (kcal/mol)	ΔG_{exp} (kcal/mol)
$1mp \rightarrow nbb$	186L	-3.38 ± 0.41	-2.3^*
$1mp \rightarrow nbb$	181L	-1.15 ± 0.28	-2.3^*
$1mp \rightarrow dcb$	181L	-1.25 ± 0.08	-2.0
$bnz \rightarrow wa1$	181L	10.49 ± 0.93	-

Table 7.5: Results of dual topology alchemical relative free energy simulations (ΔG_{calc}) between some of the T4Lys ligands shown together with the experimental relative binding free energies (ΔG_{exp}) taken as the difference between affinities shown in table 7.2. Note the experimental binding affinity of $1mp \rightarrow nbb$ has been marked (*), since the experiment will not distinguish between different protein conformations but is expected to sample the different side chain conformations that are the main definitory trait of 186L and 181L (see section 7.1.4).

The question that then arises is whether this problem is associated with the JAFS score protocol, or whether the affinity predicted by the force field used would correspond correctly with the experimental affinities. To study the influence of the

force field in this system, the relative binding affinities of dcb and nbb with respect to 1mp were calculated with dual topology alchemical simulations. The results can be seen in table 7.5, together with the relative binding affinity between bnz and wa1. The two different results for nbb are discussed below. If we pay attention to the affinity of dcb compared to that of 1mp, the consequent conclusion is that our force field seems to predict dcb binding less tightly (or 1mp more tightly) than experimental results suggest.

The two different results for the affinity of nbb hint towards the issue which will be discussed now, in relation to the lack of consistency in finding the crystal pose for nbb. First, let us look at figure 7.6. Here the difference in side chain conformation between all clusters of nbb and its crystallographic binding pose is obvious. While nbb is still successfully ranked as top binder by JAFS, some further study into the reasons behind this conformational difference was undertaken. It is important to remember here that for the JAFS score simulations, only one protein structure can be chosen. The affinity of different ligands is then effectively measured against one protein conformation. While full flexibility is assigned to the protein (the protein is free to sample both backbone and sidechain of its residues), limited changes in conformation are expected for our implementation of Monte Carlo protein moves, within our simulation timescale. While the particular implementation will vary the degree of sampling, the lack of conformational sampling is a general issue in computer simulations of biomolecules.

In figure 7.7, a section of two different protein conformations, obtained from the complex with two different ligands used in our JAFS calculations (bnz and nbb) are shown. The protein conformation taken from the complex with bnz (PDB code 181L) was used as our T4Lys protein in the JAFS score simulations. As a result of a different conformation of a Valine sidechain, the hydrocarbon chain of nbb does not have space to be located with the bnz T4lys conformation as it would in its native complex. A slightly different bend to the hydrocarbon tail is then adopted. While this does not seem to affect the affinity sufficiently to change the raking of the different ligands used, it logically changes the clusters found for nbb. In particular, when looking at the RMSD from clusters to crystal pose, while 4 out of 5 repeats present a cluster that falls within 2 Å of the crystal, on the fifth repeat a cluster is observed with RMSD = 2.1 Å. This can probably be considered

a "correct" cluster, taking into account that the exact tail conformation cannot be expected.

The results of running JAFS score on the T4Lys can hence be considered successful, due both to the ranking obtained for the different fragments by affinity and the location of their crystal binding geometries. The obtained scores of binding affinities for each fragment are consistent with the calculated relative free energies of binding with well established (alchemical transformations, see section 3.6.2) methods, when the errors in the JAFS scores are taken into account. However, lower capacity of discerning between binders with similar affinities is found with the JAFS method. It may be concluded that the force field does not perfectly capture the affinities of the different compounds towards the binding cavity, since the results of the well established simulation method to calculate relative binding affinities do not perfectly correlate with the relative binding affinities as obtained from experimental data (see table 7.5). Besides, as discussed previously in this section, simulations do not seem to provide the system with sufficient flexibility so that the different relevant configurations of the protein side-chains in the binding pocket are sampled.

A dry cavity: MUP-1

The most relevant facts on MUP-1 and its ligands which affect our studies have been presented above (see section 5.1.2). In particular, for the study of the affinity of its ligands, it was chosen for its simplicity. Two different reasons should make this a simple system to which JAFS score can be applied:

- Only two ligands must be ranked
- The lack of waters in the cavity

While more than two ligands were available, other ligands presented waters key to their binding geometry, as observed in their crystal structure.[242] As explained in the introduction of this chapter (for the reason why the JAFS score protocol has not been run on the HSP90 system) the presence of key waters increases the difficulty of performing a JAFS score simulation, and requires the use of prior knowledge on the system, which goes against the black-box approach applied throughout the

project. For this reason, only the available ligands with no key waters in their binding configuration were used in the MUP system.

The reason why a dry cavity is expected to be a simple choice of system for JAFS score is related to different hydration patterns for different ligands. In a highly hydrated cavity, where waters are bound to the protein at the same time as the ligands, the particular disposition of these waters in the bound configuration of each of these ligands will generally affect their affinity. In JAFS pose, the whole of the binding cavity is considered, and the position of waters for each potential binding configuration of the ligand is expected to be captured by the waters included as JAFS particles. However, in JAFS score, only a small region of the binding cavity is considered, and the rest of the cavity may be hydrated by solvent water molecules. However, we know that these solvent waters, treated as standard Monte Carlo solvent particles, are not likely to change disposition easily within the binding cavity. Hence the different hydration patterns for different ligands are generally not expected to be taken into account during a JAFS score calculation. Other approaches, such as GCTI, which will be discussed in section 8.2.2 as applied to pose re-scoring, are being developed which could be used in this context. Choosing a dry cavity alleviates this problem since all ligands present in their bound configuration have the same hydration pattern — none.

The other reason for the simplicity of the system is obvious in the presence of only two ligands in the simulation. A random number generation for the scores would have a 50% chance of getting the ranking correct (if we forget the associated standard errors). Further, within the JAFS score protocol, one of the potential risks is the lack of sufficient sampling in θ. Imagine, for example, a situation where one of the ligands presents a high affinity towards the target protein, binding significantly tighter than the rest of ligands. This tightest ligand is then expected to remain at high θ values for long periods of time during the simulation, potentially leaving little time for the rest of the fragment to sample their high θ values. The ranking of all but the tightest ligand might not be meaningful (likely incorrect, inconsistent between repeats, or presenting big associated errors), due to lack of sampling of both states (above and below θ threshold) which are measured to obtain the score.

The results of ranking for MUP are satisfactory, with correct assessment of which is the tightest and weakest binder, as well as presenting clearly different

Clustering MUP-1

Fragment	ipz	prz
Clusters per run	3.6	5.2
Runs with crystal pose	2/5	3/5

Table 7.6: Average number of clusters per repeat and proportion of runs presenting a cluster within 2 Å of their crystal pose for all MUP-1 ligands.

scores, separated by several standard errors (see table 7.2).

However, as can be seen in table 7.6, the success in finding the crystal binding pose is not as optimal as might be expected, taking into account the size of the JAFS box. The exact scores given for each of these ligands (table 7.2) can hint towards one of the potential reasons.

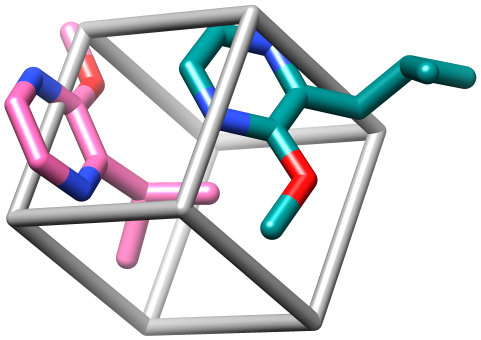


Figure 7.8: Configurations of ligands ipz (pink) and prz (blue) in snapshot 300 of the third repeat of JAFS score run on MUP. In this snapshot both ligands present a $\theta > 0.9$. The JAFS box is displayed in grey.

It is important to remember that the scores are obtained as a measure of the proportion of time spent by each ligand with their θ above a certain threshold (in this thesis the chosen threshold was always $\theta > 0.5$, see section 6.4.1 for a further study). It can be seen that both ligands present a score above 0.5, meaning both of them spent most of the simulation time at high θ values. How is this possible, and why would this correlate with the inaccuracy in finding their crystal binding pose?

Precisely because MUP presents a dry and big binding cavity, all around the small JAFS box is mostly empty, not presenting any potential Lennard-Jones clashes, were any of the parts of the JAFS ligands to end up there. It is important to remember at this point that the JAFS box constraints only the centre of geometry of the JAFS particles to remain within the box; other chemical groups or atoms may very well end up outside the box. We could then effectively say that the sampling space available for the MUP ligands is larger than that of a fully hy-

drated or smaller cavity, where besides the constraint to their centre of geometry, ligands are expected to experience limits in their Cartesian sampling due to the presence of other molecules right outside the JAFS box.

It is thanks to this extra space available that both ligands can be present at once in the simulation with high θ values. For both of them (their geometric centres, at least) to fit in the box, they are expected to end up in opposite corners of the box. Whatever their exact positions, both being at high θ values, their configurations will be conditioned by one another, they will be interacting, moving the system away from the conditions in which its is favourable for them to adopt their crystal binding pose. An example of the configuration of one snapshot with both ligands at high values of θ is shown in figure 7.8. This figure corresponds to snapshot 300 of the third repeat, and for both ligands, in that snapshot, $\theta > 0.9$. Notice both ligands are located at opposite extremes of the JAFS box.

The Most Demanding: CDK2

CDK2 was known to be the most complex of the systems simulated with JAFS score. However, results were expected to be better than those in table 7.2. Further studies were undertaken to try to understand the reason for this lack of success. Since it is especially relevant to study a system where the results may not be as good as expected, a full section is presented below with the analysis of results and conclusions on this system.

7.1.5 Studying a Problematic System: CDK2

From the systems studied with the JAFS score protocol, CDK2 was the most demanding. Nine fragments had to be ranked at once, four binders to target, and five decoys (non-binders). Furthermore, it was suggested that one of the key interactions in binding for some of the fragments might not be well captured by our force field (namely, the hydrogen from the ligand aromatic ring interacting with a protein backbone oxygen as found in ligand wcc and vtm — see figure 7.9).[239] This interaction corresponds to a non-standard hydrogen bond between a hydrogen bound to an aromatic carbon and the oxygen of the peptide-bond amide. It should be captured by the force field as a favourable electrostatic interaction. However, the charge assigned to hydrogens bound to aromatic carbons is very

small, in particular, an order of magnitude smaller than the charge assigned to hydrogens bound to amide nitrogens. This small charge will make the favourable electrostatic interaction between the negatively charged peptide oxygen and the slightly positively charged hydrogen very weak. Furthermore, hydrogens commonly involved in hydrogen bonds (such as those in hydroxyl groups) do not present Lennard-Jones parameters, hence avoiding any repulsion due to close proximity when forming hydrogen bonds. This is not the case for the hydrogen bound to aromatic rings, where Lennard-Jones will appear if unbound atoms are placed in close proximity. As a consequence of these charge and Lennard-Jones parameters, our force field will not be able to capture the key interaction between the wcc and vth ligands and the protein backbone. This is expected to be a common problem for fixed charge force fields and better treated with a more complex representation of charge-charge interactions representations[243] (such as that of polarizable force fields).

By looking at table 7.2 it becomes obvious that correct ranking was not obtained for the studied fragments in the CDK2 cavity. While distinction between affinities of binders can be difficult to capture, the protocol was expected to successfully discard the decoys and correctly predict the top binder (situation observed for T4Lys, see table 7.2).

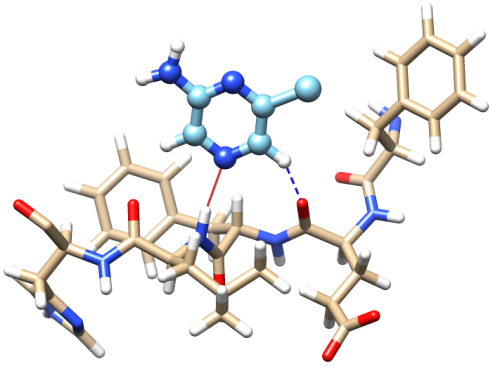


Figure 7.9: Fragment wcc in its crystallographic binding mode. Fragment is displayed in cyan and protein residues in light brown. The hydrogen bond between protein and ligand can be seen in red. Interaction of interest — hard to capture by our force field — displayed as dashed blue line.

Not only could the fragments not be ranked correctly, but the tightest binder presents the lowest score, with one of the decoys being awarded the top score. As expected, many of the scores are identical to within error. Uncertainties (standard errors of the mean score for all repeats) for many of the scores are clearly greater than those for

other systems (see table 7.2). Nevertheless, uncertainties are not sufficiently large for best and worse scores to be identical within error.

Given this sub-optimal results, further study and analysis of the simulation was performed.

Further Repeats

System: CDK2 – Set: 2

Fragment	Score	JAFS rank	Exp rank	Exp ΔG (IC_{50})
vth	0.00 ± 0.00	9	1	$120\mu M$
vta	0.47 ± 0.11	2	2	$185\mu M$
wcc	0.21 ± 0.06	5	3	$> 1mM$
vtm	0.18 ± 0.08	6	4	$> 1mM$
cd1	0.51 ± 0.05	1	9	—
cd5	0.23 ± 0.09	4	9	—
cd6	0.26 ± 0.07	3	9	—
cd8	0.09 ± 0.05	7	9	—
cd9	0.00 ± 0.00	8	9	—

Table 7.7: Results of a second set of JAFS score runs for CDK2 compared to experimental binding affinities.[239] The score obtained with JAFS, rankings for both JAFS protocol and with experiment, as well as the experimental binding affinity are shown. Note that the IC_{50} shown for wcc and vtm correspond to experimental 64%I at 1mM and 54%I at 1mM, respectively.

First, another set of 5 JAFS score repeats, with different starting structures and random seeds from the initial set was performed. The aim of these extra runs was to study consistency. While the big standard errors associated with the scores in this system indicated higher variability between repeats, a further set of repeats could suggest whether the ranking of fragments appears to be a systematic failure due to the protocol (or force field) or a sampling problem. Given that the new set of repeats start from a different set of configurations, consistency between the sets would hint towards a systematic issue with the protocol or force field, and not one of convergence.

The results obtained with the second set of repeats can be found in table 7.7. By looking at the table it can be suspected that some correlation between the rankings obtained by JAFS score on both sets of repeats is present. To analyse

whether there is a correlation, the Kendall tau that can be obtained comparing the ranking of both sets of repeats of JAFS score simulations was first calculated. This calculation was performed in *R*, using the `cor.test` functionality with the parameter `method=‘‘kendall’’`. The resulting Kendall tau was $\tau = 0.817$. Consequently, the significance of this Kendal tau was studied with the calculation of its associated p-value (actually produced by the same *R* command). The resulting $p - value = 0.00237$. It would be highly unlikely (lower than a significance level of 1%) to obtain two rankings as similar as these, if they were uncorrelated. These same results are displayed in table 7.8.

According to the results of our statistical analysis, we will proceed assuming that the ranking of both sets of runs are correlated; that is, there is a systematic issue with protocol or force field.

Convergence test

τ	p value
0.817	0.00237

Table 7.8: Results of comparing rankings produced by JAFS score on CDK2 in two difference sets five of repeats. Kendall tau and p-value shown. Original ranking shown in tables 7.2 and 7.7.

Studying Clusters

We will now proceed to study the configurations sampled by the binders and decoys of the CDK2 system while their θ was high during the JAFS score simulations. If a difference is observed in sampling between the ligands which were top ranked and those bottom ranked with JAFS score, it might provide further information on the reasons for their incorrect (not comparable to experiment) ranking.

The binding modes (obtained as cluster representatives) for both binders and decoys were studied. In figure 7.10 the representative of the most populated cluster of cd1 (the top ranked decoy) from each repeat of the initial set of repeats is shown. Different binding configurations are observed.

The numbers of clusters obtained for each fragment are shown in table 7.10. A clear correlation can be found between the average number of clusters per repeat and the position in the ranking assigned by their score (table 7.2) — higher ranking is associated with greater number of clusters. This suggests the binding affinity of the decoys assigned by our JAFS score protocol is directly related to the entropy

each of the decoys retains at high θ values. An increased sampling space (higher entropy) available to a particular ligand at high values of θ, must be correlated with an increased number of clusters (which are obtained only while $\theta > 0.9$), which we found to be correlated with the score (estimate binding affinity) used for ligand ranking in JAFS score.

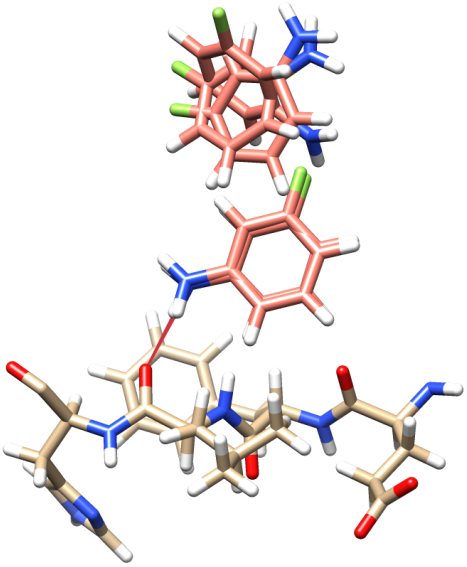


Figure 7.10: Most populated cluster of cd1 for each repeat of the first set of JAFS pose on CDK2 is shown. Fragment clusters are shown in salmon and protein backbone in light brown. A key hydrogen bond is displayed as a red line.

The total number of clusters, and the number of clusters locating the crystal pose is displayed in table 7.10. The crystallographic pose is found for all fragments except for that which presents no clusters (the ligand never presents values of $\theta > 0.5$). However, not all repeats find the crystallographic pose.

A ranked correlation between the number of clusters per run and their score can be studied — results shown in table 7.9. These data may provide some useful insight, but conclusions must be made carefully. Since the score of each fragment is the measure of the proportion of simulation time spent with $\theta > 0.5$ and the

Clustering CDK2 decoys

Fragment	cd1	cd5	cd6	cd8	cd9
Clusters per run	4.8	1.2	1.2	0.4	0.2

Clustering CDK2 binders

Fragment	vta	vth	vtm	wcc
Clusters per run	2.6	0.0	1.8	1.0
Runs with crystal pose	2/5	0/5	2/5	1/5

Table 7.10: Average number of clusters per repeat of CDK2 decoys and CDK2 binders for the initial set of five runs. The proportion of runs presenting a cluster within 2 Å of their crystal pose for each of the binders is also shown.

clusters are obtained from the configurations sampled by fragments when their $\theta > 0.9$, some correlation is necessarily expected. Clearly, the number of clusters must be zero for a particular repeat when the score is equally 0 (during this run, this fragment has not sampled $\theta > 0.5$). Nevertheless we will attempt to — carefully — gain some relevant insight from these data.

Correlating score and no. of clusters

τ	p value
0.800	0.00317

Table 7.9: Results of looking into the correlation between the number of clusters per repeat and the score for each fragment of the CDK2 JAFS score simulations (both binders and decoys). Data taken from the first set of runs. Scores can be found in table 7.2.

Both Kendal tau and p-value results presented in table 7.8, are considered consistent with those in table 7.9. Hence we will proceed assuming there is a correlation between the number of clusters per run for a particular fragment and their score. As explained above, this is not entirely surprising. However, such clear correlation indicates a trend. While it is true that a fragment must reach high θ values to increase both its score and number of clusters, we could think, as a simplification, of two different contexts in which a particular fragment might be increasing its θ at several points during a simulation.

In the first scenario, this hypothetical fragment would present one single high affinity binding pose. This being the case, the fragment is expected to sample the whole JAFS box when its θ is low, until it randomly finds a configuration close

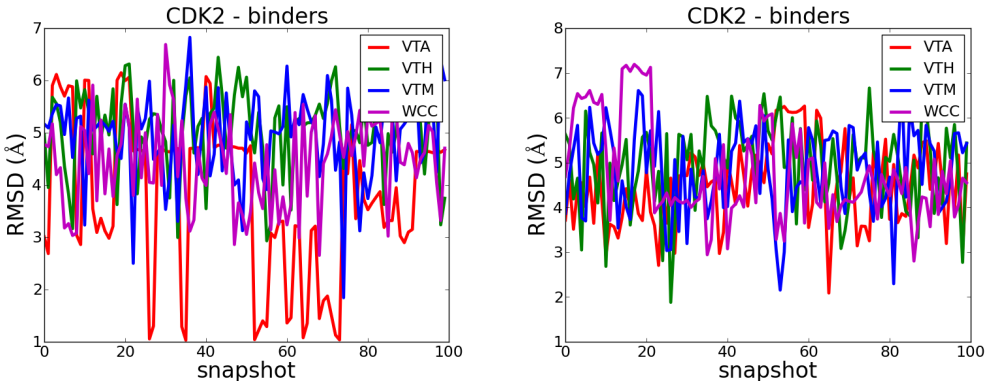
to its optimal binding pose. It would then be favourable for its θ to increase. It may stay in this condition for some time, and then out of chance, its θ would decrease and this process would start again, to sample the same highly favourable configuration every time its θ is high. Given that the affinity of this particular pose is high, this hypothetical fragment would have a high score (high proportion of the simulation time is spent at its optimal pose, with high θ), but a low number of clusters (since every time that it presents a high θ, is always sampling this single optimal configuration).

In a second, hypothetical scenario, one particular fragment may present a number of medium affinity binding poses within a JAFS box. This fragment is still expected to sample the whole box when its θ its low. However, with a given frequency (higher than that of the previous scenario), this fragment may find itself in one of its binding poses. It will then be favourable to increase its θ. The fragment is then expected to remain for some time at high θ at this particular pose, but we can expect this time to be less than that of the previous scenario due to the affinity of its pose being lower. At some point the θ will again decrease by chance and the sampling of the box with low θ should proceed again. However, by chance, we might expect this fragment to find now another one of its medium affinity poses. This process will then repeat, but with potentially different poses each time, and the fragment finding them faster, but potentially remaining in each for less time. We would then obtain both, a high score and high number of clusters for this fragment.

The correlation found in table 7.9 could indicate our CDK2 system for JAFS score is closer to the second scenario, where number of clusters and scores are highly correlated.

Studying Sampling

To further understand how this system behaves we will proceed to study the sampling for each fragment in individual repeats. While obtaining equivalent ranking results for two sets of repeats with different starting configurations could suggest the issue is not related to sampling, a further analysis is advisable. To study sampling, plots of RMSD with respect to the crystal structure have been generated for known binders.



(a) (b)

Figure 7.11: RMSD with respect to crystal binding structure for all binders to CDK2 during the last hundred snapshots of the first (a) and the second (b) repeats of JAFS score.

The RMSD sampled during the last hundred snapshots (10M moves) of the simulation for the binders to CDK2 studied for the first (a) and second (b) repeats can be found in figure 7.11. The last hundred snapshots were selected for clarity. As can be seen, none of the binders gets stuck in one particular configuration during any long period of the simulation. This is reflected in the frequently changing RMSDs.

The correlation between the RMSD sampling and θ sampling will be further studied for the first of repeats, and is shown in figure 7.12 for two of the binders (vta and vth). The two extreme examples in terms of θ values sampled during this period of simulation are shown, with vta sampling mostly high values of θ (plot (a)) and vth exclusively sampling very low values of θ (plot (b)). We can observe how the values of θ have an influence on the RMSD sampling, with more frequent changes in RMSD for the ligand which spends its time at very low values of θ (plot (d) is more variable than plot (c)). However, despite its high values of θ, vta is still often changing configuration (as reflected by the changes in RMSD seen in plot (c)). The satisfactory sampling observed in this study has reinforced our previous conclusions that the lack of (configurational or RMSD) sampling does not seem to be the issue that is causing the failure of the ranking obtained by JAFS score on the CDK2 system.

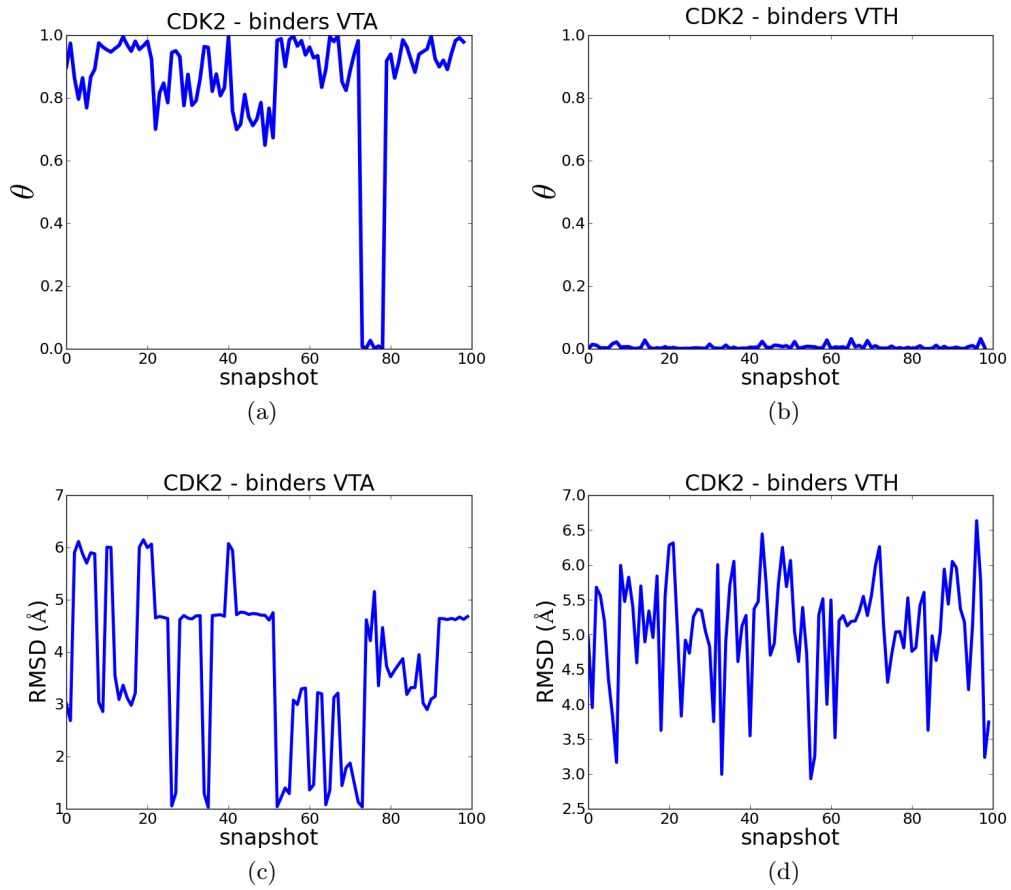
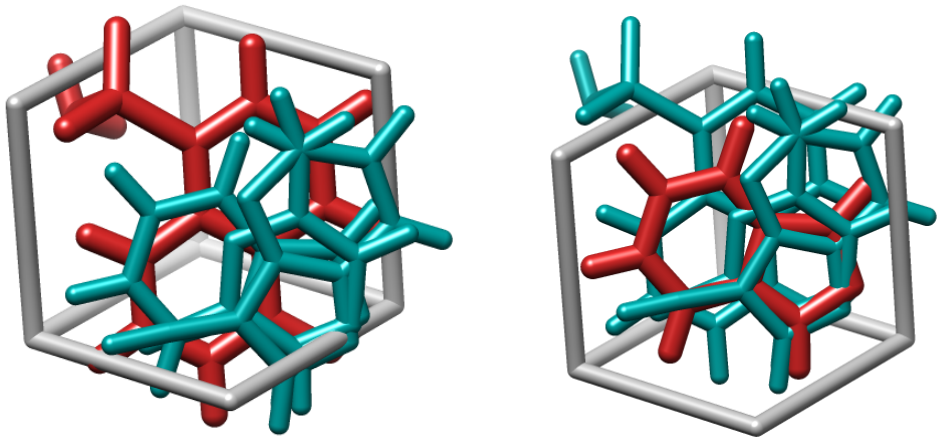


Figure 7.12: Sampling of θ ((a) and (b)) and RMSD with respect to crystal ((c) and (d)) of ligands vta ((a) and (c)) and vth ((b) and (d)) of the CDK2 system, with respect to JAFS score simulation snapshot. Only the last 100 snapshots (10M moves) are shown for clarity.

Alternative Setup

In the process of studying the reasons why a suboptimal ranking of ligands was obtained for the CDK2 system, the disposition of the crystal binding geometries within the JAFS box was studied more carefully. Note that one of the points of emphasis for the development of the JAFS methodology was the search for a black-box method. Following this approach, the setup for all systems was done consistently, without applying previous knowledge of the particular system. In particular, the dimensions of the JAFS box are identical for all JAFS score simulations (as well as for all JAFS pose runs), and its centre is determined by the geometry centre of



(a) (b)

Figure 7.13: JAFS score box setup for the CDK2 system. Box is shown in grey, the ligand which served as centering point for the box in dark red and the rest of ligands in dark cyan. The original box, centred around vth is shown in (a), while the alternative, centred around vta is shown in (b).

the ligand whose co-crystallized protein structure is used in the JAFS calculations for this particular system. The idea behind this choice is that, in a real system, we may know the binding mode for one particular ligand, but wish to study the relative affinity to another group of ligands for which we may well not even know whether they are binders or decoys. This same process was followed for CDK2, and hence the JAFS box was centred around ligand vth.

In figure 7.13a, the generated JAFS box for the JAFS score runs is shown in grey, with the crystal binding mode for vth shown in dark red and the binding modes for the rest of binders in dark cyan. It can be seen that, by centring the JAFS box around vth, the rest of the crystal binding modes end up a bit skewed to the corner. This is not a problem in terms of finding the exact binding geometry, since the centre of geometry of all fragments still lies within the JAFS box. However, we need to understand that, equally, the protein conformation used may not be identical to that in which each of these fragments binds, and as explained previously, the protein is not expected to sample much during our simulations. Taking this into account, it could be easy to conceive the possibility that some of the fragments may need to re-arrange their binding mode slightly to adapt to the slightly different

System: CDK2 – Alternative Setup

Fragment	Score	JAFS rank	Exp rank	Exp ΔG (IC_{50})
vth	0.00 ± 0.00	9	1	120 μM
vta	0.33 ± 0.15	2	2	185 μM
wcc	0.04 ± 0.03	6	3	> 1 mM
vtm	0.08 ± 0.07	5	4	> 1 mM
cd1	0.51 ± 0.19	1	9	—
cd5	0.21 ± 0.12	4	9	—
cd6	0.27 ± 0.17	3	9	—
cd8	0.00 ± 0.00	8	9	—
cd9	0.00 ± 0.00	7	9	—

Table 7.11: Results of JAFS score on CDK2 with the alternative positioning of the JAFS box compared to experimental binding affinities.[239] The score obtained with JAFS, rankings for both, JAFS protocol and with experiment, as well as the experimental binding affinities, are shown. Note that the IC_{50} shown for wcc and vtm correspond to experimental 64% I at 1 mM and 54% I at 1 mM, respectively.

conformation of the binding site. This being the case, it could well be that such re-arrangement is not possible within the limits of the provided JAFS box.

Correlating alternative setups

τ	p value
0.704	0.00878

Table 7.12: Results of looking into the correlation between the scores for JAFS score simulations on CDK2 with the original setup (see table 7.2) and the alternative box centred around vta.

To study whether this was the case, a new set of five simulation repeats was submitted with a different JAFS box, this time, centred around vta. This new box, with the crystal binding mode of vta in dark red, and the binding mode of other binders in dark cyan can be found in figure 7.13b. The results for this new set of simulations can be found in table 7.11.

Looking at table 7.11, it can be clearly appreciated that the correct ranking (as expected from experimental results) cannot be obtained with this setup either. Comparing these results with those in table 7.2, the similarities in ranking (and scores) are obvious. Nonetheless, a Kendall tau test can be performed to obtain the statistical significance of this similarity. This test is performed in an equivalent manner to that explained

previously. The results can be seen in table 7.12. Conclusions are again equivalent to those obtained for previous tests: a correlation between both rankings can be assumed. In this particular context, a small displacement of the JAFS box did not seem to improve the results of the JAFS score simulations.

Summary

The CDK2 system was originally described as the most complex of the systems studied with the JAFS score protocol. Its complications arose from the high number of ligands whose binding was studied at once as well as the presence of key interactions between some of the binders and the protein which are not expected to be captured with a fixed-charge force field (see above).

While the correct ranking of all provided ligands might be considered ambitious, selecting the group of binders as top ranked with JAFS score and the decoys (non-binders) as bottom ranked would have been considered a satisfactory result.

Throughout this result section we have seen that the ranking of CDK2 ligands by affinity did not have any correspondence with the experimental binding affinity data. Further more, several studies were performed to try to fully understand the reasons for this failure, with the associated possibility of a correction on the method or a good understanding of its limitations. In particular, as a first approach, the differentiation between a sampling problem and a systematic problem (whether related to the force field or protocol) was considered key. To discern the type of problem associated with this failure, first, another set of JAFS score repeats were run with a new set of starting conditions. The differences between the original and the new set of repeats would provide insight in whether we are facing a sampling problem (in that case both sets of repeats should provide different fragment rankings) or a systematic problem with the protocol or force field (in which case both sets of repeats would provide very similar results). Since the study of the ranking provided by both sets of repeats suggests that their differences are not statistically significant, a sampling problem was unlikely. To further test this possibility, a study of the sampling of Cartesian space (as measured by RMSD to crystal binding structure) of the ligands was performed. This study did not shown any of the binders getting stuck in any particular configuration and hence further supported the hypothesis that the issue associated with the failure of JAFS score

in this context is not related to sampling.

A study on the configurations sampled by the different ligands while their values of θ remained high was also performed. In this study, a significant correlation between the number of clusters (obtained from the configurations sampled by the ligands at $\theta > 0.9$) and the ranking obtained by JAFS score was found. Consequently, it was hypothesized that, effectively, the score obtained with JAFS score corresponded to the entropy that each ligand retained at high values of θ. No further studies could be envisioned to try to confirm this hypothesis or understand why this would be the case.

Another possibility was then tested, that the simulation setup might have an effect in the form of the exact location of the JAFS box, not leaving enough space for most binders to sample correctly their crystal binding configurations. Another set of simulations was then run, with the box centred around another of the CDK2 binders (vta rather than vth, which was the centre of the original JAFS box). This new set of simulations provided, however, again a ranking which was found to be no significantly different from that obtained with previous simulation setups.

In conclusion, after all studies performed, the reasons for the failure of JAFS score in this system are still not known. It could be concluded that the reason does not seem to be one related to sampling. Given the deficiencies in the force field representation of this system (see above, capturing non-standard hydrogen bonds), is likely that this failure is influenced by the use of a fix-charge force field.

7.2 JAFS pose

7.2.1 Specific Setup

All production runs of JAFS pose share the same simulation specifications:

- Simulations consisted of 5 million moves of equilibration and 40 million production moves
- Five repeats were run for each fragment
- The same protein structure was used for all simulations within each system
- The JAFS box was defined as a 12 Å x 12 Å x 12 Å simulation box around the centre of the ligand which was co-crystallized with the protein structure used
- Solute tempering was applied on the simulations, with 16 equally spaced replicas with solute temperatures between 25° C and 100° C
- The proportion of moves applied to all JAFS runs has been: solvent = 51%, protein = 9%, solute = 20% and θ moves = 20%

Protein	Fragment
HSP90	2dl
CDK2	vth
MUP	ipz

Table 7.13: Co-crystallized fragment with each of the protein structures used in JAFS pose production runs.

For each of the three systems presented here, the protein has been co-crystallized with one of the fragments taking part in JAFS pose simulations. This can be found in table 7.13. This could provide an advantage in finding the correct binding conformation for that particular fragment with respect to other fragments in the system, since the protein is allowed to sample throughout the simulation, but it is not expected to sample much within our setup of Monte Carlo simulations.

The number of copies of fragment and water molecules were always determined following the same method (specified in section 4.2.2 and 4.4.1), which aimed at

making consistently over-packed cavities for all simulations in terms of volume of particles per volume of the binding cavity. The resulting number of particles differed for each simulation (due to the difference size of each fragment) and is presented in table 7.14. The same number of copies of fragment and water molecules were always included.

System	Fragment	Copies	System	Fragment	Copies
HSP90	2dl	6	MUP	ipz	9
	ata	12		prz	7
	atb	9			
	atc	10	CDK2	vta	12
	atd	10		vth	7
	ate	11		vtm	10
	atf	9		wcc	12

Table 7.14: Number of copies of each JAFS particle included in the JAFS pose simulations for each fragment (the indicated number of copies of fragment and same number of water molecules).

7.2.2 General Discussion

The results of the JAFS pose simulations can be found in table 7.15. The results of two parallel clustering protocols (varying in clustering cut-off), on the same set of simulation results are shown. Independently of the clustering protocol used, for all ligands in all three systems, the crystallographic binding geometry is found at least once within the five repeats run per fragment. The number of appearances of the crystallographic pose decreases for a bigger cut-off. This could be expected, since a higher clustering cut-off correlates with a smaller total number of clusters (with fewer clusters there is a lower statistical probability of finding the crystal binding geometry). It is important to mention here that we consider the crystal geometry to be found if the RMSD between a particular cluster representative and the crystal pose is lower than 2 Å. 2 Å RMSD cut-off between the predicted and the known binding conformation is one of the common cut-offs used to assess correct prediction of the ligand conformation binding to target.[244–246]

These results prove that the JAFS pose protocol can locate the correct binding geometry for the presented systems. However, the success of the protocol cannot be assessed based exclusively on the possibility of finding this pose among those

System: CDK2

	clusters per run		runs with crystal pose	
ligands	2 Å cut-off	4 Å cut-off	2 Å cut-off	4 Å cut-off
vta	26.6	10.4	2/5	1/5
vth	12.2	8.4	2/5	2/5
vtm	22.4	10.4	5/5	4/5
wcc	24.2	9.6	3/5	1/5

System: MUP-1

	clusters per run		runs with crystal pose	
ligands	2 Å cut-off	4 Å cut-off	2 Å cut-off	4 Å cut-off
ipz	16.6	8.8	4/5	3/5
prz	13.2	6.4	4/5	4/5

System: HSP90

	clusters per run		runs with crystal pose	
ligands	2 Å cut-off	4 Å cut-off	2 Å cut-off	4 Å cut-off
2dl	10.0	7.0	5/5	5/5
ata	20.2	9.0	4/5	1/5
atb	16.8	10.0	3/5	3/5
atc	21.8	10.0	1/5	1/5
atd	17.6	9.4	3/5	2/5
ate	18.8	8.0	5/5	3/5
atf	14.2	9.0	3/5	2/5

Table 7.15: Clustering results with different cut-offs in hierarchical clustering. The same 5 repeats are analysed but with different clustering cut-offs. The number of clusters and the proportion of clusters in which the crystal pose was found (within 2 Å RMSD) is shown.

located, for each fragment. In this study, we have submitted five repeats per fragment, and in some of the examples (vta and wcc for the CDK2 system and ata and atc for the HSP90 system) in only one of the repeats the crystal pose is found when applying the more restrictive (higher) clustering threshold, while in one example (atc binding to HSP90) this is also true even in the less restrictive clustering threshold. Not only this, but currently, JAFS does not come associated with any protocol to rescore the generated poses (a protocol is currently under development in the group, see section 8). This means that the crystallographic pose cannot be picked among all generated poses. Consequently, a key factor to keep in mind when analysing the results displayed in this section, is the total number of poses (or clusters) generated with these runs. If we were to, hypothetically, generate such a number of poses that their conformations covered the whole Cartesian space of possible geometries for that fragment in the cavity, locating the crystallographic pose would have absolutely no value (that would have been generated just like any other individual possible conformation). Hence, the success of finding the correct pose must necessarily be dependent on the total number of poses found.

Different difficulties are expected for each of the systems studied. In particular, the different requirements for the treatment of water molecules in each of the cavities will be key to examine the correct behaviour of the JAFS pose protocol and its automatic treatment of water. MUP-I was considered the simplest system in JAFS score simulations. However, the fact that its binding cavity is dry is not necessarily an advantage in JAFS pose. In the JAFS pose protocol, a set of water molecules are included with the fragments in the JAFS box. The same setup was followed for all three systems (consistent with the "black-box" approach of the JAFS methodology). Hence the waters in the binding cavity must keep their θ low to produce a dry cavity, which is likely to be needed to find the correct binding pose in the MUP-1 system. This does not seem to be a problem with this protocol, since the correct binding pose is found for both fragments, for both clustering protocols, in more than half of the repeats.

In the HSP90 system, difficulties are also associated with water, but in this case, the presence of conserved crystallographic waters between all ligands which mediate most of the interactions between the ligands and the protein. In this case, for ligands to find their crystallographic binding pose, the correct position of

the bridging waters and the fragment needs to be found simultaneously. Again, this seems to be successful since the correct crystallographic pose for all ligands in this system is found. A more in depth study of the behaviour of waters in the binding cavity with JAFS can be found in section 7.2.5. In HSP90, differences in the number of repeats required to find the crystal pose are clear. In the case of the overwhelming consistency in finding the crystal pose for 2dl, it is important to consider the potential advantage for this ligand, since the protein structure used in our calculations was the 2dl co-crystal (PDB code 2XDL).

In the case of CDK2, we can also observe important differences in performance between different ligands, where the crystal pose of vtm is found in a much higher proportion of repeats (all of them for the lower clustering cut-off, four out of five for the more restrictive clustering) than the rest of the ligands. However, in this case, the same advantage found for 2dl in HSP90 does not apply, since the protein conformation used in these repeats is that co-crystallized with vth.

Symmetry in Clustering and RMSDs

Several of the studied fragments presented one or more points of symmetry. While our clustering algorithm does not take into account symmetry, nor does our calculation of RMSD, symmetry was accounted for manually. For ligands (binders) which presented points of symmetry (2dl and atb in HSP90 and nbb, 1mp, dcb and bnz in T4Lys), all possible symmetric conformations were generated, and RMSD was calculated to each of them, so that the lowest RMSD of any of the clusters to any of the symmetric target conformations was the one used. It is to be noted, however, that the number of clusters will be dependent on this symmetry not being accounted for. This may be appreciated both in the number of clusters presented in tables and visual representation of these clusters in figures.

7.2.3 Cluster Location

As can be seen in figure 7.14 for the example of ligand ata in the HSP90 system, clusters can be found throughout the whole cavity (this trend is observed for all systems studied). In fact, as can be appreciated, their added volume basically fills the whole of the binding cavity. However, while clusters occupy the whole volume, the density of clusters is not the same in all areas of the cavity.

In figure 7.15a the distribution of the density of clusters throughout the cluster cavity is studied (again for ligand ata in system HSP90). Three different lobes can be found, where the density of clusters is assessed as higher after visual analysis. These three lobes are marked by orange arrows in figure 7.15a. As is also shown in the same figure, the initial visual assessment is confirmed by clustering all clusters, "superclustering", with a high clustering cut-off (cut-off = 6 Å). The three superclusters generated are displayed in dark red in the aforementioned figure. They are considered as the representatives of these three lobes.

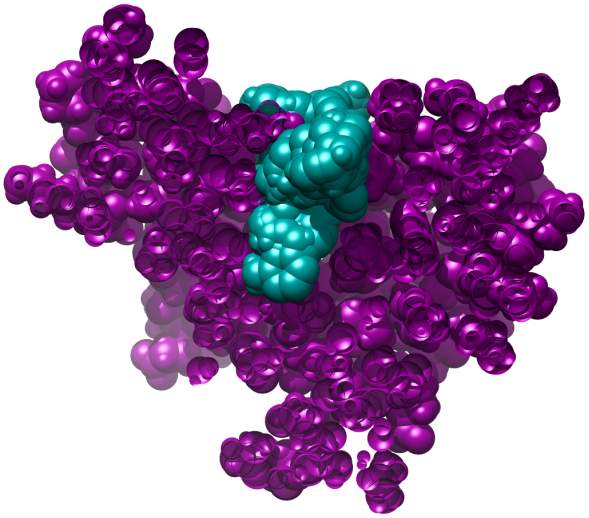
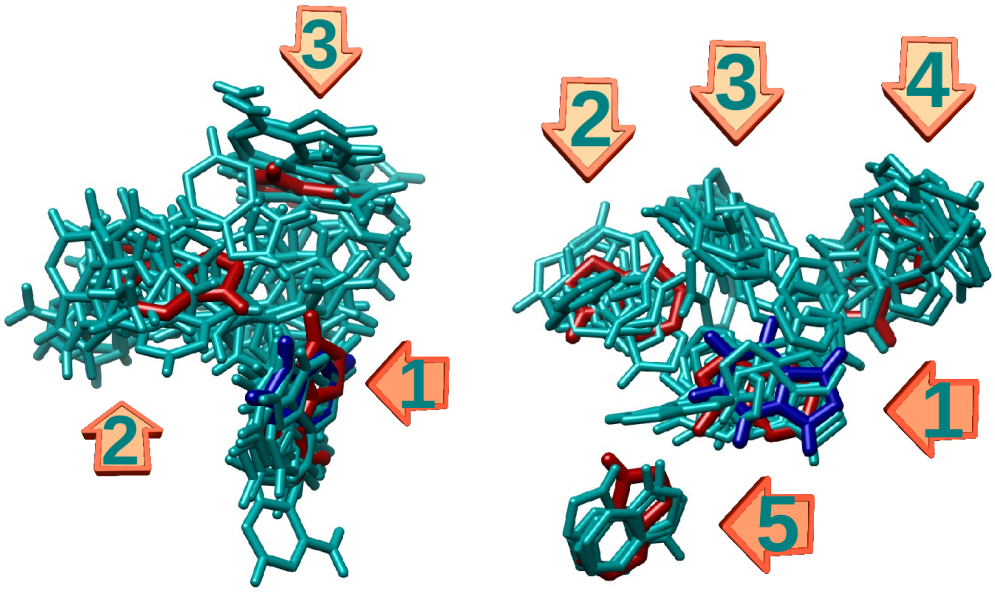


Figure 7.14: Clusters of ata for all five repeats of JAFS pose with target protein HSP90. Cluster representatives shown in dark cyan. Protein, clipped, shown in purple.

These three lobes are the particular representation for ligand ata of a trend observed in the JAFS pose simulations of systems HSP90 and CDK2 presented here (MUP cavity is occluded from solvent[247]). Namely, clusters are mostly observed in the same volume occupied by the crystallographic pose (not necessary in the same conformation), and in solvent exposed subsections of the cavity.

As has been mentioned, the same sort of disposition from that seen for HSP90 ligand ata is also found for the CDK2 system. CDK2 also presents a solvent exposed cavity, and equally, clusters can be observed in the solvent exposed region. An analysis of the disposition of clusters and cluster lobes for vta ligand in the CDK2 system can be seen in figure 7.15b. In that figure, the lobes in which all clusters can be distributed are numbered and highlighted with orange arrows. Lobe 1 in the image is that which corresponds to clusters in the same region of the cavity where the crystal binding pose is found (crystal pose in dark blue). Lobes 2 to 4 are three different solvent exposed lobes, where only two of them are

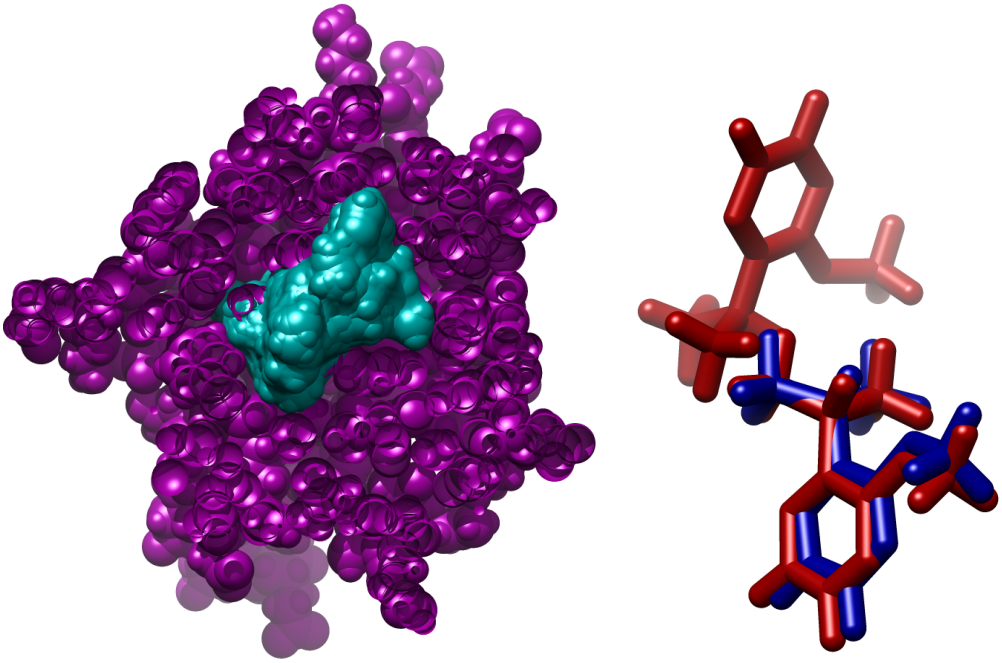
captured by "superclustering" (clustering with cut-off = 6 Å), where supercluster representatives are shown in dark red. Interestingly, another pose is found in a subcavity (5), which is much more hydrophobic and actually was not considered as part of the binding cavity during setup. However, due to the JAFS box disposition, copies of vta were capable of finding conformations where high θ could be displayed within that subcavity.



(a) (b)

Figure 7.15: Cluster representatives of ata (a) and vta (b) for all five repeats of JAFS pose (clustering threshold = 2 Å). All cluster representatives are shown in dark cyan. Crystal pose shown in dark blue. "Supercluster" (clusters of clusters, with 6 Å clustering threshold) representatives are shown in dark red.

The MUP system does necessarily present differences (i.e. no solvent-exposed clusters are found) as its cavity is occluded from solvent[247] (see figure 7.16, image a). However, the same process has been followed in clustering ipz clusters ("super-clustering") with a cut-off = 6 Å. The results are displayed in figure 7.16, image b. One of the superclusters in found in the same area as the crystal pose, while the other covers another region of the cavity, in a similar fashion as those in figure 7.15a. However, it is important to note that, in this case, one of the supercluster representatives perfectly overlays the crystal binding configuration.



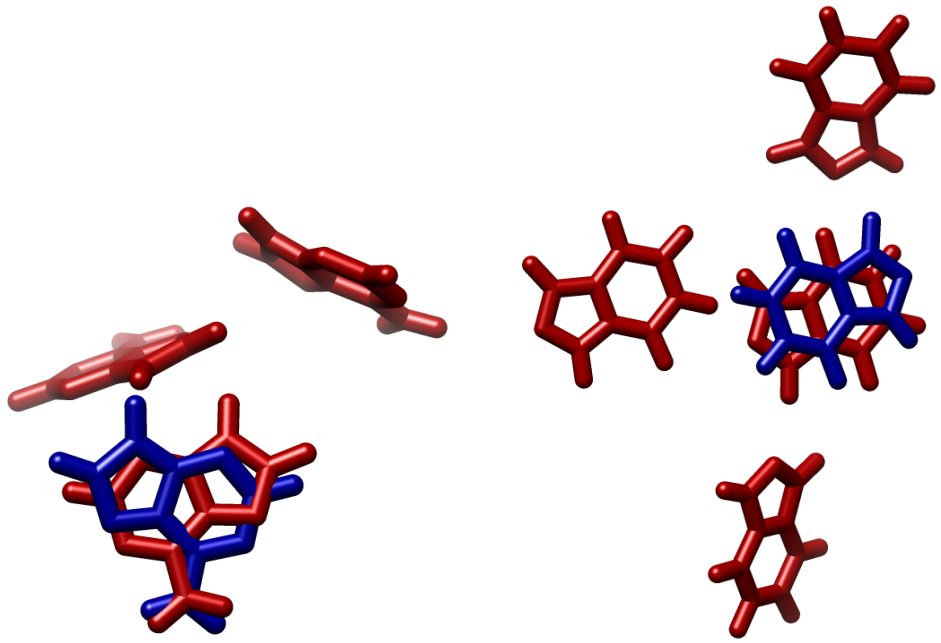
(a) (b)

Figure 7.16: (a) Cluster representatives of ipz for all five repeats of JAFS pose with target protein MUP. Cluster representatives are shown in dark cyan. Protein, clipped, shown in purple. (b) Supercluster representatives of ipz, the result of clustering all clusters from JAFS pose repeats, compared to the crystal binding geometry. Supercluster representatives shown in dark red, crystal pose in dark blue.

It may seem tempting based on the results seen for MUP to simply use "superclustering" to generate the potential binding poses. While for the other systems, one of the superclusters is also found in the same region of the cavity where the crystal pose is, it does not always overlay nicely with the crystal binding geometry (see figure 7.17).

7.2.4 Sampling

Although the sampling of both θ and Cartesian space are intrinsically related within any JAFS simulation (as explained in section 7.1.3), they will be studied separately below for clarity.



(a) (b)

Figure 7.17: Supercluster representatives of ata (a) and vta (b), the result of clustering all clusters from JAFS pose repeats with a 6 Å cut-off, compared to the crystal binding geometry. Supercluster representatives shown in dark red, crystal pose in dark blue.

Sampling of θ

There are a number of factors that influence the θ sampling of JAFS particles in a JAFS pose simulation, such as the number of particles and their size in relation to the size of the JAFS box; the presence of enhanced sampling methods such as parallel tempering or solute tempering; the strength of a particular set of interactions between a particle and its environment in a defined pose... While reasons for the (lack of) sampling and ways to improve it have been presented in sections 6.2 and 6.3.2, here we will focus on the results of sampling and associated conclusions or consequences. There are a few points that influence sampling and are important to remember here:

1. The setup has been designed in such a way so that the "crowdedness" of

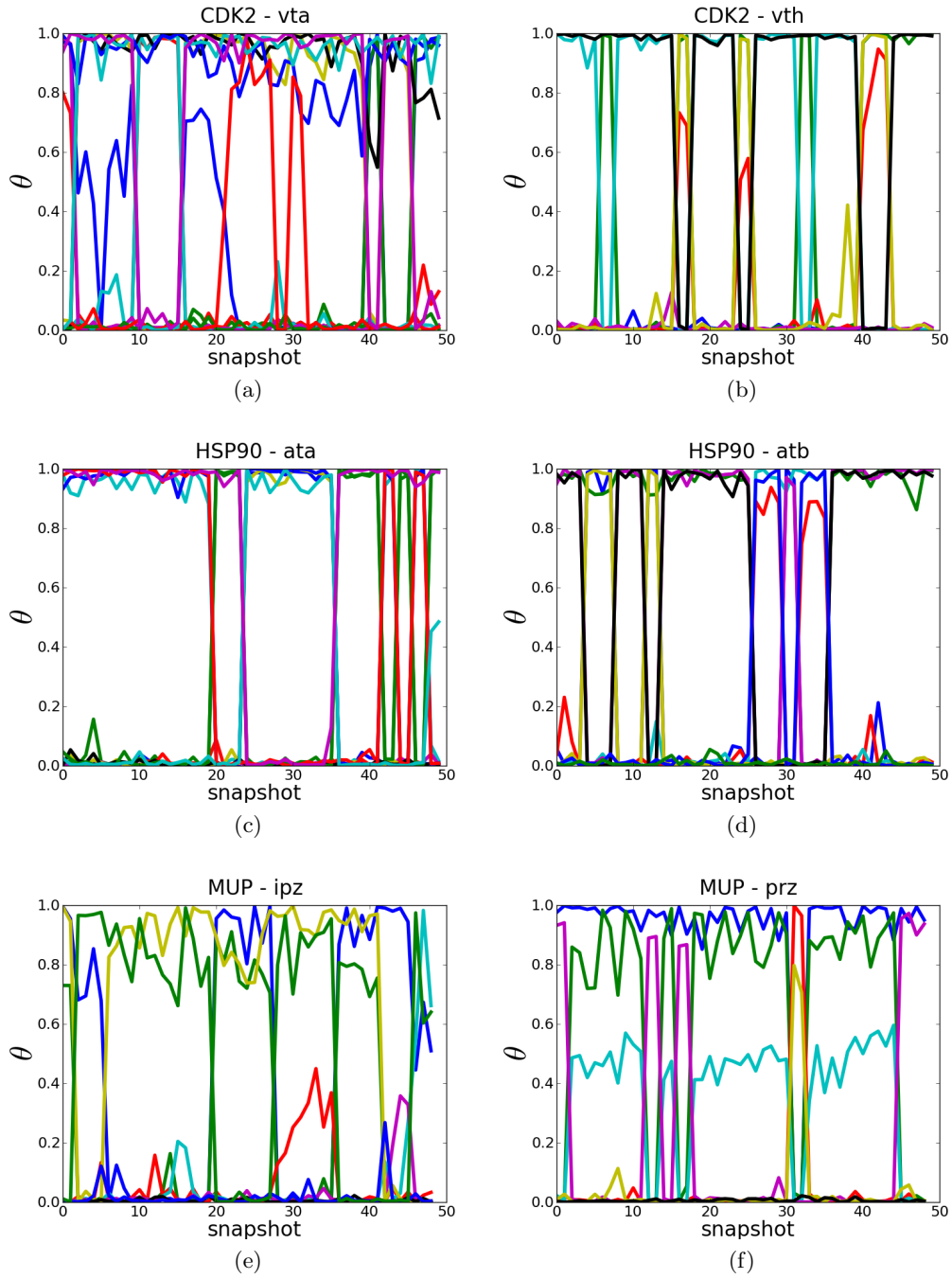


Figure 7.18: Evolution of the θ values of JAFS fragments with respect to snapshot at the end of the simulation (last 50 snapshots) of the first repeat of JAFS pose runs for fragments vta (a), vth (b) bound to CDK2, ata (c), atb (d) bound to HSP90 and ipz (e) and prz (f) bound to MUP. See chapter 5 for further information on the systems.

the cavity is approximately the same, in volume terms, for all ligands in all systems (see chapter 5 for further information on the systems and their ligands). This should hence have a minor effect on any potential differences in sampling between different systems.

2. Solute tempering is being applied on all JAFS production runs. This implies exchanges are expected between replicas at different "solute temperatures". This will influence θ sampling when studied for one particular replica (as we will do, studying the replica at lowest "solute temperature", with no added solute scaling by means of solute tempering). When a replica exchange event (between different solute tempering replicas) takes place, often a change in the θ value of the JAFS particles at one particular "solute temperature" is observed. This is due to different θ values being associated with the JAFS particles at each of the exchanged configurations.

In figure 7.18, the θ sampling of the copies of the corresponding fragment species (vta and vth for system CDK2 in plots a and b, ata and atb for system HSP90 in plots c and d, ipz and prz for system MUP in plots e and f) during the last 5 million moves (for clarity) of the first repeat of each simulation are shown. Each line in each of the plots represents a copy of the fragment in the JAFS box. Information shown is always that of the replica at the lowest "solute temperature" (with no energy scaling by solute tempering).

A few aspects must be highlighted from the observation of the plots. First, notice that most fragment copies are always present at either high ($\theta > 0.8$) or low ($\theta < 0.2$) values of theta, with generally at most one copy present at intermediate values of θ at a given time. This is relevant to the accurate representation of the realistic system with our JAFS simulations. Since in the real system, all particles will either be present or absent from the target binding cavity, and hence the existence of "partially on" or "partially present" particles (those with their interaction energy scaled down, intermediate values of θ) do not represent a real state, the existence of these intermediate particles must be minimized (see section 6.2.2). While the possibility of sampling these intermediate states of θ is essential to the whole idea of JAWS and JAFS, and key to the achieved sampling of both θ and Cartesian space (see explanation of background and ideas behind the method in section 4.2), it can be seen that, generally, particles do not remain at these intermediate θ and

instead "choose" whether to present high or low values of θ.

Another relevant result observed in figure 7.18 is the presence of exchanges between copies at high and low values of θ during the simulation. Sharp exchanges between very high and very low values of θ can be seen, for example at the end of plot c (ligand ata). These fast exchanges back and forth between the same two copies originate directly from exchanges between solute tempering replicas. In neighbouring replicas, each of the particles is respectively at high and low θ. Other, more incremental changes in θ can be observed, for example, in plots a and e (ligands vta of CDK2 system and ipz of MUP system). These incremental changes are generally the results of successive Monte Carlo moves in θ space of the fragment copy within one particular replica. The key observation, however, is that different θ are sampled for the fragment copies, hence giving them the chance to explore freely Cartesian space when their θ value is low, as well as finding favourable interactions, where presenting high θ values is favourable.

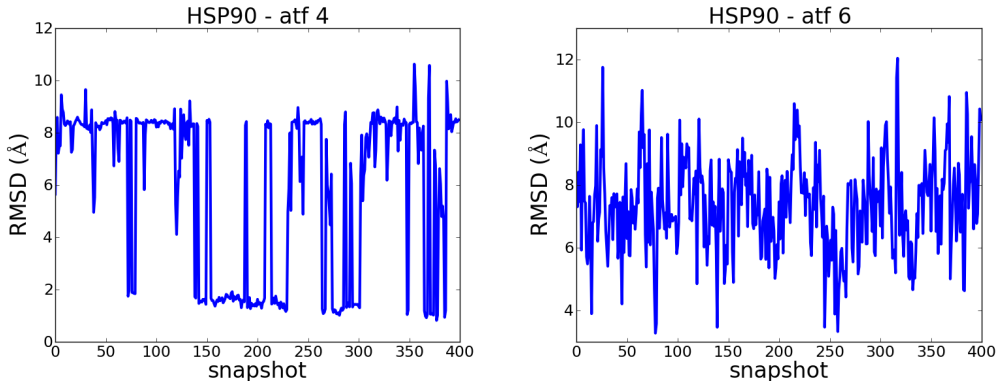
In conclusion, we observe satisfactory sampling of θ space for the fragments in the JAFS pose runs, with a dominance of states with either high or low θ rather than those of intermediate values, but also with exchanges between high and low values of θ. Success of the objectives in terms of θ sampling has hence been achieved.

Equivalent plots to those in figure 7.18 for the first repeat of all the remaining ligands used in the JAFS pose production runs can be found in section 11.3.

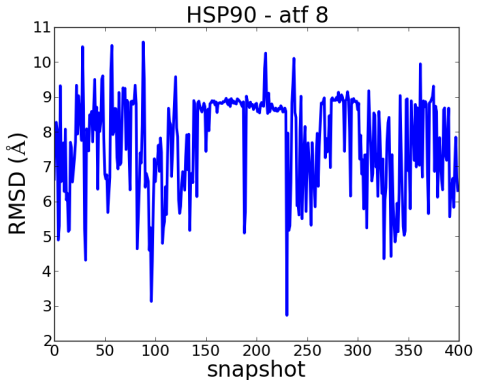
Sampling of Cartesian Space by RMSD

The RMSD with respect to the crystal structure for three copies of fragment atf on the third repeat of its JAFS pose simulation is plotted in figure 7.19. These three poses from one particular repeat have been chosen as being representative examples of the main different situations which can be found in terms of RMSD sampling. From those, copy 4 (plot a) reaches the crystal pose (see RMSD < 2 Å), while the other two copies of atf do not.

It is important to remember here the idea behind the increase of sampling that JAFS simulations are expected to provide. When a particular copy of a JAFS particle presents a low value of θ (its interaction energy close to zero), the particle is expected to sample most of the Cartesian space included in the JAFS box. When favourable interactions are found, it will be energetically favourable for this particle



(a) (b)

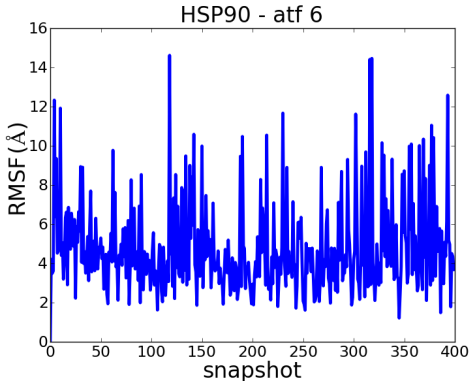


(c)

Figure 7.19: Evolution of RMSD with respect to crystal pose for three different copies of fragment atf (target protein HSP90) in its third repeat of JAFS pose.

to increase its value of θ, turning "on" its interaction energy. When its interaction energy is not scaled down, the Cartesian sampling capabilities of this particle are limited by the rest of molecules in the environment. It is hence likely that the particles get "stuck" in terms of Cartesian space while their θ value remains high, until a move that leads to the decrease of θ is accepted.

Based on that premise we can study the behaviour presented in the plots of figure 7.19. We will start by looking at plot b (copy 6 of atf) compared with plot c (copy 8 of atf). In plot b, the continuous change in RMSD with respect to crystal of copy 6, during the whole length (40 million moves) of the simulation is clear. No set of snapshots are observed during which RMSD remains constant. In other words, atf 6 seems to be exploring different regions of the JAFS box, without remaining in



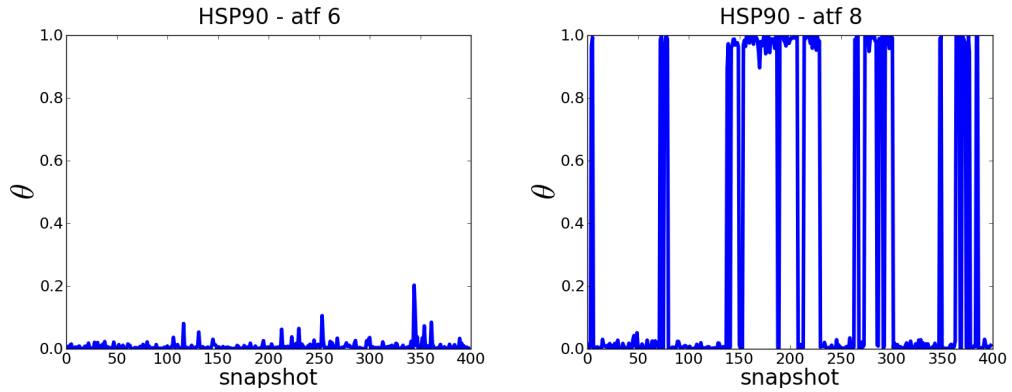
(a)

Figure 7.20: Evolution of RMSD with respect to previous snapshot (RMSF) for copy 6 of fragment atf (target protein HSP90) in its third repeat of JAFS pose.

one particular configuration for long. This can be confirmed by the plot of RMSD with respect to previous snapshot (RMSF) for this same repeat and copy of atf shown in figure 7.20. Plot c, which represents atf 8 is different in this respect. While, during most of the simulation time its RMSD is also varying constantly, during approximately 7 million moves (starting somewhere around snapshot 150) atf 8 seems to have a clear preference to keep its RMSD to crystal somewhere slightly below 9 Å. These observations would suggest copy 6 of atf remains mostly at low values of θ throughout the simulation length. However, we would expect copy 8 to present high values of θ roughly between snapshot 150 and 220 (between its 15 and 22 million moves). These suppositions are confirmed by looking at the θ sampling with respect to snapshot for each of these atf copies in figure 7.21. That pose sampled by copy 8, at RMSD slightly below 9 Å, is captured in one of the clusters (6th most populated cluster for both 2 Å and 4 Å cut-off) obtained as output of this simulation repeat.

After this first analysis of copies 6 and 8, it is probably easier to understand the behaviour of atf 4, displayed in plot a. It seems that copy 4 is sampling basically two different poses during the whole simulation. One of them with RMSD < 2 Å, considered to be the crystallographic pose. The other pose presents an approximate RMSD of 8 Å to crystal pose.

From these observations several conclusions can be made. First, we can see that the behaviour of the different copies of the fragment can be quite different in



(a) (b)

Figure 7.21: Evolution of RMSD with respect to crystal pose for three different copies of fragment atf (target protein HSP90) in its third repeat of JAFS pose.

terms of sampling of Cartesian space (here measured in RMSD). This may point to the relevance of including several copies of the fragment in each repeat of the simulation to increase sampling opportunities available to the fragment. Second, we can observe the two types of behaviour expected in terms of Cartesian sampling to be present in the simulation. We observe fragments being "stuck" in particular configurations, which can only be understood if the fragment copy is interacting with the environment. At these points θ values are high, and generally high enough to be captured as part of the configurations to be clustered in order to generate the possible binding geometries. We also observe copies of the fragment exploring large regions of the JAFS box in relatively small number of moves. The continuously changing RMSD correlate with rapidly changing configurations of the ligand. For configurations to change rapidly, the difference in energy between configurations must be low, so that a high proportion of the attempted moves are accepted, consequently moving the fragment. Most commonly, the reason for small energetic difference between different configurations is that the interaction energy of the fragment is scaled down (close to zero — low values of θ); differences in the environment as the configurations then change will hardly make any energetic difference to the fragment. That is, at low values of θ we observe frequent configurational changes while at high values of θ the particles explore defined binding configurations that are captured during the clustering process.

We can hence conclude that, while the difference between fragment copies might be more obvious than expected, both desired types of behaviour in terms of Cartesian sampling can be successfully found among fragments in JAFS pose simulations.

7.2.5 Waters in the Cavity

One of the main strengths of JAFS pose with respect to other methods of finding the binding configuration of ligands is the automatic inclusion of waters in the binding cavity. This implies no previous knowledge of the solvation state of the cavity is required, nor information on potentially conserved waters which may mediate interactions between the ligand and target protein. Both the presence or absence of waters in any particular configuration, or high variability of these, should be accounted for by the waters included in the simulation as JAFS particles. These waters can vary the scaling of their interaction energy (effectively "appear" or "disappear" from the cavity). They are equally free to sample their Cartesian space within the limits of the JAFS box. Their sampling of Cartesian space will be enhanced by their ability to sample their interaction potential, since they can, effectively, turn their interactions "off" to avoid energetic clashes.

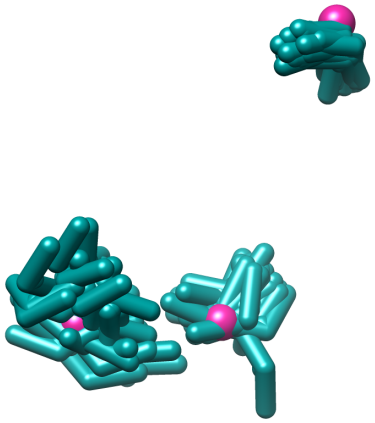


Figure 7.22: The three HSP90 crystallographic conserved waters. Waters taken from the crystal structure in complex with 2dl are shown in magenta. Cluster representatives of waters with $\theta > 0.9$ for all repeats of all HSP90 JAFS pose simulations are shown in dark cyan. Only water clusters with any of their atoms within 1.5 Å of the crystal waters are shown. All repeats bar one present one cluster within the set distance of each of the crystal waters.

But does this sampling actually take place in our simulations? Are we capable of locating crystallographic conserved waters? Is there a difference in our JAFS pose simulations between cavities which are known to be dry and those that have been seen to be highly hydrated? This has been studied and results will now be

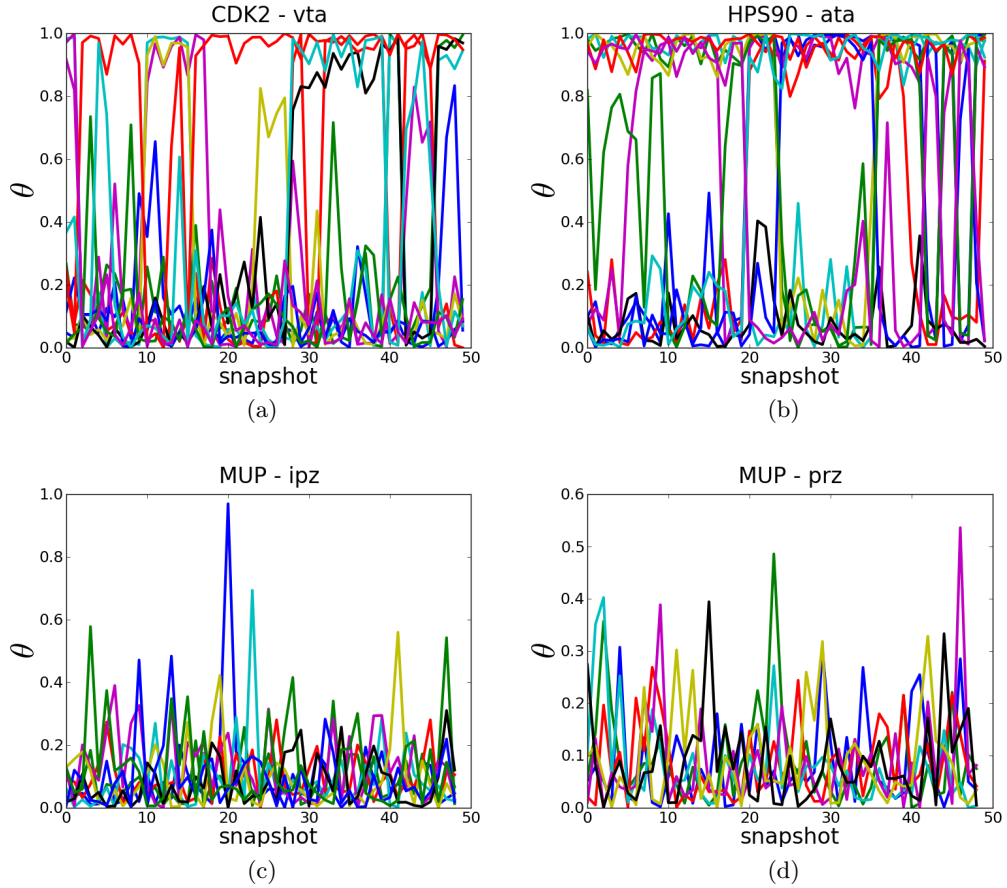


Figure 7.23: Evolution of the θ values of JAFS waters with respect to snapshot at the end of the simulation (last 50 snapshots or 50 million moves) of the first repeat of JAFS pose runs for fragments vta (a), ata (b), ipz (c) and prz (d). Only 50 snapshots of one repeat shown for clarity. The dryness of the MUP cavity (plot c, with fragment ipz and d with fragment prz) can be observed by the lack of waters at high values of θ in comparison with other cavities, in particular, the well hydrated HSP90 cavity (plot b, with fragment ata).

presented in this section.

In figure 7.22, the location of the conserved waters mediating binding interactions in HSP90 is studied (see section 5.2.1 for further information on the system). In the image, the waters from the crystal structure (magenta) — PDB code 2XDL, protein structure used in the HSP90 production runs — as well as the clusters (dark cyan) obtained for those waters in all repeats for all HSP90 ligands are shown. All molecules with any of their atoms within 1.5 Å of the crystal waters were selected. All but one of the total number of repeats found all three conserved waters within those limits. The remaining repeat failed to find one of three crystal bridging waters (W3 in figure 5.3) within the limits described. Two clusters were, however, found not far from crystal water 3 (W3), with the closest of them presenting the closest atoms at 1.9 Å, with its oxygen at 2.1 Å.

Observing these results we can confidently say that, for the HSP90 system, with the range of ligands studied, JAFS is successful in locating conserved crystallographic waters automatically, without any previous knowledge of their presence or location.

In figure 7.23 a study of the θ values of JAFS waters for the different cavities where JAFS pose has been run is shown. Plots are shown of the evolution of θ for the JAFS waters with snapshot for the last 5 million moves of the first simulation repeat (50 snapshots). Only a section of the simulation and only one repeat are shown for clarity (choosing the first repeat and last section of the simulation was arbitrary). Each line in each individual plot corresponds to one water molecule. The simulations with the two ligands in MUP are shown (plots c and d), while only one of the ligands is presented for each of the other cavities, for comparison. It can be seen how, for the case of MUP (plots c and d in figure 7.23), which is known to be a dry cavity (see section 5.1.2 for further information on the system), waters present low values of θ on average, and none of the water molecules remains at high θ. In comparison, the cavity of HSP90 (plot b in figure 7.23) which is known to be highly hydrated (see section 5.2.1) does present, in fact, several water molecules with $\theta > 0.8$ during all of the last 5 million moves of the simulation. The average θ per water molecule can be estimated as higher than that of MUP. The case of the CDK2 cavity, as presented in plot a (within figure 7.23), is an intermediate one, with at least one water continuously presents a value of $\theta > 0.8$, but the total

number of waters with high θ values is lower than that of HSP90 cavity (plot b). As an ilustrative example, measuring water molecules with $\theta > 0.9$ during the first simulation run, for the HSP90 runs with ata, an average of 4.5 "on" waters were found, while an average of 2.3 "on" waters per simulation snapshot were found in the first repeat of the CDK2 runs with vta.

By looking at figure 7.23 and the analysis presented above we can conclude that the dryness of the MUP cavity, in opposition to the abundant hydration of HSP90 is detected correctly. It is important to note that an analysis of the exact number of waters present on average in a hydrated cavity, such as HSP90, may well not provide the correct value when studied with JAFS. This is a consequence that JAFS waters do not present the same degrees of freedom as a real water molecule would, since it has the extra capability of scaling its interaction energy (appearing and disappearing). This will drive these JAFS waters to sample a different free energy landscape. However, clearly, for these systems, the approximation of reality that is our JAFS representation, is good enough to capture the difference between a well hydrated and a dry cavity, and it seems to set all the water-related conditions in place to be able to capture the binding geometries correctly.

7.2.6 Differences Between Clustering Thresholds

One of the observations taken from the data presented in table 7.15 was that the number of clusters per repeat was considered higher than optimal. Given that the JAFS pose protocol cannot select the "correct" binding pose among those generated, a posteriori re-scoring of the poses might be desired — depending on the particular application of the JAFS methodology. Problems of traditional free energy calculations for pose re-scoring, together with our use of a novel methodology, will be discussed in a later chapter (see chapter 8). However, it is important to mention here that standard free energy methods, as well as the one used in this project, deal with the difference in free energies of ligands (or poses) in a pair-wise basis — the relative binding free energy between (only) two poses is calculated from each simulation. The number of generated poses with JAFS pose is hence relevant, since a new re-scoring simulation will, in principle, be required for each new pose obtained. It is also worth mentioning that the time length of each of these re-scoring calculations will be on the same order as that of one JAFS pose

simulation.

However, the main objective of a JAFS calculation is to find the correct (crystallographic) binding pose. Statistically, it may be expected that, the lower the number of poses we generate, the less likely that the crystal pose will be among them. Further study of the two different thresholds for average hierarchical clustering was considered necessary to present this project with an optimal threshold among those used. In table 7.16, a summary of the complete clustering results (seen in table 7.15), is displayed as a form of comparison between clustering thresholds.

Clustering summary for 2 Å cut-off

No. repeats	Clusters per repeat	Clusters within 2Å
65	18.0	46

Clustering summary for 4 Å cut-off

No. repeats	Clusters per repeat	Clusters within 2Å
65	9.0	32

Table 7.16: Summary of clustering the results of the JAFS pose simulations over all systems with two different thresholds applied on the hierarchical clustering protocol (2 Å and 4Å). The complete data set is available in table 7.15.

From the data in table 7.16 we can calculate the ratio between the number of successful clusters (those with RMSD to crystal pose $RMSD < 2\mathring{A}$) and the number of total clusters (number of repeats times the number of clusters per repeat). This will be taken as the global measure of success for a particular clustering protocol. Performing this calculation we obtain:

- Measure of success for 2 Å cut-off: 0.0393
- Measure of success for 4 Å cut-off: 0.0547

Consequently, for JAFS pose, clustering with a 4 Å cut-off was considered a more optimal choice when subsequent re-scoring of poses is desired, and limited time to perform those calculations is available. A lower number of clusters is generated with the 4 Å cut-off, where each cluster has a higher chance to be the "correct" binding geometry than for clustering performed with 2 Å cut-off. However, if the "correct" binding geometry wants to be obtained and, either no re-scoring is desired, or the time spent on re-scoring is not a relevant matter, the 2 Å cut-off

is recommended, since it is more likely to provide the correct binding geometry among all its generated alternatives.

7.2.7 Comparison to random pose generation

Given these results, a logical question was raised of whether the presented results were any better than random placement of the fragment within the JAFS box delimiting the sampling region. To answer this question, clustering was performed on the initial distributions on fragments in the JAFS box, which had been randomly generated. It is important to mention that, while the position of the fragment is randomly selected within the box limits (centre of geometry of the fragment remaining within the box) the internal structure of the fragment remains identical to that provided by the crystallographic information. The different random configurations were generated with the *distribute_waters.py* ProtoMS tool (see section 4.4.1).

Randomly generated poses for all JAFS pose systems

No. repeats	Clusters per repeat	Clusters within 2Å	Clusters within 4Å
65	8.8	0	14

JAFS pose results for all systems

No. repeats	Clusters per repeat	Clusters within 2Å	Clusters within 4Å
65	9.0	32	78

Table 7.17: Results of applying clusters on the randomly generated poses used as input to the JAFS calculations. Clusters are calculated with hierarchical clustering, with a 4 Å cut-off. Displayed are the total number of repeats (simulations) adding all fragments in all systems, average number of clusters per repeat, total number of clusters which fell within 2 Å or 4 Å RMSD of their corresponding crystal structure.

A summary of the results for all systems and all fragments is shown in table 7.17, the main conclusion being that no correct crystallographic pose (RMSD $<$ 2 Å) could be captured with this random approach.

A statistical test was performed to assess whether the results of clustering randomly generated poses were significantly different, from a statistical perspective, to those generated by clustering the results of JAFS pose. A one tailed binomial test was chosen, to evaluate whether the proportion of clusters within 4 Å of their

crystallographic pose were significantly higher for the clustered results of JAFS, than for the result of clustering random poses — the 4 Å cut-off was chosen since no binding geometries within RMSD $<$ 2 Å were generated randomly, and hence no successful comparison could be performed —. This test was run with an *R* script, with the `binom.test` command and the `alternative=‘‘greater’’` parameter.

On choosing the one tailed binomial test we are initially assuming that JAFS pose generates binding geometries that are as good as random in finding the crystal binding geometry (within 4 Å RMSD). If this were the case, if JAFS pose basically generated just random poses of the ligands within the JAFS box, how likely is it that we would obtain the data that we observe in the JAFS pose results, or better? (13.3% of the clusters fall within 4 Å of their crystal binding mode, see below). We will calculate this probability with the one tailed binomial test. If the answer is that it is very unlikely that we were to obtain this proportion of clusters within 4 Å of the crystal pose if JAFS pose was generating poses as good as random, then we will understand that JAFS pose is indeed performing better than random pose generation.[248]

Within this test, the probabilities are expressed as the number of successes (clusters within 4 Å of the crystal pose) divided by number of trials (total number of clusters). The null hypothesis — corresponding to the results of running clustering on the randomly generated poses — is expressed as a probability of $14/(8.8 * 65) \simeq 0.024$, where 14 is the number of cluster representatives within 4 Å of the crystal pose, 8.8 is the average number of clusters per repeat and 65 the total number of runs, as shown in table 7.17. Accordingly, the alternative hypothesis — obtained from clustering the JAFS pose results — is expressed as a probability $78/(9.0 * 65) \simeq 0.133$, where 78 is the number of cluster representatives within 4 Å of the crystal pose, 9.0 is the average number of clusters per repeat and 65 the total number of runs, as shown in table 7.17. The obtained p-value of this one tailed binomial test is $p = 2.2E - 16$. Consequently, it can be said that our null and alternative distributions are not equivalent, applying a significance level of 1%. A summary of results can be found in table 7.18. It can hence be said that the binding geometries generated with JAFS pose are significantly more likely to find the crystal binding mode (within RMSD $<$ 4 Å) than randomly generated binding configurations. In summary, JAFS pose is doing better than random at finding the

crystal binding structure.

It is important to notice that this is a naive test, where the null hypothesis does not even take into account potential Lennard-Jones clashes with protein atoms within the JAFS box.

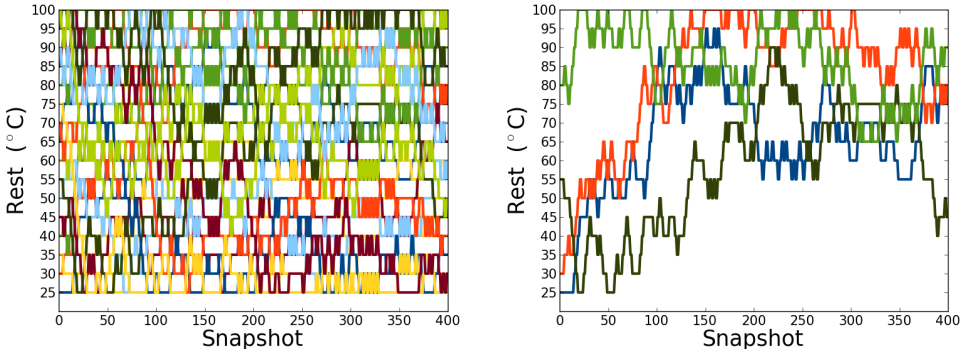
Results of one-tailed binomial test

Input probabilities		**Test results**
Null probability	Alternative probability	p value
0.024	0.133	$2.2E-16$

Table 7.18: Summary of the one tailed binomial test, data and results. The null probability (probability of the null hypothesis) corresponds to the probability of finding cluster representatives within 4 Å of the crystal pose when clustering randomly generated poses. The alternative probability (probability of the alternative hypothesis) corresponds to the probability of finding cluster representatives within 4 Å of the crystal pose when clustering the configurations sampled with JAFS pose. According to the p value resulting of the binomial test, the null and alternative distributions are considered not equivalent.

7.3 Solute Tempering

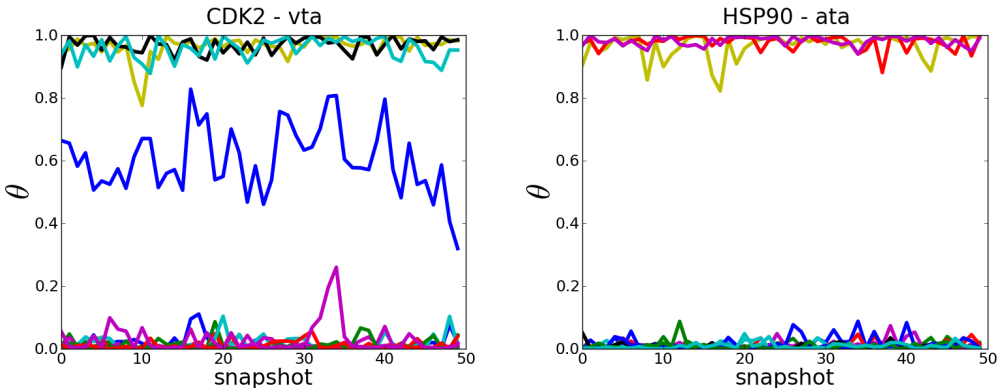
As previously explained, the enhanced sampling generated by applying the solute tempering methodology was essential for the usability of JAFS results. While the measures of sampling efficiency were focused on numbers of clusters rather than acceptance of moves between replicas, plotting the path followed by different replicas is still advisable as a sanity check on the correct behaviour of the simulation. In figure 7.24, the path of replicas for repeat one of JAFS pose on ligand ata of the HSP90 system (chosen as a representative example) is shown. Note the correct expected behaviour, with neighbouring replicas swapping in a random walk.



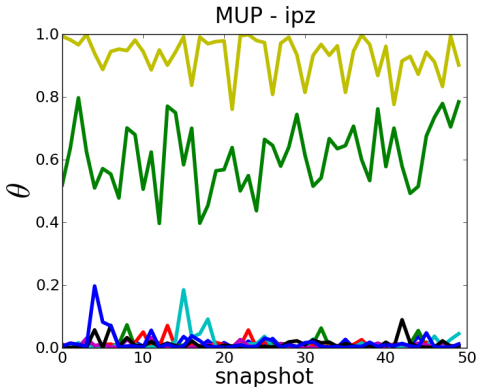
(a) (b)

Figure 7.24: Plots produced with calc_replicapath.py (see section 4.5.2) from the JAFS pose simulation results on ata (system: HSP90). Figure (a) displays all replicas, while figure (b) shows only some replicas, for clarity. The x axes corresponds to simulation snapshot and y axes shows the "solute temperature" (see section 3.7.1 and 4.4) associated with each replica.

An in-depth analysis of the effects of solute tempering, in comparison with parallel tempering and no enhanced sampling technique, as well as a discussion on the optimization of the settings for solute tempering is included in section 6.3.2. However, the direct effects on θ sampling of using solute tempering as an enhanced sampling technique in the production conditions have been further studied. To that end, simulations identical (including same initial random seed) to the first repeat of JAFS pose on ata (system HSP90), vta (system CDK2) and ipz (system MUP) but not including solute tempering have been run (one ligand of each system, and only the first repeat, chosen as representative examples). Their θ sampling results for



(a) (b)



(c)

Figure 7.25: Evolution of the θ values of JAFS fragments with respect to snapshot at the end of the simulation (last 50 snapshots) of the first repeat of JAFS pose runs for fragments vta (a), ata (b) and ipz (c) in simulations without solute tempering.

only the last 5 million moves (for clarity) can be seen in figure 7.25. By comparing these plots with those in figure 7.18, the consequences of solute tempering on θ sampling, within the production simulation conditions, are obvious.

As can be observed in figure 7.25, minimal change in the θ variable is observed in all simulations, for all ligand copies, when no solute tempering is applied. No swaps between high and low values of θ are observed for any of the ligands during these last 5 million moves of simulation. When no swaps between particles at high and low values of θ are observed the simulations present a very similar condition to a standard MC (or MD) simulation. A constant number of solutes is present in the simulation, with their interactions (almost) fully "on", and no (or very

little) sampling advantage is obtained with respect to these standard simulations — the link between θ and Cartesian sampling has been discussed in depth and for a number of examples throughout the thesis. In fact, it can be argued that this situation is possibly worse than that of standard MC (or MD) simulation for a few reasons. First, the possibility of a number of copies of the fragment being present in the simulation at once (as is the case for the HSP90 ata and CDK2 vta simulation, as can be seen in figure 7.25) which is unlikely to be the case in experimental conditions (due to the ligand concentration in solvent). The second reason is the possible presence of ligands at intermediate states of θ (as is the case in the MUP ipz and CDK2 vta simulations as shown in figure 7.25). These "half on" ligands do not represent any state which can be sampled in experimental conditions (where ligands are either present or absent from the simulations).

The essential conclusion to be taken from this section is the need for the application of solute tempering on the JAFS simulations. While the disadvantages previously described may appear at a given point during the simulation even when solute tempering is applied, the increase in sampling in the binding cavity provided by JAFS in combination with solute tempering outweighs (in our opinion) these possible drawbacks. However, these drawbacks do highlight the need to understand the JAFS method as a method to rank ligands by affinity (JAFS score) and locate their binding geometry (mostly JAFS pose, but also JAFS score) rather than a simulation where any desired thermodynamic property may be obtained.

7.4 Comparison to Docking

While for JAFS pose a (rather naive) comparison to random pose generation has been performed in section 7.2.7, to assess the scope of usability of the JAFS protocols, a comparison to current common methods is required.

Possibly the most standard choice for the location of binding geometries of small ligands to target proteins, and the ranking of these ligands, is docking and scoring. Docking and scoring techniques, further explained in section 3.2, represent what can possibly be described as the opposite extreme to the sort of advantages provided by traditional free energy techniques (such as alchemical transformations, see section 3.6.2) and (to a lesser extent) the JAFS method. They focus on the generation of fast results, with less emphasis on theoretical accuracy, and encouraging use of all

previous knowledge available for the systems of interest.

Often, when a new method is compared to its predecessors by developers of the newcomer, the use of previous methodologies is far from optimal. This is to be expected, since developers of a particular new methodology are bound to know and understand their creation more deeply and may not be experts in the use of alternative packages or procedures. Given that the study of fragments often requires specific setup of methodologies, due to the small size of the ligands, that may diverge from standards and package defaults, we considered this to be an important hindrance to our the comparison of JAFS to other docking techniques. Luckily, we have the privilege of the involvement in this project of Astex Pharmaceuticals. Their focus is the development of the fragment based approach to the generation of new therapeutical drugs. They prioritize the investment and routine application of computational methods in drug discovery, and frequently perform docking calculation of fragments, as can be seen from their publications.[10,11,25,239] For the reasons stated it was considered optimal that all docking calculations used to provide the context of our JAFS results where performed by Astex Pharmaceuticals, with their optimized docking setup (specified in the Appendix, section 11.2).

The same defining traits of docking and scoring, furthermore when using these methodologies in their optimal conditions, make it difficult to provide identical conditions to JAFS and docking protocols. The main reflection of this issue is the location of potential bridging water molecules, key to the interaction between ligand and target protein. While one of the advantages of JAFS pose is its capability to automatically locate potential water mediated interactions, docking calculations will typically make use of all previous knowledge of the system, including the presence of water mediating interactions. When the presence of key waters mediating the interaction between ligand and protein is not previously known, docking calculations will generally assume these are not present. JAFS offers the possibility of expanding the previous knowledge of the system in this respect.

7.4.1 Docking Methods

In this section we present the specifics of the docking runs performed by Astex Pharmaceuticals. They used the software package GOLD,[124] with the scoring function ChemPLP.[249] Each of the ligands (binders) used throughout this project for

the HSP90 and CDK2 systems was independently docked against the same protein conformation that was used in the JAFS runs.

Each repeat of docking produced a total of 15 resulting binding geometries (poses). Diversity is forced on these poses so that only up to three poses may fall within the same cluster of 1.5 Å threshold. For the HSP90 system, the conserved waters which mediate interaction between the ligands and the protein are included. The waters are fixed in a user-defined position and rotated to optimize the established hydrogen bonds. In the case of CDK2, one single repeat was run for each case, while three different repeats are available for the HSP90 system. See further specifications of the docking runs in the Appendix (section 11.2).

7.4.2 Discussion and Comparison

System: CDK2

Fragment	Top RMSD	Lowest RMSD	Crystal pose
vth	3.89	1.37	YES
vta	2.47	0.34	YES
vtm	0.86	0.86	YES
wcc	4.11	2.96	NO

System: HSP90

Fragment	Top RMSD	Lowest RMSD	Crystal pose
2dl	0.36	0.30	YES
ata	1.29	0.72	YES
atb	3.44	0.43	YES
atc	3.74	0.20	YES
atd	4.19	2.01	NO
ate	1.22	0.74	YES
atf	4.27	2.94	NO

Table 7.19: Docking results of one docking run on the CDK2 and HSP90 systems. The RMSD of the top scored pose and the lowest RMSD of all poses are presented for each of the fragments studied in this project. A summary column indicating whether the crystal pose was found among all poses is also shown. The target protein conformation is the same as that used in JAFS.

In this section, docking results will be presented together with their JAFS counterpart, so that effective comparison can take place. Docking results provided

by Astex Pharmaceuticals can be observed in table 7.19, while the corresponding JAFS data is shown in table 7.20. Please note that the data provided is not identical. For the docking calculations, the RMSD for the top scored pose is provided (see "Top RMSD" in table 7.19). A scoring protocol is intrinsic to the docking calculations (see section 3.2). This implies that the docking runs do provide a method of selecting the most optimal pose (while this prediction may or may not be correct in each particular case). As explained previously, JAFS does not provide the capability of ranking the different binding geometries provided by JAFS pose (further comment on this respect can be found in section 8). For this reason, the "Top RMSD" column is not available in the JAFS table of results (table 7.20). The other differences correlate with the fact that five different runs of JAFS have been performed, while only one (for CDK2) or three (for HSP90) have been completed for the docking protocol. The results of only one docking run are shown in table 7.19, while an analysis of the consistency of docking vs JAFS pose repeats for the HSP90 system is shown in table 7.21. To provide a fair comparison, the first run of JAFS was chosen and its results are displayed under the "R1" label in table 7.20. However, given that a number of repeats were available, the mean value of lowest RMSD was calculated, with their associated standard error for these (label "AV" in the same table). A column related to the finding of the crystal pose in all five runs was considered trivial (hence not included) since the crystal pose is always found within 2 Å, for all ligands studied with JAFS pose, for all systems.

When comparing the results of docking and JAFS it must be further considered that each of binding geometries generated with JAFS is necessarily unique for each repeat within our clustering threshold (since they are obtained as the cluster representatives). For docking, however (as mentioned above) up to three different poses may fall within the same 1.5 Å clustering threshold.

Two different perspectives can be found when analysing and comparing the docking data with the JAFS results, and they are related to whether we consider the successful finding of the crystal pose with JAFS when this has been found within the five repeats run, or whether we decide to analyse the first repeat individually. Looking at the results presented by the first JAFS runs in table 7.20, in comparison to those provided by the single repeat of docking in table 7.19, we can see that JAFS pose seems to perform equally (for the CDK2 system) or marginally worse (for the

System: CDK2

Fragment	Lowest RMSD - R1	Crystal pose - R1	Lowest RMSD - AV
vth	0.58	YES	2.16 ± 0.89
vta	2.62	NO	2.89 ± 1.05
vtm	0.91	YES	1.10 ± 0.09
wcc	0.53	YES	1.85 ± 1.08

System: HSP90

Fragment	Lowest RMSD - R1	Crystal pose - R1	Lowest RMSD - AV
2dl	1.166	YES	1.26 ± 0.07
ata	0.402	YES	1.05 ± 0.84
atb	3.294	NO	1.97 ± 0.74
atc	3.675	NO	3.19 ± 0.65
atd	1.604	YES	2.06 ± 0.66
ate	1.108	YES	0.92 ± 0.20
atf	4.344	NO	2.41 ± 0.91

Table 7.20: JAFS pose results of the first repeat (R1) and average of all 5 repeats (AV) on the CDK2 and HSP90 systems. The lowest RMSD of all poses is presented for each of the fragments studied in this project. A summary column indicating whether the crystal pose was found among all poses in the first repeat is also shown. The clustering cut-off applied to here was 2 Å (which increases the probability of finding the crystal binding geometry over all clusters with JAFS pose).

HSP90 system) than docking for this particular repeat. It can also be seen in table 7.20 that, while the pose is found for all ligands in all systems when we look at all five repeats of JAFS, the average lowest RMSD among all repeats often does not fall below the 2 Å RMSD threshold with respect to crystal pose. It must be noted that the "average lowest RMSD" is taken as the average over the lowest RMSD from each repeat. The fact that this value does not fall within 2 Å RMSD of the crystal pose highlights that not all repeats locate the crystal binding geometry (in some cases, not even the majority of repeats do).

It is important to remember at this point that the docking runs do not present the same conditions as the JAFS pose simulations, hence making them, arguably *not comparable*. The main difference relates to the introduction of extra knowledge about the system within the docking runs, such as the bridging waters, whose presence and location is automatically detected by the JAFS pose protocol. In computational terms, the phase space required to sample within the JAFS pose simulations is much more complex than that of docking runs, mainly due to the presence of water molecules as well as the ligand.

Consistency Between Repeats

Lowest RMSD - System: HSP90

Fragment	Docking - AV 3	JAFS - AV 3
2dl	0.32 ± 0.01	1.30 ± 0.09
ata	0.71 ± 0.15	0.47 ± 0.10
atb	0.53 ± 0.13	1.84 ± 0.96
atc	0.20 ± 0.00	2.97 ± 0.98
atd	1.14 ± 0.58	1.89 ± 0.86
ate	0.96 ± 0.14	0.95 ± 0.11
atf	2.97 ± 0.02	2.40 ± 1.28

Table 7.21: Comparison of the average result of three repeats, as well as its associated errors, between docking and JAFS pose results, for all fragments in the HSP90 system. The clustering cut-off applied here was 2 Å.

The previous comparison between docking and JAFS pose results relied on the results of a single docking repeat, hence not allowing any analysis of the consistency of docking runs. Here we present results of three repeats of docking on the HSP90 system, and show the average RMSD to crystal pose as well as its associated error.

As it can be seen in table 7.21, docking results provide a higher consistency (lower associated error) for the lowest RMSD between runs in five out of seven ligands within the HSP90 system. Similarly, the average RMSD is lower for the docking calculations, again in five out of seven repeats, while in most cases both results are within error of one another. For this case it is hence found that docking runs are generally more consistent as well as more accurate. Again, we need to take into account that the position of the bridging waters has been automatically detected within the JAFS methodology, hence not necessarily in their exact crystallographic configuration, which was used within the docking runs.

Summary and Conclusion

First, as has been noted throughout this section, the comparison of the docking runs and the JAFS simulations is a complicated one, since different conditions are applied to each methodology (i.e. the treatment of waters or the possible parametrization of the docking scoring functions against proteins of the same family as that studied, or a similar set of fragments). However, some information can still be taken from this comparison study.

As an initial observation, it is clear that JAFS pose does not consistently outperform docking calculations, despite using higher computational resources. It may hence be safely stated that JAFS must not replace docking in its routine use during FBDD. Further studying the data, it may also be noted that JAFS does not considerably under-perform docking calculations either, while a much more "black box" approach is taken with JAFS (automatic location of bridging waters and lack of calibration of energy estimations to particular protein families). Consequently, JAFS pose may be a relevant method for particular cases where a lack of information on the system is found and a black-box approach is required (i.e. if a system from an under-studied protein family is of interest).

Besides its direct usability in pharmaceutical FBDD projects, the finding that the performance of JAFS is similar to that of docking is encouraging from a development and research perspective. It suggests this technique may be found relevant for particular research projects and possibly future developments may increase its applicability and success.

Chapter 8

Rescoring Poses

The JAFS pose protocol has proven successful in finding the correct binging geometry for all presented fragments in all shown systems (as shown in section 7). However, not only the correct binding geometry, but a number of possible binding poses are generated with this protocol. Little has been discussed so far on how (or if at all possible) the correct binding geometry may be selected from all generated poses (cluster representatives).

8.1 Cluster Occupancy

An early idea in the initial development on the project (as mentioned previously in section 6.2) was to rank poses based on their cluster occupancy. The logic behind this approach is, given proper sampling of the free energy landscape, the system should remain for longer in the most favourable conformation. Consequently, more snapshots should capture the conformation of the fragment and waters which correspond to the correct binding mode. While theoretically sound, this ranking approach does not seem to produce the desired results. When ranking by cluster occupancy, the crystallographic pose can be found in any position among the generated clusters. An example of this can be found in table 8.1.

The reasons for the failure of this ranking method may be related to the imprecisions in the representation of the system. For sampling reasons, it is most convenient to introduce a number of fragment copies per simulation. Equally, all JAFS particles must interact with one another, so that interactions between frag-

ment and water can be correctly represented. However, these two traits imply two fragment copies may also be able to see one another when presenting high values of θ in different positions of the same JAFS box. They may stabilize one another, increasing their cluster occupancy after analysis. In a similar fashion, there often are two (or more) possible posses which fill (at least partially) the same space in the JAFS box (partially overlapping clusters). It is likely that the cluster occupancy of these overlapping poses is lower in proportion to their affinity, compared to a different pose which does not overlap with any other conformation. This is due to one of the overlapping poses having to present low values of θ for the other pose to display high θ.

System: HSP90 - Ligand: ate

Repeat	No clusters	Rank
1	15	4
2	13	5
3	24	15
4	16	2
5	10	10

System: MUP - Ligand: ipz

Repeat	No clusters	Rank
1	16	3
2	12	10
3	16	16
4	14	-
5	24	1

Table 8.1: Ranking obtained in each repeat by the crystal binding geometry when occupancy for the cluster is used as ranking measure. The number of clusters found in each repeat is shown. A dash (-) indicates the crystal binding conformation was not found in that repeat. Clustering cut-off of 2 Å has been used here.

8.2 Relative Free Energy of Binding

The most expensive, but (conceptually) most accurate, method to rescore different binding geometries must be the calculation of their relative binding free energy.

To proceed with this calculation, we may want to apply the previously explained dual topology alchemical transformations (see section 3.6.2). When applying dual topology to the transformation between two poses of the same ligand, we face two different problems: conformational sampling and solvation of the cavity.

8.2.1 Problems

Conformational Sampling

To assess the difference in free energy between two systems, sampling of the free energy landscape of both extremes of the transformation (as well as any intermediate states) is required. Differences in energy between two static systems could be calculated, but the result of such a calculation could never be regarded as a relative free energy, but only the difference in potential energy between those particular conformations of each system, where no entropic component is taken into account. When, for some particular calculations of interest, a section of a system must remain within a particular area (e.g. a ligand's distance to the protein) restraints can be applied. For accurate free energy calculations, these restraints need to be applied carefully, and they need to be accounted for, adding the appropriate energy terms to the final obtained results.[180]

One of the difficulties associated with calculating the relative free energy between two different poses with respect to that same calculation between different ligands relates to their possible interconversion. Without any restraints applied on them, each of the binding poses would be allowed to sample at their corresponding extreme values of λ, and they would have the capability to interconvert. Depending on the differences between both poses (and ultimately, the energy barriers between both minima), the process of interconversion may be unlikely. However, this will still be a concern for similar poses, or poorly hydrated cavities (where waters will not be in the way, providing more freedom for the fragment to exchange conformation).

The conformational sampling issue may raise some concerns, since the user would have to make sure poses have not interconverted, or reached conformations so different from the starting one that are considered different binding geometries. If ligands reach (considerably) different binding geometries from their starting configuration, the relative free energy calculated is not that between the two initial

binding conformations, nor between the final ones, since information has been collected throughout a number of different binding conformations. The end states would be ill-defined and the meaning of the calculated free energy difference would not be clear. Furthermore, the conformations sampled would be likely to differ between repeats, making the average free energy between repeats even less meaningful. While it may, in principle, be assumed that, in most cases, the atomistic Monte Carlo simulation (without the possibility of sampling in θ space) will lack enough sampling for this to be an issue, the desired behaviour (remaining within the initial binding geometries) cannot be guaranteed.

Solvation of the Cavity

There are important reasons why waters are included in the JAFS pose calculations as one of our JAFS particles. The most obvious relates to the possibility of water mediated interactions between the ligand and the protein, or the possibility that certain cavities may be well hydrated while others are dry. The effect of water distribution can also play a more subtle role in binding ligands within the protein cavity, where the presence and particular distribution of water molecules within the (potentially solvent accessible) area of the cavity may enhance or decrease the affinity of the ligands. It would be expected that, while the ligand is bound, waters in the cavity are distributed in the configuration that takes the system to the lowest possible free energy state. Most likely, two different bound configuration of a particular ligand are related to two different water distributions in the binding cavity.

The sampling expected in an atomistic Monte Carlo (or Molecular Dynamics) simulation within the restraints of the binding cavity is not high. This is the reason why we could say in the section above that, in most cases, the poses of the binding fragment are not likely to interconvert within reasonable simulation times. For the solvent conformation this lack of sampling is against us. We cannot assume that, given an initial conformation of the solvent, the water molecules will correctly rearrange around a particular conformation of the bound ligand. In the same manner, we cannot assume that the relative free energy of binding between two different ligand configurations with an arbitrary solvent configuration will be representative of the real system. Particular examples leading to this lack

of convergence include the possibility of water molecules overlapping with one of the binding geometries, the possible need for water mediated interactions for one or both binding configurations, or the subtle effects of cavity solvation.

8.2.2 Alternatives

Two different alternatives have been considered. Both focus on tackling the solvation issue. The issue of the ligand conformational sampling, which may be solved via the application of restraints — while the application of restraints is not simple, see above — is expected to be addressed later in any methodology successful in solvation terms. Neither of the alternatives have been applied to the results of the JAFS method. This section will hence present the reasonings behind them, the reasons why they have been put on hold, or how they are being continued. This section can then be understood as *future work*.

Group Dual Topology

We are faced with the issue presented in section 8.2.1, where the optimal solvation disposition is expected to differ between the binding geometries. Different dispositions of the water molecules in the cavity would then be associated with each of the binding modes. Currently, we are considering the bound ligands in different conformations as the part of the system that must be transformed with λ, while the rest of the system is left to re-arrange in response. However, the section of the system that is variable with λ can be artificially defined, as far as the end states ($\lambda = 1$ and $\lambda = 0$) correspond to the end states for which the relative free energy must be calculated. Following this line of thought, we decided to include a set of waters surrounding the bound ligand as part of the section of the system which is variable with λ — coupling water molecules into the perturbation.

While this is considered a valid alternative and no theoretical or intrinsic flaw was detected, its development stopped due to implementation issues.

In the current implementation of dual topology free energy calculations in ProtoMS, the two solutes present at $\lambda = 0$ and $\lambda = 1$ are coupled in their translational and rotational motions, hence avoiding the issue associated with the non-interacting solutes wandering out of the defined pose. This protocol relies on the similarity of the sampling space available to each of the solutes. The reason for avoiding a ligand

leaving its defined pose when in its interactions are turned "off" is the appearance of numerical instabilities in the calculation of the relative free energy when an "off" ligand is effectively in the same position as other atoms in the system. If the sampling space available to both ligands is different (or, for example, they are of a very different size) the coupling of their translational and rotational sampling does not necessarily avoid the wandering of the "off" ligand into regions occupied by neighbouring protein or solvent atoms. Hence the previously mentioned numerical instabilities could appear.

In this context, the implementation of group dual topology proved challenging. When both groups had the same number of molecules, pairing of these could be applied (one from the group fully interacting at $\lambda = 0$ with one from the group fully "on" at $\lambda = 1$). Pairing could be done automatically, based on their order within the group, or in a user defined manner. In every case, the molecules chosen for pairing should present as similar sampling space available to them as possible. When both groups had a different number of solutes the pairing became more difficult and arbitrary decisions would have to be made.

It was hence considered that, possibly, the optimal pairing algorithm should not choose between pairs of specific molecules between the dual topology groups, but rather an algorithm could be envisaged where, every time one of the dual topology molecules was to be sampled, a solute from the opposite dual topology group would be chosen at random, and its translation and rotation coupled to that of the chosen solute. Different algorithms were attempted, but in every case where a random factor is involved, solutes were displaced towards the edges of the simulated region, away from the binding cavity (the opposite effect to that desired).

Running these simulations with positional restraints in each of the solutes involved would still be possible. However, the estimation of the optimal restraint to be applied for each of the solutes, and later taking the energetics effects of the restraints into account, would not make the method particularly black-box or efficient, hence it was discarded as a potential rescoring method for JAFS calculations.

GCTI

As mentioned previously, the problem that we are trying to solve is related to the different solvation of each pose in the binding cavity. Within the research

group, the GCMC method (see section 3.5.4) has been used to predict solvation in protein cavities, and recent advances[93] have made it more reliable and widely usable. GCMC could be used to generate the optimal solvation structure for each of the states at $\lambda = 1$ and $\lambda = 0$. However the problem remains as to which of these solvation states should be applied to each of the intermediate λ, or how to transition from one state to another.

However, this issue could be resolved if GCMC were to be applied, indeed, to each of the simulations at different values of λ. For replica exchange to be applicable, both in the λ and the GCMC (B values) coordinates, an implementation similar to that of parallel tempering (see section 6.3.2) was applied.

This alternative, named Grand Canonical Thermodynamic Integration (GCTI) is theoretically sound. It was considered that the time required to fully develop this method and explore its potential applications, both in the rescoring of poses and calculation of relative binding free energies between different ligands, was such that it was better separated into a project of its own. This project is currently under development by other members of the research group. It may, at a later date, be brought back and applied to the results of the JAFS methodology.

8.3 Summary and Conclusions

A number of different methods to rescore the poses generated with JAFS pose have been covered in this section. While the first of the methods shown, ranking of binding conformation based on cluster occupancy, had an empirical basis, the rest of the section has focused on the study of relative binding free energies, its challenges, and alternative methods to undertake this calculation.

Measuring affinity by cluster occupancy was the most naive approximation. It used information already available (number of configurations per cluster) and the theoretical understanding that the most favourable binding configuration should be present during a higher proportion of the simulation time (sampling the Boltzmann distribution for the free energy landscape of the system, see section 3.3.1). However, in practice the binding geometry presenting higher cluster occupancy among those generated by analysis of the the JAFS pose simulations did not correspond to the crystal binding geometry. Inconsistency was found between repeats regarding how the crystal binding conformation was ranked among other conformations by cluster

occupancy. While a certain degree of sampling problem must be related to the inconsistency in rankings, an influence of the inaccuracies in system representation with the JAFS pose methodology (e.g. presence of molecules at intermediate values of θ, "partially on") is also expected.

Once the ranking by cluster occupancy was discarded as rescoring method, the focus was turned to the calculation of relative free energies of binding. However, the calculations of relative free energies of binding between different poses have their own attached challenges, mostly related to their drifting from the binding configuration under study and the solvation of the binding cavity. Possible alternatives to solving these challenges have been presented and were attempted within the scope of this project in the form of group dual topology and GCTI. For group dual topology, implementation challenges hindered the development of an alternative which is theoretically sound. While initial development of GCTI started within the project, the amount of work required for the full development and testing of the methodology, prior to its application to the relative binding affinity of different poses, was such that it was better undertaken as an individual project.

In conclusion, while a few interesting ideas have been highlighted in this section, an optimal method to rescore binding configurations generated with JAFS pose has not yet been found (or developed).

Chapter 9

Summary

In this section, the main traits observed in the JAFS simulation from the results of the production runs will be presented. Once these traits have been highlighted, in section 10, the conclusions obtained from them, and the project as a whole will be presented.

9.1 JAFS Score

The outcome of the results of the JAFS score production runs presented in section 7.1 may be summarized as follows:

- The simpler systems (T4Lys and MUP) provide the desired results, which more complex systems (CDK2) fail. Simplicity here is measured both in terms of the number of fragments included in a simulation and how "well-behaved" a system is deemed to be, where model systems (see section 5.1) tend to reduce complexity to a minimum.

    - An increasing number of fragments can be directly related to a greater difficulty in distinguishing their scores and ranking. Since the scores are measured as a proportion of simulation time that the fragment spends at $\theta > 0.5$, their values span the range between 0 and 1. For identical standard errors, the greater the number of fragments in a simulation, the more likely are the scores to fall within one standard error of each other, hence being statistically indistinguishable.

- While this issue may account for the examples of systems where ranking was not possible between some fragments due to an overlap in the uncertainties of their scores (T4Lys), it would not explain the more complex examples where the predicted ranking was providing statistically significant results (not all scores within one standard error of one another) which did not match in any form the experimental values provided (CDK2).

- In the most complex of our systems (CDK2) no success was achieved in terms of fragment ranking. Not even the minimum requirement of providing all non-binders with a lower score than all binders was fulfilled. Indeed, the fragment found to bind tightest experimentally was provided with a null score within JAFS, consistently, for several sets of repeats and slightly different setups.

  - No conclusive cause for this failure could be determined.
  - It is unlikely to be a sampling issue, since consistency was found between sets of repeats.
  - There is a known issue with the force field, not expected to capture one of the key binding interactions of two of the ligands (wcc and vta).
  - While the position of the JAFS box may not be considered optimally centred on the binding region of all binders, an alternative, better centred setup was attempted, where equivalent ranking results were observed.
  - There seems to be a correlation between a higher ranking and a higher number of predicted alternative binding geometries. This correlates with a higher entropy at high values of θ (as captured by JAFS). This study was based on the number of clusters per fragment at high θ.

9.2 JAFS Pose

Observing the results of the JAFS pose production runs presented in section 7.2, a few remarks can be highlighted:

- The correct binding geometry was found for every fragment in every one of the three test systems. In at least one of the five repeats run of each JAFS

pose simulation, one of the clusters (poses) generated was found to be within 2 Å RMSD of the crystallographic binding configuration. This is true for either of the thresholds (2 Å and 4 Å) used to cluster the configurations that the fragment at high θ sampled through the simulation.

- While the correct binding geometry was always found in at least one of the repeats, consistency between repeats was often poor. For the 2 Å clustering threshold, out of the 13 total number of fragments (binders to any of the systems), only three fragments were found within 2 Å RMSD of crystal binding mode in all five repeats. When looking at the results of the 4 Å clustering threshold, only one of 13 fragments was found within 2 Å RMSD of its crystal binding mode in all repeats.

- JAFS does not offer a satisfactory method to select the correct binding mode among all the generated poses. Even in the repeats where the crystal binding mode is successfully located, a number of other poses are generated as well. The only information provided by JAFS is that the correct binding geometry must be located among the generated poses, but no ranking or probability is provided to facilitate choosing one particular pose over the rest.

- The correct hydration of the cavity is automatically and correctly captured by JAFS. JAFS was presented with dry and highly hydrated systems, as well as one in which conserved crystallographic water molecules were essential to locate the correct binding configuration of the fragments. In all these aspects, JAFS provided the results expected. The dry cavity was successfully detected as such, with water molecules rarely present at high θ, while the other two cavities continuously present water molecules at high θ. Equally, the conserved crystallographic waters were correctly located in the system where these were essential for the correct binding configuration of the ligand. This can be concluded from the correct binding of the ligands, as well as from direct observation of the clusters of the configurations sampled by the water molecules at high values of θ.

    - The same protocol was applied to all the systems. No previous information on the hydration of the cavity or the position of waters mediating

interactions between the fragment and the protein was provided during setup.

Chapter 10

Conclusion

JAFS was developed as an accurate method to estimate the correct binding geometry and relative affinity of small fragment molecules to target protein cavities. It originated from the lack of consistency often found in common docking and scoring algorithms, and their requirement of previous knowledge of the system in some specific situations, such as water mediating interactions (see section 3.2).

After the development of the JAFS methods with its two protocols (JAFS score and JAFS pose) and its study on the test systems during the production runs, conclusions can be made, as well as an analysis on the success of the original objectives of JAFS.

JAFS is a method run using molecular mechanics, following the Monte Carlo sampling algorithm. While the sampling of the free energy landscape follows common sampling rules, the accuracy in capturing the free energy landscape of the real system with JAFS is what limits this method.

JAFS aims to improve the sampling ability of traditional molecular mechanics methods, where the change between different binding configurations of a ligand in a protein target, or the study of the relative affinity of several binders to the same target protein, often proves unfeasible within reasonable simulation times. To obtain this increase in sampling, the JAFS method provides an alternative path in the exchange between different JAFS particles (fragments or water molecules) at any position or conformation within a user-defined cubic region. When adding this alternative path (namely, the possibility for JAFS particles to sample the scaling of their interaction energy), the original free energy landscape is being modified

from that strictly representing the real system.

Modifications and fine tuning of the JAFS method have been carried out during the development stage of the project, to try and obtain the best combination of optimal sampling and accurate representation of the real states of the system. An analysis of the accuracy of the representation of the systems with each of the JAFS protocols, as well as the factors that may lead to the differences from optimal behaviour observed for each of the protocols, will be presented below.

10.1 JAFS Score

The main factor modulating the correlation between optimal sampling and accurate representation of experimental systems within the JAFS score protocol relates to the choice of measure for the affinity score. The score chosen to rank ligands by affinity in JAFS is calculated as the proportion of simulation snapshots where each fragment is found with a θ value above a threshold of 0.5. This choice of score prioritises sampling over accurate representation of binding affinities.

The lack of accuracy associated with our choice of score can be related to the existence of particles with intermediate values of θ. These do not represent any real state of the system (experimentally particles will either be in the cavity or not; there is no experimental intermediate state).

Experimentally, the affinity can be measured as the proportion of bound ligand to the protein with respect to the total protein and ligand available. The equivalent to the equilibrium state through huge numbers of molecules taking part in a experimental measure is obtained by a time average of the behaviour of the single system in a computational simulation. Hence computationally, given enough sampling, the affinity could be measured as the proportion of simulation time the ligand spends bound to the protein with respect to the time unbound. While in a standard molecular mechanics simulation, the binding and unbinding events cannot be expected to happen, if no extra bias is being applied, within any reasonable simulation time, a JAFS kind of setup could provide this information. In this case, the bound ligand would be the ligand at $\theta = 1$ ("on" or "present" in the cavity), and the ligand unbound that at $\theta = 0$.

As can be understood, when sampling the θ variable as a continuum, with Monte Carlo (and the same would apply to Molecular Dynamics), virtually no snapshot of

the simulation can be expected to capture any ligand at the exact value of $\theta = 1$. Some threshold needs to be applied, where an example of a strict threshold would be considering bound ligands those with $\theta > 0.99 \approx 1$. Just as no ligands can be expected at the exact value of $\theta = 1$, the bigger the range of θ values accepted as bound ligands (the lower the threshold), the larger the number of samples are expected to fall within that range. For very small ranges (strict thresholds), for which only small number of snapshots (i.e. fewer than ten) record bound ligands, it is trivial to understand that the number of samples (ligands above threshold) may not be enough to consider the binding estimates significant (i.e. one of the ligands may have reached $\theta > 0.999$ by chance at some point in the simulation, while other ligand consistently remained at high values of θ, but these, by chance, did not overcome such a strict threshold). While the idea of very strict thresholds may be extreme, it serves to illustrate that, the higher the threshold, the lower the sampling expected.

However, when lowering the threshold to improve sampling, we are including, within our definition of bound ligands, particles at intermediate values of θ, which are not representative of any state of the experimental system. We are basing our estimation of binding affinity on the affinity experienced by non-realistic states of the ligands, which may or may not correspond to that experienced by a fully interacting ligand. The lower the threshold, the bigger the approximation of what we understand as a bound ligand.

The affinity could be calculated in a more accurate manner by following equation 3.39 in section 3.8.1.[200] As expressed previously, the choice of our affinity estimate is based on sampling boosts. To obtain good enough sampling with equation 3.39 for the binding of water molecules, Bodnarchuk et al. applied sets of biases, requiring several repeats of the same calculations to adjust the bias to the affinity of water to the binding site of interest.[201,250] Coming up with a bias which would provide enough sampling with a strict threshold for all fragments simulated together could be problematic and would hinder the black-box approach of JAFS, as well as considerably increasing the simulation time required.

10.2 JAFS Pose

Some of the traits specific to the JAFS pose simulations hold the key to the balance between an accurate representation of the real system and the desired boost in sampling when compared to traditional molecular mechanics methods.

10.2.1 Intermediate θ Values

The first of these aspects has been mentioned previously when studying the different hydration penalties in section 6.3.1. The presence of JAFS particles (both fragments and waters) at intermediate values of θ can be understood as the main addition of JAFS over GCMC (see section 4.2 and 3.5.4). It allows for the inclusion of molecules bigger than water molecules (fragments) to be included in the set of molecules which can "appear" and "vanish" from the system. As explained previously, it is the possibility of taking small steps in the scaling of the interaction energy, rather than presenting a exclusive "on-off switch" behaviour, that provides JAFS with the increased acceptance ratio of moves required for the case of bigger molecules. While key to giving fragments the "vanishing" behaviour, the possibility of taking these small steps in scaling interaction energies implies that, necessarily, particles with intermediate values of this scaling parameter (θ) will be present at some points in the simulation. In fact, being strict, it makes it unlikely for particles to be fully "on" ($\theta = 1$) or fully "off" ($\theta = 0$), while most of them will be somewhere in between. While it is true that the presence of these intermediate states is required for the functioning of JAFS, the abundance and distribution of these states can be modulated. The closer each particle is to the extreme values of θ (0 and 1), the more accurately JAFS is depicting possible states of the real system, where particles may be present or absent from the binding region, but will never be "partially there".

The dominant values of θ can be modulated by altering the free energy landscape, changing the location of the minima the system mostly samples. Altering the free energy landscape at either of the extreme values of θ would modify the affinity of the system for that type of JAFS particle, hence driving us further away from a realistic representation. However, intermediate values of θ do not represent any realistic state of the system. Decreasing their presence, while leaving unchanged

the free energy at the end states should make our system more accurately represent reality. To apply this modification of the energy landscape, the shape of the hydration penalty is modified, while the same values are kept at extreme λ (see section 6.3.1).

We may look at how the presence of particles at intermediate values of θ can influence finding the correct binding geometry in more chemical (rather than physical) terms. Particles at intermediate values of θ, have their interaction energy scaled down, but are still noticeable. In JAFS pose, it is to be expected that one of the copies of the fragments may be at high values of θ while some other particles may present intermediate values of θ. Interactions between the fragment at high θ and the particle at intermediate θ may stabilize the configuration of the fragment in a geometry in which would not be otherwise likely to remain. In turn, the particle at intermediate θ may be occupying a configuration which would not be available to it while presenting a high value of θ. In this scenario, a non realistic binding configuration may have been stabilized for a fragment at high values of θ, which will later be clustered to generate a predicted binding pose.

10.2.2 Several Copies of Fragments

While not intrinsic to the design of JAFS but a choice in our JAFS pose protocol, the presence of several copies of fragments in the simulation is another factor that triggers the balance between sampling and accurate representation of reality.

There are two reasons, related to sampling, to include several copies of the fragment molecule in a JAFS pose simulation. First is the boost of sampling. When a move — in Cartesian or θ space — is going to be applied to solute molecules (JAFS particles in our simulations), one of these particles is chosen at random. With several copies of fragment present, it is more likely that the particle chosen is a copy of it. This, in turn, should increase the probability that a copy of the fragment appears at one given time at high values of θ, in any given region of the JAFS box. These configurations at high θ will then be clustered. In principle, the more simulation snapshots capturing a fragment configuration at high θ, the better the sampling of the JAFS box available to the clustering method. It must be noted that this sampling issue is related to the ProtoMS implementation of sampling and molecule types, where a different sampling rate is not applicable to two different

species of the same molecule type (i.e. waters and fragments both defined as solutes with θ sampling — JAFS particles).

While this is the most obvious, there is a second reason for several copies of fragments to be beneficial to sampling. It is related to the need to overfill the cavity. For the boost in sampling provided by the JAFS methodology to take place, JAFS particles must be at low values of θ at some time during the simulation. With their interactions scaled down, they can freely sample the whole Cartesian space of the JAFS box, hopefully finding a region of favourable interactions where the moves to increase their θ value are favourable, reaching a possible binding configuration. It can be found that, by including the optimal number of water molecules and only one copy of fragment in the cavity, JAFS particles tend to stay at high values of θ, hence decreasing the sampling potential of JAFS. Our measure of how overfilled the cavity must be is based on grid calculations which are further explained in section 4.4.1. While, in principle, the same level of overfilling could be achieved by including a large number of water molecules as JAFS particles, rather than several copies of fragment, the huge number of water molecules required would be difficult to handle by our simulation software. For this and the previously described reason, introducing several copies of fragments was considered optimal.

However, this decisions has drawbacks in terms of how well the system represents reality. Including several copies of the fragment molecule, nothing prevents several fragment copies from presenting high values of θ at the same time. And it is important to remember that JAFS particles do interact with one another (this is the reason why water mediated interactions can be found). Hence it is perfectly possible to imagine (as observed in the plots in figure 7.18) that several copies of the fragment present high values of θ, filling different regions of the cavity, and potentially stabilizing sub-optimal binding configurations as well as destabilizing those which may be optimal in the absence of other fragment copies. While it is perfectly possible for several copies of the fragment to bind at once in the binding cavity of the real system, the difference in concentration between waters and fragment molecules makes it very unlikely for most systems.

While not a simple implementation within the simulation software, the possibility of generating JAFS simulations where different fragment molecules can see the rest of the waters within the JAFS box, but not other fragments is one of the

possible alternatives to this method.

10.3 Final Remarks

The JAFS protocol provides a way of estimating binding geometries and relative affinities of small molecules to protein cavities. On the trade off between computational expense and theoretical reliability, JAFS provides a more theoretically sound perspective than that of docking and scoring. In terms of calculating relative affinities, JAFS score falls on the less computationally expensive as well as less theoretically sound end of the spectrum than that of dual topology calculations.

The use of JAFS is particularly recommended on systems with challenges appropriate for JAFS, such as systems with a lack of previous knowledge of the binding configuration of the ligand, as well as systems where the solvation structure within the binding cavity may be unknown, variable with time and / or with bound ligand. The choice of JAFS as the optimal method for calculating binding affinities and geometries will depend on the system and particular interests of the project.

The JAFS score protocol has proven successful in ranking fragments by affinity for the simpler systems, but the success has decreased as complexity of the system increases. The reasons of failure for the more complex system attempted have not been fully understood. Application of JAFS score to further systems, those with a complexity similar to that of the later system (CDK2) in this project would probably be an interesting prospect in any further developments of this method.

The JAFS pose protocol automatically detected hydration patterns and bridging waters without any previous knowledge. It has proved sufficiently theoretically sound to locate the correct binding geometry among all those generated for every ligand in every systems attempted.

While all correct binding configurations have been found through our repeats of the JAFS pose simulations, consistency has been an issue. Furthermore, no method is provided to detect the optimal pose among all generated. The development of such a method is likely to be required for the JAFS methodology to gain broad applicability.

Just as it is commonly the case in computational methods, a trade off is present in this field of finding binding affinities and geometries between computational expense and accuracy of results. The interest of each method would depend on the

resources available and accuracy required, as well as previous knowledge available.

Possible lines for further development have been described in section 8 regarding the ranking of the binding configurations generated with JAFS pose. Alternative possibilities which might increase the sampling achieved (and hence consistency between repeats) for JAFS pose have been briefly mentioned above (section 10.2). A further investigation on the reasons of failure of JAFS score on CDK2 would be desirable, where its application to other pharmaceutically relevant systems may shed some light. Furthermore, the application of JAFS score to systems such as HSP90, where water mediated interactions are involved in binding has been avoided since further decisions on implementation would be required regarding the desired treatment of the bridging waters. However, these calculations would be relevant and particularly useful in the context of JAFS, hence further developments on its implementation would be of interest. Equally, certain limitations of the implementation in ProtoMS which have been addressed by the specifics of the JAFS set-up (e.g. including a number of copies of the fragment in JAFS pose simulations) may want to be addressed from the software development perspective (e.g. allowing the possibility of choosing different sampling ratios for different particle species of the same molecule type). These software changes may allow for an equivalent level of sampling to that currently obtained, with a more accurate representation of the experimental system (by including only one copy of the fragment in the cavity). In general, while the development of the JAFS method has been undertaken during this thesis, further modifications to the presented methodology may improve its usability in the context of FBDD.

Chapter 11

Appendix

11.1 Minimization in Sander

The minimization of ligand structure was accomplished with following the script lines shown below.

Bash script lines

```
header=$(sed -n '1p' $1.pdb)
```

#Generation of required amber parameter files

```
cat << EOF > leap.in
source leaprc.gaff
loadamberprep $1.prepi
loadamberparams $1.frcmod
x=loadpdb $1.pdb
saveamberparm x prmtop prmcrd
quit
EOF
tleap -f leap.in > leap.out
```

#Specification of minimization parameters

```
cat << EOF > min.in
&cntrl
irest=0,ntx=1,
imin=1,maxcyc=100,drms=0.0001,ntmin=2,
```

```
ntc=1,ntf=1,
cut=20.0,
ntpr=100,ntwx=0,ntwv=0,ntwe=0,
ipol=0,igb=0,ntb=0,
&end
EOF
```

#Running ligand minimization

```
sander -O -i min.in -o min.out -p prmtop -c prmcrd -r mincrd
cat << EOF > ptraj.in
trajin mincrd
trajout mincrd.pdb pdb
EOF
```

#Getting pdb from the minimization output

```
ptraj prmtop ptraj.in >& ptraj.out
echo $header > $1.pdb
sed ‘‘/REMARK/d’’ mincrd.pdb.1 >> $1.pdb
```

11.2 Docking Setup

A summary of the most relevant features of the docking runs performed by Astex Pharmaceuticals can be found in section 7.4. Some further specific set-up information is shown below:

- Early termination is “off”: The docking runs are performed until specified by the rest of set-up options. No early termination is applied when the same solution (pose) is found repeatedly throughout a docking run.

- Diverse solutions is “on”, with a cluster size of 3 and RMSD of 1.5 Å: During one docking run, once the number of poses specified in cluster size (3) are included within a cluster, delimited by the RMSD (1.5 Å), no new solutions within that cluster will be accepted.

- All “population” parameters are kept as default. These apply to the genetic algorithm. Since an optimal choice for these parameters is highly correlated,

choosing one of the default sets is one of the recommended options in the GOLD configuration file user manual.[251]

- In a similar fashion as above, the "genetic operators" are left as default.
- Fitting points were generated for all solvent accessible donor and acceptor atoms, including fitting points that were not solvent accessible themselves. This is particularly relevant for kinases such as (CDK2).[251] The fitting points define the locations where the ligand may be placed during docking and are placed based on the interactions that the acceptor atoms may stablish.
- Torsion angles distributions extracted from the Cambridge Structural Data Base are used to limit the ligand conformational space sampled by the genetic algorithm
- The size of the binding site is determined to be a sphere of 10 Å of radius.
- The number of ligand binding geometries generated is set to 15.

The descriptions have been taken from definitions parameters described in the GOLD configuration file user manual.[251]

11.3 Theta Sampling for All Ligands

The θ sampling during the last 5 million moves of the first repeat of JAFS pose simulation for all ligands not included in section 7.2.4 are included below in plots 11.1 and 11.2. The conclusions from these plots are equivalent to those seen in the aforementioned section.

11.4 Fragment 2D representations

The 2D representations of the fragments used throughout this project can be found in table 11.1.

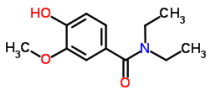
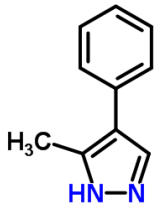
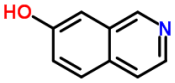
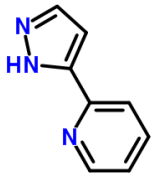
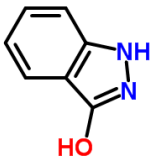
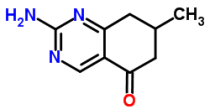
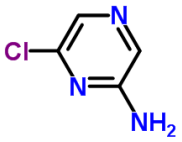
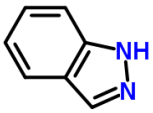
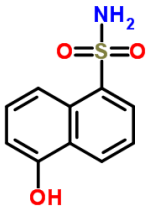
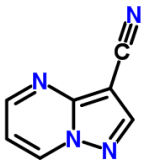
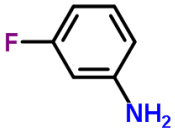
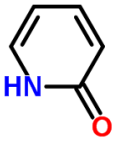
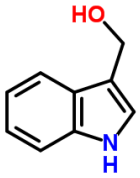
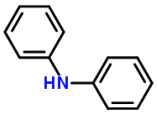
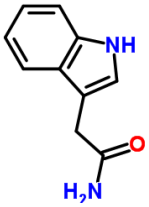
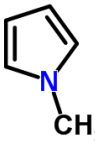
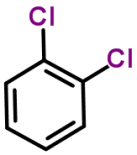
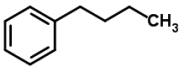
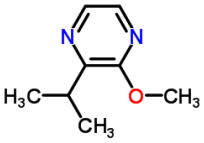
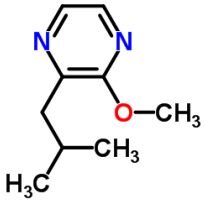
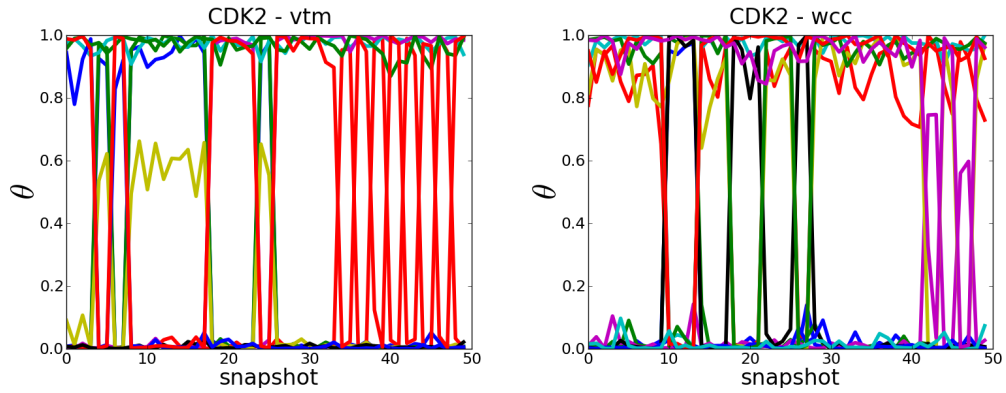
2dl 	ata 	atb 	atc 
atd 	ate 	atf 	wcc 
vta 	vth 	vtm 	cd1 
cd5 	cd6 	cd8 	cd9 
bnz 	1mp 	dcb 	nbb 
wa1 	ipz 	prz 	

Table 11.1: 2D representation of all fragments simulated in the JAFS production phase of this project.



(a) (b)

Figure 11.1: Evolution of the θ values of JAFS fragments with respect to snapshot at the end of the simulation (last 50 snapshots) of the first repeat of JAFS pose runs for fragments vtm (a), wcc (b), all binders to the target protein CDK2.

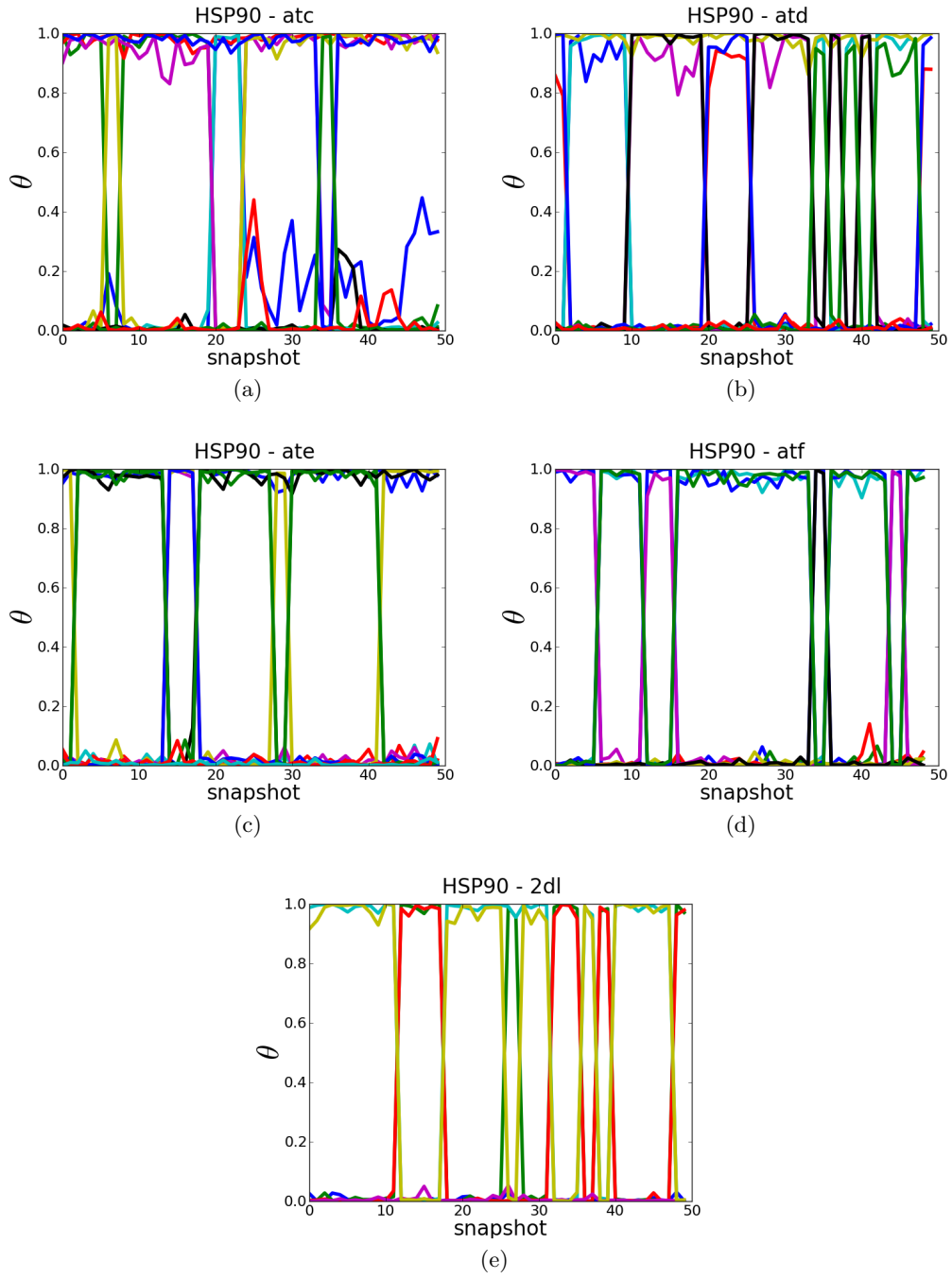


Figure 11.2: Evolution of the θ values of JAFS fragments with respect to snapshot at the end of the simulation (last 50 snapshots) of the first repeat of JAFS pose runs for fragments atc (a), atd (b), ate (c), atf (d) and 2dl (e), binders to the target protein HSP90.

Bibliography

[1] D. Guthrie, *A History Of Medicine*. London: Thomas Nelson And Sons Ltd, first ed., June 1945.

[2] M. R. Shirts, D. L. Mobley, and S. P. Brown, "Free-energy calculations in structure-based drug design," in *Drug Design*, pp. 61–86, Cambridge University Press, 2010. Cambridge Books Online.

[3] A. C. Pan, D. W. Borhani, R. O. Dror, and D. E. Shaw, "Molecular determinants of drugreceptor binding kinetics," *Drug Discovery Today*, vol. 18, pp. 667–673, July 2013.

[4] J. P. Hughes, S. Rees, S. B. Kalindjian, and K. L. Philpott, "Principles of early drug discovery.," *British journal of pharmacology*, vol. 162, pp. 1239–1249, Mar. 2011.

[5] T. Cheng, Q. Li, Z. Zhou, Y. Wang, and S. Bryant, "Structure-Based virtual screening for drug discovery: a Problem-Centric review," *The AAPS journal*, vol. 14, pp. 133–141, Mar. 2012.

[6] 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," *Computational and Structural Biotechnology Journal*, vol. 5, pp. 1–14, Feb. 2013.

[7] W. P. Jencks, "On the attribution and additivity of binding energies.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 78, pp. 4046–4050, July 1981.

[8] P. J. Hajduk, W. R. J. D. Galloway, and D. R. Spring, "Drug discovery: A question of library design," *Nature*, vol. 470, pp. 42–43, Feb. 2011.

[9] D. Joseph-McCarthy, A. J. Campbell, G. Kern, and D. Moustakas, "Fragment-Based lead discovery and design," *J. Chem. Inf. Model.*, vol. 54, pp. 693–704, Mar. 2014.

[10] D. C. Rees, M. Congreve, C. W. Murray, and R. Carr, "Fragment-based lead discovery.," *Nature reviews. Drug discovery*, vol. 3, pp. 660–672, Aug. 2004.

[11] C. W. Murray, M. L. Verdonk, and D. C. Rees, "Experiences in fragment-based drug discovery.," *Trends in pharmacological sciences*, vol. 33, pp. 224–232, May 2012.

[12] M. M. Hann, A. R. Leach, and G. Harper, "Molecular complexity and its impact on the probability of finding leads for drug discovery.," *Journal of chemical information and computer sciences*, vol. 41, no. 3, pp. 856–864, 2001.

[13] W. L. Jorgensen, "The many roles of computation in drug discovery," *Science*, vol. 303, pp. 1813–1818, Mar. 2004.

[14] S.-s. Ou-Yang, J.-y. Lu, X.-q. Kong, Z.-j. Liang, C. Luo, and H. Jiang, "Computational drug discovery," *Acta Pharmacologica Sinica*, vol. 33, pp. 1131–1140, Aug. 2012.

[15] G. Sliwoski, S. Kothiwale, J. Meiler, and E. W. Lowe, "Computational methods in drug discovery," *Pharmacological Reviews*, vol. 66, pp. 334–395, Jan. 2014.

[16] M. D. Parenti and G. Rastelli, "Advances and applications of binding affinity prediction methods in drug discovery," *Biotechnology Advances*, vol. 30, pp. 244–250, Jan. 2012.

[17] D. B. Kitchen, H. Decornez, J. R. Furr, and J. Bajorath, "Docking and scoring in virtual screening for drug discovery: methods and applications.," *Nature reviews. Drug discovery*, vol. 3, pp. 935–949, Nov. 2004.

[18] G. L. Warren, C. W. Andrews, A.-M. M. Capelli, B. Clarke, J. LaLonde, M. H. Lambert, M. Lindvall, N. Nevins, S. F. Semus, S. Senger, G. Tedesco, I. D. Wall, J. M. Woolven, C. E. Peishoff, and M. S. Head, "A critical assessment of docking programs and scoring functions.," *Journal of medicinal chemistry*, vol. 49, pp. 5912–5931, Oct. 2006.

[19] N. Homeyer, F. Stoll, A. Hillisch, and H. Gohlke, "Binding free energy calculations for lead optimization: Assessment of their accuracy in an industrial drug design context," *J. Chem. Theory Comput.*, vol. 10, pp. 3331–3344, Aug. 2014.

[20] L. Wang, Y. Wu, Y. Deng, B. Kim, L. Pierce, G. Krilov, D. Lupyan, S. Robinson, M. K. Dahlgren, J. Greenwood, D. L. Romero, C. Masse, J. L. Knight, T. Steinbrecher, T. Beuming, W. Damm, E. Harder, W. Sherman, M. Brewer, R. Wester, M. Murcko, L. Frye, R. Farid, T. Lin, D. L. Mobley, W. L. Jorgensen, B. J. Berne, R. A. Friesner, and R. Abel, "Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern Free-Energy calculation protocol and force field," *J. Am. Chem. Soc.*, vol. 137, pp. 2695–2703, Feb. 2015.

[21] M. Congreve, G. Chessari, D. Tisi, and A. J. Woodhead, "Recent developments in Fragment-Based drug discovery," *J. Med. Chem.*, vol. 51, pp. 3661–3680, July 2008.

[22] G. Marcou and D. Rognan, "Optimizing fragment and scaffold docking by use of molecular interaction fingerprints," *J. Chem. Inf. Model.*, vol. 47, pp. 195–207, Jan. 2007.

[23] J. D. Chodera, D. L. Mobley, M. R. Shirts, R. W. Dixon, K. Branson, and V. S. Pande, "Alchemical free energy methods for drug discovery: progress and challenges," *Current Opinion in Structural Biology*, vol. 21, pp. 150–160, Apr. 2011.

[24] W. Warr, "Fragment-based drug discovery," *Journal of Computer-Aided Molecular Design*, vol. 23, pp. 453–458, Aug. 2009.

[25] C. W. Murray, M. G. Carr, O. Callaghan, G. Chessari, M. Congreve, S. Cowan, J. E. Coyle, R. Downham, E. Figueroa, M. Frederickson, B. Graham, R. McMenamin, M. A. O'Brien, S. Patel, T. R. Phillips, G. Williams, A. J. Woodhead, and A. J.-A. J. Woolford, "Fragment-based drug discovery applied to hsp90. discovery of two lead series with high ligand efficiency.," *Journal of medicinal chemistry*, vol. 53, pp. 5942–5955, Aug. 2010.

[26] L. Zhao, D. Cao, T. Chen, Y. Wang, Z. Miao, Y. Xu, W. Chen, X. Wang, Y. Li, Z. Du, B. Xiong, J. Li, C. Xu, N. Zhang, J. He, and J. Shen, "Fragment-based drug discovery of 2-thiazolidinones as inhibitors of the histone reader BRD4 bromodomain.," *Journal of medicinal chemistry*, vol. 56, pp. 3833–3851, May 2013.

[27] R. A. Ward, C. Brassington, A. L. Breeze, A. Caputo, S. Critchlow, G. Davies, L. Goodwin, G. Hassall, R. Greenwood, G. A. Holdgate, M. Mrosek, R. A. Norman, S. Pearson, J. Tart, J. A. Tucker, M. Vogtherr, D. Whittaker, J. Wingfield, J. Winter, and K. Hudson, "Design and synthesis of novel lactate dehydrogenase a inhibitors by Fragment-Based lead generation," *J. Med. Chem.*, vol. 55, pp. 3285–3306, Apr. 2012.

[28] E. A. Larsson, A. Jansson, F. M. Ng, S. W. Then, R. Panicker, B. Liu, K. Sangthongpitag, V. Pendharkar, S. J. Tai, J. Hill, C. Dan, S. Y. Ho, W. W. Cheong, A. Poulsen, S. Blanchard, G. R. Lin, J. Alam, T. H. Keller, and P. Nordlund, "Fragment-Based ligand design of novel potent inhibitors of tankyrases," *J. Med. Chem.*, vol. 56, pp. 4497–4508, June 2013.

[29] A. Friberg, D. Vigil, B. Zhao, R. N. Daniels, J. P. Burke, P. M. Garcia-Barrantes, D. Camper, B. A. Chauder, T. Lee, E. T. Olejniczak, and S. W. Fesik, "Discovery of potent myeloid cell leukemia 1 (mcl-1) inhibitors using Fragment-Based methods and Structure-Based design," *J. Med. Chem.*, vol. 56, pp. 15–30, Jan. 2013.

[30] M. Frederickson, O. Callaghan, G. Chessari, M. Congreve, S. R. Cowan, J. E. Matthews, R. McMenamin, D.-M. Smith, M. Vinković, and N. G. Wallis, "Fragment-Based discovery of mexiletine derivatives as orally bioavail-

able inhibitors of Urokinase-Type plasminogen activator†," *J. Med. Chem.*, vol. 51, pp. 183–186, Jan. 2008.

[31] D. L. Cheney, J. M. Bozarth, W. J. Metzler, P. E. Morin, L. Mueller, J. A. Newitt, A. H. Nirschl, A. R. Rendina, J. K. Tamura, A. Wei, X. Wen, N. R. Wurtz, D. A. Seiffert, R. R. Wexler, and E. S. Priestley, "Discovery of novel p1 groups for coagulation factor VIIa inhibition using Fragment-Based screening," *J. Med. Chem.*, vol. 58, pp. 2799–2808, Mar. 2015.

[32] D. Stumpfe, Y. Hu, D. Dimova, and J. Bajorath, "Recent progress in understanding activity cliffs and their utility in medicinal chemistry," *J. Med. Chem.*, vol. 57, pp. 18–28, Jan. 2014.

[33] M. Congreve, R. Carr, C. Murray, and H. Jhoti, "A 'rule of three' for fragment-based lead discovery?," *Drug Discovery Today*, vol. 8, pp. 876–877, Oct. 2003.

[34] C. A. Lipinski, F. Lombardo, B. W. Dominy, and P. J. Feeney, "Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings.," *Advanced drug delivery reviews*, vol. 46, pp. 3–26, Mar. 2001.

[35] H. van de Waterbeemd and E. Gifford, "ADMET in silico modelling: towards prediction paradise?," *Nature reviews. Drug discovery*, vol. 2, pp. 192–204, Mar. 2003.

[36] A. J. Woodhead, H. Angove, M. G. Carr, G. Chessari, M. Congreve, J. E. Coyle, J. Cosme, B. Graham, P. J. Day, R. Downham, L. Fazal, R. Feltell, E. Figueroa, M. Frederickson, J. Lewis, R. McMenamin, C. W. Murray, M. A. O'Brien, L. Parra, S. Patel, T. Phillips, D. C. Rees, S. Rich, D.-M. M. Smith, G. Trewartha, M. Vinkovic, B. Williams, and A. J.-A. J. Woolford, "Discovery of (2,4-dihydroxy-5-isopropylphenyl)-[5-(4-methylpiperazin-1-ylmethyl)-1,3-dihydroisoindol-2-yl]methanone (AT13387), a novel inhibitor of the molecular chaperone hsp90 by fragment based drug design.," *Journal of medicinal chemistry*, vol. 53, pp. 5956–5969, Aug. 2010.

[37] A. L. Hopkins, C. R. Groom, and A. Alex, "Ligand efficiency: a useful metric for lead selection," *Drug Discovery Today*, vol. 9, pp. 430–431, May 2004.

[38] I. D. Kuntz, K. Chen, K. A. Sharp, and P. A. Kollman, "The maximal affinity of ligands," *Proceedings of the National Academy of Sciences*, vol. 96, pp. 9997–10002, Aug. 1999.

[39] A. L. Hopkins, G. M. Keseru, P. D. Leeson, D. C. Rees, and C. H. Reynolds, "The role of ligand efficiency metrics in drug discovery," *Nature Reviews Drug Discovery*, vol. 13, pp. 105–121, Jan. 2014.

[40] Nissink, "Simple Size-Independent measure of ligand efficiency," *J. Chem. Inf. Model.*, vol. 49, pp. 1617–1622, June 2009.

[41] M. M. Hann, A. R. Leach, and G. Harper, "Molecular complexity and its impact on the probability of finding leads for drug discovery," *J. Chem. Inf. Comput. Sci.*, vol. 41, pp. 856–864, May 2001.

[42] A. R. Leach and M. M. Hann, "Molecular complexity and fragment-based drug discovery: ten years on.," *Current opinion in chemical biology*, vol. 15, pp. 489–496, Aug. 2011.

[43] M. M. Hann and T. I. Oprea, "Pursuing the leadlikeness concept in pharmaceutical research," *Current Opinion in Chemical Biology*, vol. 8, pp. 255–263, June 2004.

[44] G. M. Makara, "On sampling of fragment space.," *Journal of medicinal chemistry*, vol. 50, pp. 3214–3221, July 2007.

[45] T. Fink and J.-L. L. Reymond, "Virtual exploration of the chemical universe up to 11 atoms of c, n, o, f: assembly of 26.4 million structures (110.9 million stereoisomers) and analysis for new ring systems, stereochemistry, physicochemical properties, compound classes, and drug discovery.," *Journal of chemical information and modeling*, vol. 47, pp. 342–353, Mar. 2007.

[46] D. Joseph-McCarthy, "Challenges of fragment screening," *Journal of Computer-Aided Molecular Design*, vol. 23, pp. 449–451, Aug. 2009.

[47] B. J. Davis and D. A. Erlanson, "Learning from our mistakes: The 'unknown knowns' in fragment screening," *Bioorganic & Medicinal Chemistry Letters*, vol. 23, pp. 2844–2852, May 2013.

[48] S. B. Shuker, P. J. Hajduk, R. P. Meadows, and S. W. Fesik, "Discovering high-affinity ligands for proteins: SAR by NMR.," *Science (New York, N.Y.)*, vol. 274, pp. 1531–1534, Nov. 1996.

[49] A. Persidis, "High-throughput screening," *Nature Biotechnology*, vol. 16, pp. 488–489, May 1998.

[50] C. A. Shepherd, A. L. Hopkins, and I. Navratilova, "Fragment screening by SPR and advanced application to GPCRs," *Progress in Biophysics and Molecular Biology*, vol. 116, pp. 113–123, Nov. 2014.

[51] A. R. Leach, M. M. Hann, J. N. Burrows, and E. J. Griffen, "Fragment screening: an introduction," *Mol. BioSyst.*, vol. 2, no. 9, pp. 429–446, 2006.

[52] M. J. Hartshorn, C. W. Murray, A. Cleasby, M. Frederickson, I. J. Tickle, and H. Jhoti, "Fragment-based lead discovery using x-ray crystallography.," *Journal of medicinal chemistry*, vol. 48, pp. 403–413, Jan. 2005.

[53] J. Barker, S. Courtney, T. Hesterkamp, D. Ullmann, and M. Whittaker, "Fragment screening by biochemical assay," *Expert Opinion on Drug Discovery*, vol. 1, pp. 225–236, Aug. 2006.

[54] A. Boettcher, S. Ruedisser, P. Erbel, D. Vinzenz, N. Schiering, U. Hassiepen, P. Rigollier, L. M. Mayr, and J. Woelcke, "Fragment-Based screening by biochemical assays," *Journal of Biomolecular Screening*, vol. 15, pp. 1029–1041, Oct. 2010.

[55] R. Manetsch, A. Krasiński, Z. Radić, J. Raushel, P. Taylor, K. B. Sharpless, and H. C. Kolb, "In situ click chemistry: enzyme inhibitors made to their own specifications.," *Journal of the American Chemical Society*, vol. 126, pp. 12809–12818, Oct. 2004.

[56] B. Huang, D. Kang, P. Zhan, and X. Liu, "Fragment-based approaches to anti-HIV drug discovery: state of the art and future opportunities," *Expert Opinion on Drug Discovery*, vol. 10, pp. 1271–1281, Dec. 2015.

[57] J. Badger, "Crystallographic fragment screening.," *Methods in molecular biology (Clifton, N.J.)*, vol. 841, pp. 161–177, 2012.

[58] I. Navratilova and A. L. Hopkins, "Fragment screening by surface plasmon resonance," *ACS Med. Chem. Lett.*, vol. 1, pp. 44–48, Apr. 2010.

[59] X. Chen, S. Qin, S. Chen, J. Li, L. Li, Z. Wang, Q. Wang, J. Lin, C. Yang, and W. Shui, "A ligand-observed mass spectrometry approach integrated into the fragment based lead discovery pipeline.," *Scientific reports*, vol. 5, 2015.

[60] E. Rühmann, M. Betz, M. Fricke, A. Heine, M. Schäfer, and G. Klebe, "Thermodynamic signatures of fragment binding: Validation of direct versus displacement ITC titrations," *Biochimica et Biophysica Acta (BBA) - General Subjects*, vol. 1850, pp. 647–656, Apr. 2015.

[61] J. J. Barker, O. Barker, S. M. Courtney, M. Gardiner, T. Hesterkamp, O. Ichihara, O. Mather, C. A. G. N. Montalbetti, A. Müller, M. Varasi, M. Whittaker, and C. J. Yarnold, "Discovery of a novel hsp90 inhibitor by fragment linking," *ChemMedChem*, vol. 5, pp. 1697–1700, Oct. 2010.

[62] N. Howard, C. Abell, W. Blakemore, G. Chessari, M. Congreve, S. Howard, H. Jhoti, C. W. Murray, L. C. A. Seavers, and R. L. M. van Montfort, "Application of fragment screening and fragment linking to the discovery of novel thrombin inhibitors," *J. Med. Chem.*, vol. 49, pp. 1346–1355, Feb. 2006.

[63] A. O. Frank, M. D. Feldkamp, J. P. Kennedy, A. G. Waterson, N. F. Pelz, J. D. Patrone, B. Vangamudi, D. V. Camper, O. W. Rossanese, W. J. Chazin, and S. W. Fesik, "Discovery of a potent inhibitor of replication protein a ProteinProtein interactions using a Fragment-Linking approach," *J. Med. Chem.*, vol. 56, pp. 9242–9250, Nov. 2013.

[64] V. Borsi, V. Calderone, M. Fragai, C. Luchinat, and N. Sarti, "Entropic contribution to the linking coefficient in fragment based drug design: A case study," *J. Med. Chem.*, vol. 53, pp. 4285–4289, May 2010.

[65] E. A. Larsson, A. Jansson, F. M. Ng, S. W. Then, R. Panicker, B. Liu, K. Sangthongpitag, V. Pendharkar, S. J. Tai, J. Hill, C. Dan, S. Y. Ho, W. W. Cheong, A. Poulsen, S. Blanchard, G. R. Lin, J. Alam, T. H. Keller, and P. Nordlund, "Fragment-Based ligand design of novel potent inhibitors of tankyrases," *J. Med. Chem.*, vol. 56, pp. 4497–4508, June 2013.

[66] L. Zhao, D. Cao, T. Chen, Y. Wang, Z. Miao, Y. Xu, W. Chen, X. Wang, Y. Li, Z. Du, B. Xiong, J. Li, C. Xu, N. Zhang, J. He, and J. Shen, "Fragment-based drug discovery of 2-thiazolidinones as inhibitors of the histone reader BRD4 bromodomain.," *Journal of medicinal chemistry*, vol. 56, pp. 3833–3851, May 2013.

[67] M. Frederickson, O. Callaghan, G. Chessari, M. Congreve, S. R. Cowan, J. E. Matthews, R. McMenamin, D.-M. Smith, M. Vinković, and N. G. Wallis, "Fragment-Based discovery of mexiletine derivatives as orally bioavailable inhibitors of Urokinase-Type plasminogen activator†," *J. Med. Chem.*, vol. 51, pp. 183–186, Jan. 2008.

[68] C. A. Berkeley, "Plexxikon announces FDA approval of zelboraf (vemurafenib) and companion diagnostic for the treatment of patients with BRAF Mutation-Positive metastatic melanoma," Aug. 2011.

[69] G. Bollag, P. Hirth, J. Tsai, J. Zhang, P. N. Ibrahim, H. Cho, W. Spevak, C. Zhang, Y. Zhang, G. Habets, E. A. Burton, B. Wong, G. Tsang, B. L. West, B. Powell, R. Shellooe, A. Marimuthu, H. Nguyen, K. Y. Zhang, D. R. Artis, J. Schlessinger, F. Su, B. Higgins, R. Iyer, K. D'Andrea, A. Koehler, M. Stumm, P. S. Lin, R. J. Lee, J. Grippo, I. Puzanov, K. B. Kim, A. Ribas, G. A. McArthur, J. A. Sosman, P. B. Chapman, K. T. Flaherty, X. Xu, K. L. Nathanson, and K. Nolop, "Clinical efficacy of a RAF inhibitor needs broad target blockade in BRAF-mutant melanoma.," *Nature*, vol. 467, pp. 596–599, Sept. 2010.

[70] D. Erlanson, "Practical fragments: Fragments in the clinic: 2015 edition." Online Blog, Jan. 2015.

[71] *Fragment-based Drug Discovery: Lessons and Outlook (Methods and Principles in Medicinal Chemistry)*. Wiley-VCH, 1 ed., Feb. 2016.

[72] P. J. Hajduk, G. Sheppard, D. G. Nettesheim, E. T. Olejniczak, S. B. Shuker, R. P. Meadows, D. H. Steinman, G. M. Carrera, P. A. Marcotte, J. Severin, K. Walter, H. Smith, E. Gubbins, R. Simmer, T. F. Holzman, D. W. Morgan, S. K. Davidsen, J. B. Summers, and S. W. Fesik, "Discovery of potent non-

peptide inhibitors of stromelysin using SAR by NMR," *J. Am. Chem. Soc.*, vol. 119, pp. 5818–5827, June 1997.

[73] G. E. de Kloe, D. Bailey, R. Leurs, and I. J. de Esch, "Transforming fragments into candidates: small becomes big in medicinal chemistry.," *Drug discovery today*, vol. 14, pp. 630–646, July 2009.

[74] J. Tsai, J. T. Lee, W. Wang, J. Zhang, H. Cho, S. Mamo, R. Bremer, S. Gillette, J. Kong, N. K. Haass, K. Sproesser, L. Li, K. S. M. Smalley, D. Fong, Y.-L. Zhu, A. Marimuthu, H. Nguyen, B. Lam, J. Liu, I. Cheung, J. Rice, Y. Suzuki, C. Luu, C. Settachatgul, R. Shellooe, J. Cantwell, S.-H. Kim, J. Schlessinger, K. Y. J. Zhang, B. L. West, B. Powell, G. Habets, C. Zhang, P. N. Ibrahim, P. Hirth, D. R. Artis, M. Herlyn, and G. Bollag, "Discovery of a selective inhibitor of oncogenic B-Raf kinase with potent antimelanoma activity," *Proceedings of the National Academy of Sciences*, vol. 105, pp. 3041–3046, Feb. 2008.

[75] L. Whitesell and S. L. Lindquist, "HSP90 and the chaperoning of cancer," *Nat Rev Cancer*, vol. 5, pp. 761–772, Oct. 2005.

[76] R. Law, O. Barker, J. Barker, T. Hesterkamp, R. Godemann, O. Andersen, T. Fryatt, S. Courtney, D. Hallett, and M. Whittaker, "The multiple roles of computational chemistry in fragment-based drug design," *Journal of computer-aided molecular design*, vol. 23, pp. 459–473, June 2009.

[77] J. Mortier, C. Rakers, R. Frederick, and G. Wolber, "Computational tools for in silico fragment-based drug design.," *Current topics in medicinal chemistry*, vol. 12, no. 17, pp. 1935–1943, 2012.

[78] T. Lengauer and M. Rarey, "Computational methods for biomolecular docking," *Current Opinion in Structural Biology*, vol. 6, pp. 402–406, June 1996.

[79] F. Jiang and S. H. Kim, ""soft docking": matching of molecular surface cubes.," *Journal of molecular biology*, vol. 219, pp. 79–102, May 1991.

[80] B. K. Shoichet, I. D. Kuntz, and D. L. Bodian, "Molecular docking using shape descriptors," *J. Comput. Chem.*, vol. 13, pp. 380–397, Apr. 1992.

[81] H.-J. Böhm, "The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure," *Journal of Computer-Aided Molecular Design*, vol. 8, pp. 243–256, June 1994.

[82] C. Levinthal, S. J. Wodak, P. Kahn, and A. K. Dadivanian, "Hemoglobin interaction in sickle cell fibers. i: Theoretical approaches to the molecular contacts," *Proceedings of the National Academy of Sciences*, vol. 72, pp. 1330–1334, Apr. 1975.

[83] Z. Zhou, A. K. Felts, R. A. Friesner, and R. M. Levy, "Comparative performance of several flexible docking programs and scoring functions: enrichment studies for a diverse set of pharmaceutically relevant targets," *J. Chem. Inf. Model.*, vol. 47, pp. 1599–1608, July 2007.

[84] X. Li, Y. Li, T. Cheng, Z. Liu, and R. Wang, "Evaluation of the performance of four molecular docking programs on a diverse set of protein-ligand complexes," *J. Comput. Chem.*, vol. 31, pp. 2109–2125, Aug. 2010.

[85] B. D. Bursulaya, M. Totrov, R. Abagyan, and C. L. Brooks, "Comparative study of several algorithms for flexible ligand docking.," *Journal of computer-aided molecular design*, vol. 17, pp. 755–763, Nov. 2003.

[86] M. Kontoyianni, L. M. McClellan, and G. S. Sokol, "Evaluation of docking performance: comparative data on docking algorithms," *J. Med. Chem.*, vol. 47, pp. 558–565, Jan. 2004.

[87] J. B. Cross, D. C. Thompson, B. K. Rai, J. C. Baber, K. Y. Fan, Y. Hu, and C. Humblet, "Comparison of several molecular docking programs: Pose prediction and virtual screening accuracy," *J. Chem. Inf. Model.*, vol. 49, pp. 1455–1474, June 2009.

[88] M. L. Verdonk, I. Giangreco, R. J. Hall, O. Korb, P. N. Mortenson, and C. W. Murray, "Docking performance of fragments and druglike compounds," *J. Med. Chem.*, vol. 54, pp. 5422–5431, Aug. 2011.

[89] M. L. Verdonk, P. N. Mortenson, R. J. Hall, M. J. Hartshorn, and C. W. Murray, "Protein-ligand docking against non-native protein conformers.," *Journal of chemical information and modeling*, vol. 48, pp. 2214–2225, Nov. 2008.

[90] Y. Li, L. Han, Z. Liu, and R. Wang, "Comparative assessment of scoring functions on an updated benchmark: 2. evaluation methods and general results," *J. Chem. Inf. Model.*, vol. 54, pp. 1717–1736, June 2014.

[91] M. Clark, F. Guarnieri, I. Shkurko, and J. Wiseman, "Grand canonical monte carlo simulation of LigandProtein binding," *J. Chem. Inf. Model.*, vol. 46, pp. 231–242, Jan. 2006.

[92] M. Clark, S. Meshkat, and J. S. Wiseman, "Grand canonical free-energy calculations of protein-ligand binding.," *Journal of chemical information and modeling*, vol. 49, pp. 934–943, Apr. 2009.

[93] G. A. Ross, M. S. Bodnarchuk, and J. W. Essex, "Water sites, networks, and free energies with grand canonical monte carlo," *J. Am. Chem. Soc.*, Oct. 2015.

[94] J. Seco, F. J. Luque, and X. Barril, "Binding site detection and druggability index from first principles," *J. Med. Chem.*, vol. 52, pp. 2363–2371, Apr. 2009.

[95] P. M. U. Ung, P. Ghanakota, S. E. Graham, K. W. Lexa, and H. A. Carlson, "Identifying binding hot spots on protein surfaces by mixed-solvent molecular dynamics: HIV-1 protease as a test case," *Biopolymers*, vol. 105, pp. 21–34, Jan. 2016.

[96] C. Faller, Raman, A. MacKerell, and O. Guvench, "Site identification by ligand competitive saturation (SILCS) simulations for Fragment-Based drug design," in *Fragment-Based Methods in Drug Discovery* (A. E. Klon, ed.), vol. 1289 of *Methods in Molecular Biology*, pp. 75–87, Springer New York, 2015.

[97] O. Guvench and A. D. MacKerell, "Computational fragment-based binding site identification by ligand competitive saturation.," *PLoS computational biology*, vol. 5, pp. e1000435+, July 2009.

[98] E. P. Raman, W. Yu, S. K. Lakkaraju, and A. D. MacKerell, "Inclusion of multiple fragment types in the site identification by ligand competitive saturation (SILCS) approach," *J. Chem. Inf. Model.*, vol. 53, pp. 3384–3398, Dec. 2013.

[99] S. K. Lakkaraju, E. P. Raman, W. Yu, and A. D. MacKerell, "Sampling of organic solutes in aqueous and heterogeneous environments using oscillating excess chemical potentials in grand canonical-like monte Carlo-Molecular dynamics simulations," *J. Chem. Theory Comput.*, vol. 10, pp. 2281–2290, June 2014.

[100] T. Imai, K. Oda, A. Kovalenko, F. Hirata, and A. Kidera, "Ligand mapping on protein surfaces by the 3D-RISM theory: Toward computational Fragment-Based drug design," *J. Am. Chem. Soc.*, vol. 131, pp. 12430–12440, Sept. 2009.

[101] C. Schubert and C. Stultz, "The multi-copy simultaneous search methodology: a fundamental tool for structure-based drug design," *Journal of Computer-Aided Molecular Design*, vol. 23, pp. 475–489, Aug. 2009.

[102] X. Kong and C. L. Brooks, "dynamics: A new approach to free energy calculations," *The Journal of Chemical Physics*, vol. 105, pp. 2414–2423, Aug. 1996.

[103] Z. Guo, C. L. Brooks, and X. Kong, "Efficient and flexible algorithm for free energy calculations using the -Dynamics approach," *J. Phys. Chem. B*, vol. 102, pp. 2032–2036, Mar. 1998.

[104] S. Banba, Z. Guo, and C. L. Brooks, "Efficient sampling of ligand orientations and conformations in free energy calculations using the -Dynamics method," *J. Phys. Chem. B*, vol. 104, pp. 6903–6910, July 2000.

[105] S. Banba and C. L. Brooks, "Free energy screening of small ligands binding to an artificial protein cavity," *The Journal of Chemical Physics*, vol. 113, pp. 3423–3433, Aug. 2000.

[106] J. L. Knight and C. L. Brooks, "-Dynamics free energy simulation methods," *J. Comput. Chem.*, vol. 30, pp. 1692–1700, Aug. 2009.

[107] K. A. Armacost, G. B. Goh, and C. L. Brooks, "Biasing potential replica exchange multisite -Dynamics for efficient free energy calculations," *J. Chem. Theory Comput.*, vol. 11, pp. 1267–1277, Mar. 2015.

[108] T. B. Steinbrecher, M. Dahlgren, D. Cappel, T. Lin, L. Wang, G. Krilov, R. Abel, R. Friesner, and W. Sherman, "Accurate binding free energy predictions in fragment optimization," *J. Chem. Inf. Model.*, vol. 55, pp. 2411–2420, Nov. 2015.

[109] Y. Fan, X. Chen, L. Yang, P. S. Cremer, and Y. Q. Gao, "On the structure of water at the Aqueous/Air interface," *J. Phys. Chem. B*, vol. 113, pp. 11672–11679, Aug. 2009.

[110] J. Kessler, H. Elgabarty, T. Spura, K. Karhan, P. Partovi-Azar, A. A. Hassanali, and T. D. Kühne, "Structure and dynamics of the instantaneous Water/Vapor interface revisited by Path-Integral and ab initio molecular dynamics simulations," *J. Phys. Chem. B*, vol. 119, pp. 10079–10086, Aug. 2015.

[111] F. Banhart, J. Kotakoski, and A. V. Krasheninnikov, "Structural defects in graphene," *ACS Nano*, vol. 5, pp. 26–41, Jan. 2011.

[112] M. Tang, L. Colombo, J. Zhu, and T. Diaz de la Rubia, "Intrinsic point defects in crystalline silicon: Tight-binding molecular dynamics studiesof self-diffusion, interstitial-vacancy recombination, and formation volumes," *Phys. Rev. B*, vol. 55, pp. 14279–14289, Jun 1997.

[113] D. Douguet, E. Thoreau, and G. Grassy, "A genetic algorithm for the automated generation of small organic molecules: Drug design using an evolutionary algorithm," *Journal of Computer-Aided Molecular Design*, vol. 14, pp. 449–466, July 2000.

[114] L. C. Blum and J.-L. Reymond, "970 million druglike small molecules for virtual screening in the chemical universe database GDB-13," *J. Am. Chem. Soc.*, vol. 131, pp. 8732–8733, July 2009.

[115] R. E. Carhart, D. H. Smith, H. Brown, and C. Djerassi, "Applications of artificial intelligence for chemical inference. XVII. approach to computer-

assisted elucidation of molecular structure," *J. Am. Chem. Soc.*, vol. 97, pp. 5755–5762, Oct. 1975.

[116] P. Ertl, S. Jelfs, J. Mühlbacher, A. Schuffenhauer, and P. Selzer, "Quest for the rings. in silico exploration of ring universe to identify novel bioactive heteroaromatic scaffolds.," *Journal of medicinal chemistry*, vol. 49, pp. 4568–4573, July 2006.

[117] D. Huang and A. Caflisch, "Library screening by fragment-based docking.," *Journal of molecular recognition : JMR*, vol. 23, pp. 183–193, Mar. 2010.

[118] A. R. Leach, B. K. Shoichet, and C. E. Peishoff, "Prediction of ProteinLigand interactions. docking and scoring: Successes and gaps," *Journal of Medicinal Chemistry*, vol. 49, pp. 5851–5855, Oct. 2006.

[119] R. Abagyan and M. Totrov, "High-throughput docking for lead generation," *Current Opinion in Chemical Biology*, vol. 5, pp. 375–382, Aug. 2001.

[120] A. Dhanik and L. E. Kavraki, "ProteinLigand interactions: Computational docking," 2001.

[121] T. Schulz-Gasch and M. Stahl, "Binding site characteristics in structure-based virtual screening: evaluation of current docking tools.," *Journal of molecular modeling*, vol. 9, pp. 47–57, Feb. 2003.

[122] R. Abagyan, M. Totrov, and D. Kuznetsov, "ICMa new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation," *J. Comput. Chem.*, vol. 15, pp. 488–506, May 1994.

[123] G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew, and A. J. Olson, "Automated docking using a lamarckian genetic algorithm and an empirical binding free energy function," *J. Comput. Chem.*, vol. 19, pp. 1639–1662, Nov. 1998.

[124] G. Jones, P. Willett, R. C. Glen, A. R. Leach, and R. Taylor, "Development and validation of a genetic algorithm for flexible docking," *Journal of Molecular Biology*, vol. 267, pp. 727–748, Apr. 1997.

[125] M. Rarey, B. Kramer, T. Lengauer, and G. Klebe, "A fast flexible docking method using an incremental construction algorithm," *Journal of Molecular Biology*, vol. 261, pp. 470–489, Aug. 1996.

[126] I. R. Craig, J. W. Essex, and K. Spiegel, "Ensemble docking into multiple crystallographically derived protein structures: An evaluation based on the statistical analysis of enrichments," *J. Chem. Inf. Model.*, vol. 50, pp. 511–524, Apr. 2010.

[127] S.-Y. Huang, S. Z. Grinter, and X. Zou, "Scoring functions and their evaluation methods for protein-ligand docking: recent advances and future directions," *Phys. Chem. Chem. Phys.*, vol. 12, pp. 12899–12908, 2010.

[128] J. Liu and R. Wang, "Classification of current scoring functions," *J. Chem. Inf. Model.*, vol. 55, pp. 475–482, Mar. 2015.

[129] A. Leach, *Molecular Modelling: Principles and Applications (2nd Edition)*. Prentice Hall, 2 ed., Apr. 2001.

[130] B. Widom, *Statistical Mechanics: A Concise Introduction for Chemists*. Cambridge University Press, 1 ed., May 2002.

[131] D. E. Shaw, "166 millisecond-long molecular dynamics simulations of proteins on a special-purpose machine," *Journal of Biomolecular Structure and Dynamics*, vol. 31, p. 108, Jan. 2013.

[132] V. A. Voelz, M. Jäger, S. Yao, Y. Chen, L. Zhu, S. A. Waldauer, G. R. Bowman, M. Friedrichs, O. Bakajin, L. J. Lapidus, S. Weiss, and V. S. Pande, "Slow unfolded-state structuring in Acyl-CoA binding protein folding revealed by simulation and experiment.," *Journal of the American Chemical Society*, vol. 134, pp. 12565–12577, Aug. 2012.

[133] X. Daura, A. E. Mark, and W. F. Van Gunsteren, "Parametrization of aliphatic CHn united atoms of GROMOS96 force field," *J. Comput. Chem.*, vol. 19, pp. 535–547, Apr. 1998.

[134] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. de Vries, "The MARTINI force field: coarse grained model for biomolecular simulations," *J. Phys. Chem. B*, vol. 111, pp. 7812–7824, July 2007.

[135] Y. Ueda, H. Taketomi, and N. G, "Studies on protein folding, unfolding, and fluctuations by computer simulation. II. a. three-dimensional lattice model of lysozyme," *Biopolymers*, vol. 17, pp. 1531–1548, June 1978.

[136] J. F. Dama, A. V. Sinitskiy, M. McCullagh, J. Weare, B. Roux, A. R. Dinner, and G. A. Voth, "The theory of Ultra-Coarse-graining. 1. general principles," *J. Chem. Theory Comput.*, vol. 9, pp. 2466–2480, May 2013.

[137] J. Kleinjung and F. Fraternali, "Design and application of implicit solvent models in biomolecular simulations," *Current Opinion in Structural Biology*, vol. 25, pp. 126–134, Apr. 2014.

[138] S. Takada, "Coarse-grained molecular simulations of large biomolecules," *Current Opinion in Structural Biology*, vol. 22, pp. 130–137, Apr. 2012.

[139] P. K. Weiner and P. A. Kollman, "AMBER: Assisted model building with energy refinement. a general program for modeling molecules and their interactions," *J. Comput. Chem.*, vol. 2, pp. 287–303, Sept. 1981.

[140] D. A. Pearlman, D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatham, S. DeBolt, D. Ferguson, G. Seibel, and P. Kollman, "AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules," *Computer Physics Communications*, vol. 91, pp. 1–41, Sept. 1995.

[141] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, "CHARMM: A program for macromolecular energy, minimization, and dynamics calculations," *J. Comput. Chem.*, vol. 4, pp. 187–217, June 1983.

[142] A. D. MacKerell, J. Wiorkiewicz-Kuczera, and M. Karplus, "An all-atom empirical energy function for the simulation of nucleic acids," *J. Am. Chem. Soc.*, vol. 117, pp. 11946–11975, Dec. 1995.

[143] W. L. Jorgensen and J. Tirado-Rives, "The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations

for crystals of cyclic peptides and crambin," *J. Am. Chem. Soc.*, vol. 110, pp. 1657–1666, Mar. 1988.

[144] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, "Development and testing of the OPLS All-Atom force field on conformational energetics and properties of organic liquids," *J. Am. Chem. Soc.*, vol. 118, pp. 11225–11236, Jan. 1996.

[145] W. R. P. Scott, P. H. Hünenberger, I. G. Tironi, A. E. Mark, S. R. Billeter, J. Fennen, A. E. Torda, T. Huber, P. Krüger, and W. F. van Gunsteren, "The GROMOS biomolecular simulation program package," *J. Phys. Chem. A*, vol. 103, pp. 3596–3607, May 1999.

[146] N. Schmid, A. P. Eichenberger, A. Choutko, S. Riniker, M. Winger, A. E. Mark, and W. F. van Gunsteren, "Definition and testing of the GROMOS force-field versions 54A7 and 54B7.," *European biophysics journal : EBJ*, vol. 40, pp. 843–856, July 2011.

[147] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, "A second generation force field for the simulation of proteins, nucleic acids, and organic molecules," *J. Am. Chem. Soc.*, vol. 117, pp. 5179–5197, May 1995.

[148] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, "Comparison of multiple amber force fields and development of improved protein backbone parameters.," *Proteins*, vol. 65, pp. 712–725, Nov. 2006.

[149] C. Oostenbrink, A. Villa, A. E. Mark, and W. F. Van Gunsteren, "A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6," *J. Comput. Chem.*, vol. 25, pp. 1656–1676, Oct. 2004.

[150] D. A. Case, T. A. Darden, T. E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A. W. Götz, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, R. M. Wolf, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M. J. Hsieh, G. Cui,

D. R. Roe, D. H. Mathews, M. G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P. A. Kollman, "Amber 12," 2012.

[151] "The energy function - CHARMM tutorial." url=http://www.charmmtutorial.org/index.php/The_Energy_Function.

[152] N. L. Allinger, Y. H. Yuh, and J. H. Lii, "Molecular mechanics. the MM3 force field for hydrocarbons. 1," *J. Am. Chem. Soc.*, vol. 111, pp. 8551–8566, Nov. 1989.

[153] J. H. Lii and N. L. Allinger, "Molecular mechanics. the MM3 force field for hydrocarbons. 2. vibrational frequencies and thermodynamics," *J. Am. Chem. Soc.*, vol. 111, pp. 8566–8575, Nov. 1989.

[154] J. H. Lii and N. L. Allinger, "Molecular mechanics. the MM3 force field for hydrocarbons. 3. the van der waals' potentials and crystal data for aliphatic and aromatic hydrocarbons," *J. Am. Chem. Soc.*, vol. 111, pp. 8576–8582, Nov. 1989.

[155] J. W. Ponder, C. Wu, P. Ren, V. S. Pande, J. D. Chodera, M. J. Schnieders, I. Haque, D. L. Mobley, D. S. Lambrecht, R. A. DiStasio, M. Head-Gordon, G. N. Clark, M. E. Johnson, and T. Head-Gordon, "Current status of the AMOEBA polarizable force field.," *The journal of physical chemistry. B*, vol. 114, pp. 2549–2564, Mar. 2010.

[156] J. Wang, W. Wang, P. A. Kollman, and D. A. Case, "Automatic atom type and bond type perception in molecular mechanical calculations.," *Journal of molecular graphics & modelling*, vol. 25, pp. 247–260, Oct. 2006.

[157] K. Vanommeslaeghe and A. D. MacKerell, "Automation of the CHARMM general force field (CGenFF) i: bond perception and atom typing.," *Journal of chemical information and modeling*, vol. 52, pp. 3144–3154, Dec. 2012.

[158] K. Vanommeslaeghe, E. P. Raman, and A. D. MacKerell, "Automation of the CHARMM general force field (CGenFF) II: assignment of bonded parameters and partial atomic charges.," *Journal of chemical information and modeling*, vol. 52, pp. 3155–3168, Dec. 2012.

[159] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, "Development and testing of a general amber force field," *J. Comput. Chem.*, vol. 25, pp. 1157–1174, July 2004.

[160] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, and A. D. Mackerell, "CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields," *J. Comput. Chem.*, vol. 31, pp. 671–690, Mar. 2010.

[161] D. A. Case, T. Darden, T. E. Cheatham, C. Simmerling, A. Roitberg, J. Wang, R. E. Duke, R. Luo, D. R. Roe, R. C. Walker, S. LeGrand, J. Swails, D. Cerutti, J. Kaus, R. Betz, R. M. Wolf, K. M. Merz, G. Seabra, P. Janowski, A. W. Götz, I. Kolossváry, F. Paesani, J. Liu, X. Wu, T. Steinbrecher, H. Gohlke, N. Homeyer, Q. Cai, W. Smith, D. Mathews, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, J. Berryman, and P. A. Kollman, "Amber 14 reference manual."

[162] M. Orsi and J. W. Essex, "The ELBA force field for Coarse-Grain modeling of lipid membranes," *PLoS ONE*, vol. 6, pp. e28637+, Dec. 2011.

[163] A. Warshel, M. Kato, and A. V. Pisliakov, "Polarizable force fields: history, test cases, and prospects," *J. Chem. Theory Comput.*, vol. 3, pp. 2034–2045, Nov. 2007.

[164] P. E. M. Lopes, J. Huang, J. Shim, Y. Luo, H. Li, B. Roux, and A. D. MacKerell, "Polarizable force field for peptides and proteins based on the classical drude oscillator," *J. Chem. Theory Comput.*, vol. 9, pp. 5430–5449, Dec. 2013.

[165] A. Hinchliffe, *Molecular Modelling for Beginners*. Wiley, 2 ed., Dec. 2008.

[166] S. Genheden, A. I. Cabedo Martinez, M. P. Criddle, and J. W. Essex, "Extensive all-atom monte carlo sampling and QM/MM corrections in the SAMPL4 hydration free energy challenge.," *Journal of computer-aided molecular design*, vol. 28, pp. 187–200, Mar. 2014.

[167] R. Eckhardt, "Stan ulam, john von neumann, and the monte carlo method." url=http://library.sciencemadness.org/lanl1_a/lib-www/pubs/00326867.pdf, 1987.

[168] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines," *The Journal of Chemical Physics*, vol. 21, pp. 1087–1092, June 1953.

[169] W. K. Hastings, "Monte carlo sampling methods using markov chains and their applications," *Biometrika*, vol. 57, pp. 97–109, Apr. 1970.

[170] D. Frenkel and B. Smit, *Understanding Molecular Simulation*. Computational Science Series, Academic Press - Elsevier, second ed., 2002.

[171] R. Baierlein, "The elusive chemical potential," *American Journal of Physics*, vol. 69, pp. 423–434, Apr. 2001.

[172] P. W. Atkins and J. De Paula, *The Elements of Physical Chemistry*. Oxford Univ Pr, 5 ed., Oct. 2009.

[173] G. C. Lynch and B. M. Pettitt, "Grand canonical ensemble molecular dynamics simulations: Reformulation of extended system dynamics approaches," *The Journal of Chemical Physics*, vol. 107, pp. 8594–8610, Nov. 1997.

[174] C. Lo and B. Palmer, "Alternative hamiltonian for molecular dynamics simulations in the grand canonical ensemble," *The Journal of Chemical Physics*, vol. 102, pp. 925–931, Jan. 1995.

[175] *Free Energy Calculations: Theory and Applications in Chemistry and Biology (Springer Series in Chemical Physics)*. Springer, Sept. 2007.

[176] R. W. Zwanzig, "HighTemperature equation of state by a perturbation method. i. nonpolar gases," *The Journal of Chemical Physics*, vol. 22, pp. 1420–1426, Aug. 1954.

[177] D. A. Pearlman, "A comparison of alternative approaches to free energy calculations," *J. Phys. Chem.*, vol. 98, pp. 1487–1493, Feb. 1994.

[178] J. Michel, M. L. Verdonk, and J. W. Essex, "ProteinLigand complexes: computation of the relative free energy of different scaffolds and binding modes," *J. Chem. Theory Comput.*, vol. 3, pp. 1645–1655, Sept. 2007.

[179] M. Zacharias, T. P. Straatsma, and J. A. McCammon, "Separationshifted scaling, a new scaling method for LennardJones interactions in thermodynamic integration," *The Journal of Chemical Physics*, vol. 100, pp. 9025–9031, June 1994.

[180] S. Boresch, F. Tettinger, M. Leitgeb, and M. Karplus, "Absolute binding free energies: a quantitative approach for their calculation," *J. Phys. Chem. B*, vol. 107, pp. 9535–9551, Sept. 2003.

[181] C. J. Woods, J. Michel, M. Bodnarchuk, S. Genheden, R. Bradshaw, G. Ross, C. Cave-Ayland, A. I. Cabedo Martinez, and J. Graham, *ProtoMS documentation ProtoMS 3.2*, 2015.

[182] M. Mezei and D. L. Beveridge, "Free energy simulations.," *Annals of the New York Academy of Sciences*, vol. 482, pp. 1–23, 1986.

[183] M. R. Shirts and J. D. Chodera, "Statistically optimal analysis of samples from multiple equilibrium states.," *The Journal of chemical physics*, vol. 129, pp. 124105+, Sept. 2008.

[184] M. Mezei, "The finite difference thermodynamic integration, tested on calculating the hydration free energy difference between acetone and dimethylamine in water," *The Journal of Chemical Physics*, vol. 86, pp. 7084–7088, June 1987.

[185] C. H. Bennett, "Efficient estimation of free energy differences from monte carlo data," *Journal of Computational Physics*, vol. 22, pp. 245–268, Oct. 1976.

[186] A. de Ruiter, S. Boresch, and C. Oostenbrink, "Comparison of thermodynamic integration and bennett acceptance ratio for calculating relative protein-ligand binding free energies.," *Journal of computational chemistry*, vol. 34, pp. 1024–1034, May 2013.

[187] P. Bruscolini, "A coarse-grained, realistic model for protein folding," *The Journal of Chemical Physics*, vol. 107, pp. 7512–7529, Nov. 1997.

[188] C. Clementi, "Coarse-grained models of protein folding: toy models or predictive tools?," *Current Opinion in Structural Biology*, vol. 18, pp. 10–15, Feb. 2008.

[189] R. H. Swendsen and J. S. Wang, "Replica monte carlo simulation of spin glasses.," *Physical review letters*, vol. 57, pp. 2607–2609, Nov. 1986.

[190] Y. Sugita and Y. Okamoto, "Replica-exchange molecular dynamics method for protein folding," *Chemical Physics Letters*, vol. 314, pp. 141–151, Nov. 1999.

[191] A. Laio and M. Parrinello, "Escaping free-energy minima," *Proceedings of the National Academy of Sciences*, vol. 99, pp. 12562–12566, Oct. 2002.

[192] R. C. Bernardi, M. C. R. Melo, and K. Schulten, "Enhanced sampling techniques in molecular dynamics simulations of biological systems," *Biochimica et Biophysica Acta (BBA) - General Subjects*, vol. 1850, pp. 872–877, May 2015.

[193] C. J. Woods, J. W. Essex, and M. A. King, "The development of Replica-Exchange-based Free-Energy methods," *J. Phys. Chem. B*, vol. 107, pp. 13703–13710, Dec. 2003.

[194] L. Wang, R. A. Friesner, and B. J. Berne, "Replica exchange with solute scaling: A more efficient version of replica exchange with solute tempering (REST2)," *J. Phys. Chem. B*, vol. 115, pp. 9431–9438, Aug. 2011.

[195] A. Barducci, M. Bonomi, and M. Parrinello, "Metadynamics," *WIREs Comput Mol Sci*, vol. 1, pp. 826–843, Sept. 2011.

[196] R. Martonak, A. Laio, M. Bernasconi, C. Ceriani, P. Raiteri, and M. Parrinello, "Simulation of structural phase transitions by metadynamics," Nov. 2004.

[197] V. S. Pande, K. Beauchamp, and G. R. Bowman, "Everything you wanted to know about markov state models but were afraid to ask.," *Methods (San Diego, Calif.)*, vol. 52, pp. 99–105, Sept. 2010.

[198] J. D. Chodera and F. Noé, "Markov state models of biomolecular conformational dynamics.," *Current opinion in structural biology*, vol. 25, pp. 135–144, Apr. 2014.

[199] I. Buch, T. Giorgino, and G. De Fabritiis, "Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations," *Proceedings of the National Academy of Sciences*, vol. 108, pp. 10184–10189, June 2011.

[200] J. Michel, J. Tirado-Rives, and W. L. Jorgensen, "Prediction of the water content in protein binding sites," *J. Phys. Chem. B*, vol. 113, pp. 13337–13346, Oct. 2009.

[201] M. Bodnarchuk, "Predicting the location and binding affinity of small molecules in protein binding sites." September 2012.

[202] G. A. Ross, G. M. Morris, and P. C. Biggin, "Rapid and accurate prediction and scoring of water molecules in protein binding sites," *PLoS ONE*, vol. 7, pp. e32036+, Mar. 2012.

[203] D. L. Mobley and P. V. Klimovich, "Perspective: Alchemical free energy calculations for drug discovery," *The Journal of Chemical Physics*, vol. 137, pp. 230901+, Dec. 2012.

[204] W. Jiang, M. Hodoscek, and B. Roux, "Computation of absolute hydration and binding free energy with free energy perturbation distributed Replica-Exchange molecular dynamics," *J. Chem. Theory Comput.*, vol. 5, pp. 2583–2588, Oct. 2009.

[205] M. M. Teeter, "Water-protein interactions: theory and experiment.," *Annual review of biophysics and biophysical chemistry*, vol. 20, pp. 577–600, 1991.

[206] A. Jakalian, B. L. Bush, D. B. Jack, and C. I. Bayly, "Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. method," *J. Comput. Chem.*, vol. 21, pp. 132–146, Jan. 2000.

[207] A. Jakalian, D. B. Jack, and C. I. Bayly, "Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. parameterization and validation.," *Journal of computational chemistry*, vol. 23, pp. 1623–1641, Dec. 2002.

[208] 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.," *Journal of computational chemistry*, vol. 25, pp. 1605–1612, Oct. 2004.

[209] Z. Zhang, Y. Li, B. Lin, M. Schroeder, and B. Huang, "Identification of cavities on protein surface using multiple computational approaches for drug binding site prediction," *Bioinformatics*, vol. 27, pp. 2083–2088, Aug. 2011.

[210] K. P. Tan, R. Varadarajan, and M. S. Madhusudhan, "DEPTH: a web server to compute depth and predict small-molecule binding cavities in proteins," *Nucleic Acids Research*, vol. 39, pp. W242–W248, July 2011.

[211] S. H. P. Oliveira, F. A. N. Ferraz, R. V. Honorato, J. A. Xavier-Neto, T. J. P. Sobreira, and P. S. L. de Oliveira, "KVFinder: steered identification of protein cavities as a PyMOL plugin," *BMC Bioinformatics*, vol. 15, pp. 197+, June 2014.

[212] M. Hendlich, F. Rippmann, and G. Barnickel, "LIGSITE: automatic and efficient detection of potential small molecule-binding sites in proteins," *Journal of Molecular Graphics and Modelling*, vol. 15, pp. 359–363, Dec. 1997.

[213] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. Cambridge University Press, 1 ed., July 2008.

[214] L. H. Weaver and B. W. Matthews, "Structure of bacteriophage t4 lysozyme refined at 1.7 å resolution," *Journal of Molecular Biology*, vol. 193, pp. 189–199, Jan. 1987.

[215] M. Matsumura, J. A. Wozniak, D. P. Sun, and B. W. Matthews, "Structural studies of mutants of t4 lysozyme that alter hydrophobic stabilization.," *Journal of Biological Chemistry*, vol. 264, pp. 16059–16066, Sept. 1989.

[216] J. H. Hurley, W. A. Baase, and B. W. Matthews, "Design and structural analysis of alternative hydrophobic core packing arrangements in bacteriophage t4 lysozyme," *Journal of Molecular Biology*, vol. 224, pp. 1143–1159, Apr. 1992.

[217] P. Billsten, M. Wahlgren, T. Arnebrant, J. McGuire, and H. Elwing, "Structural changes of t4 lysozyme upon adsorption to silica nanoparticles measured by circular dichroism," *Journal of Colloid and Interface Science*, vol. 175, pp. 77–82, Oct. 1995.

[218] I. R. Vetter, W. A. Baase, D. W. Heinz, J. P. Xiong, S. Snow, and B. W. Matthews, "Protein structural plasticity exemplified by insertion and deletion mutants in t4 lysozyme.," *Protein science : a publication of the Protein Society*, vol. 5, pp. 2399–2415, Dec. 1996.

[219] D. P. Sun, U. Sauer, H. Nicholson, and B. W. Matthews, "Contributions of engineered surface salt bridges to the stability of t4 lysozyme determined by directed mutagenesis," *Biochemistry*, vol. 30, pp. 7142–7153, July 1991.

[220] W. A. Baase, L. Liu, D. E. Tronrud, and B. W. Matthews, "Lessons from the lysozyme of phage t4.," *Protein science : a publication of the Protein Society*, vol. 19, pp. 631–641, Apr. 2010.

[221] E. Baldwin, W. A. Baase, X.-j. Zhang, V. Feher, and B. W. Matthews, "Generation of ligand binding sites in t4 lysozyme by deficiency-creating substitutions," *Journal of Molecular Biology*, vol. 277, pp. 467–485, Mar. 1998.

[222] Y. Deng and B. Roux, "Calculation of standard binding free energies: aromatic molecules in the t4 lysozyme L99A mutant," *J. Chem. Theory Comput.*, vol. 2, pp. 1255–1273, Sept. 2006.

[223] D. L. Mobley, A. P. Graves, J. D. Chodera, A. C. McReynolds, B. K. Shoichet, and K. A. Dill, "Predicting absolute ligand binding free energies to a simple model site," *Journal of Molecular Biology*, vol. 371, pp. 1118–1134, Aug. 2007.

[224] K. Haider and D. J. Huggins, "Combining solvent thermodynamic profiles with functionality maps of the hsp90 binding site to predict the displacement of water molecules.," *Journal of chemical information and modeling*, vol. 53, pp. 2571–2586, Oct. 2013.

[225] C. E. Stebbins, A. A. Russo, C. Schneider, N. Rosen, Hartl, and N. P. Pavletich, "Crystal structure of an Hsp90Geldanamycin complex: Targeting of a protein chaperone by an antitumor agent," *Cell*, vol. 89, pp. 239–250, Apr. 1997.

[226] Y. E. Kim, M. S. Hipp, A. Bracher, M. Hayer-Hartl, and F. U. Hartl, "Molecular chaperone functions in protein folding and proteostasis.," *Annual review of biochemistry*, vol. 82, pp. 323–355, 2013.

[227] G. I. Shapiro, "Cyclin-Dependent kinase pathways as targets for cancer treatment," *Journal of Clinical Oncology*, vol. 24, pp. 1770–1783, Apr. 2006.

[228] P. D. Jeffrey, A. Russo, A., K. Polyak, E. Gibbs, J. Hurwitz, J. Massague, N. Pavletich, and P., "Mechanism of CDK activation revealed by the structure of a cyclinA-CDK2 complex," *Nature*, vol. 376, pp. 313–320, July 1995.

[229] H. Chen, P. D. Lyne, F. Giordanetto, T. Lovell, and J. Li, "On evaluating Molecular-Docking methods for pose prediction and enrichment factors," *J. Chem. Inf. Model.*, vol. 46, pp. 401–415, Jan. 2006.

[230] T. A. Halgren, R. B. Murphy, R. A. Friesner, H. S. Beard, L. L. Frye, W. T. Pollard, and J. L. Banks, "Glide: a new approach for rapid, accurate docking and scoring. 2. enrichment factors in database screening," *J. Med. Chem.*, vol. 47, pp. 1750–1759, Mar. 2004.

[231] 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, pp. 1–12, 2009.

[232] D. J. Huggins, "Quantifying the entropy of binding for water molecules in protein cavities by computing correlations," *Biophysical Journal*, vol. 108, pp. 928–936, Feb. 2015.

[233] P. Setny, "Prediction of water binding to protein hydration sites with a discrete, semiexplicit solvent model," *J. Chem. Theory Comput.*, Oct. 2015.

[234] G. Gerogiokas, M. W. Y. Southey, M. P. Mazanetz, A. Hefeitz, M. Bodkin, R. J. Law, and J. Michel, "Evaluation of water displacement energetics in protein binding sites with grid cell theory," *Phys. Chem. Chem. Phys.*, vol. 17, no. 13, pp. 8416–8426, 2015.

[235] J. R. Tame, G. N. Murshudov, E. J. Dodson, T. K. Neil, G. G. Dodson, C. F. Higgins, and A. J. Wilkinson, "The structural basis of sequence-independent peptide binding by OppA protein," *Science*, vol. 264, pp. 1578–1581, June 1994.

[236] H. Lei and Y. Duan, "Improved sampling methods for molecular simulation," *Current Opinion in Structural Biology*, vol. 17, pp. 187–191, Apr. 2007.

[237] V. Spiwok, Z. Sucur, and P. Hosek, "Enhanced sampling techniques in biomolecular simulations," *Biotechnology Advances*, vol. 33, pp. 1130–1140, Nov. 2015.

[238] R. J. Bingham, J. B. C. Findlay, S.-Y. Hsieh, A. P. Kalverda, A. Kjellberg, C. Perazzolo, S. E. V. Phillips, K. Seshadri, C. H. Trinh, W. B. Turnbull, G. Bodenhausen, and S. W. Homans, "Thermodynamics of binding of 2-Methoxy-3-isopropylpyrazine and 2-Methoxy-3-isobutylpyrazine to the major urinary protein," *J. Am. Chem. Soc.*, vol. 126, pp. 1675–1681, Feb. 2004.

[239] P. G. Wyatt, A. J. Woodhead, V. Berdini, J. A. Boulstridge, M. G. Carr, D. M. Cross, D. J. Davis, L. A. Devine, T. R. Early, R. E. Feltell, E. J. Lewis, R. L. McMenamin, E. F. Navarro, M. A. O'Brien, M. O'Reilly, M. Reule, G. Saxty, L. C. A. Seavers, D.-M. Smith, M. S. Squires, G. Trewartha, M. T. Walker, and A. J. A. Woolford, "Identification of N-(4-Piperidinyl)-4-(2,6-dichlorobenzoylamino)-1H-pyrazole-3-carboxamide (AT7519), a novel cyclin dependent kinase inhibitor using Fragment-Based X-Ray crystallography and structure based drug design†," *J. Med. Chem.*, vol. 51, pp. 4986–4999, Aug. 2008.

[240] M. Kolář and P. Hobza, "On extension of the current biomolecular empirical force field for the description of halogen bonds," *J. Chem. Theory Comput.*, vol. 8, pp. 1325–1333, Apr. 2012.

[241] R. Wilcken, M. O. Zimmermann, A. Lange, A. C. Joerger, and F. M. Boeckler, "Principles and applications of halogen bonding in medicinal chemistry and chemical biology," *J. Med. Chem.*, vol. 56, pp. 1363–1388, Feb. 2013.

[242] D. E. Timm, L. J. Baker, H. Mueller, L. Zidek, and M. V. Novotny, "Structural basis of pheromone binding to mouse major urinary protein (MUP-i)," *Protein Science*, vol. 10, pp. 997–1004, May 2001.

[243] T. Zhou and A. Caflisch, "Support information high-throughput virtual screening using quantum mechanical probes: Discovery of selective kinase inhibitors," 2010.

[244] M. Repasky, R. Murphy, J. Banks, J. Greenwood, I. Tubert-Brohman, S. Bhat, and R. Friesner, "Docking performance of the glide program as evaluated on the astex and DUD datasets: a complete set of glide SP results and selected results for a new scoring function integrating WaterMap and glide," *Journal of computer-aided molecular design*, vol. 26, pp. 787–799, June 2012.

[245] T. Cheng, Z. Liu, and R. Wang, "A knowledge-guided strategy for improving the accuracy of scoring functions in binding affinity prediction," *BMC Bioinformatics*, vol. 11, no. 1, pp. 1–16, 2010.

[246] F. Ding, S. Yin, and N. V. Dokholyan, "Rapid flexible docking using a stochastic rotamer library of ligands," *J. Chem. Inf. Model.*, vol. 50, pp. 1623–1632, Sept. 2010.

[247] S. Perez-Miller, Q. Zou, M. V. Novotny, and T. D. Hurley, "High resolution x-ray structures of mouse major urinary protein nasal isoform in complex with pheromones.," *Protein science : a publication of the Protein Society*, vol. 19, pp. 1469–1479, Aug. 2010.

[248] H. J. Motulsky, *"The binomial test", GraphPad Statistics Guide*, Accessed April 2016.

[249] O. Korb, T. Stützle, and T. E. Exner, "Empirical scoring functions for advanced protein-ligand docking with PLANTS.," *Journal of chemical information and modeling*, vol. 49, pp. 84–96, Jan. 2009.

[250] M. S. Bodnarchuk, R. Viner, J. Michel, and J. W. Essex, "Strategies to calculate water binding free energies in ProteinLigand complexes," *J. Chem. Inf. Model.*, vol. 54, pp. 1623–1633, June 2014.

[251] Cambridge Crystallographic Data Centre, *GOLD configuration file user guide. A component of the GOLD Suite*, 2015.