Osaka University Knowledge Archive

Title	Complex Structure and Affinity Prediction of a Flexible Protein Receptor and its Inhibitor using Molecular Dynamics Simulations
Author(s)	Bekker, Gert-jan
Citation	大阪大学, 2017, 博士論文
Version Type	VoR
URL	https://doi.org/10.18910/61877
rights	
Note	

Osaka University Knowledge Archive : OUKA

https://ir.library.osaka-u.ac.jp/

Osaka University

Complex Structure and Affinity Prediction of a Flexible Protein Receptor and its Inhibitor using Molecular Dynamics Simulations

Gert-Jan Bekker

Laboratory of Protein Informatics (Prof. Nakamura)
Institute for Protein Research
Graduate School of Frontier Biosciences
Osaka University

March 2017

ABSTRACT

In order to predict the accurate binding configuration as well as the binding affinity for a flexible protein and its inhibitor drug, enhanced sampling with multicanonical molecular dynamics (McMD) simulations and thermodynamic integration (Therm. Int.) were combined as a general drug docking method. Cyclin-dependent kinase 2 (CDK2) is involved in the cell cycle regulation. Malfunctions in CDK2 are thought to cause tumorigenesis, and is thus a potential drug lead. Here, a long McMD simulation for docking the inhibitor CS3 to CDK2 starting from the unbound configuration was performed. McMD can explore a wide conformational space without conformational trapping at local minima by applying a bias to the system. This bias enables uniform sampling over a specific energy range, facilitating a random walk within this energy space, where each energetic state is equally probable to be sampled. From the multicanonical ensemble, the canonical, i.e. physically accepted ensemble, can be obtained via the reweighing process. Using McMD, stable binding configurations can thus be obtained, while the simulation is not trapped in local energy minima. However, due to the limited amount of unbound structures in the canonical ensemble, the affinity is difficult to determine from the McMD simulations. Therm. Int. however is very effective at calculating the affinity by measuring the average forces on the ligand along the binding/unbinding pathway, where integration of these forces produces the binding free energy. Using the multicanonical ensemble, the predicted bound complex, a potential binding/unbinding pathway connecting the predicted bound complex with the unbound one, and stable intermediary structures along this pathway were obtained. Subsequently, the forces along this pathway were sampled, starting from these stable intermediary structures. Finally, the binding free energy was readily computed by Therm. Int. Using this combination, the correct binding configuration of CS3 to CDK2 was predicted, and their affinity coincided well with the experimental value.

TABLE OF CONTENTS

1 INTRODUCTION 4
1.1 The role of Molecular Dynamics 4
1.2 Enhanced sampling 6
1.3 Affinity calculation using MD 7
1.4 The role of Cyclin Dependent Kinase 2 9
1.5 Objective 11
2 MATERIALS & METHODS 12
2.1 Molecular docking using MD simulations 12
2.1.1 (1) Estimation of the space restraining cylinder λ 12
2.1.2 (2) McMD simulations for flexible ligand docking 14
2.1.3 (3) Estimation of the reaction coordinate ξ 15
2.1.4 (4) Binding free energy calculation by Thermodynamic Integration along ξ 17
2.2 Simulation details 20
2.3 McMD simulations 23
2.4 Advanced visualization using Molmil 24
3 RESULTS 26
3.1 Selection of λ by RAMD 26
3.2 Docking of CS3 to CDK2 by McMD 26
3.3 Binding affinity calculation of CS3 to CDK2 by Thermodynamic Integration 29
4 DISCUSSION 31
5 CONCLUSION 38
6 FIGURES 39
8 REFERENCES 56
9 ACADEMIC ACCOMPLISHMENTS 66
10 ACKNOWLEDGEMENTS 68

1 INTRODUCTION

1.1 The role of Molecular Dynamics

Proteins are biological systems' main workforce; they take care of processes ranging from intracellular communication to replication. Without proteins, life as we know it would not exist. Due to the importance of proteins in biological processes, the Protein Data Bank (PDB) was founded in 1971 at Brookhaven National Laboratory ("Protein Data Bank," 1971) and started with 7 structures part of the databank. Protein Data Bank Europe (PDBe) was founded in 1995 at the European Bioinformatics Institute (Velankar et al., 2005). Similarly, in 2001 Protein Data Bank Japan (PDBj) was founded at the Institute for Protein Research at Osaka University (Standley, Kinjo, Kinoshita, & Nakamura, 2008). In 2003 these three groups united and formed the wwPDB (World Wide Protein Data Bank) (Berman, Henrick, & Nakamura, 2003), in order to maintain the PDB as a single archive. Over the years, the collaboration has expanded, however each site still maintains their own web interface to the PDB with customized services to search, explore and analyze the databank. The structures part of the PDB are obtained by experimental methods, primarily from X-ray crystallography, NMR Nuclear Magnetic Resonance (NMR) spectroscopy and Electron Microscopy (EM). These structures however, only represent a single still snapshot of what the molecule might look like as an averaged stable conformation under certain experimental conditions. However, in vitro and more importantly in vivo, proteins exist under significantly different conditions. These conditions also enable the protein to fluctuate significantly, continually changing its conformation. These effects are not included in the structures contained in the PDB, but are important to the biological function of the protein. Molecular Dynamics (MD) simulations enables us to study the dynamics of proteins and how they might behave in an environment close to physiological conditions. By using MD, it becomes possible to study small and large scale dynamics of proteins, as well as how multiple molecules might interact with each other. Thus, MD is an important tool to help us understand

how biological processes are performed by proteins and how they can be influenced by external factors such as drugs.

Molecular docking between a protein and small chemical compounds is one of the key tools in virtual screening from ligand databases. So far, many “flexible docking” methods have been developed, where only the ligand molecules are treated as flexible, leaving the protein essentially “rigid”, such as DOCK (Kuntz, Blaney, Oatley, Langridge, & Ferrin, 1982), AutoDock (Goodsell & Olson, 1990), Flex (Rarey, Kramer, Lengauer, & Klebe, 1996), GOLD (Jones, Willett, Glen, Leach, & Taylor, 1997), and myPresto/sievgene (Fukunishi, Mikami, & Nakamura, 2005). These programs attempt to make a rapid search to find the correct docking configurations for millions of ligand candidates by reducing computational costs. Since the screening result strongly depends on the structure of the proteins, the “ensemble docking” method has been developed. It is a common strategy to use MD to generate an ensemble of protein conformations (Craig, Essex, & Spiegel, 2010; Wada, Kanamori, Nakamura, & Fukunishi, 2011), after which rigid body docking is applied to generate docking poses. The docking poses are finally ranked by some empirical or physicochemical scores using shape complementarity and electrostatics. While these approaches can take protein flexibility into account, the conformational space searched by such ensemble docking methods is limited. In addition, the scores lack accuracy due to the inclusion of many approximations and empirical parameters, because they are designed for rapid screening. Thus, the quantitatively correct affinities between the protein and ligands are difficult to obtain, so that researchers have to be satisfied by the qualitative, relative affinities. Therefore, a method that can predict both the accurate protein-ligand complex and its quantitatively correct affinity is needed, where the protein is treated as flexible in addition to using flexible ligands.

1.2 Enhanced sampling

Exhaustively sampling the protein and ligand structures by conventional MD simulations to obtain the docking configuration and binding free energy may be one solution to the above problem, but it is impractical, even nowadays. The reason is that binding of a ligand to its target protein generally takes in the order of micro to milliseconds. In order to properly sample a natively bound structure, multiple binding-unbinding events are required, and so the total simulation time becomes in the order of seconds to minutes. These timescales are far outside the realm of even the fastest MD optimized supercomputer in existence (i.e. the ANTON system (Shaw et al., 2008)). The rate-limiting process is conformational trapping of a ligand at many local energy minima. In order to overcome this trapping, several biasing methods, which add weighted probabilities, have been designed to improve the sampling efficiency. This has been realized by generalized ensemble and biasing methods, such as Multicanonical MD (McMD) (Nakajima, Nakamura, & Kidera, 1997), Replica Exchange MD (REMD) (Sugita & Okamoto, 1999), Filling Potential (Fukunishi, Mikami, & Nakamura, 2003), Meta-dynamics (Laio & Parrinello, 2002), Adaptive Umbrella Sampling (AUS) (Bartels & Karplus, 1997; Higo et al., 2015) and Adaptive Biasing Force (ABF) (Darve, Rodriguez-Gomez, & Pohorille, 2008).

In the McMD method, a bias is applied to the system to enable uniform sampling over a specific energy range, facilitating a random walk within the energy space, where each energetic state is equally probable to be sampled. This bias enables sampling at higher energy regions, similar to a high temperature simulation, and at lower energy regions, similar to a low temperature simulation, resulting in exploring a wide conformational space without conformational trapping at local minima. A merit of the McMD method is that the physically accepted ensemble, i.e. the canonical ensemble, can be obtained by the reweighting procedure (Nakajima, Nakamura, et al., 1997). The McMD method can continuously switch between the higher and lower temperature regions defined by the bias, and it has been widely applied for

enhanced conformational sampling (Kamiya, Higo, & Nakamura, 2002) and docking (Kamiya, Yonezawa, Nakamura, & Higo, 2008) simulations. The McMD method first forms the multicanonical ensemble, which covers a wide temperature range. Then, the conformational ensemble at 300 K is obtained by the reweighing method, followed by calculating the Potential of Mean Force (PMF) from the probability distribution of the sampled structures. Finally, the free energy landscape (FEL) can be displayed as the PMF distribution versus one or more principal components (Kamiya et al., 2002), or any other arbitrary axes. REMD is another popular enhanced sampling method (Sugita & Okamoto, 1999) and has also been applied to find native docking configurations (Kokubo, Tanaka, & Okamoto, 2011).

1.3 Affinity calculation using MD

Besides the above mentioned sampling of the docking configurations, it is also important to calculate the binding affinity for drug development. Various methods using MD exist to calculate the binding free energy. A popular method is MM/PBSA (Kollman et al., 2000; Massova & Kollman, 1999) which can calculate the binding free energy at a reasonable computational cost, however the absolute value often differ significantly from the experimental one (Singh & Warshel, 2010; R. H. Zhou, 2003). In some cases, positive binding free energies have been obtained, which would suggest that these ligands would rather separate from the protein. The Umbrella Sampling (U.S.) method (Torrie & Valleau, 1977) is an equilibrium simulation technique that is widely used to calculate the binding free energy between a ligand and a protein and has seen some promising results, however it is difficult to find the appropriate reaction coordinate despite its expensive computation costs. The Thermodynamic Integration (Therm. Int.) method (Frenkel, 2002) is another one to compute the binding free energy, where the average force acting upon the ligand is calculated by the force field along a binding or

unbinding path. Integrating these forces produces the binding free energy. In the Smooth Reaction Path Generation (SRPG) method (Fukunishi, Mitomo, & Nakamura, 2009; Nguyen et al., 2015), bound ligand molecules are first dissociated by the Filling Potential method (Fukunishi et al., 2003) *in vacuo* to assume and generate an appropriate smooth unbinding pathway. Then, Therm. Int. along the path in aqueous solvent is performed to estimate the binding free energy. The SRPG method however, requires an experimental complex structure or a reliable model with a correctly predicted bound ligand configuration.

Both U.S. and Therm. Int. require a definition of a reaction coordinate in 3D space and initial structures spread over this reaction coordinate, preferably a single structure for every window. On the contrary, McMD and REMD are more general without requiring a spatial reaction coordinate, by using the potential energy as a reaction coordinate instead. Due to the requirements of a spatial reaction coordinate of the former methods, the choice of using the later methods, which don't require an explicit reaction coordinate, seems better for *ab-initio* docking, where the mode of binding is not yet known. Thus, one would also expect that McMD or REMD would provide the correct binding free energy (Kamiya et al., 2008). However, since both the McMD and REMD simulations give the correct canonical ensemble, the simulations essentially causes lower ligand densities at positions far from the protein. In fact, simulations that successfully sampled near the native complex structure by using the powerful McMD method, were unable to obtain accurate values for the binding free energies due to insufficient sampling in the unbound region. Some examples are flexible docking simulations between an SH3 domain and a proline-rich peptide in vacuo (Nakajima, Higo, Kidera, & Nakamura, 1997), between lysozyme and one of its inhibitors tri-NAG in explicit solvent (Kamiya et al., 2008), and between the N-terminal domain of neural restrictive silencer factor and the paired amphipathic helix domain of mSin3 in explicit solvent (Higo, Nishimura, & Nakamura, 2011).

In order to obtain an accurate binding free energy, such less probable structures are also required. However, they can only be sampled by very long simulations or by employing a bias in an additional dimension via an umbrella potential (Dickson, Ahlstrom, & Brooks III, 2016). These kind of simulations would have much more expensive computing costs. To overcome this problem, a combination of the McMD simulation with Therm. Int. to accurately predict both the binding complex and the affinity is proposed. The essential weakness in the Therm. Int. or SRPG method requiring the correct bound complex structure (Fukunishi et al., 2009; Nguyen et al., 2015) must also be overcome by providing a reliable complex model from the McMD simulation.

1.4 The role of Cyclin Dependent Kinase 2

Cyclin-dependent kinase 2 (CDK2) is involved in regulating cell cycle control. Extensive research has been done to understand the function of CDKs, where Hartwell, Hunt and Nurse received the 2001 Nobel Prize in Physiology or Medicine (Hartwell, 2002; Hunt, 2002; P. M. Nurse, 2002). CDKs, whose activity is under strict control, are regulatory proteins that drive the cell cycle transition (Hunt, Nasmyth, & Novak, 2011; Lim & Kaldis, 2013; Malumbres, 2006; P. Nurse, Masui, & Hartwell, 1998; Sherr, 1996). CDK1, CDK2, CDK4 and CDK6 directly promote cell cycle progression. Whereas CDK1 is a key determinant of mitotic progression, CDK2 is relevant for DNA replication in higher eukaryotes (also see Figure 1). CDK4 and CDK6 are responsive to numerous growth regulatory signals in metazoans, which control much of the cell cycle entry in these species. Alternatively, CDK7, CDK8 and CDK9 regulate transcription (Bregman, Pestell, & Kidd, 2000; Drapkin, LeRoy, Cho, Akoulitchev, & Reinberg, 1996; Lim & Kaldis, 2013; Nemet, Jelicic, Rubelj, & Sopta, 2014), and CDK5 is involved in post-mitotic functions in specialized tissues (Pozo et al., 2013). Since CDKs play a

crucial role in the cell cycle, cancers often disrupt their normal function in order to enhance cell growth (Malumbres & Barbacid, 2001). CDK2 is particularly frequently disrupted in certain types of cancer (Calbo, Serna, Garriga, Grana, & Mazo, 2004; Scaltriti et al., 2011; Volm et al., 1997; Y. Zhou, Wang, Gobl, & Oberg, 1999). Several safeguards exist however to modulate the activity of CDK2, and subsequently DNA replication. CDK2 is primarily controlled by cyclins, where retinoblastoma protein (RB) and the transcription factor E2F regulate the abundance of CDK2 (Herrera et al., 1996; Markey et al., 2002; Moroy & Geisen, 2004; Ren et al., 2002; Zhu, Giangrande, & Nevins, 2004). The CDK-interacting protein/kinase inhibitory protein (CIP/KIP) class of CDK inhibitors bind to CDK2-cyclin complexes to render them inactive (Eldeiry et al., 1993; Martin et al., 2005; Polyak, Kato, et al., 1994; Polyak, Lee, et al., 1994; Zhu et al., 2004), while CDK2 is also regulated by phosphorylation of Tyr14/Tyr15 in the glycine-rich region (Wu et al., 2001). Cyclin binding to CDK2 causes a significant change in the conformation of the activation loop by pulling the loop out from the active site. This enables CDK-activating kinase (CAK) to phosphorylate the threonine residue part of the loop, which stabilizes the CDK2-cylin complex further (Echalier, Endicott, & Noble, 2010; Pavletich, 1999). CDK2 has also been shown to be involved in other processes besides the cell cycle, such as regulating the migration of tumor cells by BRMS1 phosphorylation (Roesley et al., 2016) and reducing EGF-induced cell transformation by phosphorylation of ELK4, which in turn regulates c-fos expression (Peng et al., 2016). Due to CDK2's integral relationship with cancers, drugs that can modulate CDK2 activity could be crucial to combat many of the associated cancers.

1.5 Objective

In this study, I have conducted flexible docking of an aminopyrazole inhibitor (CS3) to the flexible cyclin-dependent kinase 2 (CDK2), which is involved in regulating cell cycle control. The malfunctions of CDK2 can cause tumorigenesis (Asghar, Witkiewicz, Turner, & Knudsen, 2015; Malumbres, 2014; Malumbres & Barbacid, 2009), and so its inhibitor is a highly potential drug candidate. Since the complex crystal structure of CDK2 with CS3 (PDB ID 4EK5) has been deposited at the Protein Data Bank (PDB) (Berman, Henrick, Nakamura, & Markley, 2007) and their affinity data from Community Structure-Activity Resource (CSAR) (Dunbar et al., 2013) are available, my goal is to reproduce the native complex structure and to calculate their binding affinity in a general *ab-initio* manner. The position of the ligand in a cylindrical region, which covers both the ligand binding pocket and the bulk water, is sampled by the McMD simulation starting from the unbound ligand structure, where the atoms of the protein inside the original cylinder and the surrounding buffering one were treated as free without any restraints. By combining the McMD simulations and Therm. Int., both the bound complex of CDK2 with CS3 and their affinity were reproduced correctly. This combined scheme could become a method used for a general flexible protein-ligand docking protocol, providing accurate prediction of both the bound complex and the binding free energy, starting from the unbound structure.

2 MATERIALS & METHODS

2.1 Molecular docking using MD simulations

The CDK2 structure in complex with an aminopyrazole inhibitor with the PDB ID 4EK5 (resolution 1.6 Å) was used. This complex has a known binding affinity of -8.31 kcal/mol obtained by ITC (Isothermal Titration Calorimetry) experiments which was deposited to the CSAR database (Dunbar et al., 2013). The simulations were performed using a modified version of Gromacs (Abraham et al., 2015; Pronk et al., 2013) and the system was prepared using Gromacs' built-in tools, resulting in a system with a size of 68.1 x 53.3 x 78.0 $Å^3$. Additional details regarding the system's preparation and the simulation details are written in Chapter 2.2.

The strategy to obtain the binding free energy consists of the following four steps:

1) Define a vector λ as the basis for the space restraints on the ligand.

2) Perform McMD using potential energy as a reaction coordinate to sample complex structures in the restricted space along λ and to predict the native binding configuration.

3) Generate a reaction coordinate ξ from the McMD data in order to obtain a smooth and realistic path, which can be effectively sampled by Therm. Int.

4) Calculate the binding free energy using Therm. Int.

2.1.1 (1) Estimation of the space restraining cylinder λ

In rigid body docking, the search area is pre-defined by the user and limited to inside the pocket. When docking using MD however, we want to simulate binding and unbinding events, i.e. the simulation is not limited to only the pocket, but the bulk is included as well.

However, the bulk is assumed to be infinite in size. This means that the ligand can spend a long time in the bulk due to becoming adrift. This decreases the number of binding events and thus the number of docking experiments the simulation can do in a limited amount of time. To increase the efficiency of the simulation, i.e. to maximize the number of binding events, the search space of the ligand is here assumed to be limited to a finite region that connects a bulk region to the binding-pocket region. Each region can be represented as a sphere like the cyan areas in Figure 2, where these two spherical regions can be connected and the combined area can be represented as a cylinder.

In order to determine the axis of the cylinder, dissociation experiments were performed using Random Acceleration Molecular Dynamics (RAMD) (Ludemann, Lounnas, & Wade, 2000). RAMD was implemented in Gromacs and applied to the current system for prediction of the dissociation direction of the ligand. Shortly, RAMD applies a constant magnitude of force on the center of mass (COM) of the ligand in a randomly chosen direction. In case the ligand does not move further than a predefined cutoff within a single RAMD block, then the random direction would be reset. Here, each RAMD block was alternated with an unaccelerated, i.e. canonical, MD block in order to relax the system. By performing multiple RAMD simulations, each with different random seeds, some statistics with respect to the ligand path distribution can be obtained and an average dissociation direction starting from the COM of the initial structure can be calculated. From here on, this vector is called λ. In order to restrain the translation and rotation of the protein, position restraints on the Cα atoms were used. The Cα atoms of CDK2 more than 10 Å from the atoms of CS3 were position restrained with a force constant of 1 kcal / (mol $Å^2$). Here, the X-ray complex structure, after equilibration, was used as the initial coordinates for the RAMD simulations to estimate the dissociation direction, since the complex structure was available for the system. However, this is not an imperative requirement of the

strategy, and alternative methods without using the complex structure are also available, as will be shown in Chapter 4.

2.1.2 (2) McMD simulations for flexible ligand docking

Then, the McMD simulations were performed for the ligand docking to predict binding conformation configurations. The starting conformation for McMD was obtained via RAMD simulations, where one of the dissociated ligand structures was chosen as shown in Figure 2. The ligand was restrained to stay within a cylinder defined by the direction calculated from the RAMD simulations. The cylinder was defined with a radius of 4 Å and λ spanned from -5 Å to 20 Å, where 0 Å corresponds to the COM of the ligand after equilibration. During the McMD simulation, in case the COM of the ligand leaves the boundary of the cylinder in the orthogonal direction to the axis, a harmonic potential with a force constant of 20 kcal / (mol Å^2) was used to pull it back towards the cylinder. Similarly, in case the ligand leaves the confines of the cylinder along its axis, a force constant of 1000 kcal / (mol Å^2) was used. No restraints were applied to the ligand while the COM remained inside the cylinder. This cylinder thus restrains the COM of the ligand in order to reduce the degrees of freedom, while connecting the bound and unbound regions together to enable multiple binding-unbinding experiments. Finally, the protein atoms beyond 13.5 Å from the axis of the cylinder were restrained as described in Chapter 2.2.

The McMD simulations consist of training runs to iteratively estimate the density of states using eq. (10), described in Chapter 2.3, and a successive run (i.e. the production run) to sample structure trajectories. The potential energy (E) and the COM of the ligand at each step were stored for the production run, which lasted in total 6.135 x 10^9 steps (12.27 μs). In addition, the structures were stored every 20 ps. Projection of the probability distribution of the

reweighted ensemble onto various axes including λ, were used to identify stable regions and structures.

2.1.3 (3) Estimation of the reaction coordinate ξ

After ab initio docking using McMD, Therm. Int. is to be used in order to accurately calculate the binding free energy corresponding to the predicted bound configuration. A strategy which involves picking a representative structure at the predicted global minimum, followed by picking compatible configurations along the dissociation path was used. First, a representative structure corresponding to the predicted global minimum at 300 K was picked. From the configurations with $P_c(E, 300\ \mathrm{K}) > 0.5$, where P_c is the canonical distribution obtained by reweighing the McMD distribution via eq. (14) described in Chapter 2.3, those with a λ value between -0.05 Å and +0.05 Å located at the predicted energy minimum by the McMD simulations were taken, and clustered by their relative Root Mean Square Deviations (RMSDs) as a distance within 1 Å. From the largest cluster, which would correspond to the ensemble of the most stable structures, the configuration with its λ value closest to the bin center was picked as the representative structure of the predicted native configuration. This is followed by picking structures with similar orientations along λ as described below.

The initial structures for Therm. Int. should be similar enough so that the conformational ensemble sampled in a window overlaps with the ensembles sampled in its neighboring windows. However, simply selecting configurations in the McMD ensemble based on the proximity of their COM can result in very abrupt changes in the ligand's configuration between neighboring windows. In order to prevent this, first the docking configurations from these unsuccessful docking attempts need to be detected and removed from the ensemble. Ligands inside deep pockets are generally unable to suddenly rotate about their own COM and change their orientation with respect to the native bound configuration, without first completely unbinding. However, it is possible to filter the ensemble by the orientation of the ligands with

respect to the predicted native configuration and remove those that that deviate too much. In the case of CDK2, three layers can be identified naively:

1) bound layer (λ < 10 Å)

2) intermediary layer (10 < λ < 15 Å)

3) unbound layer (λ > 15 Å)

In each of the above layers, different criteria should be used due to changes in accessible volume, where in the bound region the lack of available space prevents the ligand from rotating, while in the unbound region the ligand is completely free. To filter the ensemble, first all the configurations of the ligand, including the predicted native bound configuration, were translated, so that their centroids coincide with the origin of the coordinate system. Next, the $RMSD_L$ of each of these configurations was calculated with respect to the predicted native bound configuration. Here, $RMSD_L$ indicates the RMSD calculated between the pre-translated ligand configurations with respect to the pre-translated predicted native bound configuration. In the bound layer all configurations with an $RMSD_L$ > 2.0 Å were removed, while in the unbound layer all configurations were kept. In the intermediary layer, the $RMSD_L$ cutoff value was linearly interpolated between 2.0 Å and the maximum $RMSD_L$ (~8.6 Å) along the range of the layer. This produces an ensemble which should reflect the ligand binding or unbinding path to or from the predicted native configuration, where the orientation of the configurations is similar to the bound configuration in the deep pocket region and becoming increasingly variable as the ligand moves closer to the bulk. Finally, all configurations, whose COM positions are located more distant than 3.5 Å from the λ axis in the perpendicular direction, were also removed, in order to remove configurations close to or beyond the edge of the cylinder. The resulting ensemble has been reduced to the configurations accessible to the predicted native

bound configuration as the configuration dissociates from the pocket, similar to an ensemble produced by pulling simulations such as Steered Molecular Dynamics (Isralewitz, Gao, & Schulten, 2001).

Because the axis of the cylinder λ is a rough estimate of the binding/unbinding direction, a smooth and realistic path for binding/unbinding should be estimated in order to conduct Therm. Int. As shown in Figure 3, representative structures $\{q_n, n=43\}$ were first picked at 0.5 Å interval bins along λ, ranging from -1 Å to 20 Å from the above described filtered ensemble, by taking the structure with the lowest RMSD with respect to the representative structure from the preceding bin (q_{k-1}). A new path was traced through the COMs of these structures $\{q_n\}$ and was smoothed by using a Savitzky–Golay filter (Savitzky & Golay, 1964), which primarily smoothed the regions where there are only a few structures available, such as the bulk region. This produces a new path, which is called as ξ hereafter and represents the binding/unbinding path of CS3 to CDK2 predicted from the McMD simulations. Whereas λ only serves as a guideline to differentiate between and connect the bound and unbound regions, ξ is a path connecting physically accepted structures from the McMD ensemble at 300 K in the bound region to the bulk region. Thus, this should be an appropriate path, traveling along near the transition state without major frictions between the protein and the ligand.

2.1.4 (4) Binding free energy calculation by Thermodynamic Integration along ξ

Finally, in order to perform Therm. Int. along ξ, the path was partitioned into 0.1 Å interval windows, where the center of each window was positioned on the path and separated by 0.1 Å. For each window, a structure from the above described filtered ensemble was taken whose COM was the closest to the center of the window on ξ. In the case of the intermediary and unbound regions, neighboring bins often share the same starting structure, due to a limited

amount of structures available in the (filtered) ensemble at 300 K. Each window was equilibrated for 2 ns and subsequently sampled for 1 ns in the canonical ensemble. For each window, three independent simulations were performed, where the velocities were randomized according to a different random seed for each simulation following a Maxwell distribution. The COM of the ligand was restrained using a harmonic potential to the window center using a strong force constant of 400 kcal / (mol Å^2). The PMF was calculated by integrating the forces along ξ using the following equations (Fukunishi et al., 2009):

$$\langle \vec{F}(\vec{r}) \rangle = \frac{\int -\overrightarrow{\nabla E} e^{-\beta E} d\boldsymbol{r}}{\int e^{-\beta E} d\boldsymbol{r}} \approx \frac{\sum -\overrightarrow{\nabla E} e^{\beta V_\xi}}{\sum e^{\beta V_\xi}} \tag{1}$$

$$TI(\xi) = -\int_0^\xi \langle \vec{F}(\overrightarrow{r(\xi)}) \rangle \cdot d\overrightarrow{r(\xi)} \tag{2}$$

where $-\overrightarrow{\nabla E}$ is the force on the COM of the ligand calculated from the potential energy E, $\boldsymbol{r}$ is a position on ξ to which the ligand's COM is restrained, and V_ξ is the restrained energy to keep the ligand near the window center along the path from 0 to ξ calculated as

$$V_\xi = \frac{1}{2} k_\xi (r - r_\xi)^2 \tag{3}$$

Here, the square distance between the COM r and the anchor point r_ξ is computed with the weight $k_\xi = 400$ kcal / (mol Å^2). The integration along the path in eq. (2) was approximated by taking the sum of $\langle F(r) \rangle \cdot \Delta r$ at each small bin, the width of which was 0.1 Å, along the path ξ. Note that the effect of the restraint is compensated between the denominator and numerator in the last formula of eq. (1). Which then becomes

$$\left(\mathrm{G}(\boldsymbol{r}_{\infty}) - \mathrm{G}(\boldsymbol{r_0})\right) = TI(\xi \to \infty) \tag{4}$$

In order to compare the calculated binding free energy with the experimentally determined one, the simulation has to be corrected for the loss of entropy on binding by taking the sampled volume into account. The standard binding free energy can be determined as follows:

$$\Delta G_b^0 = -k_B T \ln\left(\frac{P_B}{P_U}\right) \tag{5}$$

where P_B is the probability of the bound form and P_U is probability of the unbound form. Here, P_B and P_U are defined as follows:

$$P_B = e^{-\beta G(\boldsymbol{r_0})} \int_{site} e^{-\beta G(\boldsymbol{r})} d\boldsymbol{r} \tag{6}$$

$$P_U = V_0\, e^{-\beta G(\boldsymbol{r}_{\infty})} \tag{7}$$

Here, $G(\boldsymbol{r})$, $G(\boldsymbol{r_0})$, V_0 and $G(\boldsymbol{r}_{\infty})$ are the PMF at position $\boldsymbol{r}$, the PMF at the global free energy minimum $\boldsymbol{r_0}$, the standard concentration at 1M (1661 Å^3) and the PMF at a reference position in the bulk $\boldsymbol{r}_{\infty}$, respectively (Fukunishi et al., 2009). The P_B term in eq. (6) can be approximated using the following equation:

$$P_B \approx e^{-\beta G(\boldsymbol{r_0})} \sum_{site} e^{-\beta G(\xi,\nu,\zeta)} \Delta\xi \Delta\nu \Delta\zeta \tag{8}$$

where ν and ζ are the axes perpendicular to ξ, and $\Delta\xi\Delta\nu\Delta\zeta$ is a small region part of the site, finally becoming

$$\Delta G_b^0 = -(\mathrm{G}(\boldsymbol{r}_{\infty}) - \mathrm{G}(\boldsymbol{r_0})) - k_B T \ln\left(\frac{1}{V_0} \sum_{site} e^{-\beta G(\xi,\nu,\zeta)} \Delta\xi \Delta\nu \Delta\zeta\right) \tag{9}$$

Thus, the binding free energy is obtained by estimating $\mathrm{G}(\boldsymbol{r}_\infty) - \mathrm{G}(\boldsymbol{r_0})$ from eq. (4) and by estimating $\sum_{site} e^{-\beta G(\xi,\nu,\zeta)} \Delta\xi\Delta\nu\Delta\zeta$ from the probabilities at the binding site obtained by the McMD simulation.

The current simulations were made on a HPC system at the Cybermedia Center PC cluster for large-scale visualization (VCC) at Osaka University consisting of 37 GPGPUs to accelerate the non-bonded calculations by the zero-dipole summation method (Fukuda, Yonezawa, & Nakamura, 2011), and it required approximately three weeks to finish all the McMD simulations including the training and production runs. The final Therm. Int. computations required another 36 hours to be completed.

2.2 Simulation details

A missing 9-residue stretch from residue 36 to 44 of the original X-ray structure of CDK2 (PDB ID 4EK5) was modeled in using the ASEDock plugin (Goto, Kataoka, Muta, & Hirayama, 2008) of the MOE software. The C-terminal 11-residue long loop region was removed in order to simplify the system. Afterwards, the system was prepared using Gromacs' (Abraham et al., 2015) build-in tools, ACPYPE (Sousa da Silva & Vranken, 2012) and the Amber Tools (J. M. Wang, Wang, Kollman, & Case, 2006). The Amber ff99SB-ILDN force field (Hornak et al., 2006; Lindorff-Larsen et al., 2010) was used for the parameterization of the protein, while the General AMBER force field (GAFF) (J. M. Wang, Wolf, Caldwell, Kollman, & Case, 2004) was used for the ligand CS3. The electrostatic potential of the ligand was determined using GAMESS (Schmidt et al., 1993) at the HF/6-31G* level, after which the atomic partial charges were obtained via RESP (Bayly, Cieplak, Cornell, & Kollman, 1993). A 69 x 54 x 79 $Å^3$ sized box around the system was created and filled with TIP3P (Jorgensen,

Chandrasekhar, Madura, Impey, & Klein, 1983) water molecules. Na^+ and Cl^- were added in order to neutralize the system and to bring it to physiological salt concentrations (0.1 M). The system was minimized twice using Gromacs; (Abraham et al., 2015) for the first minimization the positions of the protein's and ligand's heavy atoms were restrained while for the second minimization no restraint was applied. Finally, a 100 ps NVT MD simulation followed by a 100 ps NPT MD simulation were performed to equilibrate the water box, resulting in a final box size of 68.1 x 53.3 x 78.0 $Å^3$. During these simulations the positions of the protein's and ligand's heavy atoms were also restrained.

All the MD simulations were performed with a 2-fs time step. Hydrogens were constrained using the LINCS algorithm (Hess, 2008), while waters were constrained using the SETTLE algorithm (Miyamoto & Kollman, 1992). V-rescale (Bussi, Donadio, & Parrinello, 2007) was used to maintain the thermostat while Parrinello-Rahman (Parrinello & Rahman, 1981) was used to maintain the barostat during the NPT simulation. The van-der-Waals and Coulomb interaction cutoff was set to 12 Å. The electrostatics were calculated using the Zero-Dipole summation (ZD) method (Fukuda, Kamiya, Yonezawa, & Nakamura, 2012; Fukuda et al., 2011), which has been confirmed to exhibit good performance of the MD simulations as well as the Particle Mesh Ewald method (H. Wang, Nakamura, & Fukuda, 2016), having high scalability (Mashimo et al., 2013). Previously, ZD was applied to the investigation of several systems including the systems of small ions (Fukuda et al., 2011), pure water (Fukuda et al., 2012), soluble proteins (Kamiya et al., 2016), membrane proteins (Kamiya, Fukuda, & Nakamura, 2013) and highly charged double-stranded DNA in solution (Arakawa, Kamiya, Nakamura, & Fukuda, 2013). The Reaction-Field (RF) method, which has been implemented in Gromacs (Abraham et al., 2015), with the dielectric constant ε set to infinity should give the same result as the ZD method with the dampening factor $\alpha = 0.0$ ($Å^{-1}$), as long as the correction terms are calculated correctly (Fukuda & Nakamura, 2012). From Gromacs version 5.0 and

higher, a new atom-based calculation scheme (Verlet) is used to calculate the interactions, thus the electrostatic potential should be similar to those calculated directly with the ZD method (H. Wang et al., 2016). To directly confirm this, a pure water system similar to Fukuda et al (Fukuda et al., 2012) was prepared and the ratio of the electrostatic energy difference from that by the Ewald method (see eq. (24) in (Fukuda et al., 2012)) to Gromacs' RF implementation with ε set to infinity was measured. The energy error ratio as a function of the cut-off distance (r_c) yielded a similar result as compared to Fukuda et al's work (Fukuda et al., 2012) (Figure 4). This indicates that RF in Gromacs (with $\varepsilon = \infty$) is implemented to function in a similar manner as ZD (with $\alpha = 0$). Note that artifacts observed in many studies for the RF method should be due to the group-based calculation scheme and the use of a finite value for ε (Fukuda et al., 2012; Fukuda & Nakamura, 2012; H. Wang et al., 2016). These issues are physically sound based on the reaction-field concept, but lead to practical instabilities (e.g., the force continuity breaks at the cutoff length) and causes severe artifacts to dielectric properties (H. Wang et al., 2016). However, these instabilities and artifacts are drastically removed (H. Wang et al., 2016) by the use of the atom-based calculation and $\varepsilon = \infty$, which were also utilized in the current simulation on the basis of the ZD neutralization concept.

The position restrains on the protein were designed in order to minimize their effects on the ligand binding along λ, but to prevent the protein from unfolding, as well as to prevent it from diffusing away from its initial position, and thus the cylinder whose position was not updated during the simulation. The position restraints were based on the approximate size of the ligand and the axis of the cylinder, which was calculated by taking the maximum distance from the ligands center of mass (COM) to any of its atoms with an additional 2 Å buffer region plus the radius of the cylinder (4 Å), resulting in a radius of approximately 13.5 Å from the axis of the cylinder. All protein heavy atoms outside of this range from the axis of the cylinder were

position restrained. This means that all the protein atoms that could interact with the ligand were not restrained, but the remainder of the protein was restrained using a weak force constant of 1.0 kcal / (mol Å^2).

2.3 McMD simulations

Multicanonical MD simulations have previously been extensively reported (Higo et al., 2001; Ikebe et al., 2011; Ikeda & Higo, 2003; Kamiya et al., 2002; Kamiya et al., 2008; Nakajima, Higo, et al., 1997; Nakajima, Higo, Kidera, & Nakamura, 2000; Nakajima, Nakamura, et al., 1997). Here, the fundamentals of McMD simulations are briefly explained, which work by iteratively estimating the density of states of a system and biasing the system based on this estimate. The probability distribution of the potential energy of the multicanonical ensemble is given by the following equation:

$$P_{mc}(E, T_0) = \frac{1}{Z_{mc}} n(E) e^{-W(E,T_0)} = constant \tag{10}$$

where E is the potential energy of the system, T_0 is the simulation temperature, $n(E)$ is the density of states, Z_{mc} is a partition function given by:

$$Z_{mc} = \int n(E) e^{-W(E)} dE \tag{11}$$

and $W(E, T_0)$ is the weighing function to modulate the probability distribution P_{mc} in order for it to become flat, which enables the system to take a random walk along the energy range. The weighing function is given as follows:

$$W(E) = \ln n(E) = \frac{E}{RT_0} + \ln P_c(E, T_0) \tag{12}$$

Here, R is the gas constant and P_c the canonical energy distribution at T_0. By iteratively calculating and updating this weighing function using multiple McMD simulation runs, a flat distribution can be obtained from the accurately estimated value of P_c:

$$W^{i+1}(E) = W^{i}(E) + \ln P_c(E, T_0) \tag{13}$$

After calculating the flat energy distribution, the reweighing technique (Nakajima, Nakamura, et al., 1997) can be used to derive the canonical ensemble $P_c(E, T)$ at an arbitrary temperature T within the flat energy range as follows:

$$P_c(E,T) = \frac{1}{Z_c} n(E) e^{-\frac{E}{RT}} = \frac{Z_{mc}}{Z_c} P_{mc}(E) e^{W(E) - \frac{E}{RT}} \tag{14}$$

The McMD simulation was executed according to the references (Kamiya et al., 2002; Kamiya et al., 2008) and was trivially parallelized (Ikebe et al., 2011) using 37 trajectories, where the velocities were randomized following a Maxwell distribution with different random seeds for each parallel trajectory. First, 37 high-temperature canonical MD simulations at 700 K were performed for 1 ns. Next, $W(E)$, which covers wide temperature range (280 K – 700 K), was iteratively estimated. In the above training runs, 30 iterations were required (3.17×10^9 steps in total with the unit time step 2 fs, corresponding to 6.34 μs in total). Finally, the production run was executed without updating the bias to sample the binding configurations.

2.4 Advanced visualization using Molmil

An important aspect of research is to visualize the results. I have developed a molecular viewer named Molmil (Bekker, Nakamura, & Kinjo, 2016) for visualizing the protein structures using high quality graphics. Molmil makes use of WebGL, enabling it to run directly from inside modern web browsers without requiring additional plugins, even on simple devices such as smartphones and tablets. As such, it is widely used by PDBj (Kinjo et al., 2016). Molmil supports various rendering and coloring modes, which are beneficial to clearly display detailed scientific findings. Furthermore, Molmil can display animations such as MD trajectories. To visualize MD trajectories, Molmil currently supports simple PDB files with multiple models, myPresto's (Mashimo et al., 2013; Morikami, Nakai, Kidera, Saito, & Nakamura, 1992)

trajectory format and Gromacs' (Abraham et al., 2015) TRR and XTC formats. Using advanced scripting, the display and coloring modes can be configured very meticulously. Molmil was used in this dissertation for the rendering of the molecular structures.

3 RESULTS

3.1 Selection of λ by RAMD

For 100 RAMD simulations, 97 properly dissociated the ligand from the protein. The COM of each final snapshot and the average dissociation direction are shown in Figure 5. This average direction serves as the axis of the cylinder, λ, used during the docking simulation by McMD to limit the search space of the ligand. The direction and the COM of the ligand in the native bound structure are shown in Figure 5. After determining λ, an initial structure for the McMD simulation that has its COM within 4 Å of the λ axis and whose pocket is relatively undisturbed, was chosen from the 100 final structures of the RAMD simulations. The chosen initial configuration of the ligand for the McMD simulation is shown in Figure 2 along with the average dissociation direction as the dotted line, which coincides with the dotted line in Figure 5.

3.2 Docking of CS3 to CDK2 by McMD

The probability histogram of McMD obtained from this production run and the reweighted canonical distributions at 300 K, 500 K and 700 K are plotted in Figure 6. Although the distribution in the energy region higher than -75,000 kcal/mol (> 500 K) is less sampled than the lower energy region, the physiological region around 300 K is perfectly flat. The production run successfully sampled both the bound and unbound regions as indicated by Figure 7. In Figure 7A, the Solvent Accessible Surface Area (SASA) of CS3 of the sampled configurations is plotted against the RMSD values of the ligand configurations with respect to the crystal structure of CS3 at 300 K, 500 K and 700 K, respectively. Similarly, in Figure 7B, for each temperature, 1000 random samples from the trajectories was selected and the COM of the ligand configurations was plotted. In the figures, at 300 K the samples mainly occupy the

bound region, while at 500 K the ligand also ventures deep into the bulk. At 700 K, the samples are more uniformly distributed along the search space. By using McMD to randomly walk over the energy space, it becomes possible to efficiently sample both the bound and unbound states and perform multiple docking experiments.

To accurately determine the PMF, especially in the bulk region, the COM of the ligand was stored for each step during a significantly extended production run, 6.135 x 10^9 steps (12.3 μs). A PMF calculated from the reweighted McMD simulation indicates the most sampled stable regions on a given reaction coordinate. The sampled configurations were reweighted to 300 K, where these reweighted configurations were used to calculate the PMF along λ. For easy comparison, the λ axis was aligned so that the ligand in the native X-ray structure (after energy minimization) coincides with $\lambda = 0$. This PMF in Figure 8A, shows a clear global minimum (a) around $\lambda = 0$ Å, where the native bound structure is also located. Namely, the native bound structure was successfully predicted by using the free energy along the λ axis. Furthermore, 98.3% of the predicted configurations along λ from -0.5 Å to 0.5 Å had an RMSD value of less than 2 Å with respect to the X-ray structure. Other local minima were found around (b) -2 Å, (c) 4 Å and (d) 6 Å. Figure 8B shows representative structures at these minima.

The structures at the global minimum and at the local minima were as follows. First, the near-native configuration depicted in blue (Figure 9B, which coincides with the blue structure in Figure 8B) was very close to the native structure (Figure 9A) in CPK colors (whose carbon atoms are colored green), and it has a root-mean-square-deviation of 0.7 Å against the crystal structure. CS3 is hydrogen bonded to CDK2 primarily via CDK2's backbone in this near native configuration as well as in the X-ray structure. In the case of the X-ray structure, one additional hydrogen bond is made between Gln85 and CS3's amine group. This hydrogen bond however

is not present in the predicted structure. Since Gln85 and CS3's amine group are both on the surface, this hydrogen bond may be unstable in explicit water.

Second, the structure with a λ value of -2 Å in cyan Figure 8B had a significantly different configuration compared to the native conformation and has a high RMSD value of 6.8 Å against the crystal structure. This configuration was positioned perpendicular to the axis of the cylinder inside the pocket, and it could potentially be achieved by the ligand rotating from the original configuration with the carboxamide group going farther up, while the cyclopropyl group going down as in Figure 8B. Alternatively, the ligand might directly bind as this configuration without first assuming the native bound configuration as an intermediary conformation. Figure 10 shows the superposition between this configuration and the X-ray structure for the agonist ATP, which are very similar configurations and are described in further detail in Chapter 4.

The third representative structure with a λ value of 4 Å in orange in Figure 8B, was bound in the opposite direction with an RMSD of 10.2 Å, i.e. the carboxamide group was inside the pocket while the cyclopropyl was outside the pocket. This partially bound structure did not fully enter the pocket and this configuration is probably an unspecific binding configuration with respect to the native conformation.

Lastly, the structure with a λ value of 6 Å in red shown in Figure 8B was an intermediary structure between the bound and the unbound forms near the mouth of the pocket with an RMSD of 7 Å. The cyclopropyl group partially entered the pocket, while the carboxamide group was still outside the pocket, pointing into the bulk space. Noticeably, the carbonyl group was upwards in this configuration, versus downwards in the native bound complex, indicating that for successful binding from this configuration, the bond about the secondary amine group would have to rotate in order to assume the correct binding configuration.

These latter two configurations at the local minima with λ = 4 Å and 6 Å, had a significantly lower probability than the near native conformation, and so they could be much less stable than the one at the global minimum, the near native configuration.

In addition, PCA (Principal Component Analysis) of the configurations at 300 K was performed using the distance matrix for the atoms listed in table S9, and the FEL is shown in Figure 11A. Here, X denotes the correct binding configuration observed in the crystal structure, which is located close to the free energy minimum. In Figure 11B, typical structures picked from the FEL are shown. The FEL has the shape of a waxing moon, where the bottom part corresponds to the bound form and the top part to the unbound form, which was only represented sparsely, because only a few unbound structures were sampled at 300 K.

3.3 Binding affinity calculation of CS3 to CDK2 by Thermodynamic Integration

In order to calculate the binding affinity, Therm. Int. in eq. (2) was performed starting from the initial structures shown in Figure 12A to obtain the binding free energy ΔG_b^0 in eq. (9) using the probable path in Figure 13A, both of which were provided by the McMD simulation as described in Chapters 2.1.3 and 2.1.4. In Figure 13B, the one-dimensional FEL (from the reweighted McMD ensemble) along the path ξ is shown. From Figure 14, $TI(\xi)$ seems to converge at 19 Å, and taking the average value from 18 to 19 Å of $TI(\xi)$, $-(G(r_\infty) - G(r_0))$ = -13.67 ±0.0076 kcal/mol from eq. (4). The standard deviation of the $TI(\xi)$ in the final 1 Å region (10 bins) is indeed very small, indicating that $TI(\xi)$ has almost converged to the value at the bulk unbound region. Starting from the 1-dimensional FEL in Figure 13B, the bound state was defined as inside a 3-dimensional cuboid, with the center placed at the PMF minimum, where the width along ξ was 3 Å, which was determined from Figure 13B. Furthermore, all structures along the perpendicular axes ν and ζ were included by setting the cutoff value to a sufficiently

high value far beyond the radius of the cylinder at 10 Å. Then, from eq. (9) with the above Therm. Int. value, the binding free energy, ΔG_b^0 = -13.67 + 4.81 = -8.86 kcal/mol was obtained. Here, the contribution of the correction factor in eq. (9) corresponds to $k_B T \ln\left(\frac{1}{V_0}\sum_{site} e^{-\beta G(\xi,\nu,\zeta)}\Delta\xi\Delta\nu\Delta\zeta\right) = -4.81$ kcal/mol. The predicted binding affinity is thus close to the experimental value obtained by ITC -8.31 ±0.08 kcal/mol (Dunbar et al., 2013).

4 DISCUSSION

The most probable binding configuration of CS3 to CDK2 starting from an unbound state of CS3 was predicted using McMD simulations, and several configurations were provided with their probabilities. In fact, the most probable configuration coincided with that in the X-ray crystal structure of the complex of CS3 and CDK2.

In Figure 7A, the SASA of CS3 at various configurations along λ is plotted with the RMSD values of the ligand configurations against the crystal structure of CS3, The SASA corresponding to the native structure is indicated with a red circle. At 300 K, most of the ligands were located at the ligand binding site with an RMSD around 0.7 Å, while at 500 K and 700 K, the samples were more spread out over the search space. At all the temperatures, the ligand was completely solvated at RMSD values from 12 Å to 18 Å. These images indicate that the native like configuration was stable even at 700 K. However, at 700 K, the alternate binding configuration at RMSD = 6 Å was no longer stable, while the one at 7.5 Å still appeared to be stable at 700 K, suggesting that the configuration located at RMSD = 6 Å is an intermediary configuration and that the one at 7.5 Å corresponds to a configuration at a local energy minimum.

In addition, the FEL given by the McMD simulation is able to provide an accurate free energy surface around the native like configuration, so that it is not necessary to approximate the surface by harmonic functions as attempted previously (Fukunishi et al., 2009). Instead, $\sum_{site} e^{-\beta G(\xi,\nu,\zeta)} \Delta\xi \Delta\nu \Delta\zeta$ in eq. (9) can be directly computed from the FEL to account for the entropy effect in the bound state.

Although the advantage of the McMD simulation is its efficiency for enhanced sampling, it is less suitable for affinity calculation because of a limited structural ensemble in the bulk unbound region as shown in Figure 8A and Figure 13B. This phenomenon is a consequence of the physically correct sampling of CS3, which tightly binds CDK2. Thus, it is not practical to try to obtain the accurate binding free energy from the PMF given by only the McMD simulation.

In order to calculate the affinity of CS3 to CDK2, Therm. Int. computation along a newly defined pathway ξ was performed, which was generated from the ensemble given by the McMD simulation. In order to perform the sampling required for Therm. Int., there are two general issues: (1) finding the appropriate pathway, and (2) choosing the initial structure at each window of ξ. Naively determining the path and the initial structures would result in an incorrect or an unphysical affinity. Here, the physically accepted ensemble at 300 K determined from the McMD simulation was used to determine both the binding pathway and the initial structures.

A major obstacle in estimating the appropriate pathway for docking and undocking the ligand to and from its partner protein, is that the structural ensemble are all independently sampled from each other. Namely, the structures are not connected by time. Here, a method to take the orientation of the ligand into account was employed, when building the binding pathway. Although this method was effectively able to produce a properly connected binding pathway, the conformation of the protein was ignored. Since the pocket of the protein was flexible, albeit that position restraints were employed to the protein atoms distant from the pocket, some variations were observed among the structures that have similar ligand binding configurations. In fact, sudden side-chain flips were observed between neighboring windows,

which might cause small artifacts due to sudden changes in the forces imposed upon the ligand during the Therm. Int. computation.

Another issue with the method employed here, is that in the intermediate and unbound regions only a limited amount of structures were available at 300 K. This not only reduces the accuracy of the path in these regions, but also reduces the number of initial structures for the Therm. Int. In order to improve the accuracy of the Therm. Int., the path was smoothed using Savitzky–Golay filtering. Secondly, nearest neighbor interpolation of the initial Therm. Int. configurations was employed in the regions with limited stable configurations in the physically accepted ensemble obtained from the McMD simulations, such as the unbound region.

When looking at the structures sampled by the Therm. Int. simulations, the interactions between CS3 and CDK2 vary as the ligand moves along binding/unbinding path ξ. In the region around the native configuration ($0.0\ \text{Å} \leq \xi < 1\ \text{Å}$), the conformations of the ligand are relatively stable; the primarily flexibility is the carboxyl group and some in the cyclopropyl group. Next ($1.0\ \text{Å} \leq \xi < 2.5\ \text{Å}$), the hydrogen bonds formed by the native configuration are slowly broken as the ligand moves out from its native configuration and hydrogen bonding becomes very rare in the $2.5\ \text{Å} \leq \xi < 4.0\ \text{Å}$ region, where none of the native hydrogen bonds remain. In the $4.0\ \text{Å} \leq \xi < 5.0\ \text{Å}$ region, hydrogen bonding is limited, but there is an increase in entropy, as can be seen in Figure 12B which shows an increase in the average RMSF (Root Mean Square Fluctuation). Between $5.0\ \text{Å} \leq \xi < 11.5\ \text{Å}$ the number of hydrogen bonds fluctuate, along with the entropy as the ligand moves through the mouth of the pocket which is located on the surface of the protein. In the $11.5\ \text{Å} \leq \xi < 13.5\ \text{Å}$ region, as CS3 slowly unbinds from the surface of CDK2 the entropy increases. This is also apparent from Figure 12B, where the average RMSF slowly increases in this region, with some occasional spikes, indicating very random movement of the ligand. Beyond $\xi > 13.5\ \text{Å}$, the entropy reaches its peak and

stabilizes, indicating that CS3's interaction with CDK2 is very limited and indistinguishable from CS3's interaction with the surrounding bulk. This can also be seen in Figure 14, where $TI(\xi)$ barely changes from this point onwards. Finally, at $\xi > 17.5$ Å the ligand has fully dissociated from the protein.

In the current study, the X-ray coordinates of the ligand- protein complex are available, and so they were used as the initial coordinates for the RAMD simulations to determine the axis λ for the initial guess of the dissociation direction. However, there are many cases where the complex structures are not known, although the ligand binding sites are known. In those cases, it is necessary to assume the axes λ via other alternative methods.

One may opt to estimate the direction of λ in the following naive manner by simply searching for the optimal vector that connects the pocket and the bulk region. First, grid points $\{\vec{\mathrm{h}}_{\mathrm{m}}(\mathrm{R}), \mathrm{m} = 1 \ldots \mathrm{m}_{max}\}$ on a spherical surface are generated at increasing radii R at constant intervals ΔR (e.g. 1 Å) starting from the center of a given pocket. Here, $\mathrm{m}_{\mathrm{max}}$ is determined so that the density of the surface grid points is approximately 1 Å^{-2}. Second, among the grid points $\{\vec{\mathrm{h}}_{m}(\mathrm{R}), m = 1 \ldots m_{max}\}$ for each radius R, the furthest point on the grid $\vec{\mathrm{h}}(\mathrm{R})_{max}$ from all the heavy atoms of the protein is defined, i.e. the point whose distance to its nearest protein heavy atoms is the largest. Among these points, only the points $\vec{\mathrm{h}}(\mathrm{R})_{max}$ were selected such that the distance between $\vec{\mathrm{h}}(\mathrm{R})_{max}$ and its nearest heavy atom of the protein is between Rg + 2 Å and $Rg + d$ Å. Here Rg is the radius of gyration of the ligand molecule and d is a value representing the distance from the protein to the bulk. Finally, by using linear least squares fitting of the selected points $\vec{\mathrm{h}}(\mathrm{R})_{max}$ and the center of the pocket, a straight line connecting the pocket and the bulk is generated, which can be used as the initial guess of the cylinder, λ. Then, the ligand

can be positioned at the end of the cylinder in an arbitrary orientation. The procedure of this naive method is shown in Figure 15A. Applying this naive method to the current CDK2-CS3 system, a new cylinder was created, where 99.9% of the total sampled structures by the current McMD simulation was found in both the original cylinder that was generated from the RAMD simulations, as well as the naively generated cylinder, as shown in Figure 15B. Even for the sampled structures located further than 14 Å, where the orientations of the ligand were almost random due to the McMD method, 81.3% of them were found both in the original cylinder and the cylinder created by the naive method. Thus, a simple naive method could be available to estimate the initial direction of λ, where the unbound structure and any orientation of the ligand can be assumed as the initial configuration of the McMD simulation.

Position restraints in this work (described in Chapter 2.2) were primarily used for two reasons: preventing the protein from unfolding and to prevent the protein from dissociating away from the cylinder. In case distance restraints were to be used, this would not prevent the protein from dissociating away from the cylinder. By using position restraints for the distant areas of the protein, the pocket and the surrounding areas can be treated as flexible, while the protein remains largely stable and doesn't dissociate away from the cylinder. This strategy however should introduce some difficulties. For example, in Figure 10 the CDK2-CS3 complex is colored in light-red/magenta, while the CDK2-ATP is colored light-blue/cyan. The loop region, which has a darker tint, is shown to differ between the two structures. This interaction in combination with the position restraints, limit the movement of the sheet structure and thus make ATP's sub-pocket more rigid. Similarly, in case the pocket were to be near a large hinge movement region, this movement could be restrained, thus preventing from properly simulating the protein's dynamics and possibly preventing the ligand from binding to the protein. In this

situation, one could use distance restraints for the protein instead of position restraints. To prevent the protein from dissociating away from the cylinder, the position of the cylinder based on the relative position and orientation of the protein needs to be updated, or translational and rotational restraints need to be applied on the COM of the protein.

Using McMD the correct binding configuration of the ligand was predicted, however, from the PMF of λ in Figure 8A, a secondary stable configuration was discovered around λ = -2 Å. Interestingly, this configuration is similar to how CDK2's agonist, ATP, binds. Figure 10 shows the superposition between the structure at -2.25 Å (red/magenta) and the X-ray structure for the agonist ATP (blue/cyan, PDB ID=1FIN (Jeffrey et al., 1995)). The predicted configuration binds in a similar manner to ATP, but this configuration was not found by the X-ray experiment. Considering that ATP binds in a similar manner to CDK2 as this docking configuration, this could potentially also be a valid binding mode of CS3 to CDK2, even though the X-ray experiment only seems to have found the primary binding site. However, when the overlapping volume was measured and the Jaccard coefficient was calculated, a ratio of only 0.28 was detected. Here, this coefficient is defined as the volume of the intersection between the CS3 configuration and the ATP configuration divided by their union, calculating the overlap ratio between the two molecules. This low overlap ratio is primarily caused by CS3 being translated away from the sheet region with respect to ATP, thus decreasing the overlap. ATP's larger volume also decreases the Jaccard coefficient. Noticeable though is that the sheet region in the right side is folded differently to make room for ATP. A major difference between the fold of the ATP bound CDK2 and the CS3 bound CDK2, is that in the case of CS3 bound CDK2 the loop region from Phe149 – Thr165 is folded in such a way that it interacts with the sheet structure near the pocket, while in the ATP bound CDK2 the loop is interacting with a region around Val123, forming a sheet structure distant from the pocket. Although the position of the

predicted configuration differs somewhat from ATP's configuration based on the volume overlap, the powerful McMD simulation suggested a secondary binding site for CS3.

5 CONCLUSION

McMD simulations were executed in order to dock the inhibitor CS3 into CDK2 starting from the unbound configuration, followed by Therm. Int. to calculate the affinity. The majority of the stable configurations sampled by the McMD simulations were found to coincide with the native bound complex in the X-ray structure, as indicated by the FEL and the RMSD values. From the ensemble of the configurations obtained by McMD docking, a new path, ξ, was created along which Therm. Int. was performed. Then, the binding free energy was accurately predicted in accordance with the experimental result. Although the current procedure requires considerable computational time even when using a GPGPU cluster, it could provide highly reliable information to optimize chemical compounds for drug development.

6 FIGURES

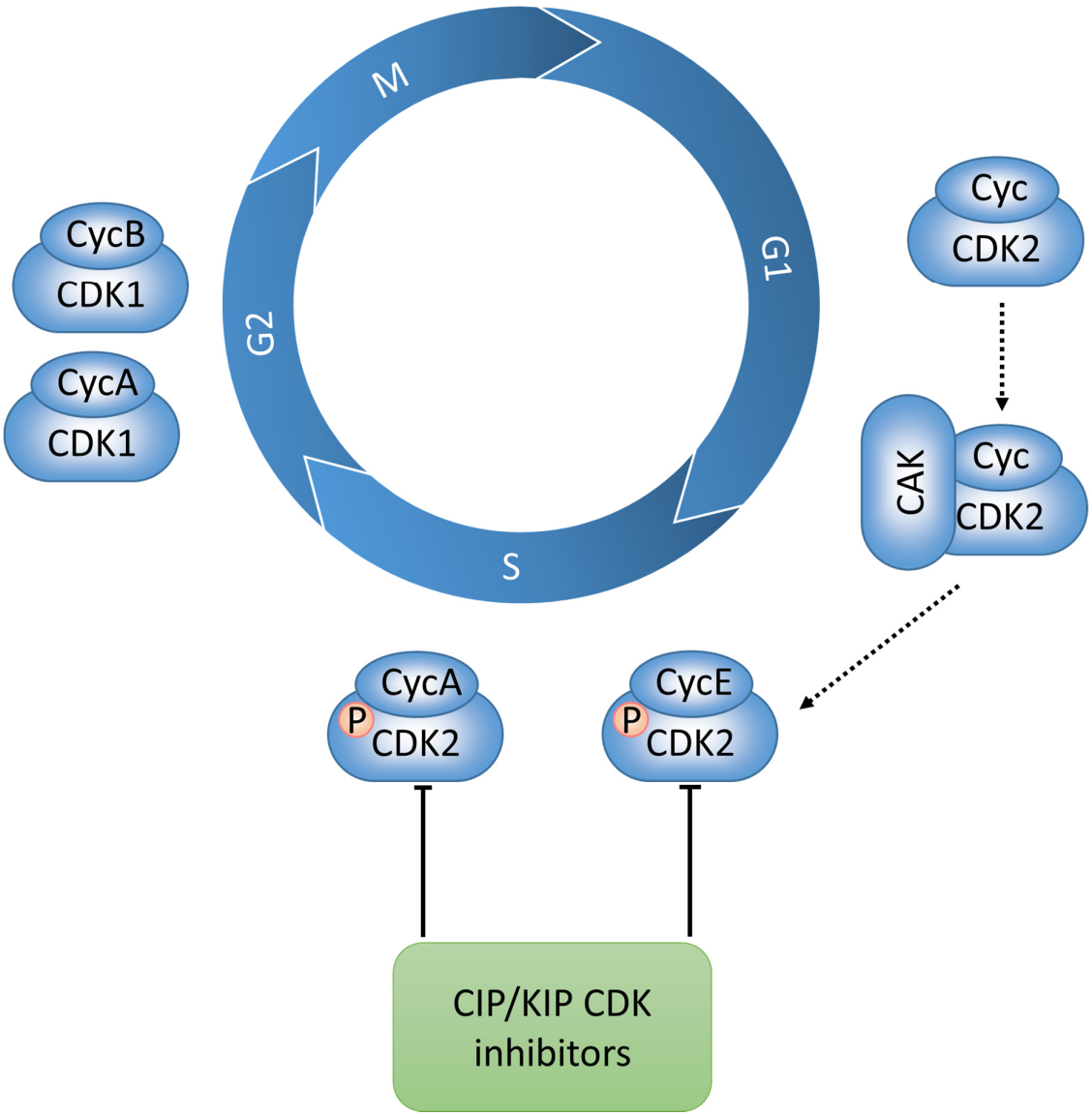

Figure 1. CDK2 and the cell cycle. Phosphorylation of CDK2 complexed with Cyclin by CAK stabilizes the complex, which can be inhibited by CDK-interacting protein/kinase inhibitory protein (CIP/KIP) CDK inhibitors. CDK2 plays an important role in the S-phase of the cell cycle, while CDK1 which is very similar to CDK2, plays an important role in the M-phase.

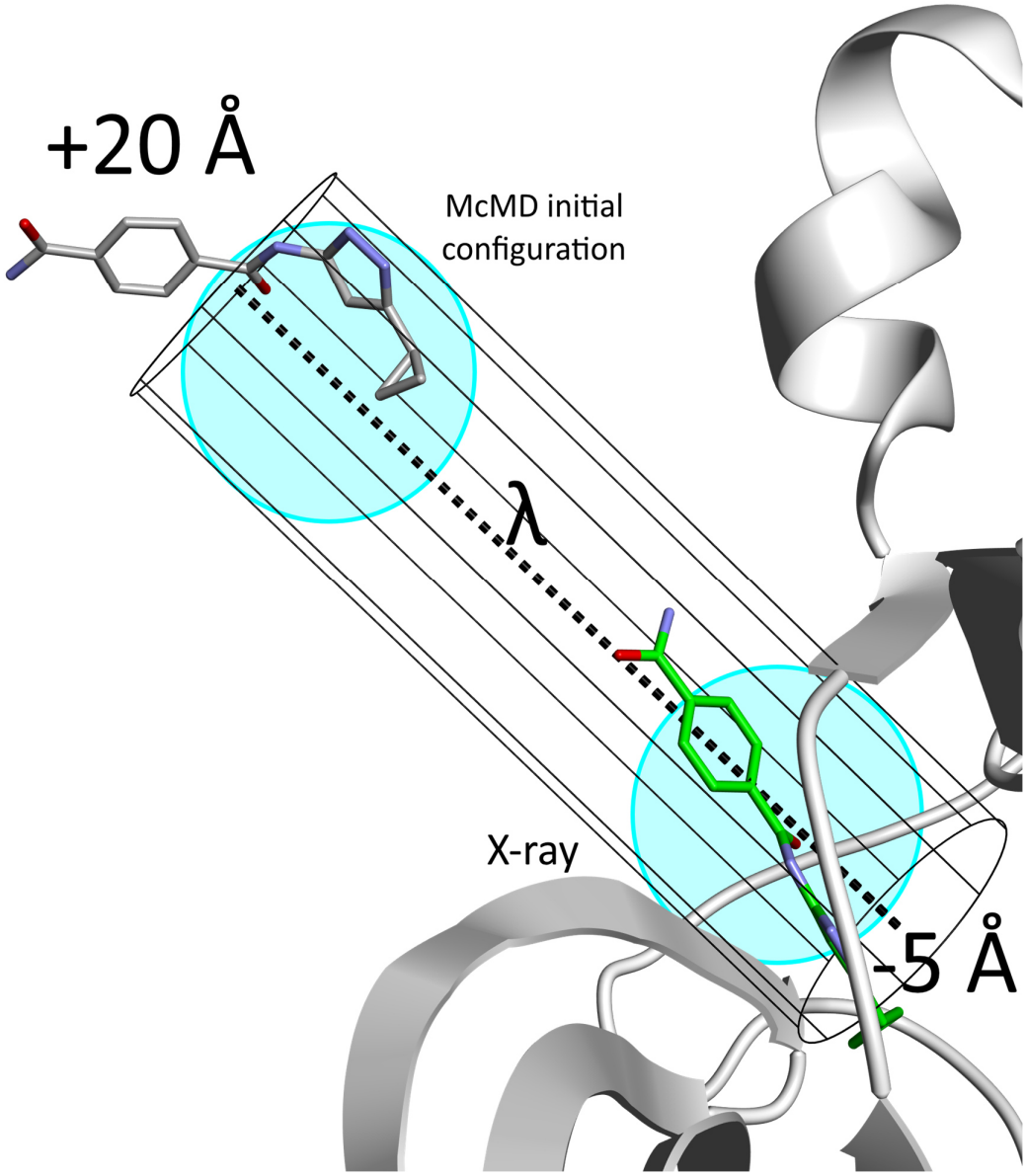

Figure 2. Space restraints and initial structure for McMD simulation. The center of mass (COM) of the ligand is restrained to stay inside the cylinder, which is defined based on the axis λ ranging from -5 Å to +20 Å as a black dotted line. The energy minimized X-ray structure is colored in CPK with the carbon atoms colored green, is denoted by "X-ray" and hereon after referred to as the native configuration or structure. Also shown is the position of the ligand before the McMD simulation denoted by "McMD initial configuration". All of the images of molecular structures were rendered using Molmil (Bekker et al., 2016).

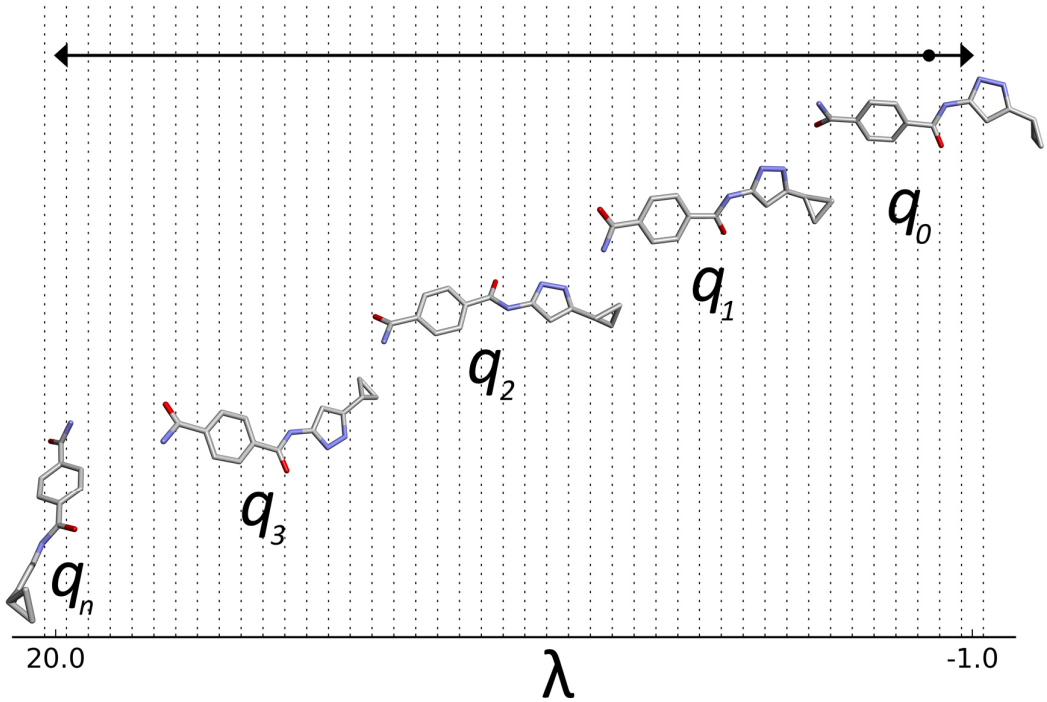

Figure 3. Schematic representation of the generation of the path ξ for Therm. Int. by binning along λ. Starting from the representative configuration q_0 at the PMF's free energy minimum and using λ as a guiding parameter, representative configurations along λ are picked from the filtered ensemble at 0.5 Å intervals, where the minimum RMSD configuration in each bin k is picked with respect to the picked configuration from the preceding bin k-1. A path is traced through the COMs of these picked configurations, where after this path is smoothed, a new path ξ is obtained.

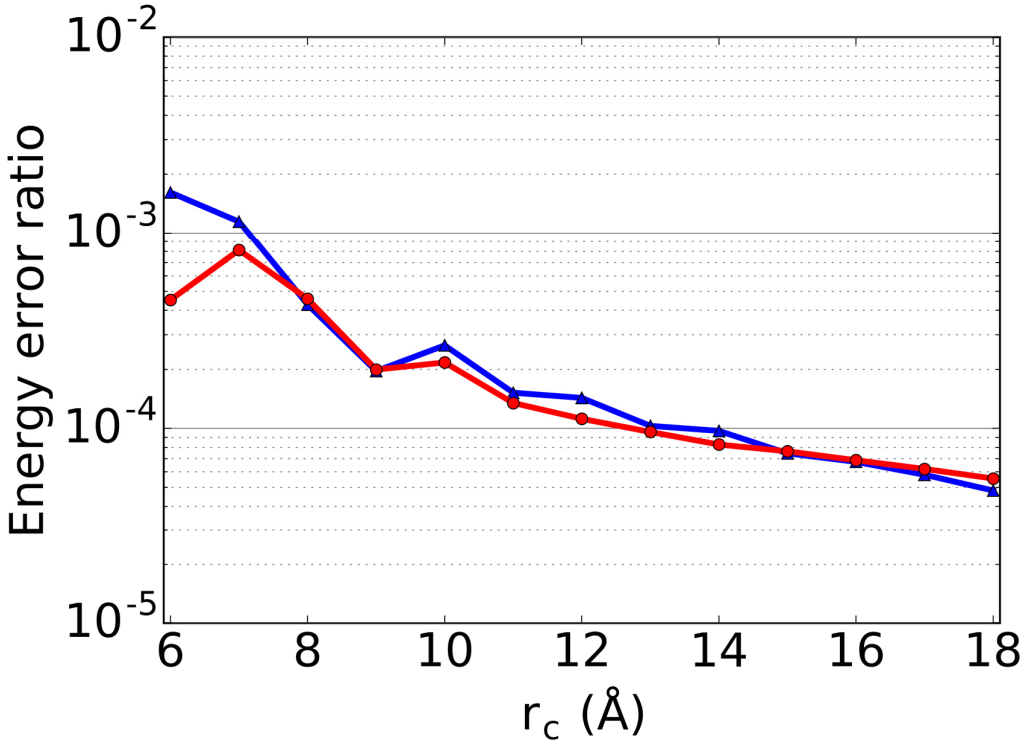

Figure 4. Comparison of the discrepancy in the electrostatic energy, in reference to the Ewald method, for the Zero Dipole method with α=0.0 (blue)(Fukuda et al., 2012) and the Reaction Field method with ε=∞ using the Gromacs program (red) as a function of r_c, as described in Chapter 2.2.

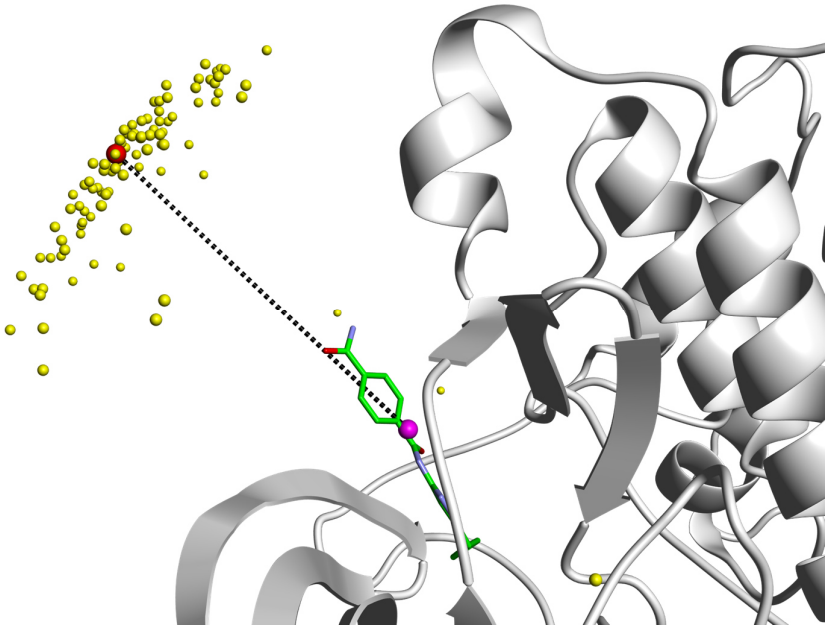

Figure 5. Distribution of the COMs of the ligand obtained from the final snapshot of each RAMD simulation. The dissociation direction of a single RAMD simulation is defined as the unit vector connecting the initial COM (magenta point) to the COM of the ligand at the end of each RAMD simulation (yellow point). The average dissociation direction is then simply the average of these unit vectors. The black dotted line is parallel to the average dissociation direction and starts from the initial COM (magenta point). By setting the length of the line to 20 Å (which corresponds to the cutoff used during the RAMD simulations), the dotted line ends at the red point.

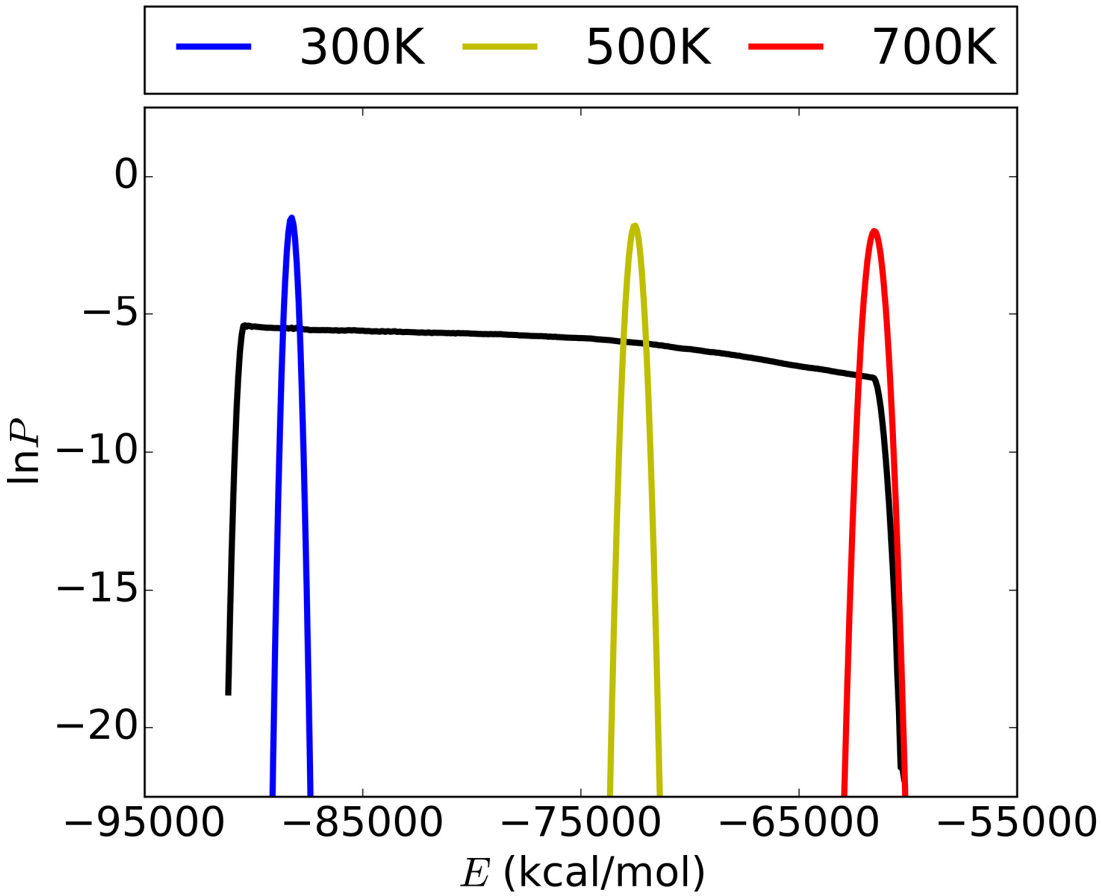

Figure 6. Energy probability distribution obtained from the McMD production run shown as a black line. Blue, yellow and red are the canonical distributions obtained by reweighing the energy distribution at 300 K, 500 K and 700 K, respectively.

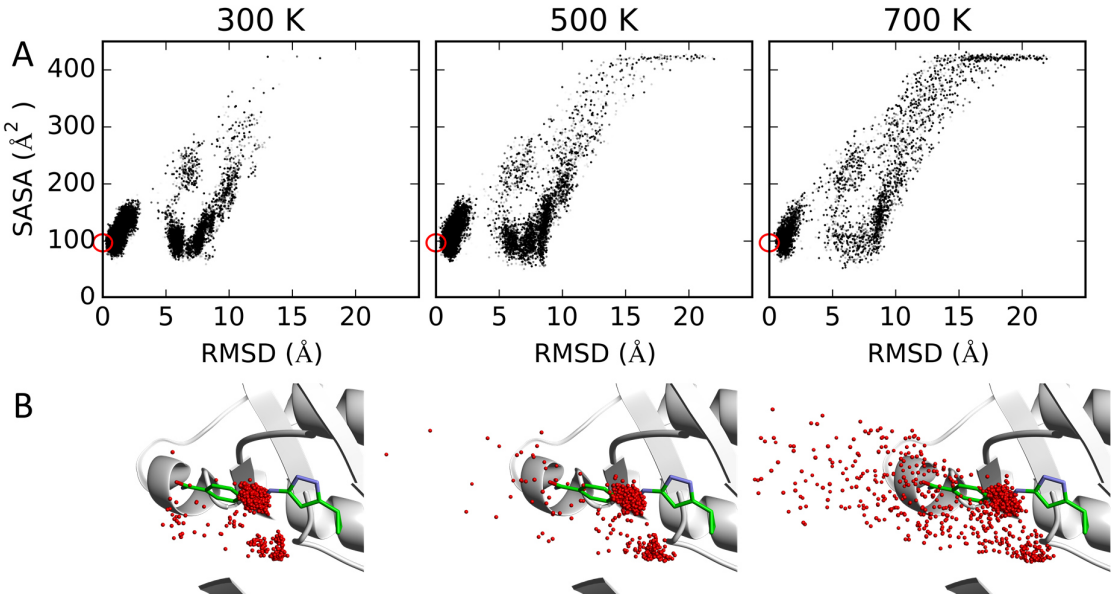

Figure 7. (A) The RMSD versus the Solvent Accessible Surface Area (SASA) of the ligand at 300 K, 500 K and 700 K. The intensity level of the points coincide with the P_c value calculated from eq. (14) in Chapter 2.3, where a darker intensity equals a higher probability. The SASA of the native structure is indicated by a red circle. The CDK2 heavy main-chain atoms (N, Cα, C, O, OXT) were first superimposed onto the X-ray structure before calculating the RMSDs of the ligand's configurations versus the native configuration. (B) Distribution of the COM of the sampled ligand conformations in 3D space. The points indicate the COM of the ligand at 300 K, 500 K and 700 K respectively. For each temperature 1000 samples were randomly selected from each ensemble. The native ligand structure is shown in CPK colors with its carbon atoms colored green.

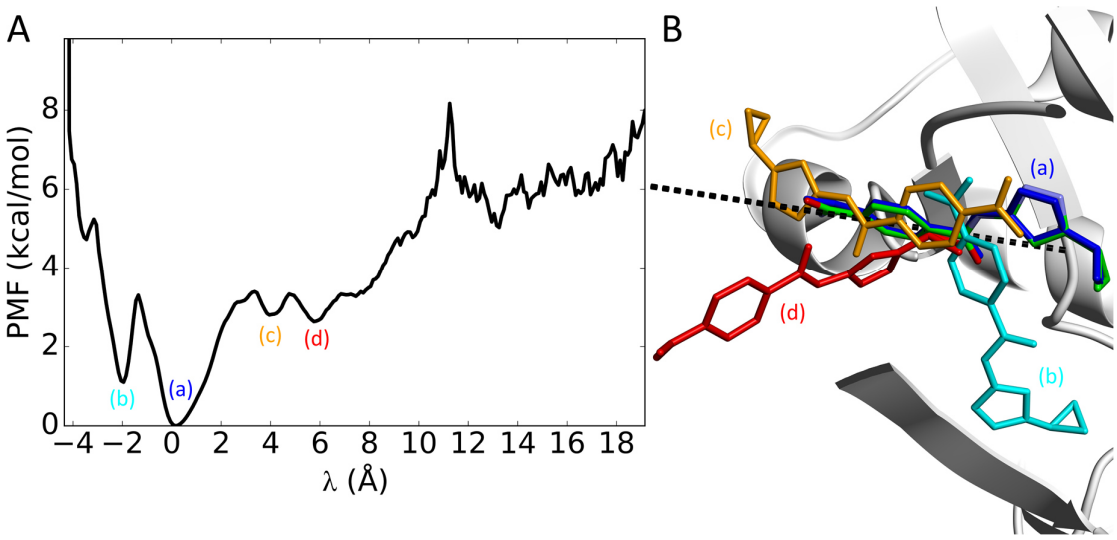

Figure 8. (A) PMF of the ligand by projection of the COMs obtained from the reweighted McMD ensemble at 300 K onto λ. The PMF was normalized so that the global free energy minimum was set to zero. The λ value was stored at each step and was used to generate the plot, with a bin-size of 0.1 Å. The positions of each minima are shown in the figure as a-d. (B) Docking configurations for structures at λ values of (a) 0 Å, (b) -2 Å, (c) 4 Å and (d) 6 Å in blue, cyan, orange and red respectively. The native structure is shown in CPK color with the carbon atoms colored green as a reference. Also shown is the axis λ as a black dotted line. Notable is the cyan configuration, which was perpendicular to the axis of the cylinder inside the pocket, and it could potentially be achieved by the ligand rotating from the original configuration with the carboxamide group going farther up, while the cyclopropyl group going down and binds in a similar manner as ATP.

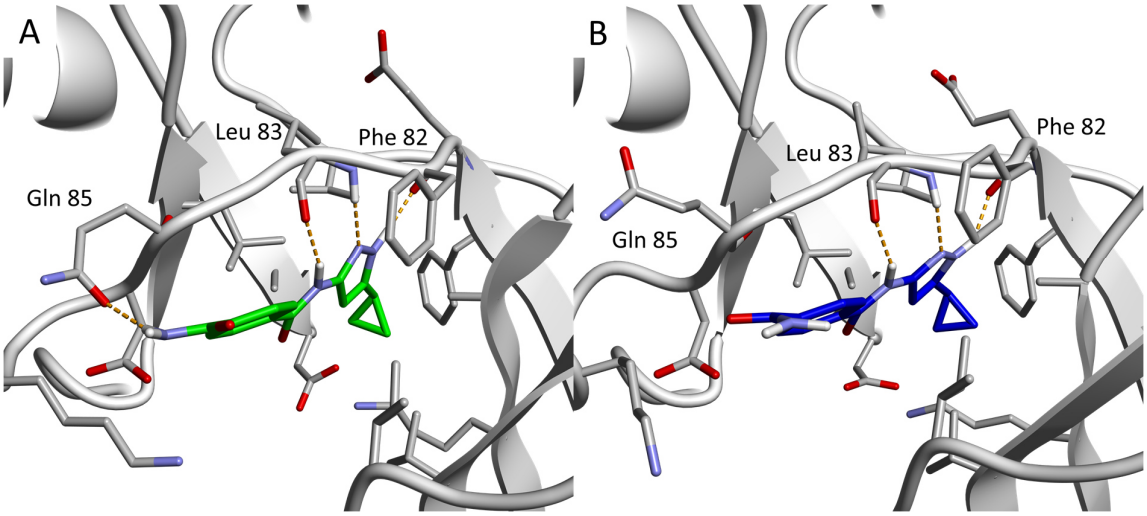

Figure 9. Structure around the bound complex for the native conformation (A) and the predicted conformation at λ = 0 Å (B). Shown are the side chains in close proximity to CS3 and the hydrogens bonds (orange dotted lines) formed between CS3 and CDK2. The carbon atoms of CS3 are colored green and blue respectively for panel A and B, and the hydrogen bond forming residues are labeled.

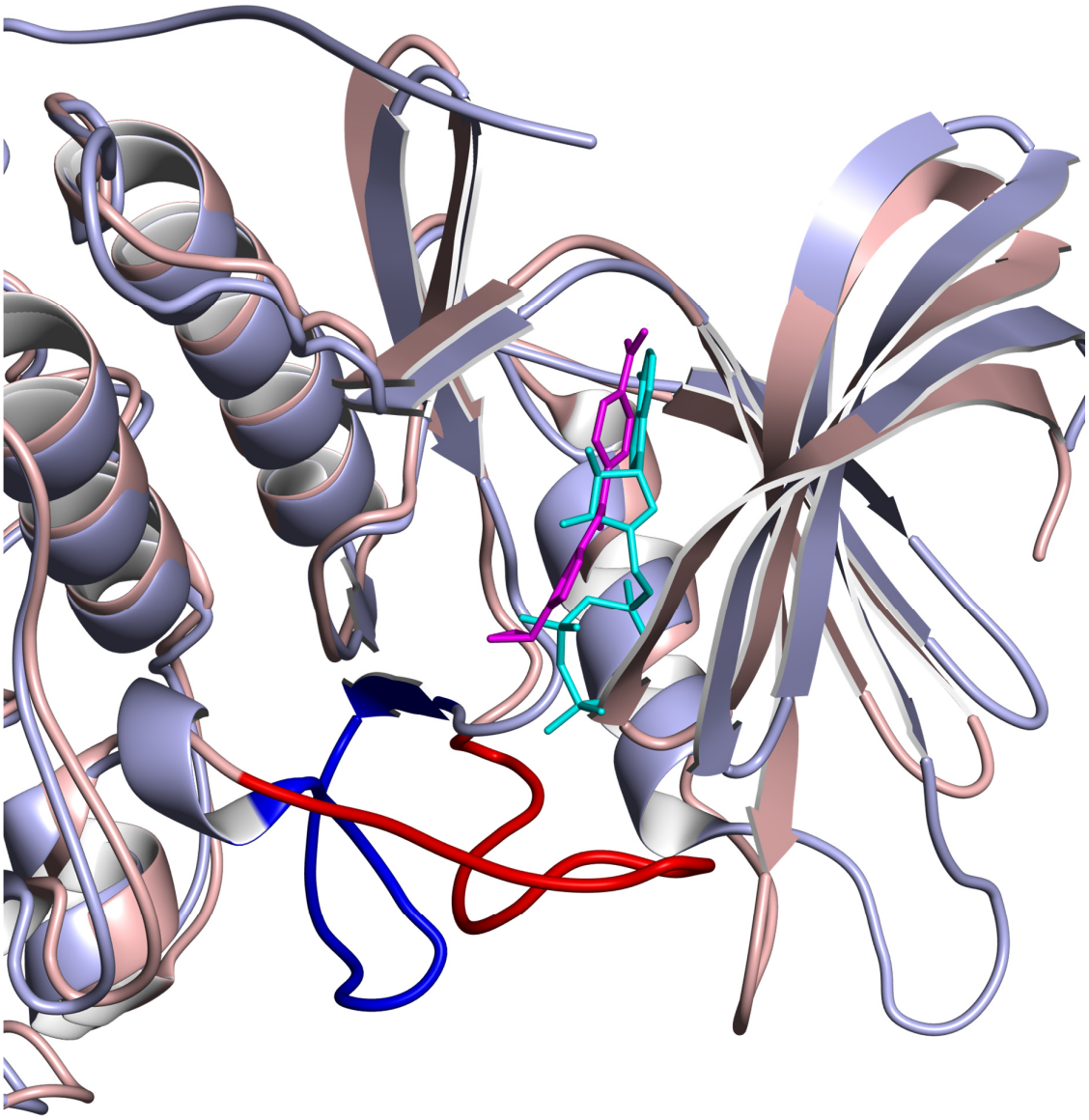

Figure 10. Superposition between the docking configuration found by McMD at λ = -2.25 Å and the agonist ATP structure (PDB ID=1FIN) in light red and light blue respectively. The lower loop regions vary significantly between the two structures and are colored darker. The ligands are colored magenta and cyan for CS3 and ATP respectively.

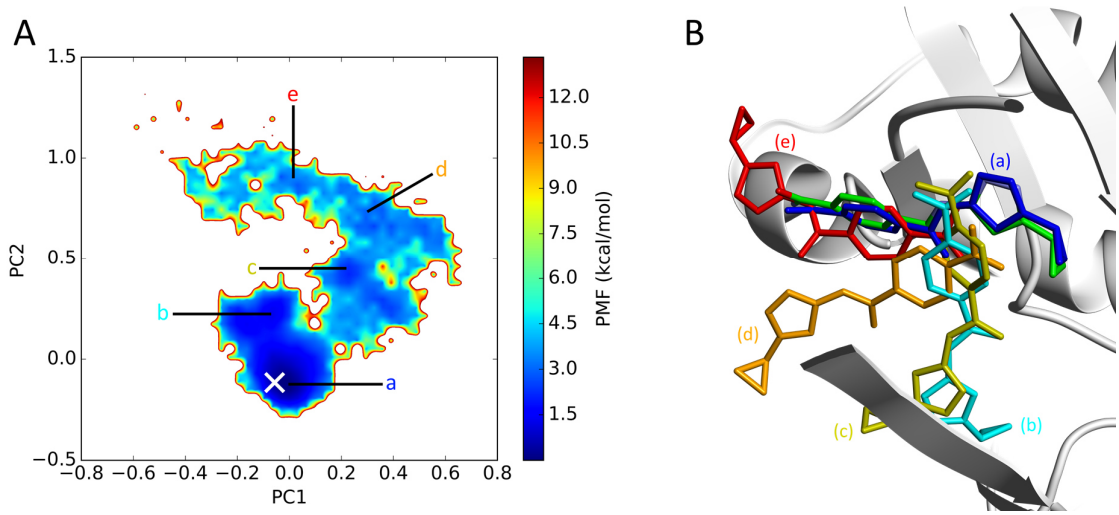

Figure 11. (A) The PMF projected onto the first and second principal components, PC1 and PC2 respectively. PC1 possesses 51.3 % of the variance, while PC2 does 28.9 %. The native structure is denoted by the white cross symbol near the energy minimum. The top left area is associated with the unbound conformations. The locations a-e on the landscape correspond to local minima on the FEL. The sidebar indicates the intensity of the PMF. (B) Representative configurations corresponding to the minima a-e on the PCA-PMF landscape, which correspond to blue (0.7 Å), cyan (7.9 Å), dark yellow (8.0 Å), orange (9.9 Å) and red (10.3 Å) respectively, where the RMSDs with respect to the X-ray structure of the representative configurations are denoted in parenthesis. The native configuration is shown in CPK colors with the carbons colored green, as a reference.

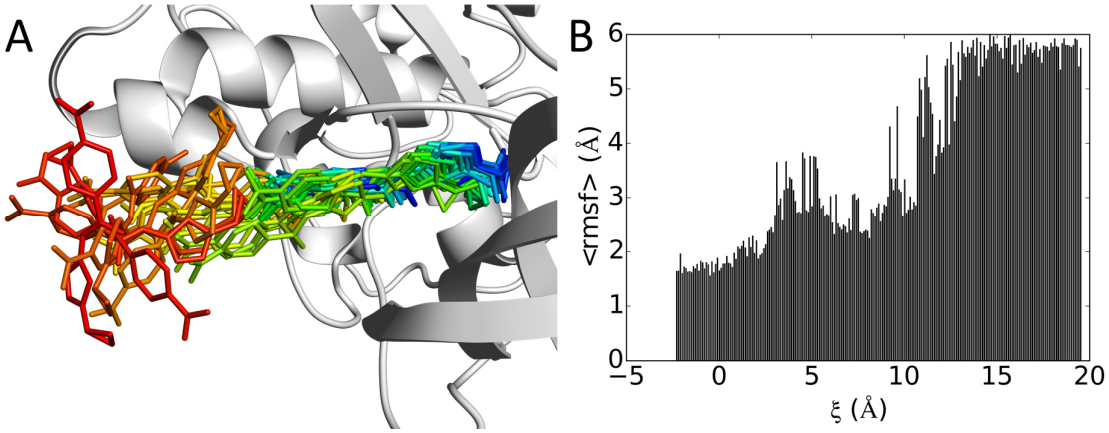

Figure 12. (A) Initial structures used by the Therm. Int. selected from the McMD docking configurations, where the configurations are colored in a blue-to-red gradient according to their respective ξ values. The protein of the native structure is shown in white. (B) The average Root Mean Square Fluctuation (RMSF) of the ligand inside each bin. At low ξ values (< 2.5 Å) the RMSF is generally low, with some peaks in the $2.5\ \text{Å} \leq \xi < 8\ \text{Å}$ region, while in the intermediary ($8\ \text{Å} \leq \xi < 12.5\ \text{Å}$) region the average RMSF values slowly rises, where at $\xi > 13.5$ Å the average RMSFs are all large and relatively stable, indicating that the ligand has almost completely dissociated from CDK2 with many random orientations, and that the interaction between CDK2 and CS3 is minimal to none.

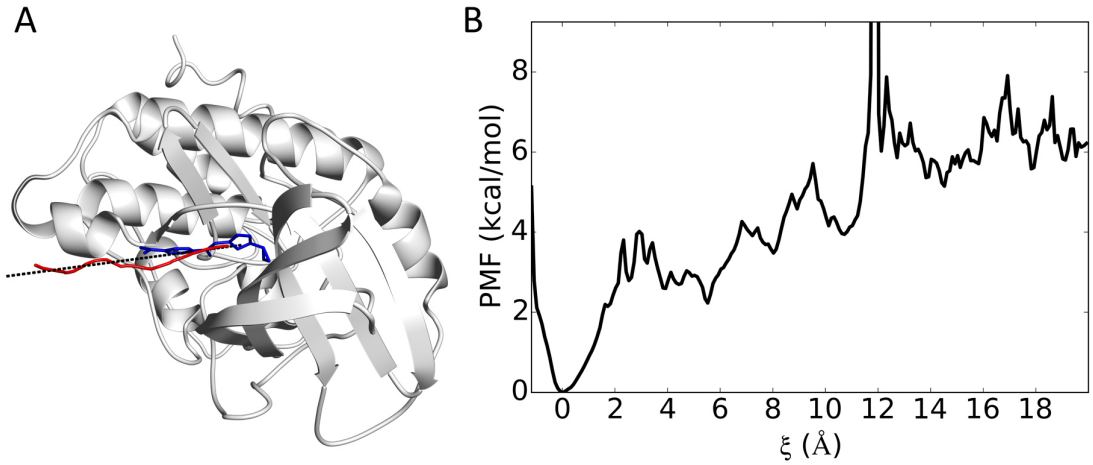

Figure 13. (A) Smoothed path ξ in red estimated from the McMD calculations along which the configurations were sampled using Therm. Int. Also shown are the axis λ as a black dotted line and the predicted bound configuration in blue. (B) PMF from the reweighted McMD ensemble at 300 K by projection of the COMs onto ξ. The PMF was normalized so that the global free energy minimum was set to zero.

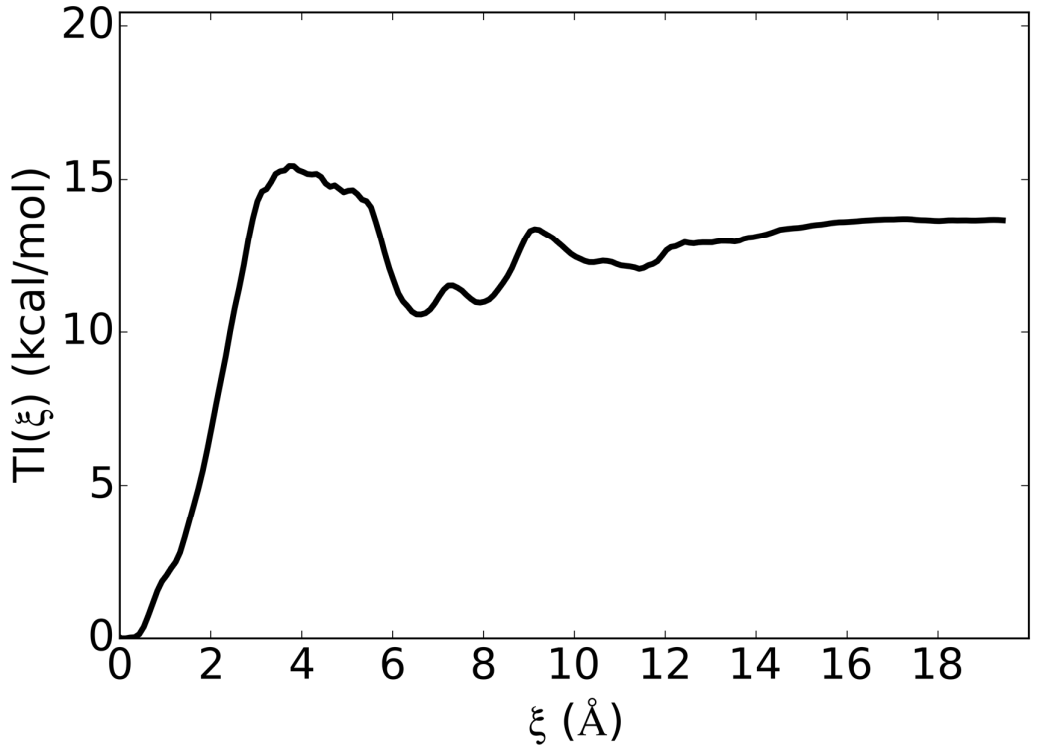

Figure 14. Therm. Int. along ξ. The average *TI(ξ)* value between ξ = 18.5 Å and 19.5 Å corresponds to 13.67 kcal/mol, with a standard deviation of ±0.0076 kcal/mol.

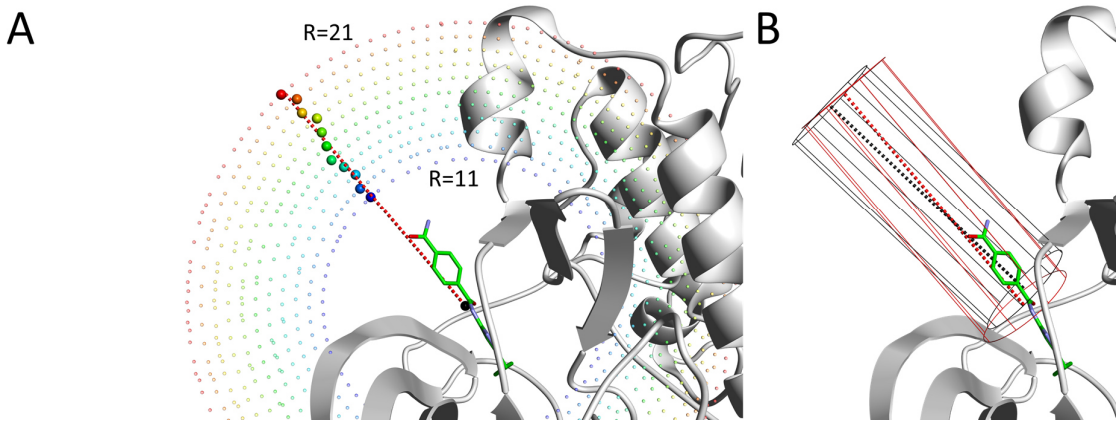

Figure 15. (A) Overview of the naive λ vector searching method. Shown are the partial spherical grids $\{\vec{\mathrm{h}}_{\mathrm{m}}(\mathrm{R}), \mathrm{m} = 1 \ldots \mathrm{m}_{max}\}$ between R=11 Å and R=21 Å in semi-transparent colors. For clarity, only the points intersecting with the plane defined by the red dotted line are shown. The selected grid points $\vec{\mathrm{h}}(R)_{max}$ for these grids are shown in opaque colors, while the given center of the pocket is shown as a black point. Here, the distance between $\vec{\mathrm{h}}(R)_{max}$ and the nearest protein heavy atom is between Rg + 2 Å and $Rg + d$ Å, where d = 10.5 is applied. Least squares fitting between all the selected grid points $\vec{\mathrm{h}}(\mathrm{R})_{max}$ and the given center of the pocket produces a line connecting the pocket and the bulk, shown as a dotted red line. (B) Cylinder generated by the naïve method (red) and by the RAMD method used in this work (black), ranging from 0 Å to 20 Å. The overlapping volume between the two cylinders is approximately 90%, where 99.9% of the configurations sampled during the McMD simulation were sampled in both the RAMD-derived cylinder, as well as inside the naively-derived cylinder.

Table 1. List of atoms used for the PCA.

Residue name	Atom name*
Ile10	CA
Ile52	CA
Asp86	CA
CS3	CAT
CS3	CAO
CS3	CAN

*The atom name corresponds to the one in the PDB file.

8 REFERENCES

Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindahl, E. (2015). Gromacs: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX, 1-2*, 19-25. doi:10.1016/j.softx.2015.06.001

Arakawa, T., Kamiya, N., Nakamura, H., & Fukuda, I. (2013). Molecular Dynamics Simulations of Double-Stranded DNA in an Explicit Solvent Model with the Zero-Dipole Summation Method. *Plos One, 8*(10). doi:10.1371/journal.pone.0076606

Asghar, U., Witkiewicz, A. K., Turner, N. C., & Knudsen, E. S. (2015). The History and Future of Targeting Cyclin-Dependent Kinases in Cancer Therapy. *Nat. Rev. Drug Discov., 14*(2), 130-146. doi:10.1038/nrd4504

Bartels, C., & Karplus, M. (1997). Multidimensional Adaptive Umbrella Sampling: Applications to Main Chain and Side Chain Peptide Conformations. *Journal of Computational Chemistry, 18*(12), 1450-1462. doi:10.1002/(Sici)1096-987x(199709)18:12<1450::Aid-Jcc3>3.0.Co;2-I

Bayly, C. I., Cieplak, P., Cornell, W. D., & Kollman, P. A. (1993). A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges - the Resp Model. *Journal of Physical Chemistry, 97*(40), 10269-10280. doi:10.1021/j100142a004

Bekker, G. J., Nakamura, H., & Kinjo, A. R. (2016). Molmil: A Molecular Viewer for the Pdb and Beyond. *Journal of cheminformatics, 8*. doi:10.1186/s13321-016-0155-1

Berman, H., Henrick, K., & Nakamura, H. (2003). Announcing the Worldwide Protein Data Bank. *Nature Structural Biology, 10*(12), 980-980. doi:10.1038/nsb1203-980

Berman, H., Henrick, K., Nakamura, H., & Markley, J. L. (2007). The Worldwide Protein Data Bank (Wwpdb): Ensuring a Single, Uniform Archive of Pdb Data. *Nucleic Acids Research, 35*, D301-D303. doi:10.1093/nar/gkl971

Bregman, D. B., Pestell, R. G., & Kidd, V. J. (2000). Cell Cycle Regulation and Rna Polymerase Ii. *Frontiers in Bioscience, 5*, D244-D257. doi:10.2741/Bregman

Bussi, G., Donadio, D., & Parrinello, M. (2007). Canonical Sampling through Velocity Rescaling. *Journal of Chemical Physics, 126*(1). doi:10.1063/1.2408420

Calbo, J., Serna, C., Garriga, J., Grana, X., & Mazo, A. (2004). The Fate of Pancreatic Tumor Cell Lines Following P16 Overexpression Depends on the Modulation of Cdk2 Activity. *Cell Death and Differentiation, 11*(10), 1055-1065. doi:10.1038/sj.cdd.4401481

Craig, I. R., Essex, J. W., & Spiegel, K. (2010). Ensemble Docking into Multiple Crystallographically Derived Protein Structures: An Evaluation Based on the Statistical Analysis of Enrichments. *J. Chem. Inf. Model., 50*(4), 511-524. doi:10.1021/ci900407c

Darve, E., Rodriguez-Gomez, D., & Pohorille, A. (2008). Adaptive Biasing Force Method for Scalar and Vector Free Energy Calculations. *Journal of Chemical Physics, 128*(14). doi:10.1063/1.2829861

Dickson, A., Ahlstrom, L. S., & Brooks III, C. L. (2016). Coupled Folding and Binding with 2d Window-Exchange Umbrella Sampling. *Journal of Computational Chemistry, 37*(6), 587-594. doi:10.1002/jcc.24004

Drapkin, R., LeRoy, G., Cho, H., Akoulitchev, S., & Reinberg, D. (1996). Human Cyclin-Dependent Kinase-Activating Kinase Exists in Three Distinct Complexes. *Proceedings of the National Academy of Sciences of the United States of America, 93*(13), 6488-6493. doi:10.1073/pnas.93.13.6488

Dunbar, J. B., Jr., Smith, R. D., Damm-Ganamet, K. L., Ahmed, A., Esposito, E. X., Delproposto, J., . . . Carlson, H. A. (2013). Csar Data Set Release 2012: Ligands, Affinities, Complexes, and Docking Decoys. *J. Chem. Inf. Model., 53*(8), 1842-1852. doi:10.1021/ci4000486

Echalier, A., Endicott, J. A., & Noble, M. E. M. (2010). Recent Developments in Cyclin-Dependent Kinase Biochemical and Structural Studies. *Biochimica Et Biophysica Acta-Proteins and Proteomics, 1804*(3), 511-519. doi:10.1016/j.bbapap.2009.10.002

Eldeiry, W. S., Tokino, T., Velculescu, V. E., Levy, D. B., Parsons, R., Trent, J. M., . . . Vogelstein, B. (1993). Waf1, a Potential Mediator of P53 Tumor Suppression. *Cell, 75*(4), 817-825. doi:10.1016/0092-8674(93)90500-P

Frenkel, D. S., B. (2002). Thermodynamic Integration *Understanding Molecular Simulation: From Algorithms to Applications* (Second Edition ed., pp. 168-172).

Fukuda, I., Kamiya, N., Yonezawa, Y., & Nakamura, H. (2012). Simple and Accurate Scheme to Compute Electrostatic Interaction: Zero-Dipole Summation Technique for Molecular System and Application to Bulk Water. *Journal of Chemical Physics, 137*(5), 054314. doi:10.1063/1.4739789

Fukuda, I., & Nakamura, H. (2012). Non-Ewald Methods: Theory and Applications to Molecular Systems. *Biophys. Rev., 4*(3), 161-170. doi:10.1007/s12551-012-0089-4

Fukuda, I., Yonezawa, Y., & Nakamura, H. (2011). Molecular Dynamics Scheme for Precise Estimation of Electrostatic Interaction Via Zero-Dipole Summation Principle. *Journal of Chemical Physics, 134*(16), 164107. doi:10.1063/1.3582791

Fukunishi, Y., Mikami, Y., & Nakamura, H. (2003). The Filling Potential Method: A Method for Estimating the Free Energy Surface for Protein-Ligand Docking. *Journal of Physical Chemistry B, 107*(47), 13201-13210. doi:Doi 10.1021/Jp035478e

Fukunishi, Y., Mikami, Y., & Nakamura, H. (2005). Similarities among Receptor Pockets and among Compounds: Analysis and Application to in Silico Ligand Screening. *J. Mol. Graphics Model., 24*(1), 34-45. doi:10.1016/j.jmgm.2005.04.004

Fukunishi, Y., Mitomo, D., & Nakamura, H. (2009). Protein-Ligand Binding Free Energy Calculation by the Smooth Reaction Path Generation (Srpg) Method. *J Chem Inf Model., 49*(8), 1944-1951. doi:10.1021/ci9002156

Goodsell, D. S., & Olson, A. J. (1990). Automated Docking of Substrates to Proteins by Simulated Annealing. *Proteins-Structure Function and Bioinformatics, 8*(3), 195-202. doi:10.1002/prot.340080302

Goto, J., Kataoka, R., Muta, H., & Hirayama, N. (2008). Asedock-Docking Based on Alpha Spheres and Excluded Volumes. *J. Chem. Inf. Model., 48*(3), 583-590. doi:10.1021/ci700352q

Hartwell, L. H. (2002). Nobel Lecture. Yeast and Cancer. *Biosci Rep, 22*(3-4), 373-394.

Herrera, R. E., Sah, V. P., Williams, B. O., Makela, T. P., Weinberg, R. A., & Jacks, T. (1996). Altered Cell Cycle Kinetics, Gene Expression, and G1 Restriction Point Regulation in Rb-Deficient Fibroblasts. *Molecular and Cellular Biology, 16*(5), 2402-2407.

Hess, B. (2008). P-Lincs: A Parallel Linear Constraint Solver for Molecular Simulation. *Journal of Chemical Theory and Computation, 4*(1), 116-122. doi:10.1021/ct700200b

Higo, J., Dasgupta, B., Mashimo, T., Kasahara, K., Fukunishi, Y., & Nakamura, H. (2015). Virtual-System-Coupled Adaptive Umbrella Sampling to Compute Free-Energy Landscape for Flexible Molecular Docking. *Journal of Computational Chemistry, 36*(20), 1489-1501. doi:10.1002/jcc.23948

Higo, J., Ito, N., Kuroda, M., Ono, S., Nakajima, N., & Nakamura, H. (2001). Energy Landscape of a Peptide Consisting of Alpha-Helix, 3(10)-Helix, Beta-Turn, Beta-Hairpin, and Other Disordered Conformations. *Protein Science, 10*(6), 1160-1171. doi:10.1110/ps.44901

Higo, J., Nishimura, Y., & Nakamura, H. (2011). A Free-Energy Landscape for Coupled Folding and Binding of an Intrinsically Disordered Protein in Explicit Solvent from Detailed All-Atom Computations. *J. Am. Chem. Soc., 133*(27), 10448-10458. doi:10.1021/ja110338e

Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., & Simmerling, C. (2006). Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins-Structure Function and Bioinformatics, 65*(3), 712-725. doi:10.1002/prot.21123

Hunt, T. (2002). Nobel Lecture. Protein Synthesis, Proteolysis, and Cell Cycle Transitions. *Biosci Rep, 22*(5-6), 465-486.

Hunt, T., Nasmyth, K., & Novak, B. (2011). The Cell Cycle Introduction. *Philosophical Transactions of the Royal Society B-Biological Sciences, 366*(1584), 3494-3497. doi:10.1098/rstb.2011.0274

Ikebe, J., Umezawa, K., Kamiya, N., Sugihara, T., Yonezawa, Y., Takano, Y., . . . Higo, J. (2011). Theory for Trivial Trajectory Parallelization of Multicanonical Molecular Dynamics and Application to a Polypeptide in Water. *Journal of Computational Chemistry, 32*(7), 1286-1297. doi:10.1002/jcc.21710

Ikeda, K., & Higo, J. (2003). Free-Energy Landscape of a Chameleon Sequence in Explicit Water and Its Inherent Alpha/Beta Bifacial Property. *Protein Science, 12*(11), 2542-2548. doi:10.1110/ps.03143803

Isralewitz, B., Gao, M., & Schulten, K. (2001). Steered Molecular Dynamics and Mechanical Functions of Proteins. *Current Opinion in Structural Biology, 11*(2), 224-230. doi:10.1016/S0959-440x(00)00194-9

Jeffrey, P. D., Ruso, A. A., Polyak, K., Gibbs, E., Hurwitz, J., Massague, J., & Pavletich, N. P. (1995). Mechanism of Cdk Activation Revealed by the Structure of a Cyclina-Cdk2 Complex. *Nature, 376*(6538), 313-320. doi:10.1038/376313a0

Jones, G., Willett, P., Glen, R. C., Leach, A. R., & Taylor, R. (1997). Development and Validation of a Genetic Algorithm for Flexible Docking. *Journal of Molecular Biology, 267*(3), 727-748. doi:10.1006/jmbi.1996.0897

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., & Klein, M. L. (1983). Comparison of Simple Potential Functions for Simulating Liquid Water. *Journal of Chemical Physics, 79*(2), 926-935. doi:10.1063/1.445869

Kamiya, N., Fukuda, I., & Nakamura, H. (2013). Application of Zero-Dipole Summation Method to Molecular Dynamics Simulations of a Membrane Protein System. *Chemical Physics Letters, 568-569*, 26-32. doi:10.1016/j.cplett.2013.03.014

Kamiya, N., Higo, J., & Nakamura, H. (2002). Conformational Transition States of a Beta-Hairpin Peptide between the Ordered and Disordered Conformations in Explicit Water. *Protein Science, 11*(10), 2297-2307. doi:10.1110/ps.0213102

Kamiya, N., Mashimo, T., Takano, Y., Kon, T., Kurisu, G., & Nakamura, H. (2016). Elastic Properties of Dynein Motor Domain Obtained from All-Atom Molecular Dynamics Simulations. *Protein Eng., Des. Sel., 29*(8), 317-325. doi:10.1093/protein/gzw022

Kamiya, N., Yonezawa, Y., Nakamura, H., & Higo, J. (2008). Protein-Inhibitor Flexible Docking by a Multicanonical Sampling: Native Complex Structure with the Lowest Free Energy and a Free-Energy Barrier Distinguishing the Native Complex from the Others. *Proteins-Structure Function and Bioinformatics, 70*(1), 41-53. doi:10.1002/prot.21409

Kinjo, A. R., Bekker, G. J., Suzuki, H., Tsuchiya, Y., Kawabata, T., Ikegawa, Y., & Nakamura, H. (2016). Protein Data Bank Japan (Pdbj): Updated User Interfaces, Resource Description Framework, Analysis Tools for Large Structures. *Nucleic Acids Res*. doi:10.1093/nar/gkw962

Kokubo, H., Tanaka, T., & Okamoto, Y. (2011). Ab Initio Prediction of Protein-Ligand Binding Structures by Replica-Exchange Umbrella Sampling Simulations. *Journal of Computational Chemistry, 32*(13), 2810-2821. doi:10.1002/jcc.21860

Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S. H., Chong, L., . . . Cheatham, T. E. (2000). Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res., 33*(12), 889-897. doi:10.1021/ar000033j

Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R., & Ferrin, T. E. (1982). A Geometric Approach to Macromolecule-Ligand Interactions. *Journal of Molecular Biology, 161*(2), 269-288. doi:10.1016/0022-2836(82)90153-X

Laio, A., & Parrinello, M. (2002). Escaping Free-Energy Minima. *Proceedings of the National Academy of Sciences of the United States of America, 99*(20), 12562-12566. doi:10.1073/pnas.202427399

Lim, S. H., & Kaldis, P. (2013). Cdks, Cyclins and Ckis: Roles Beyond Cell Cycle Regulation. *Development, 140*(15), 3079-3093. doi:10.1242/dev.091744

Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., & Shaw, D. E. (2010). Improved Side-Chain Torsion Potentials for the Amber Ff99sb Protein Force Field. *Proteins-Structure Function and Bioinformatics, 78*(8), 1950-1958. doi:10.1002/prot.22711

Ludemann, S. K., Lounnas, V., & Wade, R. C. (2000). How Do Substrates Enter and Products Exit the Buried Active Site of Cytochrome P450cam? 1. Random Expulsion Molecular Dynamics Investigation of Ligand Access Channels and Mechanisms. *Journal of Molecular Biology, 303*(5), 797-811. doi:10.1006/jmbi.2000.4154

Malumbres, M. (2006). Therapeutic Opportunities to Control Tumor Cell Cycles. *Clin Transl Oncol, 8*(6), 399-408.

Malumbres, M. (2014). Cyclin-Dependent Kinases. *Genome Biology, 15*(6). doi:10.1186/gb4184

Malumbres, M., & Barbacid, M. (2001). To Cycle or Not to Cycle: A Critical Decision in Cancer. *Nature Reviews Cancer, 1*(3), 222-231. doi:10.1038/35106065

Malumbres, M., & Barbacid, M. (2009). Cell Cycle, Cdks and Cancer: A Changing Paradigm. *Nat. Rev. Cancer, 9*(3), 153-166. doi:10.1038/nrc2602

Markey, M. P., Angus, S. P., Strobeck, M. W., Williams, S. L., Gunawardena, R. W., Aronow, B. J., & Knudsen, E. S. (2002). Unbiased Analysis of Rb-Mediated Transcriptional Repression Identifies Novel Targets and Distinctions from E2f Action. *Cancer Research, 62*(22), 6587-6597.

Martin, A., Odajima, J., Hunt, S. L., Dubus, P., Ortega, S., Malumbres, M., & Barbacid, M. (2005). Cdk2 Is Dispensable for Cell Cycle Inhibition and Tumor Suppression Mediated by P27(Kip1) and P21(Cip1). *Cancer Cell, 7*(6), 591-598. doi:10.1016/j.ccr.2005.05.006

Mashimo, T., Fukunishi, Y., Kamiya, N., Takano, Y., Fukuda, I., & Nakamura, H. (2013). Molecular Dynamics Simulations Accelerated by Gpu for Biological Macromolecules with a Non-Ewald Scheme for Electrostatic Interactions. *Journal of Chemical Theory and Computation, 9*(12), 5599-5609. doi:10.1021/ct400342e

Massova, I., & Kollman, P. A. (1999). Computational Alanine Scanning to Probe Protein-Protein Interactions: A Novel Approach to Evaluate Binding Free Energies. *J. Am. Chem. Soc., 121*(36), 8133-8143. doi:Doi 10.1021/Ja990935j

Miyamoto, S., & Kollman, P. A. (1992). Settle - an Analytical Version of the Shake and Rattle Algorithm for Rigid Water Models. *Journal of Computational Chemistry, 13*(8), 952-962. doi:10.1002/jcc.540130805

Morikami, K., Nakai, T., Kidera, A., Saito, M., & Nakamura, H. (1992). Presto(Protein Engineering Simulator) - a Vectorized Molecular Mechanics Program for Biopolymers. *Computers & Chemistry, 16*(3), 243-248. doi:10.1016/0097-8485(92)80010-W

Moroy, T., & Geisen, C. (2004). Cyclin E. *International Journal of Biochemistry & Cell Biology, 36*(8), 1424-1439. doi:10.1016/j.biocel.2003.12.005

Nakajima, N., Higo, J., Kidera, A., & Nakamura, H. (1997). Flexible Docking of a Ligand Peptide to a Receptor Protein by Multicanonical Molecular Dynamics Simulation. *Chemical Physics Letters, 278*(4-6), 297-301. doi:10.1016/S0009-2614(97)01074-9

Nakajima, N., Higo, J., Kidera, A., & Nakamura, H. (2000). Free Energy Landscapes of Peptides by Enhanced Conformational Sampling. *Journal of Molecular Biology, 296*(1), 197-216. doi:10.1006/jmbi.1999.3440

Nakajima, N., Nakamura, H., & Kidera, A. (1997). Multicanonical Ensemble Generated by Molecular Dynamics Simulation for Enhanced Conformational Sampling of Peptides. *Journal of Physical Chemistry B, 101*(5), 817-824. doi:10.1021/jp962142e

Nemet, J., Jelicic, B., Rubelj, I., & Sopta, M. (2014). The Two Faces of Cdk8, a Positive/Negative Regulator of Transcription. *Biochimie, 97*, 22-27. doi:10.1016/j.biochi.2013.10.004

Nguyen, H., Tran, T., Fukunishi, Y., Higo, J., Nakamura, H., & Le, L. (2015). Computational Study of Drug Binding Affinity to Influenza a Neuraminidase Using Smooth Reaction Path Generation (Srpg) Method. *J. Chem. Inf. Model., 55*(9), 1936-1943. doi:10.1021/acs.jcim.5b00319

Nurse, P., Masui, Y., & Hartwell, L. (1998). Understanding the Cell Cycle. *Nature Medicine, 4*(10), 1103-1106. doi:10.1038/2594

Nurse, P. M. (2002). Nobel Lecture. Cyclin Dependent Kinases and Cell Cycle Control. *Biosci Rep, 22*(5-6), 487-499.

Parrinello, M., & Rahman, A. (1981). Polymorphic Transitions in Single-Crystals - a New Molecular-Dynamics Method. *J. Appl. Phys., 52*(12), 7182-7190. doi:10.1063/1.328693

Pavletich, N. P. (1999). Mechanisms of Cyclin-Dependent Kinase Regulation: Structures of Cdks, Their Cyclin Activators, and Cip and Ink4 Inhibitors. *Journal of Molecular Biology, 287*(5), 821-828. doi:DOI 10.1006/jmbi.1999.2640

Peng, C., Zeng, W., Su, J., Kuang, Y., He, Y., Zhao, S., . . . Chen, X. (2016). Cyclin-Dependent Kinase 2 (Cdk2) Is a Key Mediator for Egf-Induced Cell Transformation Mediated through the Elk4/C-Fos Signaling Pathway. *Oncogene, 35*(9), 1170-1179. doi:10.1038/onc.2015.175

Polyak, K., Kato, J. Y., Solomon, M. J., Sherr, C. J., Massague, J., Roberts, J. M., & Koff, A. (1994). P27(Kip1), a Cyclin-Cdk Inhibitor, Links Transforming Growth-Factor-Beta and Contact Inhibition to Cell-Cycle Arrest. *Genes & Development, 8*(1), 9-22. doi:10.1101/gad.8.1.9

Polyak, K., Lee, M. H., Erdjumentbromage, H., Koff, A., Roberts, J. M., Tempst, P., & Massague, J. (1994). Cloning of P27(Kip1), a Cyclin-Dependent Kinase Inhibitor and a Potential Mediator of Extracellular Antimitogenic Signals. *Cell, 78*(1), 59-66. doi:10.1016/0092-8674(94)90572-X

Pozo, K., Castro-Rivera, E., Tan, C. F., Plattner, F., Schwach, G., Siegl, V., . . . Bibb, J. A. (2013). The Role of Cdk5 in Neuroendocrine Thyroid Cancer. *Cancer Cell, 24*(4), 499-511. doi:10.1016/j.ccr.2013.08.027

Pronk, S., Pall, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., . . . Lindahl, E. (2013). Gromacs 4.5: A High-Throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinformatics, 29*(7), 845-854. doi:10.1093/bioinformatics/btt055

Protein Data Bank. (1971). [Press release]

Rarey, M., Kramer, B., Lengauer, T., & Klebe, G. (1996). A Fast Flexible Docking Method Using an Incremental Construction Algorithm. *Journal of Molecular Biology, 261*(3), 470-489. doi:10.1006/jmbi.1996.0477

Ren, B., Cam, H., Takahashi, Y., Volkert, T., Terragni, J., Young, R. A., & Dynlacht, B. D. (2002). E2f Integrates Cell Cycle Progression with DNA Repair, Replication, and G(2)/M Checkpoints. *Genes & Development, 16*(2), 245-256. doi:10.1101/gad.949802

Roesley, S. N. A., Suryadinata, R., Morrish, E., Tan, A. R., Issa, S. M. A., Oakhill, J. S., . . . Sarcevic, B. (2016). Cyclin-Dependent Kinase-Mediated Phosphorylation of Breast Cancer Metastasis Suppressor 1 (Brms1) Affects Cell Migration. *Cell Cycle, 15*(1), 137-151. doi:10.1080/15384101.2015.1121328

Savitzky, A., & Golay, M. J. E. (1964). Smoothing + Differentiation of Data by Simplified Least Squares Procedures. *Anal. Chem., 36*(8), 1627-1639. doi:10.1021/ac60214a047

Scaltriti, M., Eichhorn, P. J., Cortes, J., Prudkin, L., Aura, C., Jimenez, J., . . . Baselga, J. (2011). Cyclin E Amplification/Overexpression Is a Mechanism of Trastuzumab Resistance in Her2(+) Breast Cancer Patients. *Proceedings of the National Academy of Sciences of the United States of America, 108*(9), 3761-3766. doi:10.1073/pnas.1014835108

Schmidt, M. W., Baldridge, K. K., Boatz, J. A., Elbert, S. T., Gordon, M. S., Jensen, J. H., . . . Montgomery, J. A. (1993). General Atomic and Molecular Electronic-Structure System. *Journal of Computational Chemistry, 14*(11), 1347-1363. doi:10.1002/jcc.540141112

Shaw, D. E., Deneroff, M. M., Dror, R. O., Kuskin, J. S., Larson, R. H., Salmon, J. K., . . . Wang, S. C. (2008). Anton, a Special-Purpose Machine for Molecular Dynamics Simulation. *Commun. Acm., 51*(7), 91-97. doi:10.1145/1364782.1364802

Sherr, C. J. (1996). Cancer Cell Cycles. *Science, 274*(5293), 1672-1677. doi:10.1126/science.274.5293.1672

Singh, N., & Warshel, A. (2010). Absolute Binding Free Energy Calculations: On the Accuracy of Computational Scoring of Protein-Ligand Interactions. *Proteins-Structure Function and Bioinformatics, 78*(7), 1705-1723. doi:10.1002/prot.22687

Sousa da Silva, A. W., & Vranken, W. F. (2012). Acpype - Antechamber Python Parser Interface. *BMC. Res. Notes, 5*, 367. doi:10.1186/1756-0500-5-367

Standley, D. M., Kinjo, A. R., Kinoshita, K., & Nakamura, H. (2008). Protein Structure Databases with New Web Services for Structural Biology and Biomedical Research. *Briefings in Bioinformatics, 9*(4), 276-285. doi:10.1093/bib/bbn015

Sugita, Y., & Okamoto, Y. (1999). Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chemical Physics Letters, 314*(1-2), 141-151. doi:10.1016/S0009-2614(99)01123-9

Torrie, G. M., & Valleau, J. P. (1977). Non-Physical Sampling Distributions in Monte-Carlo Free-Energy Estimation - Umbrella Sampling. *Journal of Computational Physics, 23*(2), 187-199. doi:10.1016/0021-9991(77)90121-8

Velankar, S., McNeil, P., Mittard-Runte, V., Suarez, A., Barrell, D., Apweiler, R., & Henrick, K. (2005). E-Msd: An Integrated Data Resource for Bioinformatics. *Nucleic Acids Research, 33*, D262-D265. doi:10.1093/narl/gki058

Volm, M., Koomagi, R., Stammler, G., Rittgen, W., Zintl, F., & Sauerbrey, A. (1997). Prognostic Implications of Cyclins (D1, E, a), Cyclin-Dependent Kinases (Cdk2, Cdk4) and Tumor-Suppressor Genes (Prb, P16(Ink4a)) in Childhood Acute Lymphoblastic Leukemia. *International Journal of Cancer, 74*(5), 508-512. doi:10.1002/(Sici)1097-0215(19971021)74:5<508::Aid-Ijc5>3.0.Co;2-7

Wada, M., Kanamori, E., Nakamura, H., & Fukunishi, Y. (2011). Selection of in Silico Drug Screening Results for G-Protein-Coupled Receptors by Using Universal Active Probes. *J. Chem. Inf. Model., 51*(9), 2398-2407. doi:10.1021/ci200236x

Wang, H., Nakamura, H., & Fukuda, I. (2016). A Critical Appraisal of the Zero-Multipole Method: Structural, Thermodynamic, Dielectric, and Dynamical Properties of a Water System. *J. Chem. Phys, 144*(11), 114503. doi:10.1063/1.4943956

Wang, J. M., Wang, W., Kollman, P. A., & Case, D. A. (2006). Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graph. Model., 25*(2), 247-260. doi:10.1016/j.jmgm.2005.12.005

Wang, J. M., Wolf, R. M., Caldwell, J. W., Kollman, P. A., & Case, D. A. (2004). Development and Testing of a General Amber Force Field. *Journal of Computational Chemistry, 25*(9), 1157-1174. doi:10.1002/jcc.20035

Wu, C. L., Kirley, S. D., Xiao, H., Chuang, Y. N., Chung, D. C., & Zukerberg, L. R. (2001). Cables Enhances Cdk2 Tyrosine 15 Phosphorylation by Wee1, Inhibits Cell Growth, and Is Lost in Many Human Colon and Squamous Cancers. *Cancer Research, 61*(19), 7325-7332.

Zhou, R. H. (2003). Free Energy Landscape of Protein Folding in Water: Explicit Vs. Implicit Solvent. *Proteins-Structure Function and Bioinformatics, 53*(2), 148-161. doi:10.1002/prot.10483

Zhou, Y., Wang, S., Gobl, A., & Oberg, K. (1999). Inhibition of Cdk2, Cdk4 and Cyclin E and Increased Expression of P27kip1 During Treatment with Interferon-Alpha in Carcinoid Tumor Cells. *Journal of Biological Regulators and Homeostatic Agents, 13*(4), 207-215.

Zhu, W. C., Giangrande, P. H., & Nevins, J. R. (2004). E2fs Link the Control of G1/S and G2/M Transcription. *Embo Journal, 23*(23), 4615-4626. doi:10.1038/sj.emboj.7600459

9 ACADEMIC ACCOMPLISHMENTS

Papers:

Bekker, G. J., Kamiya, N., Araki, M., Fukuda, I., Okuno, Y., & Nakamura, H. Accurate Prediction of Complex Structure and Affinity for a Flexible Protein Receptor and Its Inhibitor. *Under revision*.

Kinjo, A. R., Bekker, G. J., Suzuki, H., Tsuchiya, Y., Kawabata, T., Ikegawa, Y., & Nakamura, H. (2016). Protein Data Bank Japan (PDBj): Updated User Interfaces, Resource Description Framework, Analysis Tools for Large Structures. *Nucleic Acids Res*. doi:10.1093/nar/gkw962

Bekker, G. J., Nakamura, H., & Kinjo, A. R. (2016). Molmil: A Molecular Viewer for the PDB and Beyond. *Journal of Cheminformatics, 8*. doi:10.1186/s13321-016-0155-1

Presentations:

Kamiya, N., Bekker, G. J., Araki, M., Fukuda, I., Okuno, Y., & Nakamura, H. (2017). *Prediction of Complex Structure and Affinity between Cyclin-Dependent Kinase 2 and Its Inhibitor Using Multicanonical Molecular Dynamics and Thermodynamic Integration Simulations*. Poster presented at the 61st Annual Meeting of the Biophysical Society.

Bekker, G. J., Kamiya, N., Araki, M., Fukuda, I., Okuno, Y., & Nakamura, H. (2016). *Flexible Docking between Cyclin-Dependent Kinase 2 and Its Inhibitor Using Multicanonical MD*. Poster presented at the The 54th Annual Meeting of the Biophysical Society of Japan, Tsukuba.

Kamiya, N., Bekker, G. J., & Dasgupta, B. (2016). *Protein-Protein Binding Free-Energy by Advanced Md Simulations Facilitating Drug Design (Poster)*. Poster presented at the The 3rd project report of HPCI, Tokyo.

Bekker, G. J. (2016). *Molmil: A Molecular Viewer for the PDB and Beyond*. Oral presentation presented at the NII Shonan Meeting On Web-based Molecular Graphics, Shonan.

Bekker, G. J., Kamiya, N., & Nakamura, H. (2016). *An Efficient Methodology for Path Elucidation and Binding Free Energy Calculation Using iso-GMD and Umbrella Sampling*. Poster presented at the The 16th Annual Meeting of the Protein Science Society of Japan, Fukuoka.

Kamiya, N., Bekker, G. J., Araki, M., & Okuno, Y. (2016). *Flexible Docking Simulation between Cyclin-Dependent Kinase 2 and Its Inhibitor Using Multicanonical MD Method*. Poster presented at the The 16th Annual Meeting of the Protein Science Society of Japan, Fukuoka.

Kamiya, N., Bekker, G. J., Shiraki, T., & Nakamura, H. (2016). *In-Vitro and in-Silico Studies of Ligand Binding to the Nuclear Receptor PPARgamma Using FRET and MD*. Poster presented at the 29th Annual Symposium of the Protein-Society, Barcelona.

Bekker, G. J. (2015). *Molmil: An Efficient, High Quality Molecular Viewer for the Web*. Oral presentation presented at the Informatics in Biology, Medicine and Pharmacology 2015 Kyoto.

Bekker, G. J., Kinjo, A. R., & Nakamura, H. (2015). *Molmil: An Efficient, High Quality Molecular Viewer for the Web*. Poster presented at the Informatics in Biology, Medicine and Pharmacology 2015, Kyoto.

Bekker, G. J., Kamiya, N., & Nakamura, H. (2015). *A Novel Method to Determine the Optimal Unbinding Path between Receptor and Ligand Using Advanced Molecular Dynamics Simulations*. Poster presented at the The 53rd Annual Meeting of the Biophysical Society of Japan, Kanazawa.

Kamiya, N., Bekker, G. J., Shiraki, T., & Nakamura, H. (2015). *Experimental and Theoretical Studies of Ligand Binding to the Nuclear Receptor, PPARγ Using FRET and RAMD*. Poster presented at the The 15th Annual Meeting of the Protein Science Society of Japan, Tokushima.

10 ACKNOWLEDGEMENTS

First and foremost, I would like to express my gratitude to my supervisor, Professor Dr. Haruki Nakamura for his kind support and guidance. I would also like to thank the members of my defense committee; Professor Dr. Keiichi Namba, Professor Dr. Atsushi Nakagawa, Professor Dr. Junichi Takagi and Professor Dr. Kenji Mizuguchi for their insight and helpful comments and discussion. I would also like to thank the members of Nakamura-lab who have helped me both professionally, as well as with my daily life in Japan. And I would like to thank Professor Dr. Yasushi Okuno and Dr. Mitsugu Araki from Kyoto University, Dr. Kanji Oshima from Kaneka Corporation and Dr. Noriko Shimba from Panasonic Corporation for their collaboration. Finally, I would like to give a special thanks to Professor Dr. Narutoshi Kamiya, who guided me throughout my Ph.D. as my unofficial supervisor, as well as in life as my friend.

My work was supported by Grant-in-Aid for Scientific Research on Innovative Areas "Transcription cycle" (24118008) from the Ministry of Education, Culture, Sports, Science and Technology-Japan (MEXT). The computational resources at the Cybermedia Center, Osaka University were provided by the HPCI Research Project (hp150025, hp150146, hp150272, hp160010 and hp160213).