Computational Modeling of Ca^{2+} and Zn^{2+} Competition in Calbindin D_{28k}, Implications for Altering Calcium Homeostasis

By

Sara Ibrahim Omar

A thesis submitted to the
Faculty of Graduate Studies
of the University of Manitoba
in partial fulfilment of
the requirements for the degree of
Master of Science

Department of Chemistry
Faculty of Science
University of Manitoba
Winnipeg, Manitoba

December 2012

Abstract

Neurodegeneration in Alzheimer's disease is characterised by multiple pathologies including disrupted calcium homeostasis and elevated Zn^{2+} levels. Calbindin D_{28k} (CB-D_{28k}), which buffers Ca^{2+} and can bind Zn^{2+}, was suspected to be involved in these abnormalities. The PDB structure of this EF-hand protein shows that not all hands are well formed. Docking and molecular dynamics calculations were employed to achieve the two goals in this project. The first goal was to get a better structure of CB-D_{28k} to improve our understanding of its behavior. Calculations improved the structure protein: helix-loop-helix sequences were formed in all hands and most canonical interactions were formed in the four functional hands. The second goal was to test the Ca^{2+} binding capacity of Zn^{2+}-bound CB-D_{28k}. Analysis of calculated structures showed that the Ca^{2+} binding capability of Zn^{2+} bound protein was significantly compromised, permitting only half of the correct canonical interactions with the loop residues.

Acknowledgments

I would like to thank the following people who were always there to help and support me:

My beloved supervisor, Professor Kathleen Gough

My friends at the Gough group:
Margaret Rak, Richard Wiens, Fatemeh Farazkhorasani, Catherine Liao

The members of my committee:
Professor Benedict Albensi, Professor Joe O'Neil, Professor Georg Schreckenbach

Everyone in the Department of Chemistry

My family

Financial Support:
University of Manitoba Graduate Fellowship
NSERC

Acknowledgements for Resources:

1. This research has been enabled by the use of computing resources provided by WestGrid and Compute/Calcul Canada.
2. NAMD was developed by the Theoretical and Computational Biophysics Group in the Beckman Institute for Advanced Science and Technology at the University of Illinois at Urbana-Champaign.

Dedication

To my parents, who are always happier by my own success than I am.

Contents

List of Tables

List of Figures

List of Symbols and Abbreviations

GENERAL TERMS	
Abbreviation	Full Meaning
AD	Alzheimer's Disease
ADI	Alzheimer's Disease International Federation
ADT	AutoDockTools
AICD	Amyloid precursor protein intracellular domain
APP	Amyloid Precursor protein
Aβ	Amyloid-beta
CB-D_{28k}	Calbindin D_{28k}
CB-D_{9k}	Calbindin D_{9k}
CBP	Calcium Binding Protein
CDM	Cationic Dummy Method
CHARMM	Chemistry of Harvard Molecular Mechanics
DBI	Davier-Bouldin Index
DPF	Docking Parameter File
ER	Endoplasmic Reticulum
GA	Genetic Algorithm
GALS	Genetic Algorithm Local Search
GPF	Grid Parameter File
GUI	Graphical User Interface
HPC	High Performance Computing
IMPase	Myo-Inositol Monophosphatase
K_D	Dissociation constant
LGA	Lamarckian Genetic Algorithm
LS	Local Search
MD	Molecular Dynamics
MM	Molecular Mechanics
NAMD	Not-just-Another-Molecular-Dynamics
OZC	Ordinary Zinc Cation
PBC	Periodic Boundary Conditions
PDB	Protein Data Bank
PDBQT	Protein Data Bank- Charge- Atom type
PME	Particle Mesh Ewald
pSF	pseudo F-statistic

PSF	Protein Structure File
QM	Quantum Mechanics
Ran-BPM	Ran-binding protein M
RMSD	Root Mean Squared coordinate Deviation
SA	Simulated Annealing
SW	Solis and Wets
TIP3P	Transferable Intermolecular Potential 3 Point
TZDC	Tetrahedral Zinc Dummy Cation
vdW	Van der Waals
VMD	Visual Molecular Dynamics
AMINO ACID ABBREVIATIONS	
Ala	Alanine
Arg	Arginine
Asn	Asparagine
Asp	Aspartic acid
Cys	Cysteine
Gln	Glutamine
Glu	Glutamic acid
Gly	Glycine
His	Histidine
Lys	Lysine
Phe	Phenylalanine
Ser	Serine
Thr	Threonine
Tyr	Tyrosine

Chapter 1

Introduction

The purpose of this project was to explore the possibility that excess Zn^{2+} affects the calcium buffering properties of the protein Calbindin D_{28k} (CB-D_{28k}), thereby contributing to calcium dysregulation in Alzheimer's Disease (AD). This disease is the most common form of dementia, wherein there is a significant loss in cognitive ability (1). Prevalence of AD is high and increasing; according to the Alzheimer's Disease International federation (ADI), about 35.6 million people worldwide are living with dementia (2). This number is expected to reach 66 million and 115 million in the years 2030 and 2050 respectively (2). In another study, it has been estimated that 1 in 85 people worldwide will be living with AD by 2050 (3) . These numbers are on the rise, since no cure is available for AD (4) and both world population and lifespans are increasing. With its high incidence, the direct costs including hospice services and health care for dementia patients are forecasted to reach $200 billion in the US for the year 2012 (5). Dementia, including AD, not only affects patients but also their families and caregivers. It has been estimated that in the year 2011, caregivers in the US have provided about 17.4 billion hours of unpaid care, an equivalent of $210.5 billion (5). Therefore, a significant amount

of research has been directed towards understanding, preventing and curing this progressive neurodegenerative disease.

Pathological hallmarks of AD include neurofibrillary tangles and amyloid plaques (6). In the former hallmark, tau proteins (involved in stabilisation of microtubules and promotion of their assembly) are abnormally hyperphosphorylated. This causes them to accumulate in the neurons, forming neurofibrillary tangles (7). Amyloid plaques, on the other hand, are proteinaceous deposits made of amyloid beta (Aβ) peptides (abnormally cleaved amyloid precursor protein, APP) (8).

Calcium ions mainly act as secondary messengers in cell signalling (9). Its homeostasis, however, is disrupted in AD; a part of this may be due to Aβ oligomers since they sometimes form channels or activate channels that increase Ca^{2+} entry into cells (6). Another product of the abnormally cleaved APP is the amyloid precursor protein intracellular domain (AICD). AICD increases the expression of specific types of Ca^{2+} channels called ryanodine receptors, which are in the membrane of endoplasmic reticulum (ER). This causes increased efflux of Ca^{2+} from the ER, which in turn increases intracellular Ca^{2+} levels; resting levels of this cation in cells are in the 10^{-8} M range (1). One of the theories is that this recompartmentalisation of Ca^{2+} levels is the root cause of memory and learning decline in AD patients, and leads to cell apoptosis (neuronal loss) (1, 10, 11). Other biochemical changes include elevated Zn^{2+} (12), which has catalytic and structural roles in some proteins (9). CB-D_{28k} levels are also altered in AD (6, 13, 14). Several studies have shown that the levels of CB-D_{28k} producing cells or CB-D_{28k} expression, in more apoptosis-resistant neurons, are lower in AD (13-17). Another recent study, however, has shown that CB-D_{28k} levels in PS1-M146V mutant mice increase

instead (6). It has been proposed that CB-D_{28k} levels could depend on the age and/or the progression of the disease (6).

CB-D_{28k} functions as a calcium buffer and sensor and thus plays a part in calcium homeostasis. A study by Bauer et al. has demonstrated that CB-D_{28k} can also bind Zn^{2+} (18). Combined with the finding that Zn^{2+} concentration is elevated in AD (12), CB-D_{28k} is a good candidate for computational modeling with the goal of assessing the possibility that Zn^{2+} may compete with or otherwise interfere with the protein's normal calcium binding functions.

1.1 Structure and Function of Calbindin D_{28k}

CB-D_{28k} gets its name from some of its properties. It is a calcium binding protein and its expression is vitamin D-dependent (except in brain) (19). The suffix, 28k (28-kilo daltons), is the first estimated mass of CB-D_{28k} isolated from chick intestinal mucosa by gel filtration chromatography (20). The molecular mass of the protein is now known to be around 30 kilo daltons (21, 22); however, its name has not been modified. It is present in different locations in the body including kidneys, intestines, bone and other endocrine tissues (19). It is also abundant in many regions in the brain. However, it has higher concentrations in the cerebellum, caudate nucleus and the hippocampus, especially the dentate gyrus and the CA1 regions (23, 24). Studies have shown discrepancies in the distribution of CB-D_{28k} within cells (25, 26); more recent studies indicate that, in the brain, CB-D_{28k} is a cytosolic protein (27, 28).

1.1.1 Structure of Calbindin D_{28k}

Calcium binding proteins are made of building units called EF-hands. These hands are composed of helix-loop-helix structural motifs that were first labeled by Kretsinger in 1973 (29). EF-hands tend to occur in pairs although some exceptions exist (30). As shown in Figure 1.1, the two helices are almost perpendicular to each other and are connected by a few amino acids forming a loop. Helices in this motif are responsible for packing interactions between other hands in the protein or interactions with other target proteins (30).

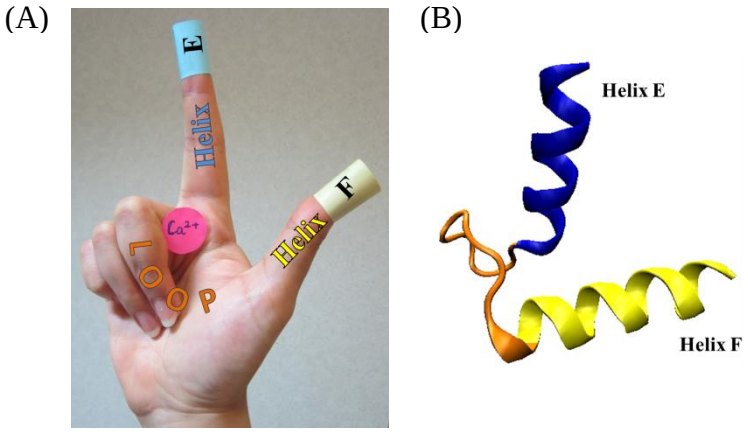


Figure 1.1: Inspired from Berg, Tymoczko & Stryer, 2006 **(28)**. (A) S. Omar, Image of right hand, illustrating the typical form of an EF-hand. (Photograph by S. I. Omar, November 2012). (B) Schematic of a typical EF-hand showing the N-terminal E-helix (blue) and the C-terminal F-helix (yellow). The loop residues (orange) form the calcium-binding site.

Depending on the sequence and length of the loop, it can be either canonical or non-canonical, as reviewed by Gifford et al. (30). The latter are functional (i.e. can bind Ca^{2+}) but vary in length or composition from more typical loops. Most EF-hand proteins, including CB-D_{28k}, have canonical loops and, for this thesis, only these are discussed. Canonical loops have 12 amino acids that are generally conserved in composition and

sequence (31). Figure 1.2 shows a schematic of typical canonical binding. The loop residues form a coordination sphere around Ca^{2+}, enabling seven binding interactions. As illustrated, interacting residues are at positions 1, 3, 5, 7, 9 and 12. The residue at position 12 is usually bidentate: two points interact with Ca^{2+}. Water can sometimes form a bridge between residue sidechains and Ca^{2+}, especially at position 9. Interacting residues are typically aspartate at positions 1, 3, 5 and 9 while positions 6 and 12 are usually glycine and glutamate respectively. However, other amino acids are also common, examples: asparagine at position 3, serine or asparagine at position 5 and glycine or glutamine at position 9.

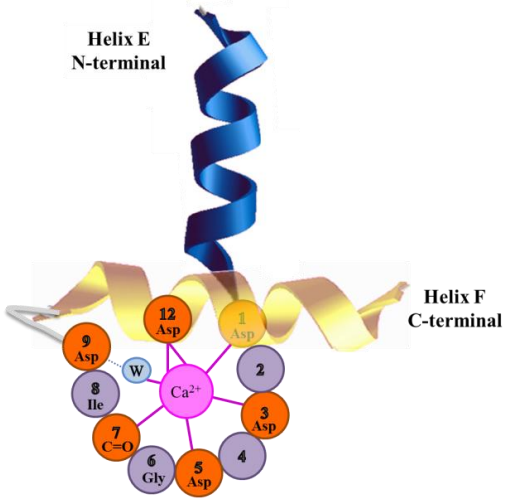


Figure 1.2: Illustration of canonical loops, inspired from Gifford, Jessica L. (2007), showing typical canonical binding. Helix E is connected to helix F by loop residues that interact with Ca^{2+} (pink). These interactions are represented by pink sticks. Note that residue 1 is at the end of the helix E, marking the first residue in the loop, while site 12 is within helix F. Shown in blue is a bridging water molecule. Note: Helix F is in front of helix E and has been rendered semi-transparent so that helix E and the loop residues can be seen clearly.

In 2006, the Cavanagh group published the NMR solution structure of the Ca^{2+}-bound form of the protein (21). In NMR spectroscopy, restraints (such as distance and

dihedral angles) are obtained from experimental data; then, models compatible with these restraints are generated. NMR experiments always yield an ensemble of proposed models; protein structures may fluctuate in solution and, when compared to X-ray crystallography, NMR data tend to be less precise (32). As a result, different models could be consistent within the experimental restraints. This situation is reflected in the structure of CB-D_{28k} deposited in the Protein Data Bank (PDB). The protein has two entries, 2G9B and 2F33 (with and without water refinement respectively), each of which has 10 proposed models (21).

CB-D_{28k} is made up of 261 residues, constituting three pairs of EF-hands. This makes CB-D_{28k} the largest known EF-hand protein. Although it has six EF-hands, it can only bind a maximum of four calcium ions (Ca^{2+}) with high affinity at hands 1, 3, 4 and 5 (21). Binding of Ca^{2+} to CB-D_{28k} is successive: it binds to loops of EF-hands 1, 4 and 5 then lastly 3 (31, 33). The fully bound protein is denoted as the holo form while the structure with no Ca^{2+} bound is called the apo form.

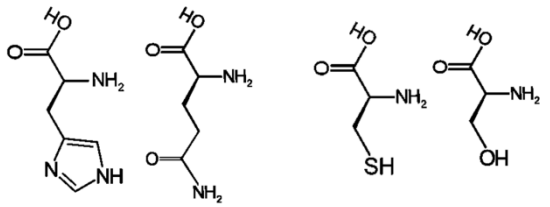
1.1.2 Functions of Calbindin D_{28k}

CB-D_{28k} is a calcium buffer, in the sense that it maintains the levels of intracellular Ca^{2+} within specific limits (34). It also functions as a calcium sensor, where it plays a regulatory role towards other proteins (34). Binding of Ca^{2+} has the effect of causing a large conformational change in calcium sensor proteins. This change in conformation enables the interaction between some segment of the sensor protein and other target proteins such as Ran-binding protein M (RanBPM), procaspase-3/caspase-3 and myo-

inositol monophosphatase (IMPase), thereby modulating their activity (22). For this thesis, the focus is on the calcium buffering capability of the protein.

1.1.3 Calbindin D_{28k} binding Zinc ions

A study by Bauer et al. has demonstrated that CB-D_{28k} can bind at least three Zn^{2+} ions (18); however, the Zn^{2+} binding locations in CB-D_{28k} could not be specified. Experimental results in that study demonstrated that there is a high affinity site (association constant (lg K_a= 5.9) in residues 1-86 that constitute EF-hands 1 and 2. The other two sites had lower affinities, with association constants (lg K_a) 4.9 and 5.1. The first of these weaker binding sites was deduced to be in region of EF-hands 3 & 4 (residues 87-175). The last site was crudely estimated to be somewhere within the N-terminal half of the molecule. Since histidine and cysteine amino acids are the most common Zn^{2+} coordinating residues, they were mutated in an attempt to discover their significance in Zn^{2+} binding to CB-D_{28k} (18).



Histidine Glutamine Cysteine Serine

Figure 1.3: Structures of the amino-acids histidine, glutamine, cysteine and serine respectively. Generally, amino acids do not substitute well for histidine in terms of structure **(29)**. However, compared to other amino acids, glutamine seems the closest in terms of polarity and geometric space. In the case of cysteine, serine seems a good replacement since SH and OH are bioisosteres **(30)**.

Association constants of mutated CB-D_{28k}, having all its cysteine residues mutated to serine, were similar to the wild type CB-D_{28k}. This indicated that cysteine is

not significant for Zn^{2+} binding in the protein. To assess the importance of histidine, residues 5, 22, 80 and 114 were mutated to glutamine, one at a time. Results, reflected in the values of the association constants, showed that histidine 80 is very important for Zn^{2+} binding to CB-D_{28k}. Decrease in affinity associated with other histidine residues' mutation (5, 22 and 114) was less significant. This indicates that histidine residues 5, 22 and 114 are less important for Zn^{2+} binding to CB-D_{28k}.

1.2 Goals of thesis

The first goal of this project is to get a good representative structure of the holo form of the protein (fully bound to Ca^{2+}). The ensemble of models provided in the PDB reflects non-specificity of the structure of CB-D_{28k}. Moreover, the structure of the protein deviates from what is expected in the sense that some parts of the protein are not well formed. As an example, EF-hands 4 and 6 both have a helix and a long loop only, but are lacking the helix F, as shown in Figure 1.4. Moreover, the ensemble of proposed models of the holo protein lacks the bound Ca^{2+} ions. The **first hypothesis** is that the use of computational tools could help obtain a more accurate structure of holo CB-D_{28k} and thereby provide more insight as to why hands 1, 3, 4 and 5 are functional while EF-hands 2 and 6 are not.

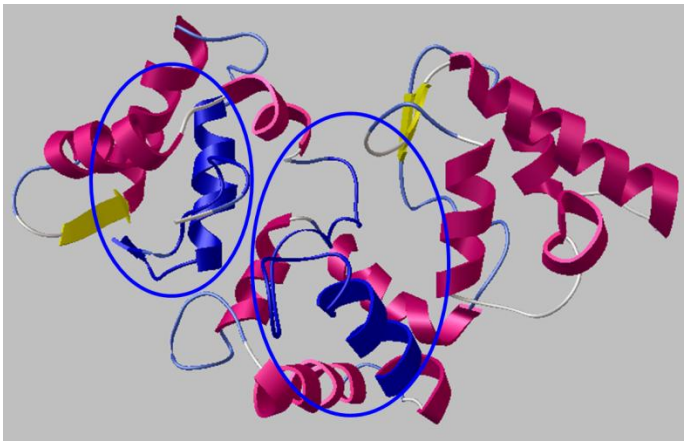


Figure 1.4: The first model in the 2G9B entry in the Protein Data Bank. The secondary structure of the protein reveals that EF-hands 4 and 6 (colored and circled in blue) have Helix-E and a very long loop but no Helix-F.

The second goal is to test the possible interference of Zn^{2+} binding with the Ca^{2+} binding capability of the protein. As mentioned earlier, Zn^{2+} levels are elevated in AD. Since CB-D_{28k} was found to bind Zn^{2+}, there could be some sort of competition, or

interference, between these ions and CB-D_{28k}. If true, this could contribute to altered Ca^{2+} homeostasis in the disease. The **second hypothesis** of this project is that the binding of Zn^{2+} to CB-D_{28k} will reduce its Ca^{2+} binding capacity.

1.3 Means to achieving goals of thesis

In order to achieve the goals defined above, computational modeling of CB-D_{28k} was employed. Different computational methods based on molecular mechanics and molecular dynamics were used; details are provided in the Methods section. Docking calculations were run to predict the cation-protein complex structure. Molecular dynamics (MD) simulations were used to allow the relaxation of the protein complex, allowing it to evolve with the passage of time. In an effort to validate the proposed docking methodology, Autodock 4.2 was used to predict the binding conformation of the protein calbindin D_{9k} (CB-D_{9k}), a protein with a single pair of EF-hands that each bind a Ca^{2+}, for which structure and ion locations have been solved (30). Relaxed structures provided other hypothetical representative structures for the calcium-bound form of CB-D_{28k}. These hypothetical structures were assessed for improvement in Ca^{2+} binding, which would indicate they should a more accurate presentation of the holo protein.

Docking calculations of Zn^{2+} into CB-D_{28k} yielded possible positions of Zn^{2+} binding to the protein. Molecular dynamics simulations on CB-D_{28k}, after docking Zn^{2+}, provided hypothetical structures of the Zn^{2+}-bound form of CB-D_{28k}. Computational modeling of the possible binding interference between Zn^{2+} and Ca^{2+} to CB-D_{28k} was therefore feasible, and these calculations were conducted.

Chapter 2

Tools and Methods

Molecules can be computationally modeled using different levels of theory. These theories can be classified into *ab initio* Quantum Mechanics (QM) and Molecular Mechanics (MM). '*Ab initio'* is a Latin phrase that means 'from the beginning', which refers to this level of theory being from first principles (35).

In QM, a system is described by a wavefunction, which it is a function of the coordinates of all the electrons in the system, and their interactions with nuclei. A wavefunction contains all the information about the system (36). Given a wavefunction of a system, the time dependent Schrödinger equation (a partial differential equation) can be used to calculate the future wavefunction at a specific time. The solutions to the equation for such systems provide different parameters of the system including energy as well as molecular structure. The Schrödinger equation has been solved for a single electron atom and can provide the exact values of parameters for this system; the equation has not been solved exactly for any larger systems. In order to solve the equation for larger systems, different approximations have been employed, including the

Born-Oppenheimer approximation that ignores the motion of nuclei since their movement is much slower than that of electrons. These approximations enable calculations on systems with tens of atoms but not larger ones. This is because QM calculations are very computationally expensive, with calculations being on the N^4 scale for the typical Hartree-Fock method where N is the number of basis functions (a set of linearly summed functions that together form molecular orbitals). Since proteins are large systems, usually with hundreds or thousands of atoms, QM cannot be applied to such systems.

Another simpler level of theory for molecular modeling is MM. This type of modeling is based on classical mechanics. Unlike QM, within this model, a system is described as a function of its chemical bonding in which atoms are treated as particles and bonds are treated as springs. The potential energy in a MM system is calculated using a force field. Force fields define energy functions, spring force constants and the equilibrium interaction values of bonded interactions. These values are usually calibrated to agree with experimental results. Empirical factors are usually obtained from experiment to account for non-bonded interactions such as van der Waals, electrostatic and hydrogen bonding.

Although MM is less computationally expensive and can sometimes provide good results, it cannot replace QM for certain calculations especially those involving bond formation/breakage or electron distribution. Dynamic simulations of protein can be achieved with a different level of calculation known as molecular dynamics (MD) simulations. Solving Newton's laws of motion in MM is the equivalent of solving the Schrödinger equation in QM. These laws relate the derivative of the potential energy, calculated from the force field, to the change in atomic positions, as a function of time.

This creates a series of new states over the simulation time called frames. Note that these simulations provide a prediction to the evolving path of the molecular motions, rather than accurately defining them. In fact, since each new frame calculation is dependent on the previous one, which itself is a prediction, long simulations could actually diverge away from the real path of the molecular motion.

Since CB-D_{28k} is a big system and cannot be handled by QM, MM was used for calculating the energies of the system in the Ca^{2+} and Zn^{2+} docking calculations as well as the equilibration simulation of the protein. MD simulations were employed in an effort to improve the results.

2.1 Computational Modeling Tools

As explained above, different computational techniques had to be employed in order to accomplish the different goals of this project. Docking was necessary to introduce the ligands into the receptor to get the protein-ligand complex structure. This was achieved by using a suite of programs called AutoDock4.2 (37). Molecular dynamics (MD) were also employed to simulate the behaviour of the protein with the passage of time. MD simulations software called NAMD (38, 39) was utilised for this calculation. Since these simulations provide a vast range of possible conformations of the protein, representative structures of the different forms of the protein were required; an analysis utility called PTRAJ (40) is used for this purpose. This is particularly important to account for the flexibility of the protein, since proteins in both their bound and unbound forms are not static. This section will describe these different software suites in detail.

2.1.1 AUTODOCK 4.2

The first goal of this project is to get a good representative structure of the holo form of CB-D_{28k}. Success in achieving this goal is measured by the fulfillment of two requirements. First, Ca^{2+} ions should be introduced into their correct positions in their binding sites: the functional canonical loops 1, 3, 4 and 5. Second, the different amino acids in the protein should conform to their correct secondary structures, that is, six helix-loop-helix sequences composing six EF-hands should be formed. The second goal of this project is to test the possible interference of Zn^{2+} binding with the Ca^{2+} binding capability of the protein. For this test, the locations of Zn^{2+} binding to CB-D_{28k} had to be identified.

The docking computational technique was used to obtain the protein-cation complex structure. In essence, this technique involves predicting the most favorable binding position of a ligand to its receptor. Choice of our docking software was based on a recent review of various docking software. The article (41) shows that the most cited docking software in 2006 was Autodock (37). Further investigation showed that this software offered all the required capability; it was well supported with tutorial information, and was readily accessible through the computational resources (see 2.3). This software was used to find the ligand (Ca^{2+} or Zn^{2+} ions) binding positions to their receptors in the protein (CB-D_{28k}). Autodock 4.2 is a suite of three programs namely AutoTors, AutoGrid and Autodock (42). The function of each of these is explained in detail in the Autodock User Manual (43) and is described briefly below, in the order in which they were used.

The graphical user interface (GUI) called AutoDockTools4 (ADT) (44) was also utilised. It is designed to facilitate the use of Autodock4.2. To elaborate, it creates an interactive computer display that allows users to visualise ligand and protein molecules, prepare the ligand and receptor files, as well as setup parameter files required for the docking calculation. In addition, it reads docking results, allowing docked poses to be visualised and analyzed by the user (37). ADT also includes other programs that help in further computations such as their Babel program, which adds missing hydrogen atoms to molecules and calculates atom partial charges using the Gasteiger method (45).

The two starting elements required for a docking calculation are the PDB files of the protein and ligand, obtained from the PDB website (46). To start with, Gasteiger charges are computed for the ligand using the Babel program in ADT. The AutoTors program adds calculated partial charges on non-polar hydrogen atoms to their bonded carbon atoms and merges them to form a united atom representation. Removal of non-polar hydrogen atoms is beneficial in reducing computational time and memory space (42, 47). AutoTors also defines rotatable bonds in the ligand, if any, and assigns the 'root' and 'branches' parts of the molecule. The last step in ligand preparation is the assignment of Autodock 4.2 atom types. These are abbreviations, composed of one or two letters, which classify atoms based on their element and bonding environment. As an example, 'N' is the atom type for non-hydrogen bonding nitrogen whereas 'NA' denotes a nitrogen atom that forms one hydrogen bond. The output of these manipulations is a ligand PDBQT file. This file format is similar to PDB files with the addition of partial charges (Q), atom types (T) and torsions, if any, and hence its name.

Similarly, a protein PDBQT file is required. Initially, any missing hydrogen atoms are detected by Babel and added to the protein. Gasteiger charges are computed, then a united atom representation is employed as described above. Finally, Autodock 4.2 atom types are assigned. Autodock4.2 allows only a selected portion of the receptor to be flexible. The user-defined flexible residues are saved separately from the rest of the protein. Thus, the protein is now divided into two PDBQT files that are treated independently: a flexible receptor and a rigid portion constituting the rest of the protein. At this point in the calculations, ADT has been used to generate three PDBQT files for the ligand, flexible and rigid parts of the protein. This is necessary since Autodock4.2 requires that both the ligand and protein be in the PDBQT format (37).

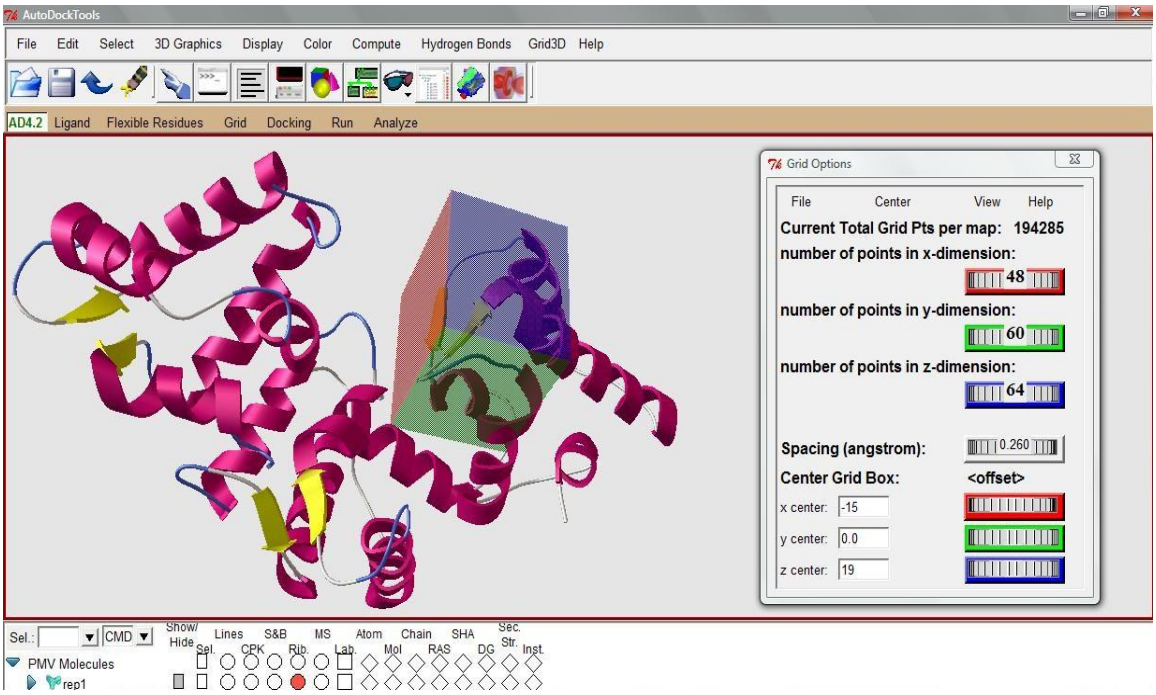


Figure 2.1: A snap shot showing a grid box covering the loop of EF-hand 1 in Calbindin D_{28k} in ADT viewer. The box is in three colors; red in the x-dimension, green in the y-dimension and blue in the z-dimension. The window labelled "Grid Options" to the right shows the variables that can be changed to control grid box location and size.

Next in use is the program AutoGrid; it calculates the energy of a probe atom at regularly spaced points in a 3D mesh, forming a box, placed around the receptor. Three user-defined variables control the size and location of the box: the number of points in each dimension in the box, the spacing between the grid points and its center point. The grid box should be placed such that it covers the binding site as well as the flexible receptors assigned in the protein. Note how this is confining the more detailed calculation (which is time consuming) to the known region of interest. The best guess at the correct placement of the grid box is achieved by manipulating the variables and examining the results in the ADT viewer as shown in Figure 2.1 above.

AutoGrid places a probe atom at a point in the grid and calculates its interaction energies with the rigid protein, using the AMBER force field (48), described later in this section. This is repeated for all atom types in the ligand and flexible receptor over all the points in the grid box. The calculated interaction energies for every atom type are saved in a grid map file. In addition, desolvation and electrostatics grid maps are also calculated. Grid map files enhance the speed of docking calculations since they serve as look-up tables during docking. Thus a calculation with the order of N^2 (N= number of interacting atoms) becomes of order N only (37, 42). These advantages come at the expense of flexibility because only a limited portion of the protein can be treated as flexible.

```
npts 48 60 62                          # num.grid points in xyz
gridfld rigid.maps.fld                 # grid_data_file
spacing 0.2                            # spacing(A)
receptor_types A C HD N NA OA SA       # receptor atom types
ligand_types Ca A C OA HD N            # ligand atom types
receptor rigid.pdbqt                   # macromolecule
gridcenter -14.0 -0.5 18.7             # xyz-coordinates or auto
smooth 0.5                             # store minimum energy w/in rad(A)
map rigid.Ca.map                       # atom-specific affinity map
map rigid.A.map                        # atom-specific affinity map
map rigid.C.map                        # atom-specific affinity map
map rigid.OA.map                       # atom-specific affinity map
map rigid.HD.map                       # atom-specific affinity map
map rigid.N.map                        # atom-specific affinity map
elecmap rigid.e.map                    # electrostatic potential map
dsolvmap rigid.d.map                   # desolvation potential map
dielectric -0.1465                     # <0, AD4 distance-dep.diel;>0,
constant
```

Figure 2.2: An example of a Grid Parameter file. Fields highlighted in yellow are the main variables that change for different calculations. Sentences after the hash symbol are not read by AutoGrid; they provide a brief description of the command line.

Commands to run AutoGrid including specifications for the calculation are all written in a Grid Parameter File (file extension .gpf), shown in Figure 2.2. The commands for setting the variables of the grid box are 'npt', 'spacing' and 'gridcenter'. These commands refer to the number of grid points in the dimensions x, y and z, the spacing between these points and the coordinates of the center point of that box, respectively. The command 'gridfld' specifies the creation of a field file (file extension .maps.fld). Field files are required for AutoDock (described later) to verify that all the maps generated by AutoGrid are compatible with the maps AutoDock reads for a specific docking calculation. It also specifies a file with extension .maps.xyz, which contains the maximum and minimum values of the x, y and z, coordinates, giving the extent of the grid box. Field files also have a list of the grid maps calculated by AutoGrid that are required for docking calculations by AutoDock. The command 'receptor_types' is used to specify the atom types in the rigid part of the receptor. The flexible part of the protein as well as the ligand's atom types are all listed after the command 'ligand_types'. A

command called ‘receptor’ is used to indicate the PDBQT file of the rigid part of the protein created earlier. The force field used by AutoDock is optimised such that the van der Waals and hydrogen bonding potentials are smoothed by 0.5 Å; this is indicated by the ‘smoothing’ parameter. Grid maps containing grid point energies between each ‘ligand_type’ atom type and the rigid part of the receptor are calculated using the ‘map’ command. Each of these grids is saved separately in a file with a ‘.atom_type_symbol.map’ extension. Electrostatic and desolvation potential energy grid maps are calculated using the commands ‘elecmap’ and ‘dsolvmap’ respectively and are saved in files with extensions ‘.e.map’ and ‘.d.map’. The command ‘dielectric’ specifies that a distance dependent dielectric function be calculated. AutoDock calculations are calibrated to use a dielectric value of -0.1465 (37). After running AutoGrid, grid map files discussed above are generated. These are used by AutoDock for the docking calculation.

The third of this suite of programs is AutoDock. This program enables the exploration of the conformational space of a ligand in the region specified by the grid box, using a search algorithm. Autodock finds different possible docked conformations of the ligand and clusters similar conformations together. In essence, the program tries to dock the ligand to a set of grids that describe the target (rigid part) protein. Execution of AutoDock has three main requirements. First are the PDBQT files of the ligand and the flexible part of the protein generated by ADT. In addition, the different grid maps previously generated by AutoGrid are also required. Lastly, a parameter file that sets the docking parameters as well as specifies files for the docking calculation is also needed.

All these requirements have been discussed earlier except the parameter file, called docking parameter file (extension .dpf).

```
autodock_parameter_version 4.2  # used by autodock to validate parameter set
outlev 1                        # diagnostic output level
intelec                         # calculate internal electrostatics
seed pid time                   # seeds for random generator
ligand_types Ca A C HD OA N     # atoms types in ligand
fld rigid.maps.fld              # grid_data_file
map rigid.Ca.map                # atom-specific affinity map
map rigid.A.map                 # atom-specific affinity map
map rigid.C.map                 # atom-specific affinity map
map rigid.HD.map                # atom-specific affinity map
map rigid.OA.map                # atom-specific affinity map
map rigid.N.map                 # atom-specific affinity map
elecmap rigid.e.map             # electrostatics map
desolvmap rigid.d.map           # desolvation map
move ligand.pdbqt               # small molecule
flexres flexible.pdbqt       # file containing flexible residues
about 0.0 0.0 0.0            # small molecule center
tran0 random                 # initial coordinates/A or random
axisangle0 random            # initial orientation
dihe0 random                 # initial dihedrals (relative) or random
tstep 1.0                    # translation step/A
qstep 25.0                   # quaternion step/deg
dstep 25.0                   # torsion step/deg
torsdof 1                    # torsional degrees of freedom
rmstol 1.0                   # cluster_tolerance/A
extnrg 1000.0                # external grid energy
e0max 0.0 10000              # max initial energy; max number of retries
ga_pop_size 300              # number of individuals in population
ga_num_evals 30000000        # maximum number of energy evaluations
ga_num_generations 28000     # maximum number of generations
ga_elitism 1        # number of top individuals to survive to next generation
ga_mutation_rate 0.02        # rate of gene mutation
ga_crossover_rate 0.8        # rate of crossover
ga_window_size 10            #
ga_cauchy_alpha 0.0          # Alpha parameter of Cauchy distribution
ga_cauchy_beta 1.0           # Beta parameter Cauchy distribution
set_ga                       # set the above parameters for GA or LGA
sw_max_its 300               # iterations of Solis & Wets local search
sw_max_succ 4                # consecutive successes before changing rho
sw_max_fail 4                # consecutive failures before changing rho
sw_rho 1.0                   # size of local search space to sample
sw_lb_rho 0.01               # lower bound on rho
ls_search_freq 0.06          # probability of performing local search on individual
set_psw1                     # set the above pseudo-Solis & Wets parameters
unbound_model bound          # state of unbound ligand
ga_run 100                   # do this many hybrid GA-LS runs
analysis                     # perform a ranked cluster analysis
```

Figure 2.3: An example of a docking parameter file. Fields highlighted in yellow are different for every docking calculation. Sentences after the hash symbol are comments that briefly describe the command line but are not read by AutoDock.

Figure 2.3 shows an example of a docking parameter file. The first line in this file indicates the force field parameter file to be used by AutoDock during the docking calculation. Information written about state variables is set by the command ‘outlev’. A value of 1 (default) dictates that the minimum, maximum and mean of each state variable be written at the end of each generation. The command ‘intelec’ is included by default to calculate the internal electrostatic energies of the ligand. The energy of the unbound ligand is specified by the command ‘unbound_model’. By default, this is equal to the energy of the protein-bound form of the ligand.

A list of atom types in the ligand and the flexible part of the protein come after ‘ligand_types’; this should be identical to those listed in the grid parameter file. The keywords ‘fld’, ‘map’, ‘elecmap’ and ‘desolvmap’ are used to specify the grid data field file, the different grid maps for the various atom types, as well as the electrostatic and desolvation grids previously calculated by AutoGrid respectively. These grid maps represent the rigid part of the protein. The flexible portion of the protein, previously assigned a separate PDBQT file, is indicated by the command ‘flexres’.

A few commands then follow to set various parameters for the ligand. The command ‘move’ is used to specify the name of the ligand’s PDBQT file. The center of the ligand about which rotations are made, with reference to the ligand PDBQT file, is specified by the command ‘about’; in this case, the ligand is centered at (0.0, 0.0, 0.0). The initial coordinates of the center of the ligand, orientation and relative dihedrals are specified by the commands ‘tran0’, ‘axisangle0’ and ‘dihe0’ respectively; in this case, the ligand is set to start at a random position. The calculation of the energy cost for losing torsional freedom upon binding requires definition of the number of rotatable bonds in

the ligand; ‘torsdof’ (torsional degrees of freedom) specifies this value: 1 for Ca^{2+} and Zn^{2+}. The keyword ‘extnrg’ is used to define the energy of any atoms, in kcal/mol, that fall outside the grid during docking. This value is usually high, 1000 kcal/mol, and thus conformations outside the box would be excluded from the docking results.

As discussed earlier, AutoDock clusters similar conformations when the command ‘analysis’ is specified. It does so by arranging the energies of the docked conformations in ascending order. The root mean squared deviations (RMSD) of all atoms in the ligand for the various docked poses are calculated and conformations within a user specified value are binned together in a cluster. Defining the RMSD cut off is set with the keyword ‘rmstol’ (root mean squared tolerance); set to 1 Å in this case.

Autodock is used to explore possible docked conformations using one of its search algorithms: Genetic Algorithm (GA), Simulated Annealing (SA) or Lamarckian Genetic Algorithm (LGA), and the latter was employed for this research. In GA, a random population of individuals is generated and the more fit of these individuals are selected. The fittest individuals reproduce by mutation and crossover creating offspring that form the new population. This process is iterated over a number of generations in an attempt to get better-fit individuals. Another search algorithm in AutoDock is the Monte Carlo SA method. In this approach, the system starts at a random state. This system is heated, which allows it to take new random conformations. The system is then cooled down so it reaches a local minimum. This process is iterated until an optimum conformation is achieved. Four SA specific parameters appear in the DPF in Figure 2.3 (not used in our calculation). The first three are the commands ‘tstep’, ‘qstep’ and ‘dstep’ which define the maximum step size that can be taken by a ligand in terms of translation,

rotational and torsion in the first SA cycle. The parameter ‘e0max’, also specific to SA, is used to define acceptable values of the ligand’s initial energy; ligands with higher energies are discarded. It also defines the maximum number of retries.

The third search algorithm is the Lamarckian Genetic Algorithm (LGA). It is similar to GA but has an additional local search step and is therefore sometimes called the hybrid Genetic Algorithm Local Search (GALS) method. This search method gets its name from the Lamarckian evolution theory proposed by Jean Baptiste de Lamarck. Although discredited as a viable evolution theory, the idea proposed was that an individual’s adaptive traits could be passed onto its offspring during its lifetime. This search method has shown enhanced performance over the other search methods, GA and SA (49), and was therefore used for this research.

The algorithm, parts and steps, are typically described using the language of genetics. The following is a simplified list of terms, defined in the genetic context, to aid the understanding of LGA.

Gene: An ordered sequence of nucleotides in a particular position of a chromosome that determines the individual’s properties (Cartesian, orientation and torsion values).

Genotype: The specific set of genes that make up an individual (State variables).

Chromosome: A genetic structure that contains an individual’s genes.

Phenotype: The observed physical characteristics of an individual (Atomic coordinates).

Reproduction: A biological process by which parents create new offspring.

Fitness: A property describing the likelihood of an individual in survival and reproduction (energy of the docked conformation).

Crossover: A process in which chromosomes exchange genetic material.

Mutation: A process leading to a change in the nucleotide sequence of a gene.

Allele: A variant of a specific gene in a chromosome (e.g. coordinates of a conformation).

LGA involves the basic steps of GA search method and a local search then follows. At the start of this algorithm a random population of individuals (different docked poses) having different genes is created. The size of the population is defined by the command 'ga_pop_size'. To generate these genes at random, AutoDock requires two integers that act as seeds to generate a series of random numbers. The seed variables chosen were 'pid' and 'time', set by the command 'seed', and they refer to the process ID and the current time during calculations, respectively. For the population generated, each individual's chromosome is made of three main parts: a gene for the three Cartesian coordinates of the ligand's translation, a gene specifying the ligand's orientation described by the quaternion number system and a gene describing the conformation of the ligand where the ligand's torsions are described. These values are with respect to the protein and together constitute the genotype of the individual, also called the state variables of a ligand. As with GA, the fitness of the randomly generated individuals is evaluated i.e. the energies of the proposed docked poses are calculated. Fitness of an individual is judged from its phenotype; in this case, the energy of the proposed poses is evaluated from the atomic coordinates of the ligand. It is therefore necessary to transform the individual's genotype to its phenotype. In other words, the ligand's state variables should be transformed to its atomic coordinates; this is done by a process called 'mapping'. More fit individuals, that is, ligand poses with better than average energies, are selected such that they are favored over others for reproduction. The average energy

is calculated over a number of previous generations, specified by the command ‘ga_window_size’, in this case the last 10 generations.

Chosen individuals mate through the genetic operator of ‘crossover’. A two-point crossover is made within a chromosome, breaking it into its individual three genes. New hybrid offspring are created replacing their parents; thus, the population size does not increase from one generation to the other. The frequency of crossover in a population is set by the command ‘ga_crossover_rate’. Another genetic operator performed on genes is mutation. In this case, mutation of a state variable (gene) is performed by the addition of a random real number. The random numbers added have a Cauchy distribution (a continuous probability distribution function). Cauchy distributions have two main variables alpha (α) and beta (β), which roughly denote the mean and variance of the distribution. The two variables α and β are specified by the commands ‘ga_cauchy_alpha’ and ‘ga_cauchy_beta’ respectively. Cauchy distributions are favored since they have a tendency towards smaller deviates, yet they can generate larger ones occasionally. Note that the genetic operator of mutation is especially important, since it can regenerate alleles that could be lost in the fitness/selection step.

Next, after mutation and crossover, is the local search step. AutoDock uses a variation on the local search (LS) optimization proposed by Solis and Wets (50) called the pseudo Solis and Wets method. In a typical Solis and Wets optimization, the dimensional variances for both translation and rotation are the same. This would not be very practical for docking, since a 1˚ rotation might not be as significant as a 1 Å translation. Therefore, a pseudo Solis and Wets method is used, where different dimensional variances are used in AutoDock; they have been set to 0.2 Å for translations

and 5˚ for orientation and torsional rotations respectively. By default, the local search is performed on 6% of the population, set by the command 'ls_search_freq'. It has been shown that increasing this frequency does not significantly enhance results. The command 'sw_rho' specifies the value of rho. Rho (ρ) is a parameter defined as the size of the local space to sample and the initial variance; it is set to one by default. After four consecutive successes or failures, the value of ρ is doubled or halved respectively. The numbers of these successive successes or failures are set by the commands 'sw_max_succ' and 'sw_max_fail' respectively. The local search was set to stop when the number of iterations had reached 300 or when the value of ρ had reached 0.01, whichever happened first. The commands specifying the number of iterations and the lower bound on ρ were 'sw_max_its' and 'sw_lb_rho' respectively. All the proposed poses (individuals) are ranked according to their energies (fitness) and only the most elite survive to the next generation. The command to specify the number of surviving elite individuals is 'ga_elitism'. LGA is iterated, until either the maximum number of energy evaluations or the maximum number of generations is reached, whichever comes first. These two parameters are specified by the commands 'ga_num_evals' and 'ga_num_generations'. Note that AutoDock uses LGA when the commands 'set_ga' (for Genetic Algorithm) and 'set_psw1' (Pseudo Solis and Wets) are specified. To get a statistically sound representation of the possible binding poses this procedure was set to run 100 times; specified by the command 'ga_run'.

As described so far, the output from AutoGrid as well as the prepared ligand and protein in the docking simulation have been used in AutoDock. The last essential component of docking calculations is the energy evaluation method, used in LGA to

evaluate the fitness of individuals. Autodock 4.2 uses a semi-empirical force field for energy calculations (51). It is called semi-empirical since it uses molecular mechanics principles for calculating the different interactions and multiplies them by empirical weights. A set of well-known experimental binding constants has been used to set weighting factors that would calibrate calculated energies to that obtained from experimental data. In broad terms, the free energy of binding (ΔG) is estimated to be the difference between the energy of the protein-ligand complex and the combined energies of the unbound ligand and unbound protein. This calculation is broken down into the following form:

$$\Delta G = (V_{bound}^{L-L} - V_{unbound}^{L-L}) + (V_{bound}^{P-P} - V_{unbound}^{P-P}) + \left(V_{bound}^{P-L} - V_{unbound}^{P-L} + \Delta S_{conf}\right)$$

Equation 2.1: Equation used by Autodock 4.2 to calculate the binding free energy of the docked conformations. V=pair-wise energy evaluations, L= ligand, P= protein

To put the above equation in words, the binding free energy is estimated by calculating the intramolecular interaction energy cost of transforming the conformations of the ligand and protein from their unbound to bound conformations individually. This is added to the intermolecular penalty of forming the ligand-protein complex. The penalty for conformational entropy loss, $\Delta S_{conf,}$ is also accounted for. This entropy term has been calculated from parameters such that it is in energy units (kcal/mol). Note that the energy evaluation (V) for the interaction between the unbound protein and ligand, second term in the third part of the above equation, is zero since they are distant from each other in their unbound form. Also, note that the intramolecular energies calculated in the first two parts of Equation 2.1 are calculated over most pairs of atoms, excluding those that are connected covalently, or are separated by at most two contiguous covalent bonds.

As described earlier, all interactions calculated by the force field used in Autodock 4.2 are multiplied by a weighting factor. To start with, the simpler energy component accounting for the conformational entropy loss is dependent on the total number of torsional degrees of freedom. It is calculated by Equation 2.2 below:

$$\Delta S_{conf} = W_{conf} N_{tors}$$

Equation 2.2: Equation for calculation of the entropic loss upon binding due to loss of torsional freedom. W_{conf} is the weighting factor calibrated from empirical data and N_{tors} is the number of torsional degrees of freedom.

The pair-wise energy evaluations (V) are composed of four main calculations that account for the van der Waals interactions (vdW), electrostatic interactions, hydrogen bonding and desolvation effects. The first of these interactions, vdW, is calculated using the 12-6 Lennard-Jones potential. Electrostatic interaction calculations are based on Coulomb's law, and Solmajer and Mehler method (52) where the dielectric constant used in the calculation is distance dependent. The third term, the hydrogen bonding term, is a 12-10 potential with Goodford hydrogen bonding directionality (53). Hydrogen bonding is sometimes difficult to model, especially when possible hydrogen bonds are close in distance. A hydrogen atom would sometimes form a bifurcated hydrogen bond with two sites, instead of two hydrogen atoms forming two parallel hydrogen bonds. The use of hydrogen bond directionality, however, improves the performance of Autodock 4.2. It limits the number of hydrogen bonds available at a grid point to the number of hydrogen bonds that could be formed from a chemical and sterical perspective, even if more bonds (e.g. bifurcated hydrogen bond) would be energetically more favorable. A term to account for desolvation effects is also included. It is a variation of the Stouten method

(54) in which the solvation free energy is calculated per unit volume. These different types of interactions together can be represented by Equation 2.3:

$$V = V_{vdW} + V_{elec} + V_{h-bond} + V_{desol}$$

Equation 2.3: General equation for calculating the energies of molecules from van der Waals, electrostatic interactions, hydrogen bonding and desolvation effects.

Autodock 4.2 also calculates the dissociation constant (K_D) of the ligand as part of its output. This constant represents the concentration of the ligand at which half the binding sites in the receptor are occupied. It is calculated from the free energy of binding (ΔG) using Equation 2.4 below:

$$K_D = e^{\frac{\Delta G}{RT}}$$

Equation 2.4: The equation used for the calculation of K_D from the free energy of binding of the ligand. R= Universal gas constant (kcal K^{-1} mol^{-1}), T= temperature (K). Note that the usual (-) sign in the exponent is already incorporated in this algorithm.

As mentioned earlier in this subsection, ADT aids the visualisation of the docking calculation results generated by Autodock 4.2, by the creation of a 3D rotatable computer graphic of the protein and ligand. In addition, it also shows the interactions between the ligand and the protein, through the display of meshes around atoms that are in close contact. To elaborate, when a pair of atoms is closer than the sum of their vdW radii, it displays spherical meshes around each of them; the meshes represent the vdW radius for each atom. The atomic vdW radii for different atom types are defined in the Autodock parameter file.

2.1.2 NAMD and VMD

As explained earlier in section 1.3, relaxation of the protein could possibly provide a more accurate representation of the protein structure. This was achieved using Molecular Dynamics (MD), a computational method designed to simulate conformational changes in a molecule. The software Not-just-Another-Molecular-Dynamics (NAMD) program (38, 39) is a CHARM++ parallel programming code and it was used to allow the equilibration of the protein. This code solves Newton's equations of motion using the CHARMM force field to calculate the path of atoms in space as a function of time: atomic trajectories. The implementation of NAMD was integrated with a GUI called Visual Molecular Dynamics (VMD) (55, 56), a molecular graphics software suite that displays molecules and animates MD trajectories, including those generated by NAMD. It is used to set up the desired protein environment and creates the required input files to run NAMD. Definitions for the terms used in this subsection as well as the exact calculation steps can be found in the NAMD and VMD tutorial manuals (57, 58). Systems being modeled in both VMD and NAMD are specified by two types of files: Protein Data bank (PDB) file and Protein Structure File (PSF). The former is a file listing all the atoms in the system and their coordinates; the latter specifies the structure of the protein, in other words, it lists the atom bonds, angles, dihedrals, impropers (dihedral force terms used to maintain planarity) and cross-terms.

To start with, the PDB file of the protein was obtained from the PDB website (46). The PSF of this file is generated by VMD using a plugin called Automatic PSF builder. Note that this plugin not only builds the protein structure, but it fixes the protein

by addition of the missing hydrogen atoms in the PDB file. The Automatic PSF builder requires a topology parameter file. The CHARMM force field file, top_all27_prot_lipid.inp, was used to define the atom connectivity in the PDB file. Atom and residue names are aliased to the topology file. This is followed by assigning default histidine protonation. Since some residues may be missing hydrogen atoms, residue definitions from the topology file were used to guess them. The output is the PSF containing structural information about the protein and a PDB file with all atomic coordinates including the guessed hydrogen atoms.

To make the simulation more realistic, the system was also solvated and neutralised. These manipulations were performed using two features of VMD, namely: Solvate and Autoionize. The Solvate plugin was utilised to put CB-D_{28k} in a water box. This box had a water layer of 15 Å in each dimension, from the atom furthest from the origin in that dimension (about 16000 water molecules). Once the dimensions of the box were calculated, the PSF of the solvated protein was generated via the water model called Transferable Intermolecular Potential 3 Point (TIP3P) water, which is the standard solute model in the Solvate package of VMD. Neutralisation of the system was performed through the Autoionize plugin. This plugin adds sodium chloride ions to the solvated system until the net charge becomes zero. Although ion addition is random, there is a minimum to the ion-to-ion/protein distance. The outputs from this plugin are the PSF and PDB file of the solvated and neutralised system, two required important elements for NAMD calculations. A CHARMM force field parameter file is necessary for providing the energy functions and numerical values of atomic parameters. It also calculates the

forces and energies in the system. The last requirement for starting a MD calculation in NAMD is a configuration file which lists the specifications of the simulation.

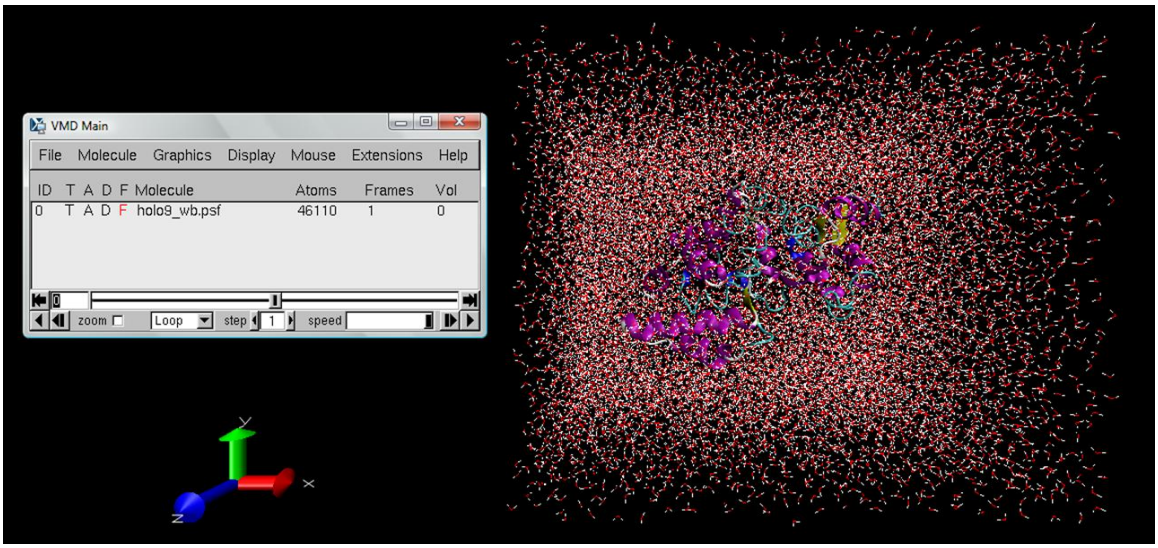


Figure 2.4: An image of CB-D_{28k} in a neutralized water box prepared using VMD.

Along with the standard conditions for running MD calculations in NAMD, other simulation parameters are also set in the configuration file. As described earlier, the subject of the simulation is a solvated neutralised system (Figure 2.4) with guessed hydrogen atom positions. Adjustment of this system is therefore necessary since the hydrogen atoms might not be at their optimum positions, or the water molecules added might not be in an equilibrated state; their network of hydrogen bonds would not be well formed. It is therefore necessary to run a few minimisation steps so that the system would not be starting from an unstable structure, which could produce artifacts in the simulation. Equilibration calculations of the protein would start on the minimised system. The configuration file included other simulation specifications, including temperature, Periodic Boundary Conditions, Langevin dynamics, as well as the Particle Mesh Ewald (described later).

The temperature of the system was set to be that of body temperature; 310 Kelvin. Random velocities are assigned according to the Boltzmann distribution for the initial velocities of all the atoms in the system, to simulate the desired temperature. If the simulation was only of a protein molecule in a box of water suspended in vacuum the simulation would not be realistic and surface interactions would arise. Therefore, the system is set to run using Periodic Boundary Conditions (PBC); under these conditions, the system would be surrounded by identical virtual cells that would interact with the real system. To make the calculation more realistic, simulation of the solvent's viscosity is important. Simply put, Langevin dynamics simulates a frictional force in the system that would account for solvent jostling. However, since the temperature of the system could increase due to friction or other effects, Langevin dynamics include a parameter that would dampen the energy of the system thus acting like a thermostat. NAMD force field parameters include a cut-off for calculating the electrostatic energy. This, however, is not accurate, especially since long range interactions would be neglected. A fast method for calculating long range electrostatic effects is therefore employed, called the Particle Mesh Ewald (PME) method. This method is particularly used when PBC are used.

The simulation is run for the desired duration of time; in this project, simulations were 50 ns long (See 2.2). These durations provided 25,000,000 frames or protein structures. The desired conformation of the protein with the hypothetically better structure was sought for in the last 2500 frames of the simulation, as the system would have probably equilibrated by then.

2.1.3 PTRAJ

The MD calculations run on NAMD for 50 ns produced 25,000,000 frames and it would be impractical to deal with all of them. It is therefore necessary to reduce these frames to a few most characteristic conformations.

Part of Ambertools11, PTRAJ (40) is a utility designed to process and analyse different MD simulation-generated files including coordinate and CHARMM DCD trajectory files. The main options used were the RMS and Cluster routines. The root-mean-squared coordinate deviation (RMSD) of the alpha carbon atoms in the protein over time from the first structure of the protein was chosen from among the RMS options, as being the best indication that equilibrium had been achieved. When this value becomes steady, the system is thought to have equilibrated.

An equilibrated system would typically alternate among a few main conformations. To get representative structures of the equilibrated protein, the last 2500 frames of the MD simulation were clustered using the Cluster option, an average-linkage algorithm (59). In this type of "bottom-up" clustering, each point is first assigned to its own cluster. Distances between each point in a cluster are calculated, as well as the distances between the different points in different clusters. The two clusters with the lowest average distances would then be merged to form a new larger cluster. This process is iterated until the required number of clusters is reached. Atoms used in clustering comparison for the holo form were the loop residues (residues 24-35, 66-77, 111-122, 155-166, 199-210, 240-251), while all atoms were used for comparison in the case of the

apo form. Representative conformations to be used in further docking calculations were chosen from those closest to the centroid of the cluster.

The number of clusters has to be predetermined by the user. This makes clustering a very subjective matter; however, some clustering metrics can be used to indicate the best number of clusters, even though they do not always provide a unanimous number. The metrics used were the Davies-Bouldin index (DBI), the SSR/SST ratio and the pseudo F-statistic (pSF). DBI is a measure of the well separation and compactness of clusters (59). Effectively, it is the average (over all clusters) of the maximal values of the pairwise sum of average distances of each point in the cluster from its centroid, divided by the intercluster separation. Lower DBI values are therefore an indication of better clustering. The second metric, SSR/SST is a ratio expressing the percentage variance, where SSR= the sum of squares regression and SST= the total sum of squares. The pSF metric is a comparison of intracluster and residual variances; calculated using Equation 2.5. Generally, higher pSF values are an indication of a better clustering.

$$pSF = \frac{SSR/g - 1}{SSE/(n - g)}$$

Equation 2.5: Equation for the calculation of pSF. SSR= sum of squares regression, SSR= sum of squares regression, SSE= sum of squares error or residual variation, n= total number of points, g= number of clusters

A good number of clusters is suggested to be one that has a local peak maximum in the pSF value, a plateau in the SSR/SST ratio and a local minimum in the DBI value. After deciding on the optimum number of clusters, the representative structure of each (structure in the centroid of the cluster) was used as a conformation of the equilibrated protein, thus accounting for the protein's flexibility.

2.2 Work Flow

Although well-calibrated functions and parameters are extremely important for molecular modeling, the structure of the system to be modeled is the foundation of any calculation. The structure of CB-D_{28k} chosen was the first model in the 2G9B PDB entry in the Protein Data Bank website. Structures representing the system are always the starting element of calculations; hence, an accurate representation is required, otherwise all calculations built on it would be unlikely to produce meaningful results. Of course, this presented a challenge in the calculations, since there were several models proposed for the structure of holo CB-D_{28k} in the 2G9B PDB file and they were missing the bound Ca^{2+}. With regard to the second goal of the project, in which the interference between Ca^{2+} and Zn^{2+} binding to CB-D_{28k} was being tested, there was no was accurate structure for the protein. In fact, there is no available structure of the Zn^{2+}-bound form of CB-D_{28k}. Therefore, it was necessary to make a plan, using the different tools previously described, that could possibly provide better starting points for the simulations and thus increase the probability that the goal could be achieved.

Since the first goal of the project required a good representative structure of holo CB-D_{28k}, it was necessary to introduce Ca^{2+} into the structure of the protein obtained from the PDB. Autodock 4.2 was first used to predict the possible positions of Ca^{2+} in the EF-hands of CB-D_{28k}. The amino acids in the loop portion of EF-hands 1, 3, 4 and 5 were set as flexible; these were residues 24 to 35, 111 to 122, 155 to 166 and 199 to 210. LGA parameters were set to 30,000,000 energy evaluations, 28,000 generations each with 300 individuals. A docked conformation was chosen as the best candidate for the MD

simulation. VMD was used to prepare and set the neutralised and solvated protein complex in a NAMD compatible form. The protein was first minimised for 500 steps then allowed to relax for 50 ns using MD calculations. Durations of MD simulations generally depend on the process being simulated as well as the size of the system; an MD calculation on the protein calmodulin (~16 kDa) was run for 40 ns (60) while another simulation which involved the folding of rhodanese (~33 kDa) was carried out for an equivalent of 200 ns (61). Considering the size of CB-D_{28k} (28 kDa) and the nature of the simulation, it was decided that the MD calculations in this project were to be run for 50 ns, as an intermediate, best compromise value.

To test whether the system had equilibrated, RMSD was calculated using PTRAJ. In addition, the last 2500 frames of the simulation were clustered into 2 to 10 clusters using the average linkage algorithm. The clustering metrics were calculated to help assess the best number of clusters. The representative structure of each bin was used in further calculations; if a good choice was made on the number of clusters to use, their representative structures would depict the main conformations visited by the protein during the last part of the simulation, thus accounting for protein flexibility. Calcium ions were removed from these models and docking calculations were run again on these equilibrated protein structures. A schematic of this work plan is illustrated in Figure 2.5 below.

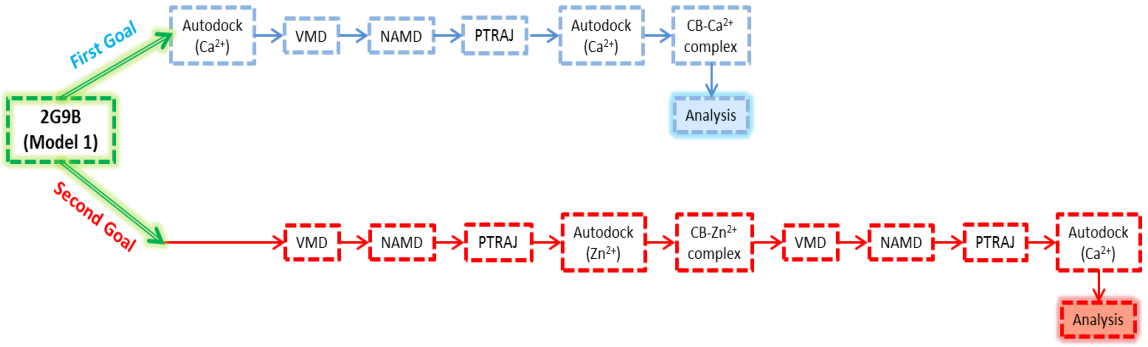


Figure 2.5: A schematic of the workflow designed to achieve the goals of the project. The steps carried out for the first goal are illustrated in blue. The steps carried out to accomplish second goal are shown in red.

The second goal of the project was to test for the possible interference between Ca^{2+} and Zn^{2+} binding to CB-D_{28k}. Since the Zn^{2+} bound form of CB-D_{28k} is unknown, a sequence of calculations was used to try to get a good hypothetical representative structure of the Zn^{2+}-bound form of the protein. Since the available CB-D_{28k} structure represents the holo form of protein, it was solvated and neutralised in VMD to be relaxed using a MD simulation in NAMD. Since the protein did not contain Ca^{2+}, the intent of the calculation was to permit the relaxation of the protein so that the structure was less like the holo form. The implications of this step are addressed in 5.2.1.

Clustering and RMSD calculations were performed using PTRAJ and the structures representing the different conformations of the protein were selected. Autodock 4.2 was used to predict the possible binding sites of Zn^{2+} to CB-D_{28k}. As explained in the last chapter, Zn^{2+} binds to CB-D_{28k} at three sites (18). Results from this study were not conclusive with regard to the Zn^{2+} binding sites, but residue 80 showed the most promise of being involved in Zn^{2+} binding. Other binding residues in CB-D_{28k} to Zn^{2+} are unknown which added to the challenge of getting a reasonable CB-Zn^{2+} complex. In proteins, zinc ions tend to be bound by cysteine, histidine, glutamic and

aspartic acid (62). Since experimental results have shown that cysteine was not significant for Zn^{2+} binding in CB-D_{28k} (18), it could be excluded from our search. Visual inspection of the protein in the region around residue 80, which lies within EF-hand 2, revealed two glutamic acid residues (residues number 57 and 77 on helices E and F respectively), which were within close proximity. The side-chains of these residues were therefore treated as flexible and the grid box was set to cover them. In addition, the side-chains of residues 78 and 81 were also treated as flexible. This was intended to allow the residues of interest (57, 77 and 80) to move more freely since rigidifying the residues surrounding them would limit their possible conformational space. Note that residue 79 is the amino-acid alanine. Since Autodock uses the united atom method, no flexibility can be set to the terminal methyl group of Ala79. The search algorithm, LGA, was set to perform 50,000,000 energy evaluations on 28,000 generations with 300 individuals each. After predicting the possible Zn^{2+} binding positions in CB-D_{28k} at EF-hand 2, MD calculations were run for an equivalent of 50 ns. Execution of the usual PTRAJ calculations, to find the best number of clusters that would best represent the last 2500 frames, was difficult since several models were subjected to NAMD calculations. To keep the number of structures manageable for the next calculation, the last 2500 frames were binned into one cluster and the centroid structure of that cluster was carried forward to the docking calculations. Structures obtained from this sequence of calculations were unlikely to be that of the Zn^{2+}-bound form of CB-D_{28k}; especially since only one zinc ion is bound to the protein. However, testing of the possible interference between Zn^{2+} and Ca^{2+} binding at EF-hand 1 was possible. Note that testing was only performed for Ca^{2+}

binding to EF-hand 1 since this is the closest site to the bound Zn^{2+} at residue 80. Results from these calculations are shown in section 5.3.1.

2.3 Calculations on WestGrid

The molecular modeling calculations in this project required long computer processing times. The resources at the High Performance Computing (HPC) consortium, WestGrid, were used to run these calculations (63). WestGrid is part of Compute/Calcul Canada, a national HPC organisation.

WestGrid involves several systems, each of which has its own operation specifications and software. The system used for running VMD, NAMD and PTRAJ was 'Lattice'; this system is designed for large (gigabyte range) scale parallel calculations. MD calculations in this project were run on 8 nodes, each of which had 8 processors. The physical time (wall time), required for the 50 ns simulations ranged from about 126 to 165 hours. Docking calculations were run on another system in WestGrid called 'Glacier'. This system is mainly intended for use by serial, small size (megabyte range) calculations. The physical time taken for docking calculations mainly depended on the number of torsions of the assigned flexible residues; these times ranged from 90 to 155 hours.

Chapter 3

Validations

One of the main steps in this project was the introduction of Ca^{2+} into CB-D_{28k}, via the docking suite Autodock 4.2 (as outlined in Section 2.1.1). The preliminary docking calculations on CB-D_{28k}, however, did not yield satisfactory results. Although this could be related to protein deformation at some sites, it was still deemed necessary to confirm the validity of the docking calculations as well as assess their performance. To do this, Autodock 4.2 was used to predict the binding conformation of the protein calbindin D_{9k} (CB-D_{9k}), an EF-hand protein that binds Ca^{2+} as a ligand, for which the complex-bound structure has been resolved and the bound ligands are in place. If the docking calculations were successful, AutoDock calculations should reproduce the experimental results. In this chapter, the calculation procedures as well as the docking results are described.

3.1 Calculations on Calbindin D_{9k}

CB-D_{9k} is the smallest monomeric EF-hand protein, and contains one pair of EF-hands (30) . In addition to being a calcium binding protein (CBP), it has a canonical and a non-

canonical loop; EF-hand 1 has a loop made up of 14 residues instead of the typical 12-residue canonical loop (28, 30) . This makes CB-D_{9k} a good candidate for the validation step, since successful results on this protein would verify the validity of docking calculations for both canonical and non-canonical type loops, as in the cases of EF-hands 1, 3, 4 and 5 and EF-hand 6 of CB-D_{28k}, respectively. In addition, there is a published structure for CB-D_{9k} (64) with Ca^{2+} bound to the loops. Thus, the calcium ions can be removed from the CB-D_{9k} complex then docked back. It is generally accepted that a successful docking result is one in which the predicted ligand position has a maximum RMSD of 2.0 Å (41, 65, 66) to 2.5 Å (51) from that in experiment. Since Ca^{2+} is a relatively small ligand, an RMSD value of 2.0 Å will be used.

The PDB file of CB-D_{9k} (64), downloaded from the PDB website (46), is the MD refined NMR structure of the protein. The 1b1g file of the protein has ten models; the first model was the one used for validation. The two calcium ions were removed from the protein complex and two separate PDBQT files were created using AutoDockTools for the protein and Ca^{2+}. Since the loop residues in CB-D_{28k} calculations were set to be flexible, loop residues in the validation calculations on CB-D_{9k} were also made to be flexible. To run AutoGrid calculations, the grid box was set to cover the loops, that is, Ca^{2+} binding sites, in CB-D_{9k}. The docking parameters were set to perform a maximum of 50,000,000 evaluations and create a maximum of 28,000 generations, each generation having 300 individuals.

3.2 Validation results

The following subsections show results for EF-hand 1 and 2 of CB-D_{9k}. Each docking calculation on an EF-hand was assigned to run 100 times. This would increase the probability of successful docking as well as provide statistical significance to the results. Docking calculations in AutoDock provide a predicted location of the ligand and flexible side-chains, in the case of flexible protein docking. The scoring function in AutoDock also provides a calculated binding energy, which it converts into the dissociation constant. AutoDockTools also lists the residues in the protein that are in close contact to the ligand. These three factors will all be described for each docking calculation below.

3.2.1 EF-hand 1 results

The 100 runs in AutoDock provided different predictions for where Ca^{2+} could bind to the loop of EF-hand 1 (region covered by the grid). These are all presented in Figure 3.1, the red sphere being the experimentally-determined position of Ca^{2+} while the green spheres represent the docking calculations results. Using a clustering RMSD of 1.0 Å, 10% of the conformations were less than 1.0 Å away from the original PDB calcium ion position, which represents the experimentally-determined position of Ca^{2+}. The docking results were also re-clustered using an RMSD value of 2.0 Å shown in Figure 3.2.

Concerning the binding energies, Autodock does not perform very well in energy calculations, especially the electrostatic aspect of the binding energies, discussed in 3.3. The conformations in the clusters of interest (the red bars in Figure 3.2) exhibit a range of binding energies. For the cluster obtained on an RMSD value of 1.0 Å, energies range

from -2.12 kcal mol^{-1} to -1.13 kcal mol^{-1}, corresponding to dissociation constants of 27.84 mM to 149.36 mM respectively. When an RMSD value of 2.0 Å was used, the range of energies was -5.95 kcal mol^{-1} to -0.04 kcal mol^{-1}, which is equivalent to dissociation constants of 43.81 μM to 940.29 mM. Note that the bars in clustering histograms are plotted at the x-axis value of the least binding energy pose in the cluster, while the y-axis represents the number of conformations in that cluster.

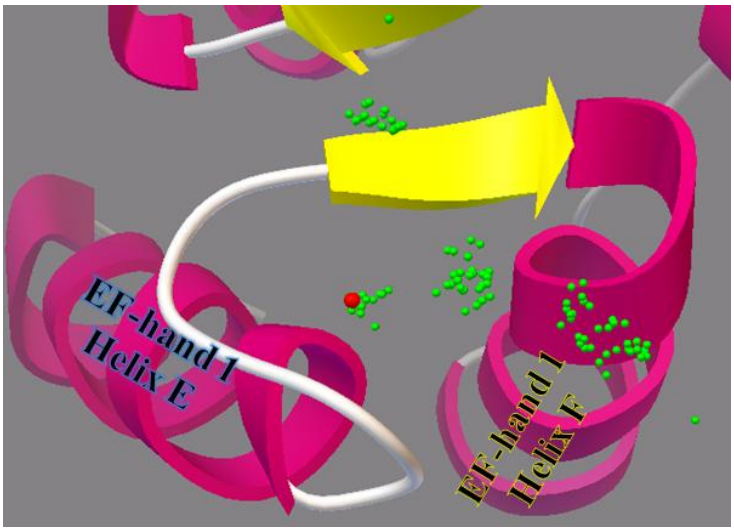


Figure 3.1: A picture of the predicted Ca^{2+} positions in the loop of the non-canonical EF-hand 1 of CB-D_{9k}. The red sphere represents the experimentally determined position of Ca^{2+}, while the spheres in green represent the centers of the 100 predicted Ca^{2+} binding positions to the loop residues.

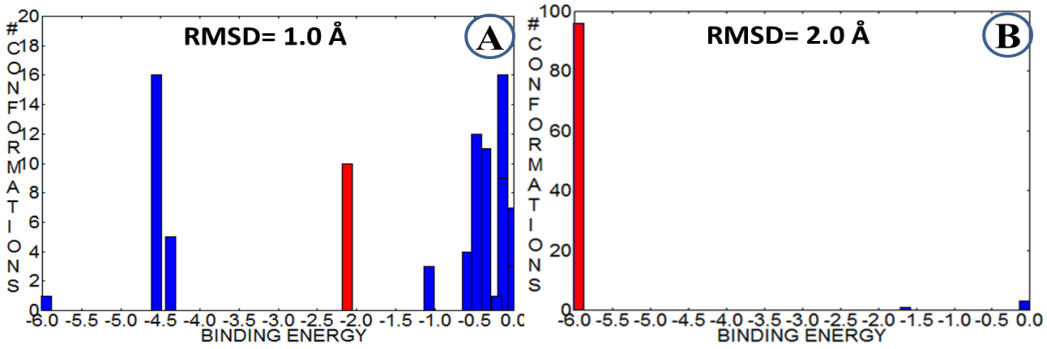


Figure 3.2: A figure showing two histograms of the clustering bins for RMSD values of 1.0 Å (A) and 2.0 Å (B). Note that the bins containing the 10 conformations that were close to the experimentally determined Ca^{2+} position are colored in red.

As mentioned earlier, AutoDockTools creates a 3D rotatable computer graphic to facilitate visualisation of the docked conformation, as well as a list of the residues that are in close contact with the ligand. The calculated conformation that was the closest to the experimentally-determined position of Ca^{2+} in the loop of EF-hand 1 is 0.280 Å away from the original Ca^{2+} position in the 1b1g PDB structure. At this position, it is in close contact with the backbone carbonyl of Ala14, an oxygen atom of Glu17 carboxyl group, the backbone carbonyl of Asp 19, both the backbone carbonyl of Gln22 and its side chain amide group and finally the carboxyl group of Glu 27, as shown in Figure 3.3 below. At this configuration, the binding energy of Ca^{2+} is -1.55 kcal mol^{-1} with a contribution of -2.86 kcal mol^{-1} from electrostatic energy and +1.01 kcal mol^{-1} from van der Waals, hydrogen bonding and desolvation effects. Loss of torsional free energy in Ca^{2+} also had a penalty of 0.3 kcal mol^{-1}.

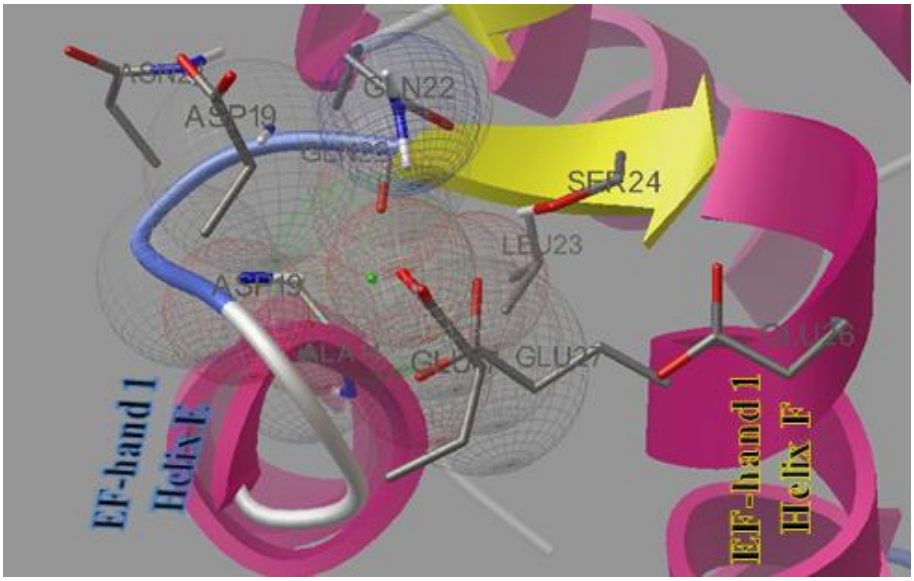


Figure 3.3: An image of the best-predicted position of Ca^{2+} (green sphere) in EF-hand 1 in CB-D_{9k}. Spherical meshes represent the interaction ranges of their respective atoms. ADT displays residues that have interaction meshes that overlap with interaction sphere of Ca^{2+}. In this case, Ca^{2+} interacts with residues Ala14, Glu17, Asp19, Gln22 and Glu27.

3.2.2 EF-hand 2 results

A similar analysis was performed on the 100 results obtained from docking calculations done for EF-hand 2 in CB-D_{9k}. Figure 3.4 shows the 100 predicted positions of Ca^{2+} at the loop of EF-hand 2. The red and green spheres represent the experimentally-determined Ca^{2+} position and the AutoDock calculated Ca^{2+} positions respectively. Figure 3.5 shows the clustering results when RMSD values 1.0 Å and 2.0 Å were used. AutoDock was successful in predicting 16 conformations that were less than 2.0 Å away from the experimentally-determined position of Ca^{2+}. These were in the first cluster (red bar in Figure 3.5) when clustering at an RMSD value of 1.0 Å. The 16 conformations that are within successful result range have a binding energy range from -1.91 kcal mol^{-1} to -1.39 kcal mol^{-1}, which are equivalent to dissociation constants from 40.09 mM to 95.66 mM. On a 2.0 Å clustering, the bin that contains these 16 successful conformations has an energy range from -1.91 kcal mol^{-1} to 0.77 kcal mol^{-1}; a positive value means that it is an unfavourable binding position.

The conformation that was the closest to the calculated position of Ca^{2+} was 0.459 Å away from the actual position of Ca^{2+} in the CB-D_{9k} PDB entry. At this position, Ca^{2+} was in close contact to the carboxyl group of Asp54, the backbone carbonyl of Gly57, the carbon of the carboxyl group of Glu60, the side-chain carbon of Val61 and one of the oxygen atoms of Glu65 carboxyl group, as shown in Figure 3.6. At this position, the calcium ion had a binding energy of -1.73 kcal mol^{-1}; the breakdown of this energy was a contribution of electrostatic energy equal to -2.83 kcal mol^{-1} and 0.3 kcal mol^{-1} for the

loss of torsional free energy. Van der Waals, hydrogen bonding and desolvation effects contributed to an energy of +0.80 kcal mol^{-1}.

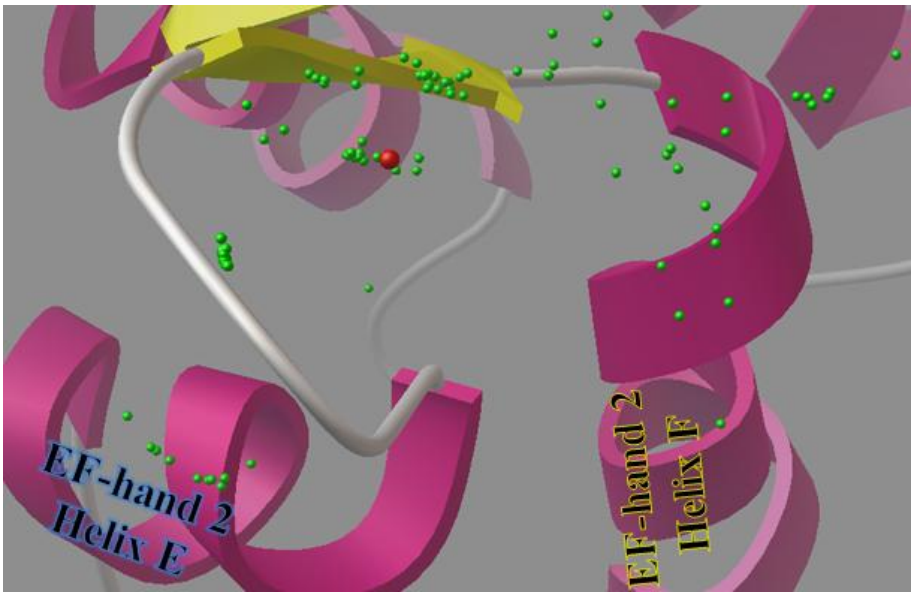


Figure 3.4: A picture of the predicted Ca^{2+} positions in the loop of the canonical EF-hand 2 of CB-D_{9k}. The red sphere represents the experimentally determined position of Ca^{2+}, while the spheres in green represent the centers of the 100 predicted Ca^{2+} binding positions to the loop residues.

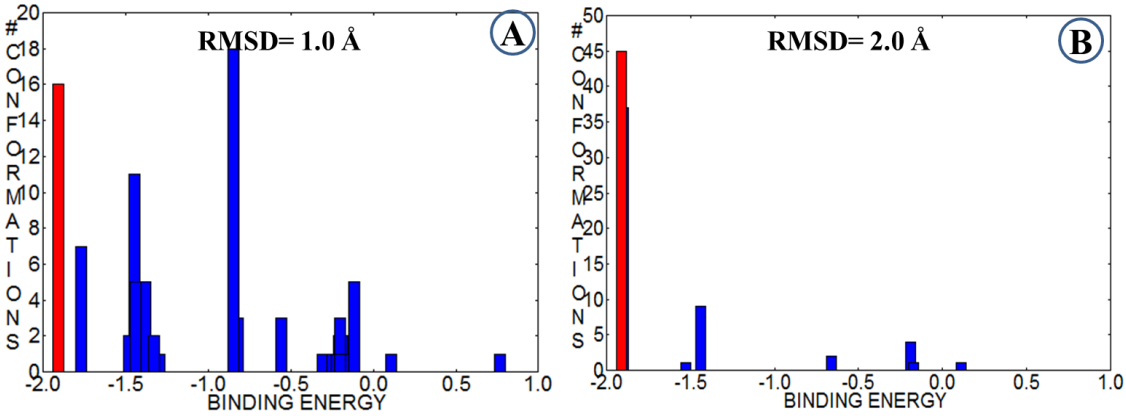


Figure 3.5: A figure showing two histograms of the clustering bins for RMSD values of 1.0 Å (A) and 2.0 Å (B). The x-axis represents the binding energy of the first pose in the cluster, while the y-axis represents the number of conformations in that cluster. Note that the bins containing the 16 conformations that were close to the experimentally determined Ca^{2+} position are colored in red.

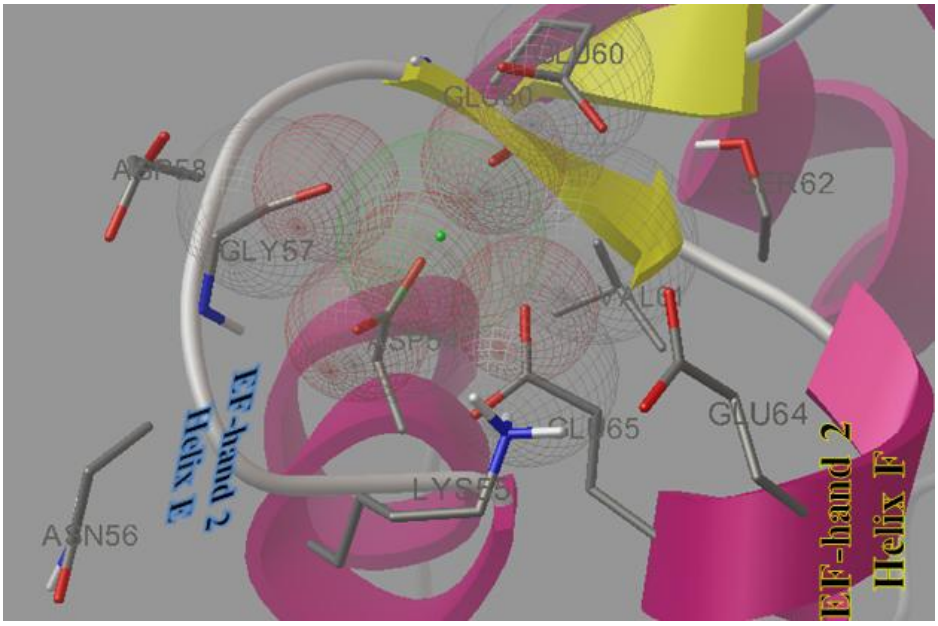


Figure 3.6: The best predicted position of Ca^{2+} (green sphere) in EF-hand 2 of CB-D_{9k}. The interaction spherical mesh of Ca^{2+} overlaps with that of residues Asp54, Gly57, Glu60, Val61 and Glu 65.

3.3 Discussion of Validation Results

Two main conclusions could be made about the energies and clustering values from the docking results described above. The scoring function of Autodock 4 (described in subsection 2.1.1) was developed using a database of compounds that mostly had electrostatic interactions that involved hydrogen bonding (51). Therefore, a big part of the electrostatic contributions is accounted for in the hydrogen-bonding coefficient in the scoring function. In addition, there is a bias in the scoring function towards molecules with higher molecular weights (67). Thus, smaller ligands, such as the calcium ions here, are prone to poor energy results. In addition, Autodock usually calculates energies with standard error of ±2.35 kcal mol^{-1} (51). In the case of the best-docked conformation for EF-hand 1, the range of error, ±2.35 kcal mol^{-1} is more than three and a half times the

calculated binding energy of this conformation, 1.55 kcal mol^{-1}. In the case of the second EF-hand, which has a binding free energy of 1.73 kcal mol^{-1}, the error is 136% of the calculated energy. Thus, it can be seen that the relative error in the results above is very high. This relatively high error can be related back to the dataset that was used to calibrate the scoring function of AutoDock. In addition, the dissociation constants in AutoDock are calculated from the binding energies and since the error in binding energy is very high, the values of the constants are unreliable too. This is also reflected by their very large values of 73.08 mM and 40.09 mM for EF-hands 1 and 2 respectively; CBP usually have dissociation constants in the low micromolar to nanomolar range (68).

Clustering was first performed at an RMSD value of 1.0 Å, since cations within this limit usually tend to interact with the same residues. At this value, AutoDock was successful at binning together the Ca^{2+} locations that fall within the standard acceptable limits as previously demonstrated in Figure 3.2 and Figure 3.5. On using an RMSD value of 2.0 Å, however, the clusters became bigger and contained other invalid conformations. Note that it was concluded that an RMSD value of 2.0 Å includes invalid conformations because the exact locations of Ca^{2+} in CB-D_{9k} were already known. However, this is not the case for CB-D_{28k} and therefore clustering at 2.0 Å was still used. This use is supported by the fact that a deviation of a maximum of 2.0 Å in the predicted ligand's position, from that of experiment, is still considered theoretically valid.

EF-hand 1 of CB-D_{9k} is non-canonical and therefore does not bind Ca^{2+} in the same pattern as the more typical, canonical EF-hands. It has been determined experimentally that Ca^{2+} interacts with the backbone carbonyl of residues Ala14, Glu17, Asp19 and Gln22 as well as the carboxyl group of Glu27 and a water molecule in the

non-canonical loop of CB-D_{9k} (69). In the results explained above (Figure 3.3), AutoDock was successful at predicting these residues with the exception of Glu17 and Gln22. For the former, one of the carboxyl oxygen atoms of Glu17 was interacting with Ca^{2+} instead of the carbonyl group of that amino acid. In the latter case, an extra interaction with the amide group of Gln22 was also predicted. In addition, since the specific water molecule that participated directly in the Ca^{2+} binding (Figure 1.2) was not used in the docking calculation, this interaction site could not be predicted. It could be possible that the incorrect interaction of Ca^{2+} with the lone pair of the amide group on Gln22 was to compensate the missing interaction with water.

Since the second hand of CB-D_{9k} is canonical, it follows the typical binding of canonical CBP (30, 69). Since water plays an important role in canonical binding, it was anticipated that the predictions by Autodock would not be accurate. Ideally, Ca^{2+} in the loop of EF-hand 2 would interact with one of the oxygen ions in the carboxyl group of both Asp54 and Asp58. In addition, it should interact with the carbonyl group in the side-chain of Asn56. The best-docked pose, which was closest in distance to the experimentally-determined position of Ca^{2+}, showed interactions with both oxygen atoms in the carboxyl group of Asp54 as well as an interaction with the carbonyl of Gly57. There are another four experimentally-determined interactions between Ca^{2+} and the loop residues, namely: backbone carbonyl interaction with Glu60, an interaction with the carboxyl group of Glu65 and an interaction with a water molecule. The chosen conformation has an interaction with the carbonyl of Glu60 but shows an interaction with only one of the oxygen atoms in the carbonyl of Glu65. The side chain carbon in Val61 as well as the carbon of the carboxyl group in Glu60 is shown to be in close contact with

Ca^{2+} at this position; however, these close contacts cannot to be part of the electrostatic interactions formed in canonical binding of calcium ions.

3.4 Conclusion

Several conclusions can be made from the results of the validation calculations. These conclusions set the standards for the expectations from docking results of Ca^{2+} into CB-D_{28k} in the next chapters. First, Autodock 4.2 has been proven to be reasonably successful for prediction of the binding position of Ca^{2+} in CB-D_{9k}, thus it can be used for docking Ca^{2+} into other EF-hand proteins such as our protein of interest: CB-D_{28k}. Although the predicted conformation did not include all the correct interacting residues, this was expected since the docking calculation did not include the structural water molecule, which is one of the interacting residues for both hands of CB-D_{9k}.

Second, the binding energies calculated by Autodock 4.2 are known to be unreliable for a ligand such as Ca^{2+}. This is attributed to its scoring function, as explained in section 3.3. The binding energies of the docking calculations in the next chapters will therefore be disregarded, as they do not have a significant meaning. Another important factor in docking calculations is the dissociation constant. Since this constant is calculated from the binding energy, it was also disregarded in the docking calculations in the following chapters.

Chapter 4

Calbindin D_{28k}- Ca^{2+} Complex

The holo form of CB-D_{28k} (Ca^{2+}-bound form of the protein) was resolved in 2006, and 10 models of the NMR solution structure were published under the PDB entry: 2G9B. However, these models are lacking their bound calcium ions. Visual inspection has also revealed that some parts of the published protein structure do not conform to the ideal helix-loop-helix secondary structure. These problems were the trigger to creating the first goal of the project, which aimed at getting a better structure of the holo protein. The achievement of this goal would also provide insight into why hands 2 and 6 are non-functional while hands 1, 3, 4 and 5 are functional. Computational modeling tools were used to introduce the four Ca^{2+} into the functional hands CB-D_{28k}. The obtained protein complex was relaxed using MD simulations, as outlined in Figure 4.1 on the next page. In this chapter, the results for the first goal will be shown and discussed.

4.1 Preliminary Docking

To achieve the first goal of this project, the first step in the sequence of calculations run was the addition of the calcium ions at the loops of CB-D_{28k} EF-hands using Autodock 4.2 (shown in yellow in Figure 4.1). Since EF-hands 1, 3, 4 and 5 are functional, while hands 2 and 6 are non-functional (i.e. they do not bind Ca^{2+} at physiological concentrations), the preliminary docking calculations were performed on the functional hands only. The protein complex structure obtained from these calculations would have Ca^{2+} in the functional EF-hands. This step was intended to provide a relatively good starting structure for the MD simulations, which are discussed in the next section.

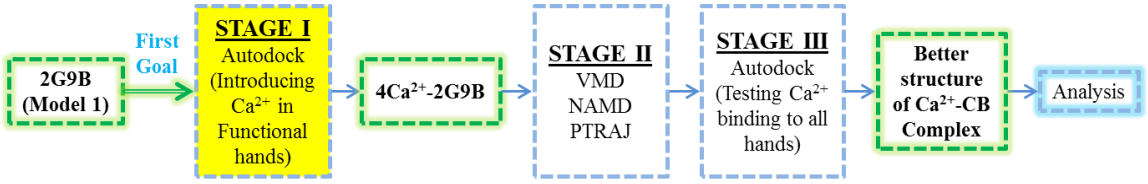


Figure 4.1: The sequence of calculations employed to achieve the first goal of the project. Autodock 4.2 was first used to introduce Ca^{2+} in the functional EF-hands. CB-Ca^{2+} complex was allowed to relax for 50ns, using MD calculations by NAMD. PTRAJ was used to get representative structures of the protein for the last 2.5 ns. Ca^{2+} was docked again in the MD modified CB-D_{28k} structure.

4.1.1 Results of Ca^{2+} Docking in Functional EF-Hands

Autodock 4.2 was used to set up and run four docking calculations with the specifications outlined in subsection 2.1.1. Results from these calculations should help obtain a structure for CB-D_{28k} with four Ca^{2+} docked into its functional loops. The search space for each of the four docking calculations was set to cover one of the functional EF-hands and the loop residues of each hand were made flexible. Each calculation was set to run

100 times, thus there were 100 proposed docking poses for each of the four calculations. The results of these calculations were inspected individually and the best-docked pose for each functional EF-hand was chosen. At this stage, the preference for a docked pose was based on the knowledge of the typical canonical binding. Thus, the pose that was more centered in the loop and was in close contact with its expected interacting residues was chosen in each case. As with the validation calculations, docking results were clustered at RMSD values of 1.0 Å and 2.0 Å.

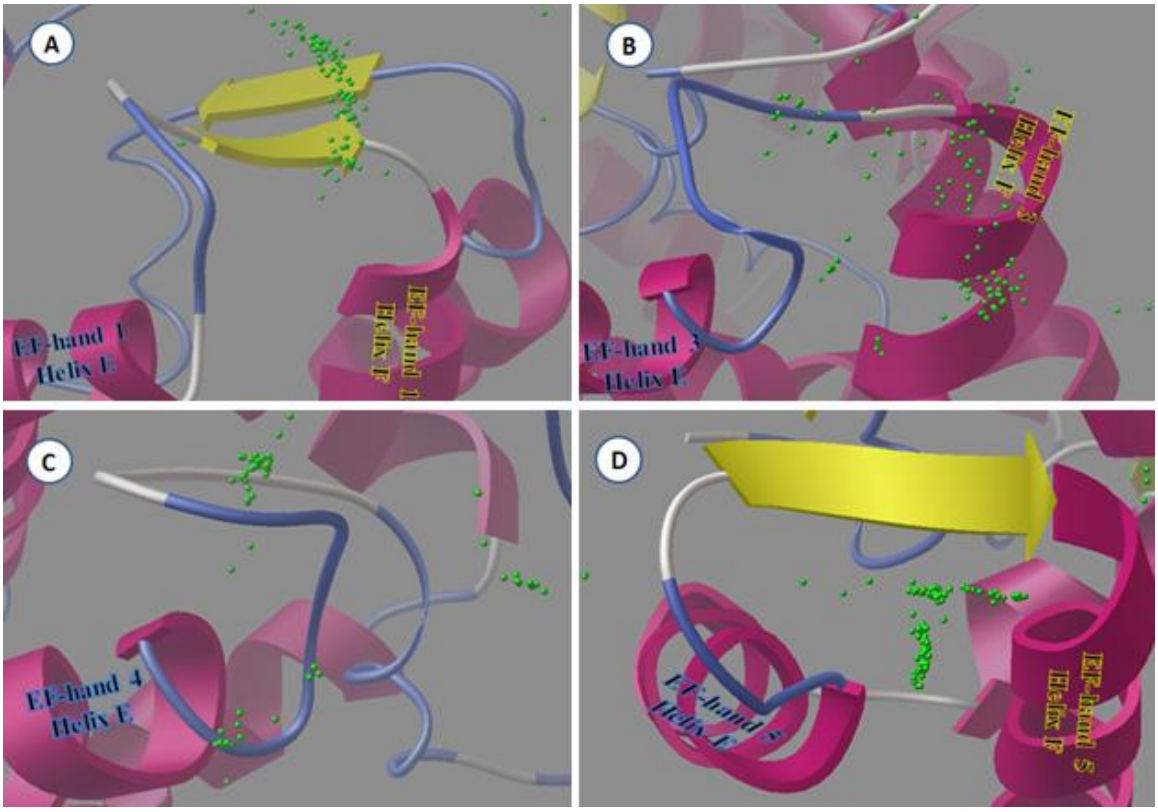


Figure 4.2: Images A to D show the calculated locations of Ca^{2+} (green spheres) in the loops of EF-hands 1, 3, 4 and 5 respectively.

Figure 4.2 above shows the 100 predicted locations of Ca^{2+} for each of the four functional EF-hands displayed in ADT. The conformation in which Ca^{2+} was more centered in the loop, and had more interactions with the canonical loop residues, was

chosen to represent the Ca^{2+} binding pose of that hand. A structure of $4Ca^{2+}$-CB-D_{28k} complex was obtained from the preliminary docking calculations and was carried forward for further calculations.

Table 4.1: The chosen conformations from each of the preliminary docking calculations. The binding energy is calculated as the sum of A= vdW, hydrogen bonding and desolvation effects, B= electrostatic energy, C= penalty of torsional loss. Note that there is no dissociation constant (K_D) for Ca^{2+} at EF-hand 5 since the binding energy of the chosen conformation at that site was positive. Underlined residues are those found in typical canonical interactions.

EF-hand #	Binging Energy =A+B+C (kcal mol^{-1})	K_D (mM)	Residues in close contact with Ca^{2+}
1	-0.29 =0.47-1.05+0.3	616.51	**Asp26**: Two oxygen atoms of side-chain carboxyl **Glu32**: Oxygen atom of side-chain carboxyl **Glu35**: Two oxygen atoms of side-chain carboxyl
3	-0.53 =0.26-1.09+0.3	406.62	**Thr112**: Carbon atom of side-chain methyl **Glu119**: Two oxygen atoms of side-chain carboxyl **Glu121**: Oxygen atom of side-chain carboxyl
4	-0.66 =1.01-1.98+0.3	326.15	**Asp155**: Oxygen atom of side-chain carboxyl **Ser156**: Nitrogen of peptide bond **Asp159**: Two oxygen atoms of side-chain carboxyl **Lys161**: Nitrogen and oxygen of peptide bond
5	0.29 =0.46-0.47+0.3	NA	**Asp201**: Oxygen atom of side-chain carboxyl **Asp207**: Oxygen atom of side-chain carboxyl **Asn209**: Nitrogen of side-chain amide **Glu210**: Two oxygen atoms of side-chain carboxyl

Table 4.1 above lists the binding energies and dissociation constants of Ca^{2+} at the chosen conformations. The table also lists the residues in CB-D_{28k} that are in close contact with the Ca^{2+} at the chosen positions. Snapshots of the chosen conformations, showing the calcium ion positions and the residues with which it is in close contact, are shown in Figure 4.3 to Figure 4.6 for EF-hands 1, 3, 4 and 5 respectively.

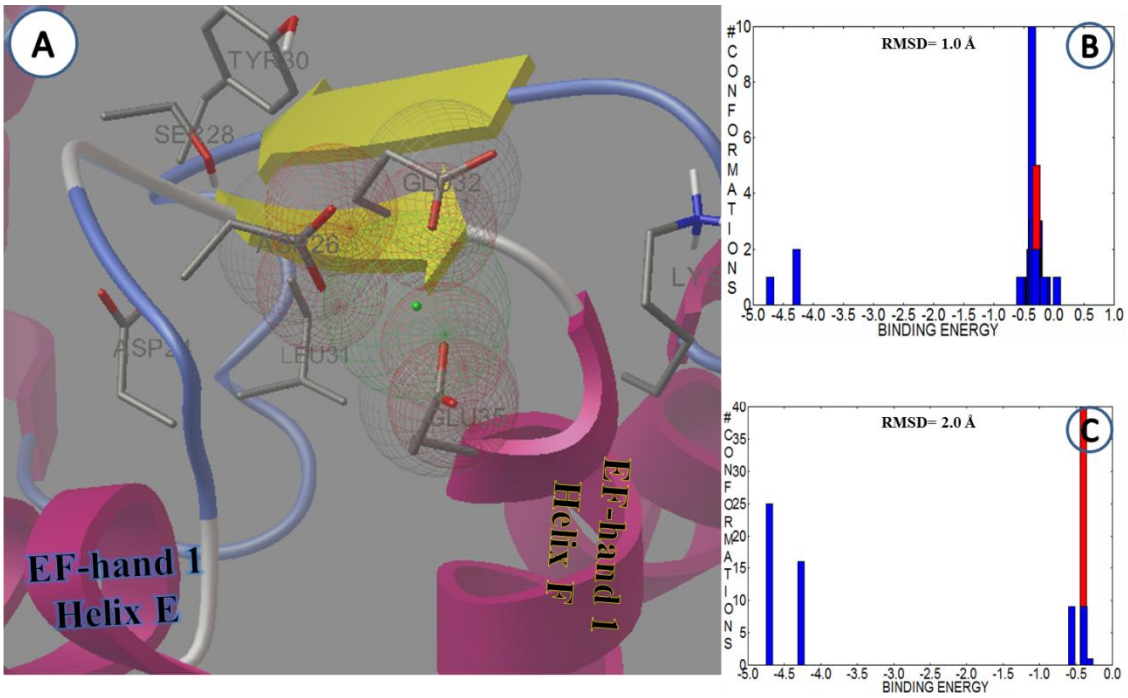


Figure 4.3: A) A snapshot of EF-hand 1 with Ca^{2+} (green sphere) docked into its loop. Ca^{2+} is in close contact with residues Asp26, Glu32 and Glu35. The docked conformations were clustered using RMSD cut-offs of **B)** 1.0 Å and **C)** 2.0 Å respectively. The histograms show that the chosen conformation is in the most populated cluster (red bar).

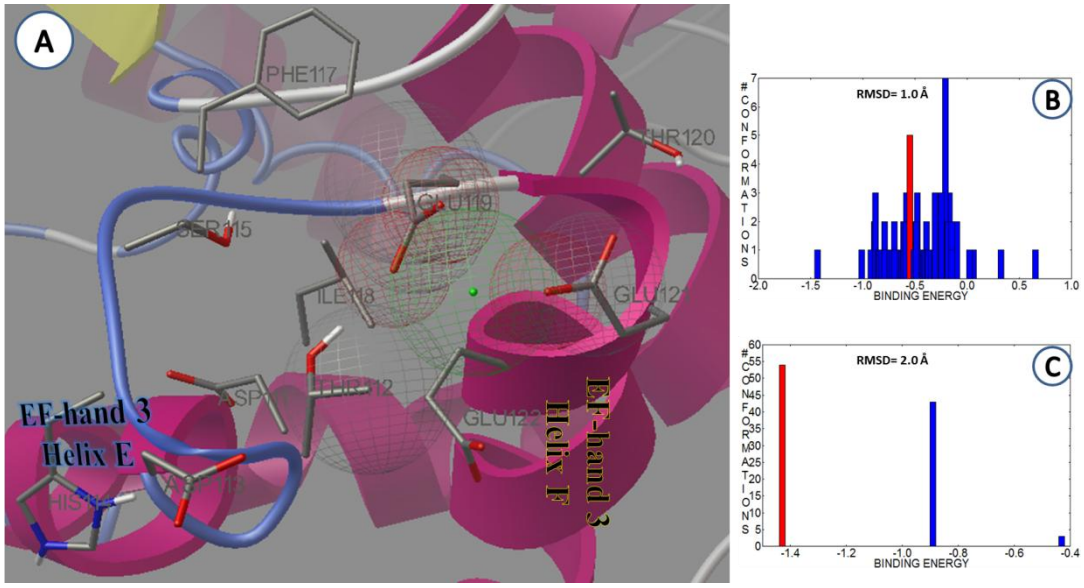


Figure 4.4: A) Calcium ion (green sphere) docked into the loop of EF-hand 3. Sphere meshes show that Ca^{2+} is in close contact with Thr 112, Glu 119 and Glu 121. **B)** The predicted conformations are clustered together with a 1.0 Å cut-off. **C)** At a 2.0 Å cut-off, the chosen conformation is the most populated cluster (in red).

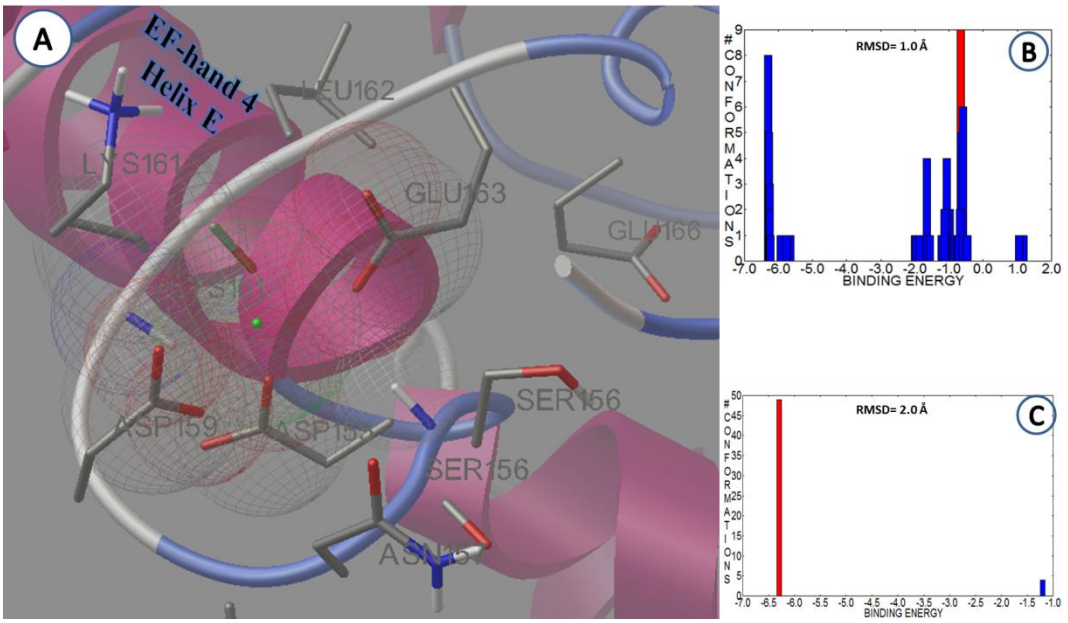


Figure 4.5: A) Image showing the chosen position for Ca^{2+} in the loop of EF-hand 4, in which it is in close contact with Asp155, Asp159 and Glu163. Also notice that helix F is not formed. **B)** The chosen conformation belongs to the biggest cluster at an RMSD cut-off of 1.0 Å and **C)** RMSD cut-off of 2.0 Å.

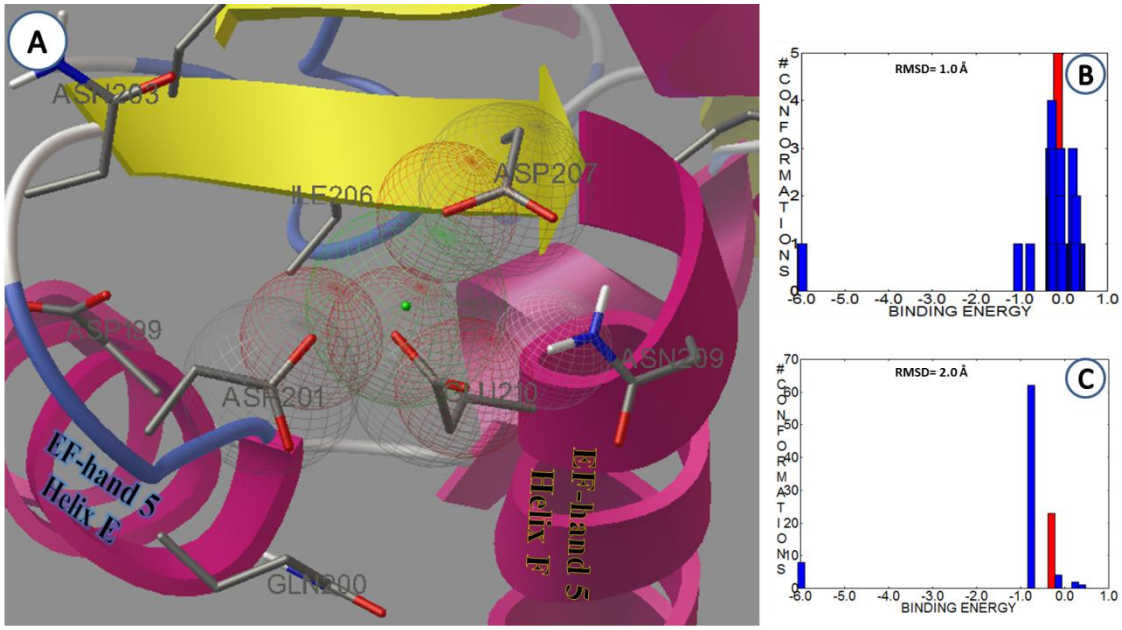


Figure 4.6: A) The chosen docked pose for Ca^{2+} in EF-hand 5. **B)** The chosen conformation belongs to the biggest cluster at an RMSD cut-off of 1.0 Å **C)** but not at an RMSD cut-off of 1.0 Å.

4.1.2 Discussion of Preliminary Docking Results

The main aim of the previous task was to get a preliminary structure of CB-D_{28k} with its four Ca^{2+} bound to their respective functional EF-hands. This structure was chosen to be relaxed in the subsequent MD simulation. Since the quality of MD calculations hugely depends on the starting protein structure, the accuracy of this structure was of significant importance. It was important to analyse the preliminary docking results in attempt to ensure the best starting point for the following calculations.

EF-hand 1: Generally, Ca^{2+} at the docked conformations did not interact with all the typical residues of canonical binding. In case of EF-hand 1, Figure 4.2 (A) shows that the predicted Ca^{2+} positions were inside the loop. The main interactions at these sites,

however, were with residues Asp26, Glu32 and Glu35 only, which is half the number of interactions it should have with the loops of EF-hand 1.

EF-hand 3: Figure 4.2 (B) shows that there were no "hot spots" where Ca^{2+} would bind in this hand. While the general picture shows that predicted locations are scattered inside the loop, many conformations placed the Ca^{2+} at helix-F of EF-hand 3. The chosen conformation that was thought to be the best starting point for Ca^{2+} inside the loop of EF-hand 3 showed only one correct interaction: with Glu119.

EF-hand 4: For this hand, Figure 4.2 (C) also shows that helix-F is not properly formed. Instead, the residues form a long loop. It was therefore difficult to choose a conformation in which Ca^{2+} was centered in the loop. The structure was chosen based on the number of correct interactions it could form, which was four, out of the six expected interactions. Another point to note is that a more recent figure of the structure of the holo protein has been published (31) and did in fact have a properly formed EF-hand 4, with the helix-loop-helix sequence. This structure, however, has not been reported in the PDB.

EF-hand 5: Predictions for the last functional EF-hand showed several more dense regions at which Ca^{2+} could possibly bind (Figure 4.2 D). Most of these conformations were centered at the loop. The maximum number of correct interactions in the predicted conformations, according to the expected canonical binding, was three out of six. The chosen conformation was one that had three of the expected interactions and was more centralised to the loop, compared to the other conformations.

The affinities of the chosen conformations were all in the millimolar range, which by definition indicate poor binding affinity. However, as seen from the results in the previous chapter, predicted energies and dissociation constants for Autodock have proven

unreliable for docking of small ligands, especially those like Ca^{2+} that depend on electrostatic interactions. With the exception of EF-hand 5, it is also important to note that the chosen conformations, which followed the choice criteria, were all in the highest populated cluster when an RMSD cut-off of 2.0 Å was used. This fact increases the likely validity of the chosen conformations. For EF-hand 5, the conformation in which Ca^{2+} had more correct interactions with the loop residues and was more centered in the loop did not follow this pattern but was still preferred over other conformations. As stated earlier, docking calculations for EF-hands 2 and 6 were not required at this stage since they are experimentally known not to bind Ca^{2+}.

4.2 Protein Relaxation and Re-Docking

A good MD calculation should be able to simulate the behaviour of the protein over time and hence could possibly provide a better approximation to the structure of holo CB-D_{28k}. The docking results from the previous calculations were merged together such that a new CB-D_{28k} PDB file was created. This new file contained the coordinates of the protein, in which the coordinates of the loop residues were slightly tweaked from the original 2G9B published file, and four calcium ions. This new PDB file constituted the new preliminary structure of the holo form of CB-D_{28k} for the next set of calculations.

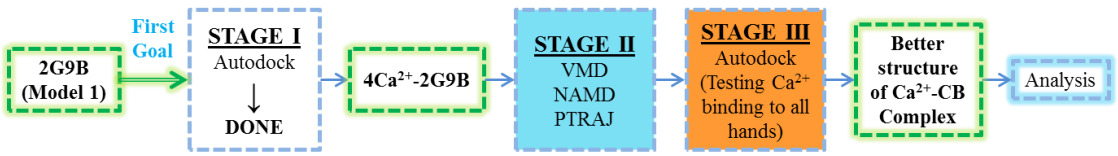


Figure 4.7: Flowchart showing the next two stages (four calculations) that were run on CB-D_{28k} to achieve the first goal of the project.

4.2.1 MD Results and Analysis

Stage II involved four calculations, using VMD, NAMD and PTRAJ (blue box in Figure 4.7). As described in section 2.1.2, the new protein structure was solvated and neutralised using VMD, then allowed to relax for 50 ns using MD calculations. This produced 25,000,000 new frames of the system, which contained 25,000,000 calculated conformations of the protein structure. The PTRAJ utility was used calculate the RMSD of these protein conformations over the simulation period, as shown in Figure 4.8 below.

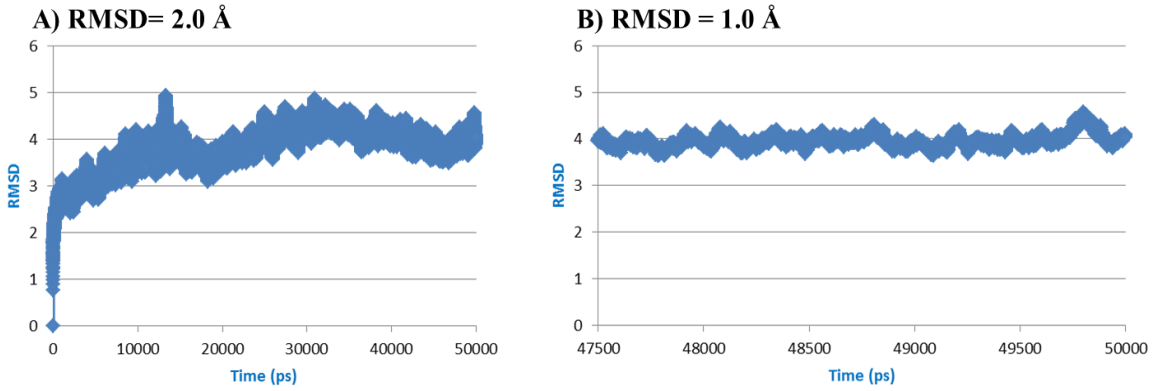


Figure 4.8: Graphs showing the RMSD of the protein position over: **A)** Whole simulation time= 50 ns, **B)** the last 2.5 ns, i.e. from time 47.5 ns to 50 ns. RMSD for each frame is that of the frame structure compared to the structure at time 0.

This utility was also used to calculate the DBI, pSF and SSR/SST ratios when the number of clusters used to represent the protein structure were tested at 2 to 10, as well as 15 and 20. Figure 4.9 shows three plots of the different numbers of clusters versus each of the clustering metrics. Based on the guidelines outlined in Section 2.1.3, it was decided that the last 2500 frames of the simulation would be divided into six clusters; six representative structures of each cluster were carried forward for the last docking calculations.

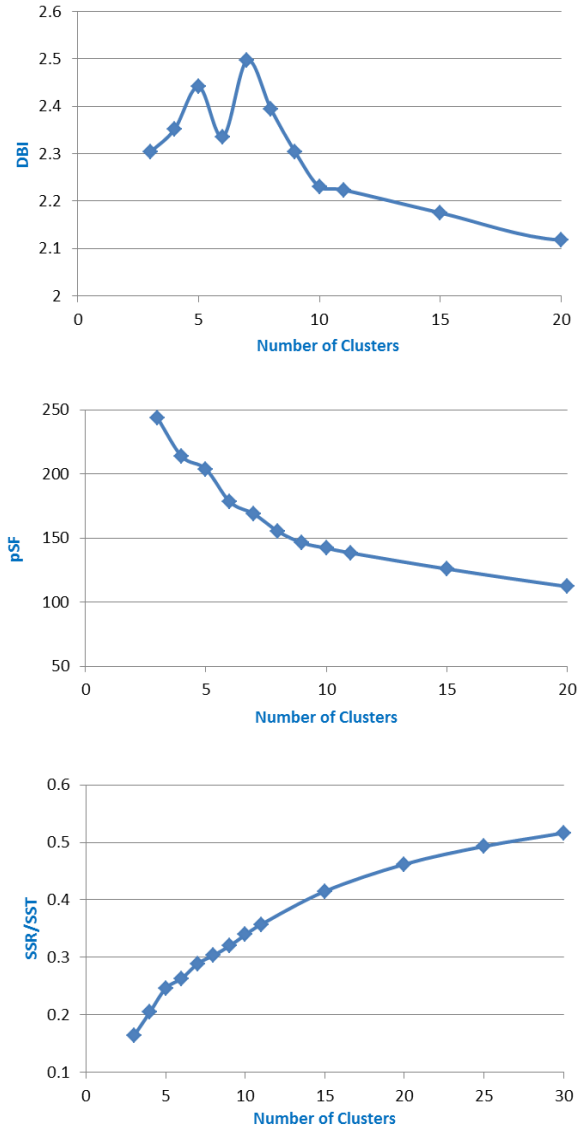


Figure 4.9: Plots of the different clustering metrics, from the top: DBI, pSF, SSR/SST, vs. Number of clusters (2 to10, 15, 20).

4.2.2 Discussion of MD Results and Analysis

Generally, the positions of atoms in a protein change over a simulation period since proteins are flexible. Therefore, it would be unrealistic to extract only one conformation to represent the protein structure over the simulation period. Theoretically, the best way to account for the flexibility of the protein during a simulation is to use all the structures

produced by the MD calculations, however, this is completely impractical. Therefore, a limited number of structures must be chosen carefully such that they provide the best sampling of the different conformations visited by the protein during the MD simulation. As previously explained in Section 2.1.3, the number of structures that would best represent the different protein conformations is very subjective, but the decision can be guided by the clustering metrics. The plot for DBI shows a clear local minimum at six clusters. In the pSF graph, there is a local maximum at clusters 4 and 6. The plot for the SSR/SST ratios does not plateau at the values shown but there is a slight kink in the graph at cluster numbers 4 and 6. Based on these subjective measures, the number of clusters was chosen to be six. The centroid structure was deemed to be the most representative structure of each cluster and was carried forward to the final docking calculations (stage III in orange, Figure 4.7). Note that these six representative structures will be referred to as RepC1-C6 in the rest of this thesis.

The 2G9B protein model obtained from the PDB had structural defects, specifically at EF-hands 4 and 6 that did not have the helix F. All the six EF-hands in the structures obtained from the MD calculations, however, have the helix-loop-helix sequence of EF-hands including hands 4 and 6 (shown later in Figure 4.16 and Figure 4.20). The formation of helix-F through protein relaxation can be related to the fact that EF-hands are anticipated to be helices and therefore should be more stable than loops. Since protein structures tend to relax in MD simulations by adopting a lower energy conformation, it is likely that it would adopt a helical conformation.

4.2.3 Ca^{2+} Re-Docking Results

The six structures obtained from the previous calculations represent the main conformations of CB-D_{28k} during the last 2.5 ns of the MD simulation. These structures were prepared for docking calculations using Autodock 4.2 (orange box in Figure 4.7). First, the calcium ions, which were repositioned during the MD simulations, were now removed from each of RepC1 to RepC6. These representative structures, lacking Ca^{2+}, were prepared for docking calculations; their PDBQT files were created, the loop residues of all hands (both functional and non-functional) were set to be flexible and grid maps were calculated using AutoGrid. Docking calculations were set and run as outlined in section 2.1.1. The predicted Ca^{2+} positions for each structure in all six EF-hands are shown in Figure 4.10, 4.12, 4.13, 4.16, 4.18 and 4.20.

The results for EF-hand 1 (Figure 4.10), shows that the predicted Ca^{2+} positions are within the loop and are mostly centered at the loop, although some are higher up, above the loop level.

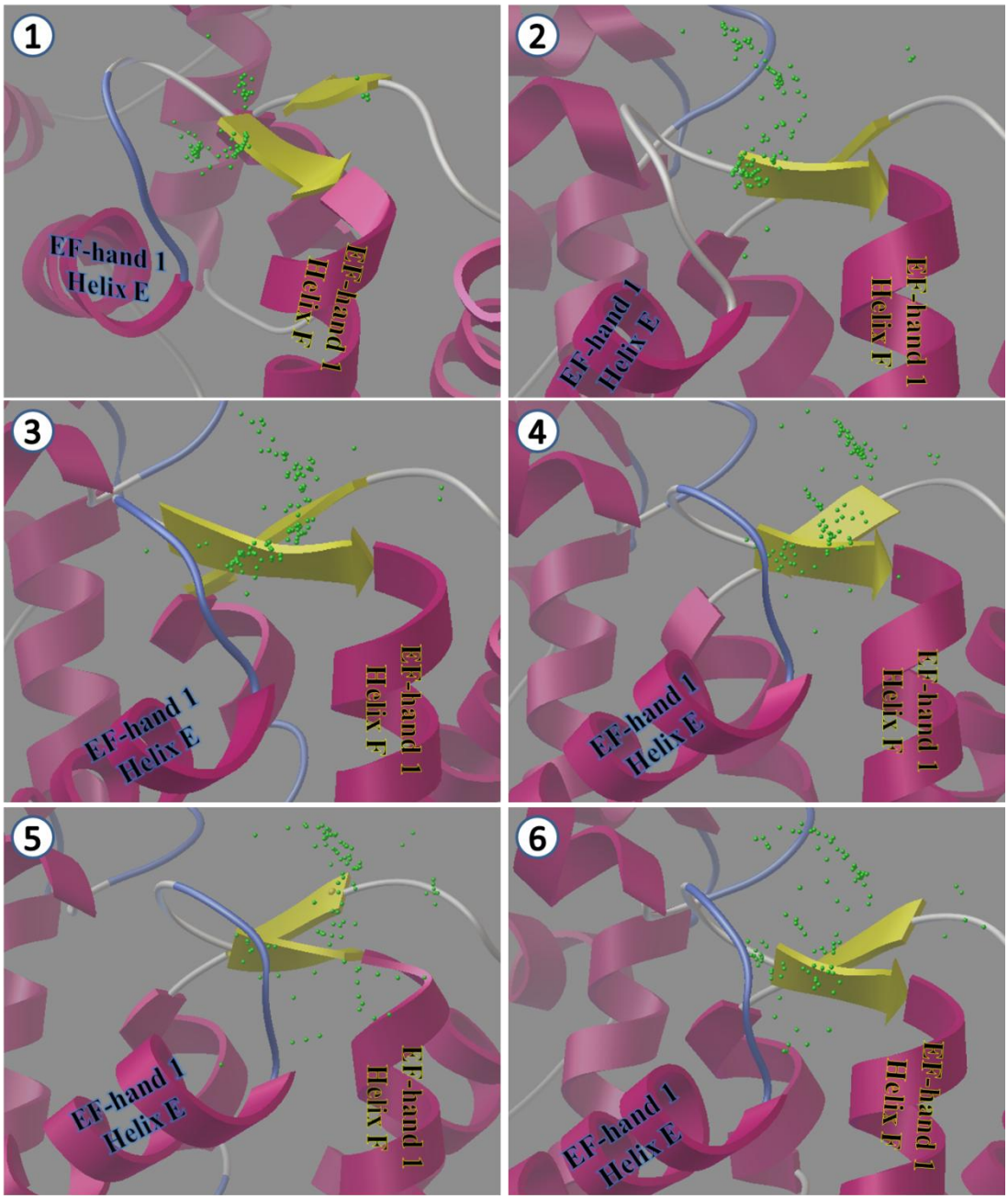


Figure 4.10: The predicted Ca^{2+} positions when the docking grid was set to cover the loop of EF-hand 1. Docked calcium ions are presented as green spheres for each of the six representative structures (RepC1 to RepC6).

In almost all of the docked conformations, the docked calcium ion was in close contact with the carboxyl groups of Asp24, Asp26, Glu35 and the carbonyl group of Glu30. Most conformations also showed an additional interaction with the backbone nitrogen of Glu 32. Figure 4.11 is an expanded view that shows the interaction spheres between Ca^{2+} and the canonical residues in this docked conformation, namely Asp24, Asp26, Glu30, Glu32 and Glu35.

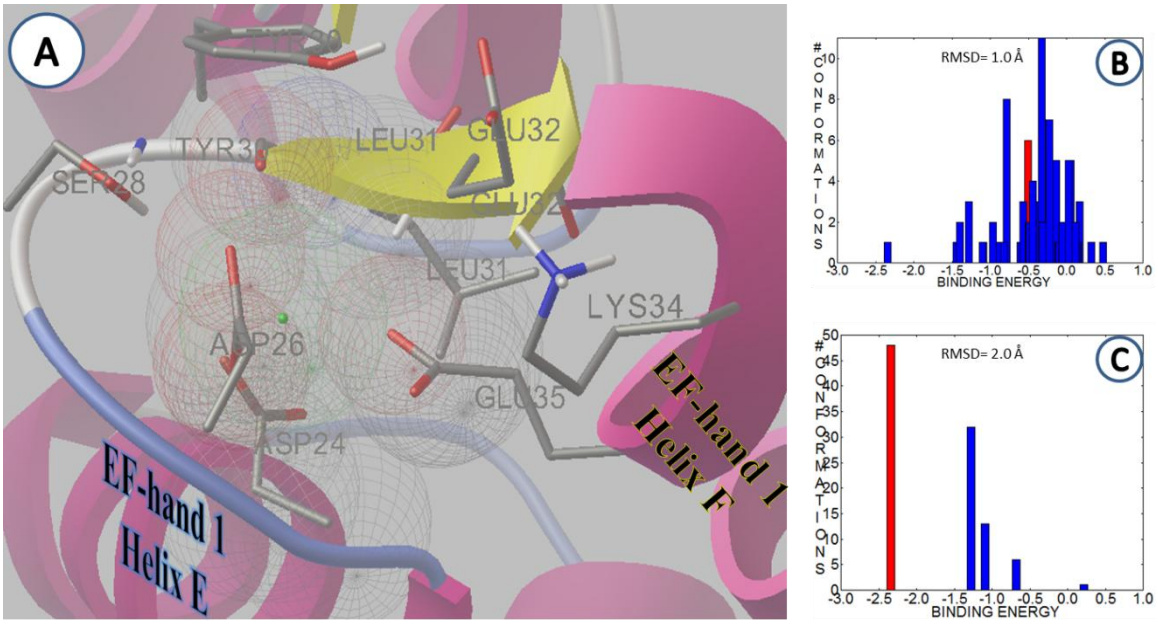


Figure 4.11: A docked conformation of Ca^{2+} at the loop of EF-hand 1 in RepC5 showing the most common interaction predicted by AutoDock. Ca^{2+} is seen in close contact with the backbone carbonyl of Tyr30 as well as the carboxyl groups of Asp24, Asp26 and Glu35. This conformation also shows an additional interaction with the backbone nitrogen of Glu32.

Figure 4.12 clearly shows that the structure of EF-hand 2 deviates significantly from the typical form of an EF-hand; the two helices are seen to be nearly parallel and the loop joining them forms an arc. Overall, none of the representatives for EF-hand 2 roughly fit the conformation of the right figure of the perfect EF-hand shown in Figure 1.2. Most docked conformations for RepC1, RepC2, RepC5 and RepC6 were at the loop of EF-hand 1. In RepC3, Ca^{2+} was mostly at interaction points with Glu77 (residue at the end of EF-hand 2 loop). Docked calcium ions in RepC4 showed similar behaviour, but Ca^{2+} was also commonly found in close contact with Asp70, which is a part of the loop of EF-hand 2. Note that the perspectives in Figure 4.12 are from different views, chosen to best demonstrate the Ca^{2+} locations.

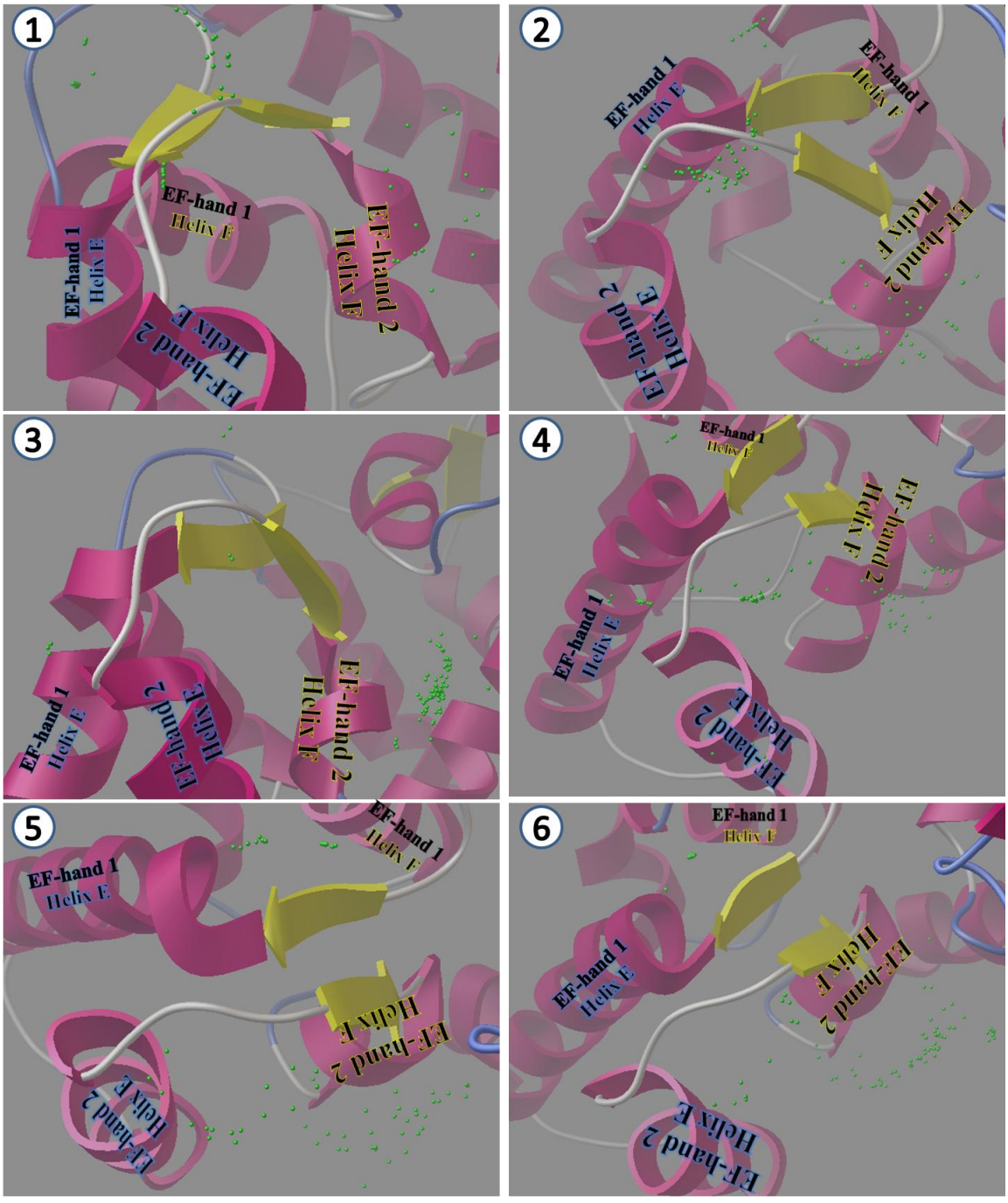


Figure 4.12: The predicted Ca^{2+} positions when the docking grid was set to cover the loop of EF-hand 2. Docked calcium ions are presented as green spheres for each of the six representative structures. Note that labels have been added to EF-hand 1 since some predicted Ca^{2+} positions were at the loop of that hand.

The docked calcium ions at EF-hand 3 were mainly at the loop, yet were not as centered as they were expected to be, as shown in Figure 4.13. In fact, the predicted Ca^{2+} positions were either at the start or at the end of the loop.

The most common predicted Ca^{2+} position was towards the first half of the loop, so that it was closer to the start. At such positions, it was in close contact with Asp111, Asp113 and Ser115, as shown in Figure 4.14. Note in the figure that 'start' and 'end' mark the start and end of the loop. Another prominent Ca^{2+} position was at the other end of the loop; at that position Ca^{2+} was in close contact with the carbonyl of Phe117 and the carboxyl groups of residues Glu119 and Glu122.

Figure 4.15 shows a more significant docked conformation for Ca^{2+}. At this position, there are electrostatic interactions between Ca^{2+} and the carboxyl groups of Asp111, Glu119 and Glu122. In addition, it interacts with the backbone carbonyl of Phe117. Although this position was predicted only once by Autodock, as shown by the red bar in part (B) of the figure, it enables Ca^{2+} to interact with four out of the six canonical residues.

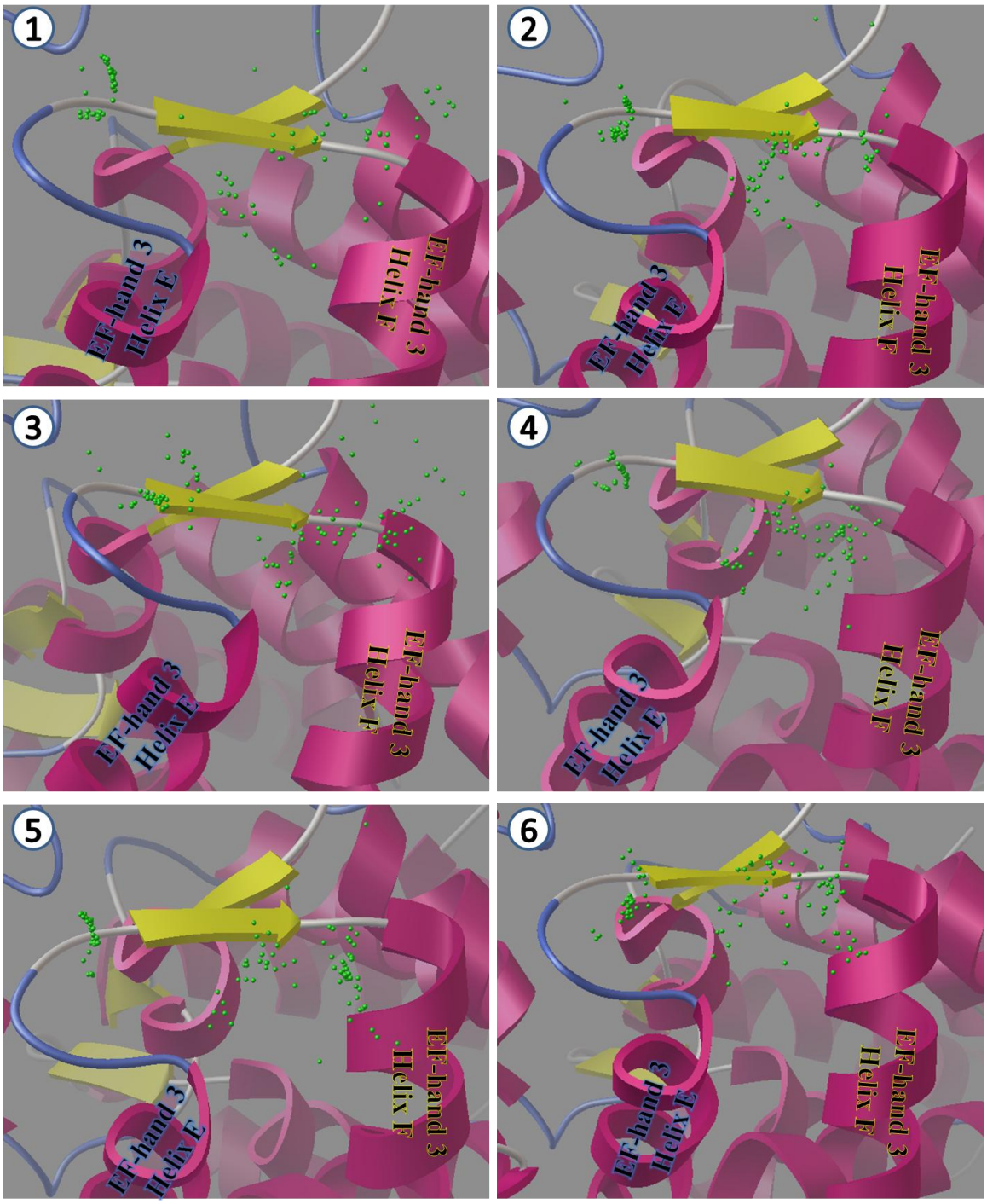


Figure 4.13: The predicted Ca^{2+} positions when the docking grid was set to cover the loop of EF-hand 3. Docked calcium ions are presented as green spheres for each of the six representative structures.

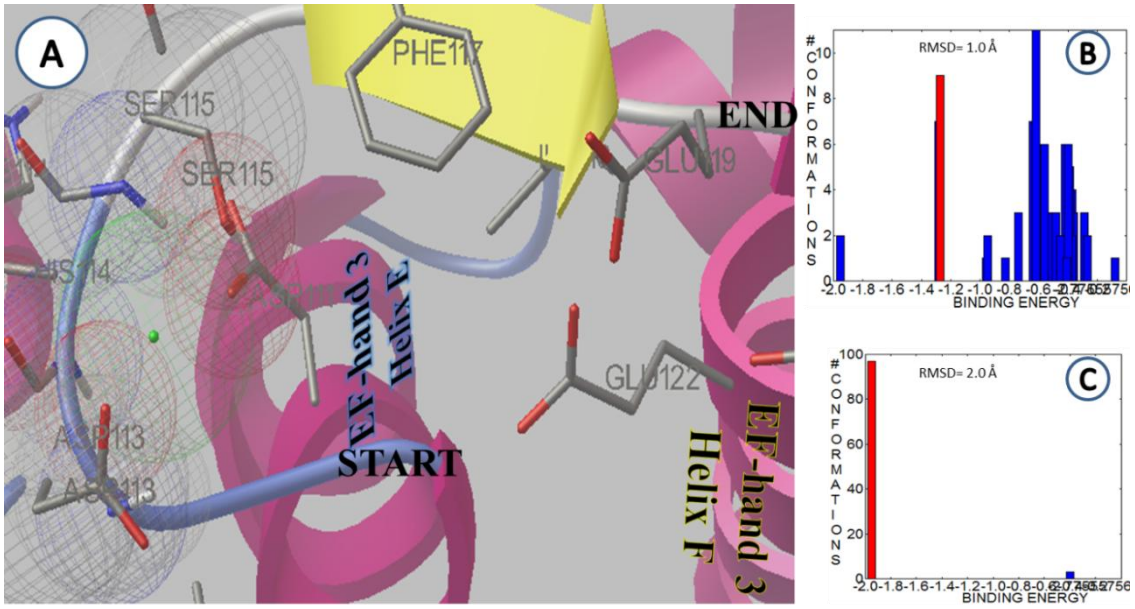


Figure 4.14: A docked conformation of Ca^{2+} at the loop of EF-hand 3 in RepC4 showing the most common interaction predicted by AutoDock. Ca^{2+} is in close contact with the carboxyl groups of Asp111, Asp113 and the hydroxyl group of Ser115. 'Start' and 'End' mark the start and end of the loop. Note that Helix goes back into the frame and is forming the expected helix-loop-helix of an EF-hand.

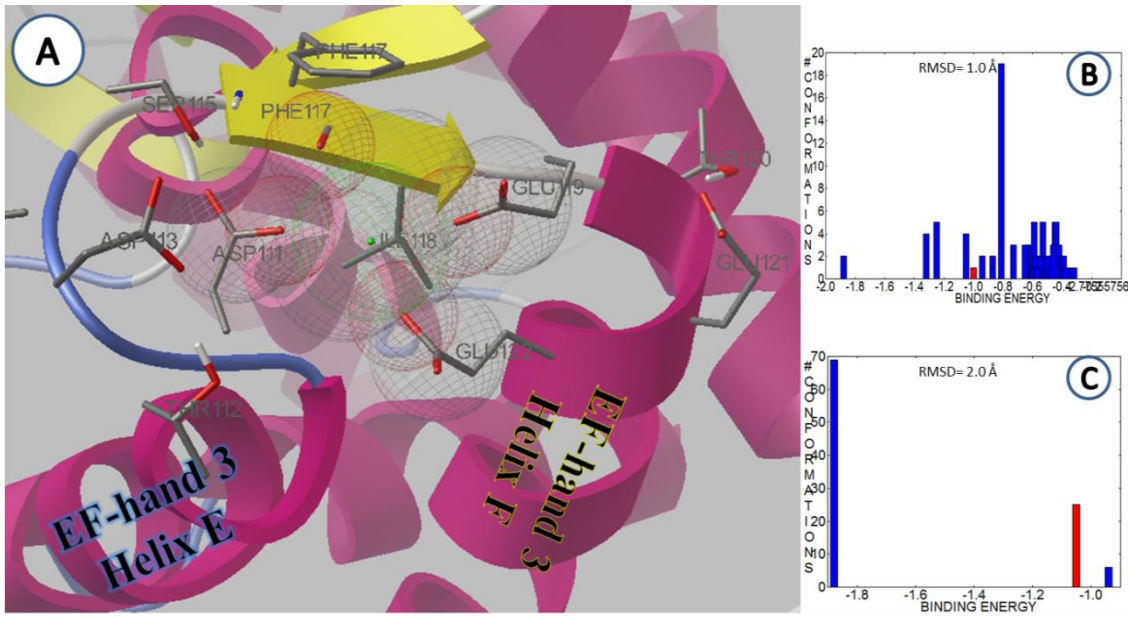


Figure 4.15: A docked conformation of Ca^{2+} at the loop of EF-hand 3 in RepC3. Although this position was predicted only once, it is significant since it allows Ca^{2+} to interact with more residues that are involved in canonical binding. As shown, Ca^{2+} is in close contact with the carboxyl groups of Asp111, Glu119, Glu122, the backbone carbonyl group of Phe117 and the side-chain of Ile118.

Results for EF-hand 4, in Figure 4.16 below, show that Ca^{2+} was most frequently docked at the center of the loop. Note that in all the representative structures, the helix-loop-helix sequence of the hand was formed.

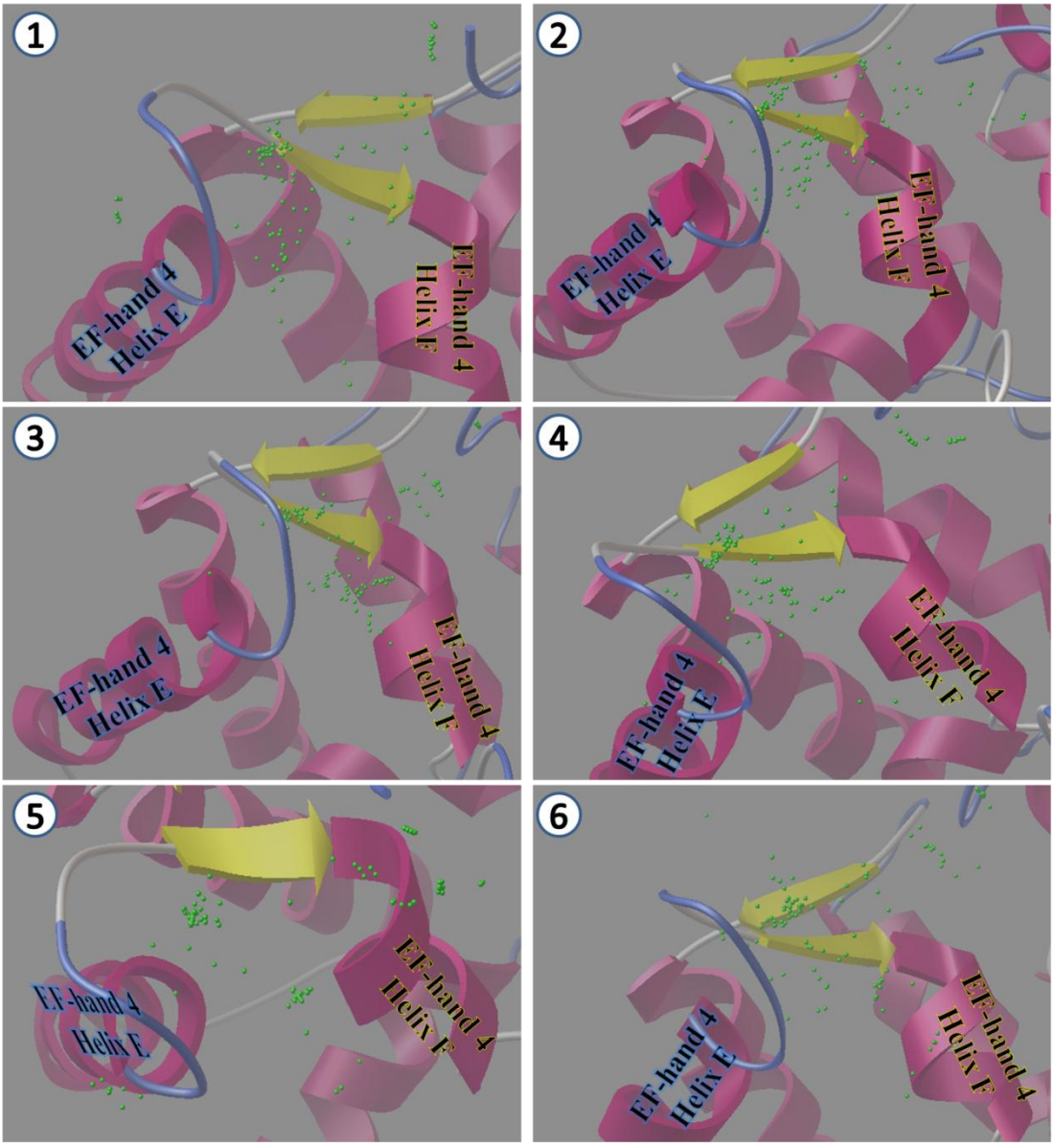


Figure 4.16: The predicted Ca^{2+} positions when the docking grid was set to cover the loop of EF-hand 4. Docked calcium ions are presented as green spheres for each of the six representative structures. Also, note that helix F is now formed for all the structures, which were obtained from MD simulations.

All the predicted conformations positioned Ca^{2+} in a manner that allowed it to have electrostatic interactions with the carboxyl groups of Asp155, Asp159, Glu163 and the backbone carbonyl of Lys161, as shown in Figure 4.17. With the exception of some conformations in RepC1, Glu166 was never a part of the interacting residues with the docked Ca^{2+}. In addition, some predicted poses involved interactions with the side-chain nitrogen of Asn157 instead of the Lys161 carbonyl group.

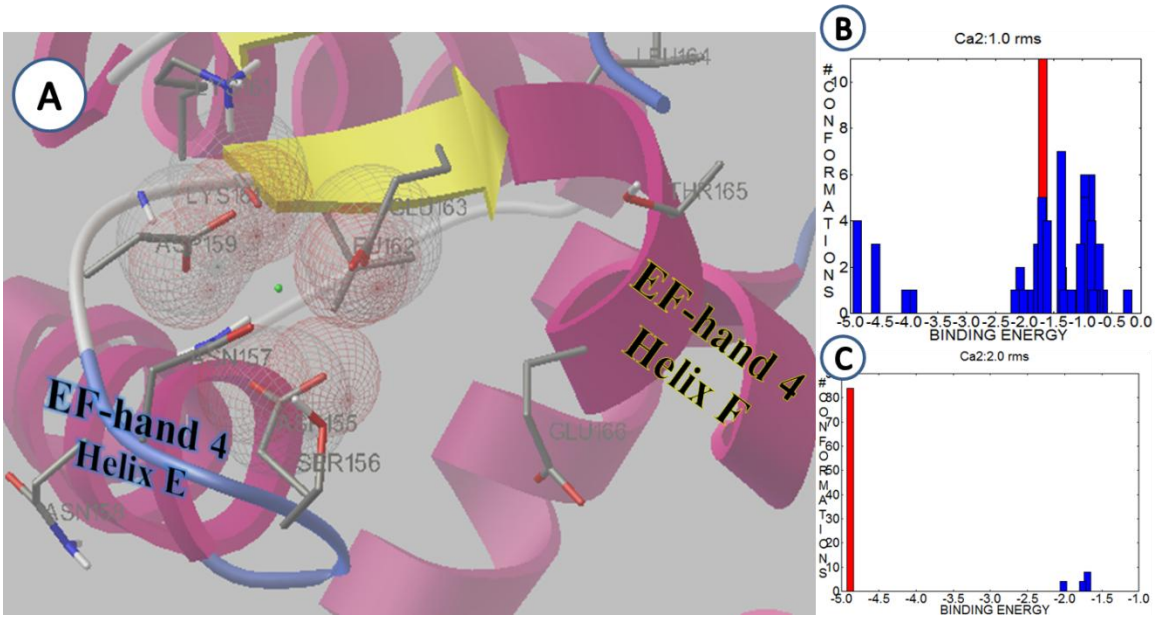


Figure 4.17: A docked conformation of Ca^{2+} at the loop of EF-hand 4 in RepC2. This conformation represents the most common Ca^{2+} position for this hand. The cation is in close contact with the carboxyl groups of Asp155, Asp159, Glu163 as well as the backbone carbonyl of Lys161.

The predicted docked conformations in the loop of EF-hand 5 had similar interactions between Ca^{2+} and the loop residues. As shown in Figure 4.18, all the predicted positions of Ca^{2+} were inside the loop. They were mostly centered but deviated slightly towards the right (the end of the loop). At the predicted positions, Ca^{2+} interacted with the carboxyl groups of residues Asp201, Asp207 and Glu210. In addition, it interacted with the hydroxyl group of Tyr205 as shown in Figure 4.19.

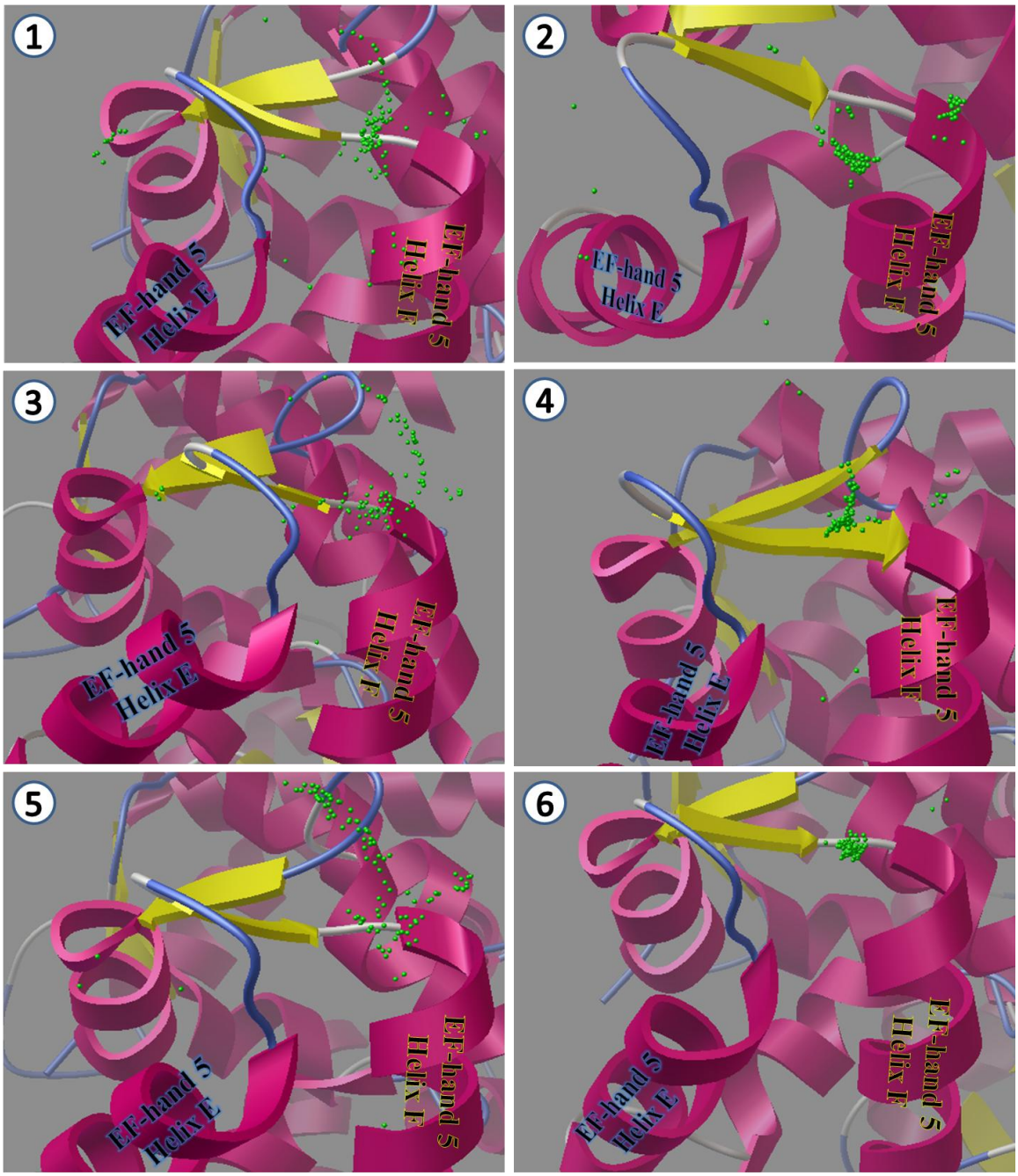


Figure 4.18: The predicted Ca^{2+} positions when the docking grid was set to cover the loop of EF-hand 5. Docked calcium ions are presented as green spheres for each of the six representative structures.

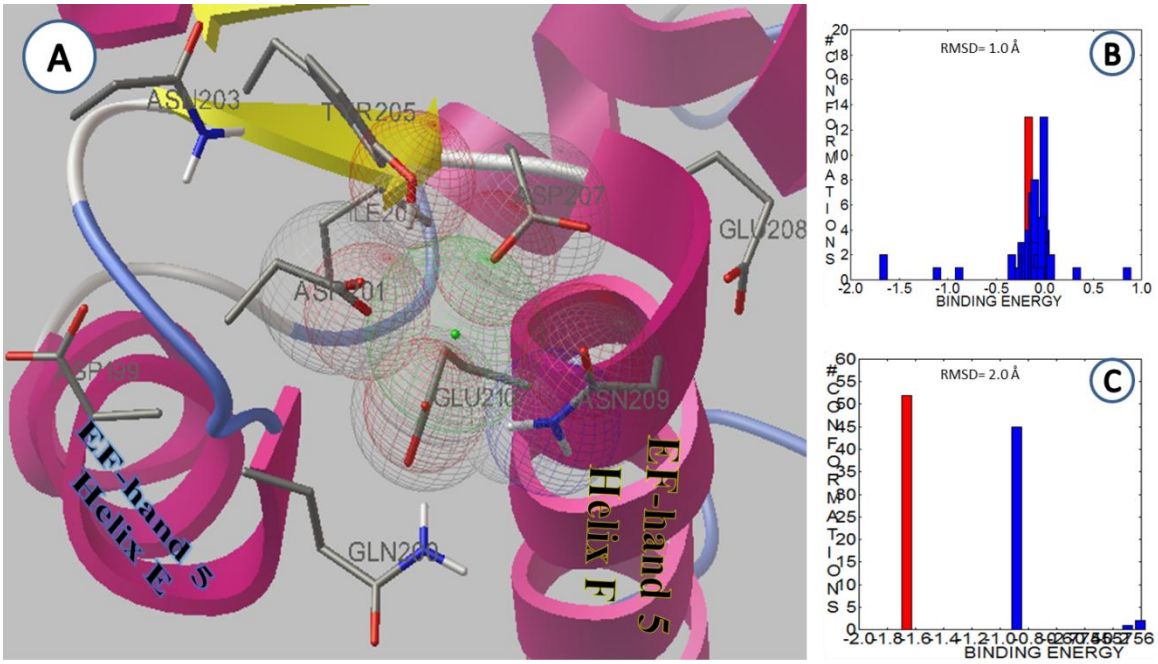


Figure 4.19: A docked conformation of Ca^{2+} at the loop of EF-hand 5 in RepC1. This conformation represents the most common Ca^{2+} position for this hand. The cation is in close contact with the carboxyl groups of Asp201, Asp207, Glu210 as well as the side-chain hydroxyl of Tyr205.

The last hand in CB-D_{28k} is the non-functional EF-hand 6. Docking results for RepC1 to RepC6 for this hand are shown in Figure 4.20. In RepC1, Ca^{2+} was most commonly docked at positions wherein it interacted with residues Asp111 and Asp113 of the loop in EF-hand 3. RepC2 results also showed interactions, as in RepC1, but also had a prominent conformation at the loop of EF-hand 5, in which it interacted with Tyr205, Asp207 and Glu210. The predicted Ca^{2+} positions in the other four structures (RepC3-RepC6) were mostly at the loop of EF-hand 6. At these positions, Ca^{2+} was in close contact with Asp243, which is residue number 4 of the loop. Some results in these representative structures also showed conformations that had Ca^{2+} interacting with Asp113 of EF-hand 3 loop plus Asp251 of EF-hand 6 loop.

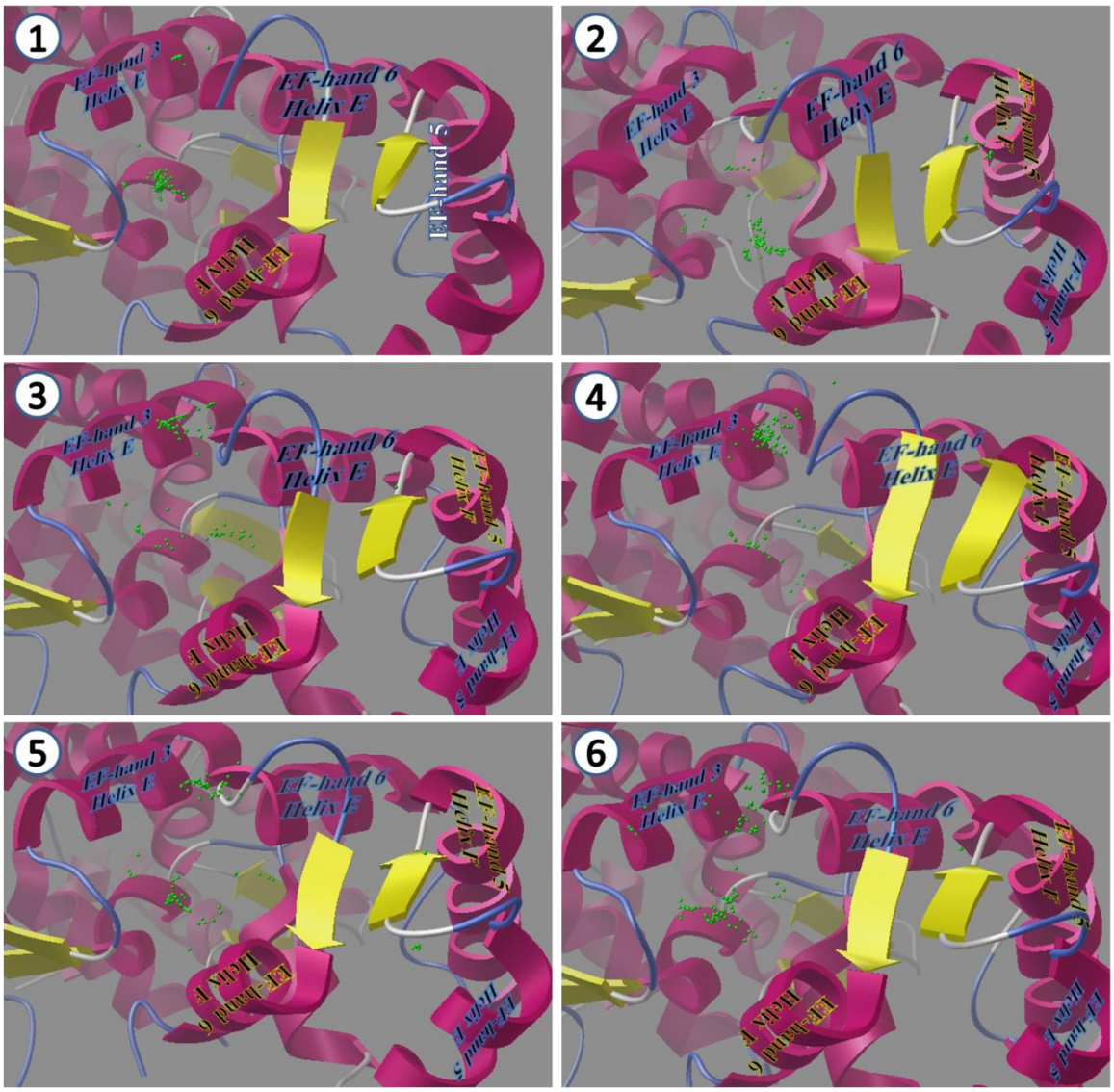


Figure 4.20: The predicted Ca^{2+} positions when the docking grid was set to cover the loop of EF-hand 6. Docked calcium ions are presented as green spheres for each of the six representative structures.

4.2.4 Discussion of Re-Docking Results

The results above show the predicted Ca^{2+} conformations in all the EF-hands of CB-D_{28k}, for the six structures created by the MD calculations for each hand. A successful docked conformation would be one that can reproduce the typical canonical binding of Ca^{2+} to loop residues (previously described in Chapter 1).

Compared to results from preliminary docking, the selected new docked conformations for EF-hand 1 had two additional interactions with the backbone carbonyl of Tyr30 as well as the nitrogen of Gly32. This was also the case for EF-hands 4 and 5, since the selected new docked conformations showed additional interactions with Lys161 for the former and Tyr205 for the latter. Results from EF-hand 3 were not as successful although they still showed improvement. The preliminary Ca^{2+} interactions in EF-hand 3 were with the amino-acid Glu119. In the new docking calculations, however, Ca^{2+} seemed to interact with residues at either the first half of the loop: Asp111, Asp113, Ser115, or the second half, with residues Phe117, Glu119, Glu122.

Although none of the predicted locations replicated the exact canonical binding, there is an obvious improvement in the interaction between Ca^{2+} and the loop residues for functional hands 1, 3, 4 and 5 over the interactions from the preliminary docking stage. Actually, it should be expected that Autodock would not be able to reproduce the canonical binding of Ca^{2+} to the loop residues of CB-D_{28k} since the docking calculations did not involve the structural water molecule at position 9. In typical canonical binding, the residue at position 9 does not interact directly with Ca^{2+}, but rather through a bridging water molecule, as shown in Figure 1.2. Absence of this intermediary molecule would

therefore result in inaccurate docking predictions, yet docking results showed success to an extent.

It is interesting that in most of the docked conformations Ca^{2+} was involved in interactions with the residue at position 9 and lacking the interaction with the residue at position 5. Since the search algorithm tries to avoid positions with ‘steric hindrance’, it could be that the interaction with the residue at position 9 is favored since it is two amino acids away from the residue at position 12, while the residue at position number 5 is only one amino acid away from both residues at positions 3 and 7. In depth analysis of the results showed that this phenomenon was observed in the best-docked conformations of all EF-hands, except EF-hand 4. Note that residue 9 in the loops of the four functional EF-hands has a carboxyl side-chain, while the residues at position 5 have hydroxyl (Ser) or carboxamide (Asn) side-chains (Table 4.2), except the loop of EF-hand 4 which has a carboxyl side-chain at position 5. Since carboxyl groups are more electronegative than hydroxyl or carbonyl groups, this can explain why position 9 was favored over 5 in the docking results of EF-hands 1, 3 and 5.

Table 4.2: This table shows the common residues of canonical EF-hands (30). Highlighted cells indicate that the residue does comply with the typical residue type at that position. Note that interactions at position 9 are mediated through a structural water molecule bridging Ca^{2+} to the loop residue.

Loop Position	1	3	5	C=O at 7	9	12
Common residues	**Asp**	**Asp, Asn**	**Asp, Ser, Asn**	**Tyr, Phe, Lys**	**Asp, Glu, Gly**	**Asp, Glu**
EF-hand 1	Asp24	Asp26	Ser28	Tyr30	Glu32	Glu35
EF-hand 2	Tyr66	Arg68	Asp70	Lys72	Gly34	Glu77
EF-hand 3	Asp111	Asp113	Ser115	Phe117	Glu119	Glu122
EF-hand 4	Asp155	Asn157	Asp159	Lys161	Glu163	Glu166
EF-hand 5	Asp199	Asp201	Asn203	Tyr205	Asp207	Glu210
EF-hand 6	Ala240	Ser242	Gly244	Lys246	Tyr248	Asp251

In Table 4.2 above, it can be seen that some residues in EF-hands 2 and 6 deviate from the typical residues involved in canonical binding. The computational results for these non-functional EF-hands also showed that they do not bind Ca^{2+}. This inability to bind the cation was evident from the fact that predicted binding residues frequently did not even belong to these hands. In the other predicted conformations, the calcium ions were bound by simple one-unit sites, which is not credible. This could be clearly seen in the tertiary structure of EF-hand 2, which does comply with the typical form of an EF-hand. Close examination of EF-hand 6 also shows that it is more closed than other EF-hands and is sigmoid-shaped at the start of the loop.

4.3 Conclusion

The 2G9B structure of CB-D_{28k} obtained from the protein data bank is slightly inaccurate. This is supported principally by three pieces of evidence. First, there is an identifiable error in the secondary structure of the protein. Specifically, EF-hands 4 and 6 have a helix and a long loop instead of the typical helix-loop-helix sequence of EF-hands. The preliminary docking of Ca^{2+} into the loops of the functional EF-hands showed the second piece of evidence since the 2G9B structure did not allow Ca^{2+} to interact with the loop residues in a canonical fashion. Third, there is a published figure of the holo protein (31) that has properly formed EF-hands; the structure for that figure has not been deposited in the PDB.

In general, the accuracy of MD simulations depends on their starting structure, so it is understood that the protein conformations obtained from the MD calculations might not exactly be that of CB-D_{28k}. However, the relaxed structures seemed to provide better solutions for the first two problems listed above. The secondary structure of the relaxed protein conforms to the typical form of an EF-hand, i.e. all hands have a helix-loop-helix sequence, including the formerly deformed hands 4 and 6. To address the second issue, six structures were extracted from NAMD calculations and used for re-docking Ca^{2+}. The results from the docking calculations could not produce the typical canonical binding poses of Ca^{2+} with the loops of the functional EF-hands. However, this could be related to the missing water molecule, which is crucial for interaction at position 9. Nevertheless, the docked conformations for EF-hand 1 showed significantly improved results over the preliminary docking calculations and EF-hands 3, 4 and 5 were somewhat improved.

Thus, it could be concluded that MD simulations were useful in achieving a more accurate structure of CB-D_{28k}.

The docking results presented here, taken along with facts already known, provide statistical evidence to explain the non-functionality of EF-hands 2 and 6. First, several of the amino acid residues of the loops of these two EF-hands are much different from the typical sequence for this class of hands as shown in Table 4.2. Second, the docking results show that Ca^{2+} either interacted with other loop residues on the neighbouring hands, or interacted with single site carboxyl groups. The third evidence is the structure of these hands. EF-hand 2 in the six representative structures was made of two helices joined together by a loop that formed an arc (Figure 4.12); thus, the two helices were somewhat parallel. For EF-hand 6, the loop was sigmoid shaped (Figure 4.20) unlike other hands, especially EF-hand 1 (Figure 4.10) that showed the best results and binds Ca^{2+} first and most strongly (33).

It is hypothesised that this sigmoid shape does not allow the full exposure of the loop residues to the cation. Support for this hypothesis can be found in the calculated structures of EF-hand 4 shown in Figure 4.16. The loops of these structures tend to be sigmoid shaped also. Interestingly, almost none of the docked conformations for EF-hand 4 showed interactions with Glu166, residue at position 12, the canonical binding residue at the end of the loop. It could be that the first two interacting residues (positions 1 and 3), at the sigmoid part of the loop, skewed the canonical ‘coordination sphere’ (Figure 1.2) away from the end of the loop. However, this was not observed in EF-hands 2 and 6, since they lack the canonical residues at these positions (Table 4.2).

Chapter 5

The Zinc-Calcium Conundrum

The second hypothesis of this project is that CB-D_{28k} will have reduced Ca^{2+} binding capacity if it has already bound Zn^{2+}. The second goal is thus to computationally model this scenario and evaluate Ca^{2+} binding capacity. This goal is very computationally challenging, especially given that the basic requirement for this computational model is not available; the structure of CB-D_{28k} bound to Zn^{2+} has not been resolved. Another factor that contributes to this challenge is the fact that the residues in CB-D_{28k} that bind Zn^{2+} have not been fully identified, although some deductions about possible binding residues have been made (18). A series of calculations, described in Section 5.2, have been carried out to address these challenges.

5.1 Zinc Modeling Challenge

Zinc ions are most commonly tetra-coordinated in proteins (62, 70-72). In fact, it has been argued that the coordination of zinc ions with more ligands is a misconception caused by limited experimental resolution in resolving the structure of some proteins

(72). The tetrahedral nature of zinc ion coordination by ligands has been supported by its electronic structure, in which it forms the four coordinates through its $4s^1 4p^3$ vacant orbitals (72).

Molecular mechanical modeling sometimes fails to replicate the four-ligand coordination of zinc ions. A model called the "Cationic Dummy Method" (CDM) was also developed to overcome this issue (70, 71). In this model, zinc cations are represented by five atoms collectively called "tetrahedral zinc divalent cation" (referred to as TZDC in this thesis); four dummy atoms are in a tetrahedral arrangement around a central atom. Note that the four peripheral atoms have 0.0 Å radii and hence their name: dummy. The distance between zinc and each dummy atom is set to be 0.9 Å, while the dummy-zinc-dummy atom angle is 109.50˚. The parameters assigned for the five atoms are shown in Table 5.1. In this model, the central atom represents the zinc atom while the surrounding four atoms represent the vacant $4s^1 4p^3$ orbitals. These four atoms are each assigned a charge of +0.5, thus the overall charge in the model is +2, the charge of a zinc ion.

Table 5.1: Parameters of the atoms in the tetrahedral zinc divalent cation. R*= sum of vdW radii of two like atoms.

Parameters / Atom	Mass (amu)	R* (Å)	Є (kcal mol^{-1})	Charge
Zn	65.380	3.1	$1x10^{-6}$	0
Dummy	0.1	0	0	+0.5

Since this model has not been tested on EF-hand proteins, it was necessary to validate its use for CB-D_{28k}. This validation can be done by docking TZDC into an EF-

hand protein whose bound structure has been resolved. A protein called S100 calcium binding protein B (S100B) is an EF-hand protein that is known to bind both Ca^{2+} and Zn^{2+} ions. It consists of one pair of hands, but usually occurs as a dimer (30). The x-ray resolved structure in the protein data bank has a PDB ID 3D10 (73). In this structure, the protein binds two zinc ions, each of which is tetrahedrally coordinated (74) at the interface between the two dimers. The zinc ions are coordinated by the residues His15 and His25 from the first monomer, and His85 and Glu89 from the second monomer, and vice versa. The two sites at which zinc is coordinated in S100B are referred to as sites A and B in this thesis. If docking calculations on S100B using CDM could replicate the tetrahedral binding of Zn^{2+} in S100B, it should be a good model for use in docking calculations on CB-D_{28k}.

The structure of S100B was obtained from the PDB and prepared for docking calculations as previously outlined in section 2.1.1. First, the PDBQT file of the protein was created, AutoGrid calculations were performed, then AutoDock was run to predict the ligand locations. The search algorithm was set to perform a maximum of 25,000,000 energy evaluations on 28,000 generations, each with 300 individuals. Four docking calculations were performed at sites A and B, as summarised in Table 5.2. In the first calculation, the tetrahedral zinc divalent cation, prepared as specified in CDM, was set as the ligand. This cation was docked into rigid S100B using Autodock 4.2. For comparison, the same calculation was also set using ordinary Zn^{2+} as a ligand. For clarity, ordinary Zn^{2+} cations used in the assessment calculations will be referred to as OZC in Subsections 5.1.1 and 5.1.2. Since docking calculations on CB-D_{28k} will involve the flexible protein, another set of calculations were run on flexible S100B using ordinary

Zn^{2+} and TZDC respectively. For flexible docking, the residues His15, His18, Asp23, Lys24 and His25 in the first monomer were set to be flexible and residues His85, Glu86 and Glu89 in the second monomer were set to be flexible.

Table 5.2: A summary of the calculations performed on S100B, using ordinary zinc cation (OZC) and TZDC (tetrahedral zinc divalent cation) as ligands to assess the performance of the Cationic dummy method in EF-hands.

Site	S100B flexible residues	Ligand
A	None (i.e. Rigid docking)	OZC
		TZDC
	15, 18, 23, 24, 25, 85', 86', 89'	OZC
		TZDC
B	None (i.e. Rigid docking)	OZC
		TZDC
	15', 18', 23', 24', 25', 85, 86, 89	OZC
		TZDC

5.1.1 Results of CDM Assessment

The results from the calculations described above were analysed in ADT. Rigid docking was very successful and the ligands were positioned into their binding sites where they interacted with all the four residues, namely: His15, His25, His85' and Glu89', for both calculations on site B. Site A was successful in the case of OZC only, while the TZDC

docked poses did not show interactions with His15. Flexible docking, however, was not successful at any of the sites for either OZC or TZDC ligands.

The results for docking OZC into site A are shown in Figure 5.1. They are presented as green spheres, while the experimentally-determined position of Zn^{2+} is represented by the yellow sphere. The closest OZC conformation to the experimentally-determined position of Zn^{2+} was 4.418 Å away. In addition, most conformations showed interactions with Glu86 and Glu89, which were assigned to be flexible along with the other residues listed in Table 5.2.

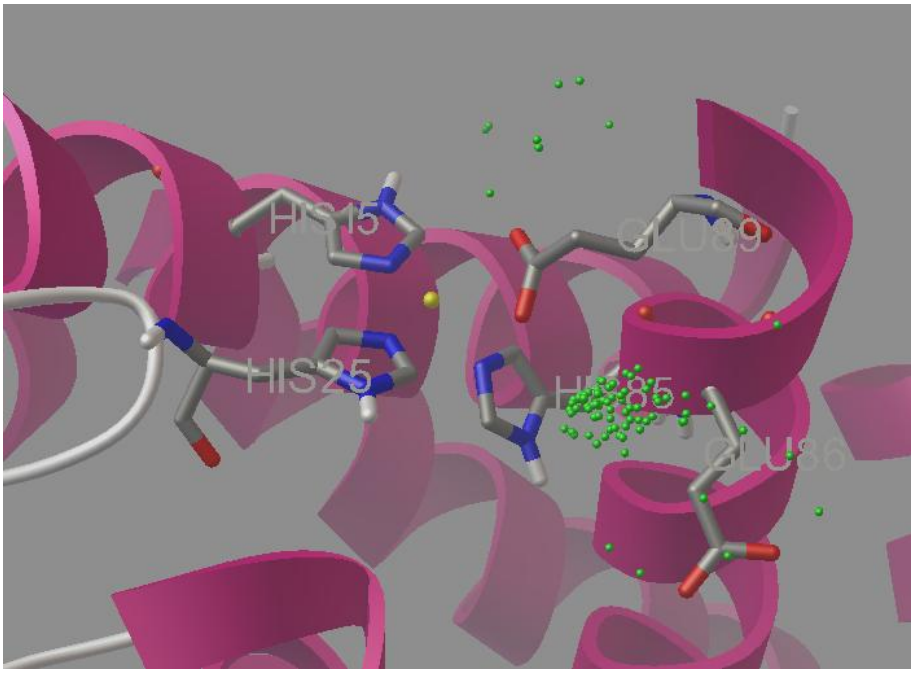


Figure 5.1: A figure showing site A in S100B. The experimentally determined position of Zn^{2+} is represented by the yellow sphere. The results from docking OZC as a ligand are displayed as green spheres, which represent the center of the predicted ligand position. Most of these predicted conformations showed interactions with residues Glu86 and Glu89, which were flexible during the docking calculations.

Results did not show much improvement when TZDC was used as a ligand instead of OZC at site A. Most of the docked conformations again showed interactions with Glu86 and Glu89, while a few showed an interaction with Asp12, which is not a valid interaction site. It can be seen in Figure 5.2 that most results were far away from the experimentally-determined position of Zn^{2+}. However, the closest TZDC predicted position to that of experiment was measured to be 2.459 Å away.

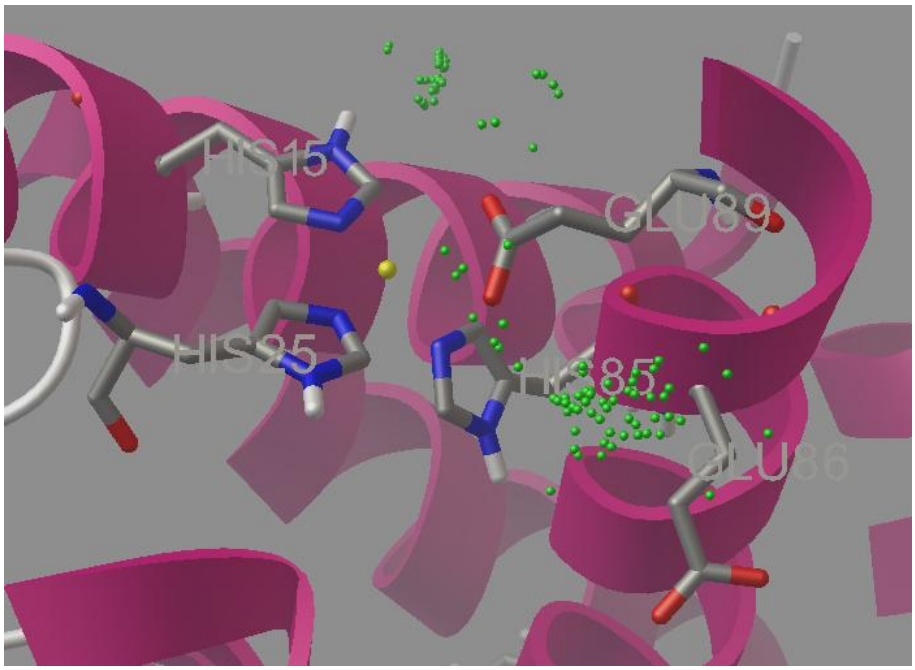


Figure 5.2: A figure showing the centers of the predicted TZDC positions (green spheres) at site A of the protein. The yellow sphere represents the experimentally determined position of the zinc ion. Most the most predicted interaction was with residues Glu86 and Glu89.

Figure 5.3 shows the docked OZC positions at site B of S100B. The results for this calculation also fail in predicting the experimentally-determined position of Zn^{2+}. In most of the docked positions, OZC was in close contact with Glu86 and Glu89. The difference between the experimentally-determined Zn^{2+} position and the closest OZC position to it was 4.128 Å.

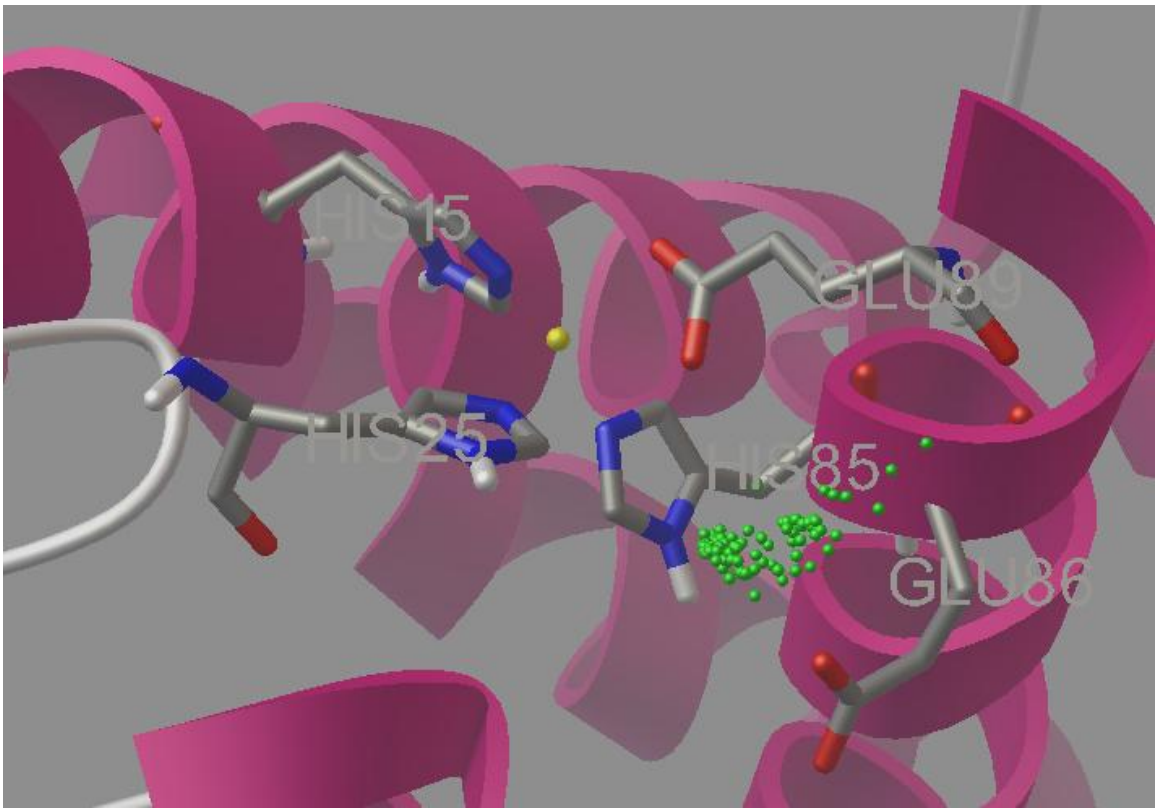


Figure 5.3: A figure showing Autodock predicted OZC positions (green spheres) at site B. Notice that all the OZC positions are away from the residues that should coordinate them, namely: His15, His25, His85' and Glu89'. In addition, the predicted conformations away from the experimental Zn^{2+} position (yellow sphere).

The docked conformations of TZDC at site B are very similar to those obtained when OZC was used as a ligand. Again, the results in Figure 5.4, for TZDC at site B, show that the predicted conformations greatly deviate from that of experiment. Most of the docked conformations were in close contact with Glu86 and Glu89, although a few showed an additional interaction with His85. The closest docked conformation of TZDC to the experimentally-determined position of Zn^{2+} was 4.380 Å away from it.

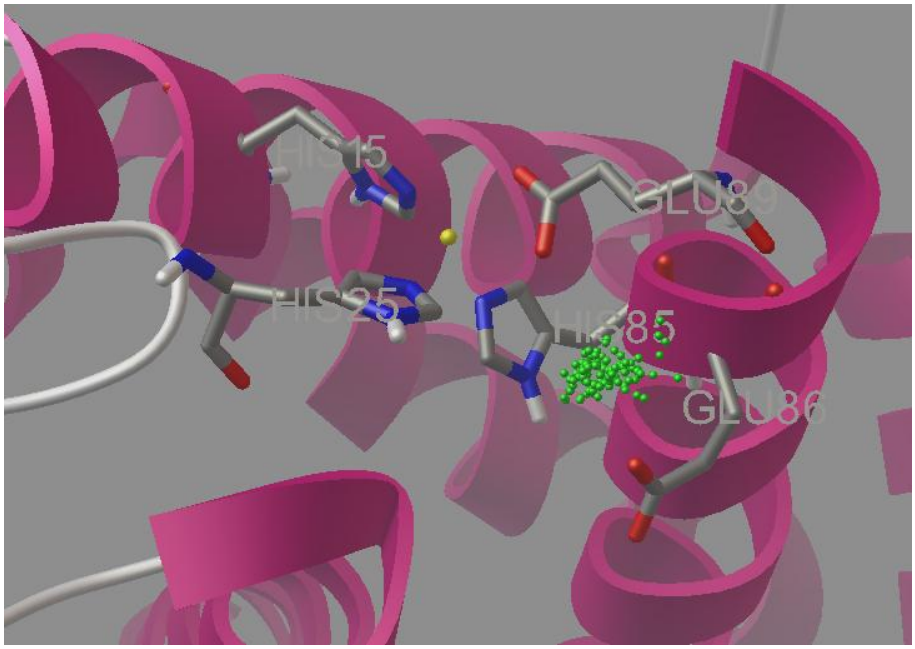


Figure 5.4: A figure showing the centers of the predicted TZDC positions (green spheres) at site B of the protein S100B. The yellow sphere represents the experimentally determined position of the zinc ion. Most the most predicted interaction was with residues Glu86 and Glu 89. His85 was also predicted in a few conformations.

5.1.2 Discussion of CDM Assessment

It was necessary to assess the performance of CDM in modeling zinc cations against the ordinary zinc model. To achieve this, docking calculations were performed on an EF-

hand protein that binds Zn^{2+} as well as Ca^{2+}. From the results in Section 5.1.1, it can be seen that the model did not show improved performance when compared to OZC. The results from site A show that the best docked conformations were 4.418 Å and 2.459 Å away from the experimental position of Zn^{2+} when OZC and TZDC were used as ligands, respectively. These values are both more than the accepted limits. In addition, instead of being in close contact with the three histidine residues 15, 25, 85' and Glu89', the ligands interacted with the carboxylates of Glu86' and Glu89'. Although at site A TZDC was closer to the experimental position than OZC, it was still not close enough. In addition, it was not successful at predicting the correct interacting residues. Owing to the structure of the protein dimer, sites A and B are very similar; they are at the interface between the two EF-hands and coordinate zinc by residues His15, His25, His85' and Glu89'. TZDC at site B was 4.380 Å away from the experimental Zn^{2+}, compared to 2.459 Å at site A. In addition, the few conformations in which it interacted with His85 were not present when it was docked into site A.

The structure of TZDC has its dummy atoms spaced in a perfect tetrahedron. Thus, it assumes perfect tetrahedral coordination between the ligands and residues and the cation, which is not necessarily always the case. In addition, rigid docking of TZDC at site A showed that it could not predict the interaction with His85. Therefore, it can be concluded that CDM is not suitable for Zn^{2+} modeling in EF-hand proteins. Another important point is that the weighting factors of the scoring function (Section 2.1.1) are based on the atomic parameters set in Autodock 4.2. In fact, the use of new parameters would ideally require calibration of the weighting factors to better suit the new assigned parameters. Therefore, ordinary Zn^{2+} (OZC) was used in the Zn^{2+} binding calculations.

5.2 Zn-CB Complex

In addition to its calcium binding ability, CB-D_{28k} was reported to bind at least three zinc ions, as discussed in Section 1.1.3. Although experimental results were not conclusive as to which residues in CB-D_{28k} coordinate Zn^{2+}, there was strong evidence that His80 was involved in Zn^{2+} binding (18). The study also included a crude conclusion that the second site lies within the residues 87 to 175 and the last site is at the N-terminal half of the protein.

Docking calculations in Autodock 4.2 require that a binding site be defined to lie within the grid box (Section 2.1.1), and only a limited number of rotatable bonds is allowed by the program. It was determined, therefore, that the most reasonable course of investigation now should be to bind one Zn^{2+} to CB-D_{28k}, at the most reliably known site localised near His80.

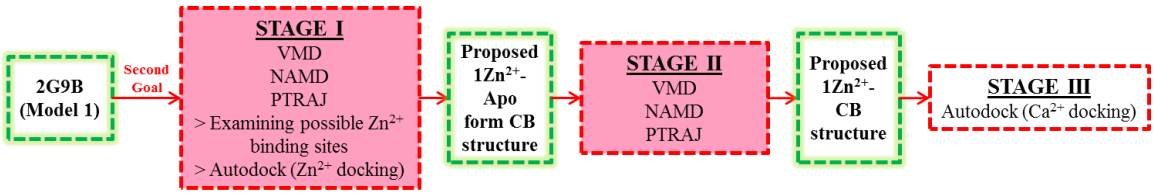


Figure 5.5: A flow-chart outlining the sequence of calculations performed to achieve the second goal of the project

Ideally, to test for the possible interference between Zn^{2+} and Ca^{2+} binding abilities of the protein, calcium ions would be docked into zinc-bound CB-D_{28k}. However, the test in the case of CB-D_{28k} is far more complicated. The first obstacle to achieve the zinc-bound form of the protein was that the protein structure available is that of holo CB-D_{28k} (75). Therefore, this structure would naturally be biased to binding Ca^{2+} over Zn^{2+}. A series of calculations was performed on the holo form of CB-D_{28k} in an

attempt to achieve its zinc-bound form, as outlined in Figure 5.5. First, the 2G9B model of the protein was relaxed for 50 ns. The purpose of this step was to get a structure that would more likely represent the apo form of the protein. This structure could be a better binding structure to Zn^{2+} than the holo form of the protein. The proposed apo structures were examined for possible regions that could bind Zn^{2+}. Docking calculations of Zn^{2+} were run on the hypothesised zinc-binding region. The new complex obtained from the docking calculations was further relaxed so that it would better conform to CB-D_{28k} with one bound Zn^{2+}. At this stage, the structure of CB-D_{28k} at hand is that of the protein with one Zn^{2+} bound to it. Calcium docking calculations were then performed on EF-hand 1 of the new CB-D_{28k} structure to test whether the zinc-bound form of the protein loses its Ca^{2+} binding ability at that site. Results for Stages I and II (pink boxes in Figure 5.5) will be discussed in the following subsections.

5.2.1 MD Results and Analysis

The first step in Stage I was relaxation of the 2G9B model obtained from the PDB. This model represents the NMR-resolved structure of the holo protein that lacks its bound Ca^{2+} ions. The macromolecule was placed in a neutralised water box using VMD. This system was allowed to relax via MD calculations for an equivalent of 50 ns. The relaxation of the protein without its bound ions should ideally provide the apo structure of the protein i.e. the structure of CB-D_{28k} when no Ca^{2+} ions are bound to it.

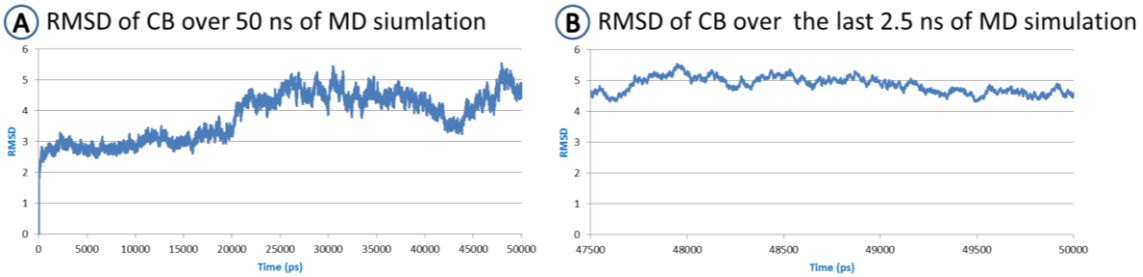


Figure 5.6: Two graphs representing the calculated RMSD of the relaxed protein over (A) the 50 ns of the simulation, (B) the last 2.5 ns of the simulation.

To ensure that the system was given enough time to equilibrate, PTRAJ was used to calculate the RMSD of the protein positions over the simulation period. The graphs of the protein RMSD over the 50 ns as well as the last 2.5 ns of the simulation are shown in Figure 5.6 above. PTRAJ was also used to calculate the clustering metrics for a number of different clusters, using the last 2,500 frames of the simulation. Again based on the previously discussed outlines (Section 2.1.3) it was decided that the last 2.5 ns of the simulation, that produced 2500 frames, are best divided into four clusters. The centroid structure of each cluster, known as the representative structure, was carried forward for the next docking calculations.

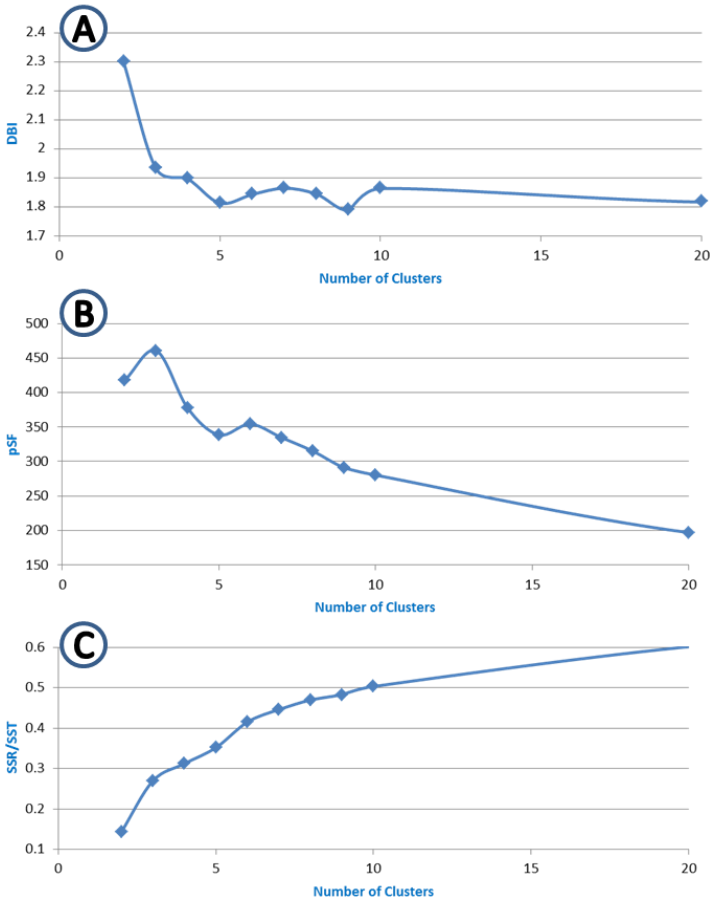


Figure 5.7: Graphs showing the values of the clustering metrics: (A) = DBI, (B) = pSF, (C) = SSR/SST, vs. the number of clusters the last 2500 frames were divided: 2-10, 20.

5.2.2 Discussion of MD Results and Analysis

It can be seen in Figure 5.6 that the protein structure fluctuated greatly during the simulation. This fluctuation is evident by the RMSD values that change greatly during the calculation, more obviously seen at around 20,000 and 40,000 ps. This fluctuation is expected since the protein is changing from its holo form to its apo form. At the time from 47,500 ps to 50,000 ps, the RMSD values do not fluctuate as much, indicating that the protein is better equilibrated at this later stage. The starting structure used at time 0 ps

is that of the holo form of the protein. Since this structure was lacking the bound Ca^{2+}, then relaxing the structure long enough should provide the structure of the protein when it is not bound to Ca^{2+} ions: the apo form. The time of 50 ns was chosen so that the protein would have enough time to relax, while it was recognised that infinitely many conformations could be produced by MD simulations. Longer times were avoided so that the calculated structure would not significantly deviate from its probable structure, a concern given that MD calculations involve numerous approximations.

Proteins do not have static structures. To account for this flexibility, PTRAJ was used to divide the last 2,500 frames of the protein into different clusters. Each cluster should ideally contain structures that are calculated as 'similar', according to the clustering algorithm used. The ideal number of clusters is one that divides the different protein conformations visited during the simulation into distinct groups. The representative structure of each cluster would be used to represent the protein. Thus, the representative structures from all clusters would collectively represent all the conformations of the protein during the simulation; in this case the last 2.5 ns of the simulation.

The three clustering metrics were used, as before, and the plots for DBI, pSF and SSR/SST against the numbers 2 to 10 and 20 clusters are shown in Figure 5.7. Unfortunately, the graph does not show a single option that has all three qualities: a local minimum in DBI, local maximum in pSF, and the start of a plateau in SSR/SST values. The values of 3 and 4 looked the best. Although the plot at 3 shows a local maximum in the pSF plot, it does not show a minimum at the DBI plot. It was decided that the frames would be divided into four clusters as a compromise between the clustering metrics

criteria. Note that this decision is very subjective and, in fact, the clustering metrics do not always provide the right indication to the number of clusters. The representative structures of the four clusters will be referred to as RepZ1-RepZ4 in the rest of this thesis.

5.2.3 The Search for Possible Zn^{2+} binding Sites

While the exact Zn^{2+} binding positions in CB-D_{28k} have not been identified, there is strong evidence that His80 is involved in Zn^{2+} binding to CB-D_{28k}. Since Zn^{2+} is usually coordinated by four residues in proteins, other residues in CB-D_{28k} must also coordinate with the cation. The most common Zn^{2+} binding residues in proteins are histidine and cysteine (62). Glutamic and aspartic acids are also known to coordinate zinc ions in proteins (62). Since experimental evidence has shown that cysteine was not involved in Zn^{2+} binding to CB-D_{28k} (18), it was ruled out as a possible binding residue.

The MD simulations in NAMD and their analysis using PTRAJ yielded a relaxed protein structure that was now assumed to represent apo-CB-D_{28k}. The structures RepZ1-RepZ4 were examined for glutamic and aspartic acid residues that were spatially close to His80. Figure 5.8 shows four glutamic acid residues that are close to His80. However, it can be seen that the residues fall into two distinct regions, one involving Glu57, Glu77 and His80 and the other involving His80, Glu85 and Glu86. The latter region seems to be very unlikely to coordinate Zn^{2+} with His80 since they are too far from it. In addition, the Glu85 and Glu86 residues, being adjacent, are unlikely to both be able to move into a favourable Zn^{2+} binding position along with His80, even with the allowance of chain flexibility. Strain and steric considerations of the two residues would limit successful repositioning to only one residue, at most.

On the other hand, it can be seen (Figure 5.8) that Glu57, Glu77 and His80 have the potential to create a tetrahedron around a zinc ion. Based on these deductions, Zn^{2+} was docked into RepZ1-RepZ4, in an attempt to achieve the Zn^{2+}-bound structure in the region of Glu57, Glu77 and His80 residues.

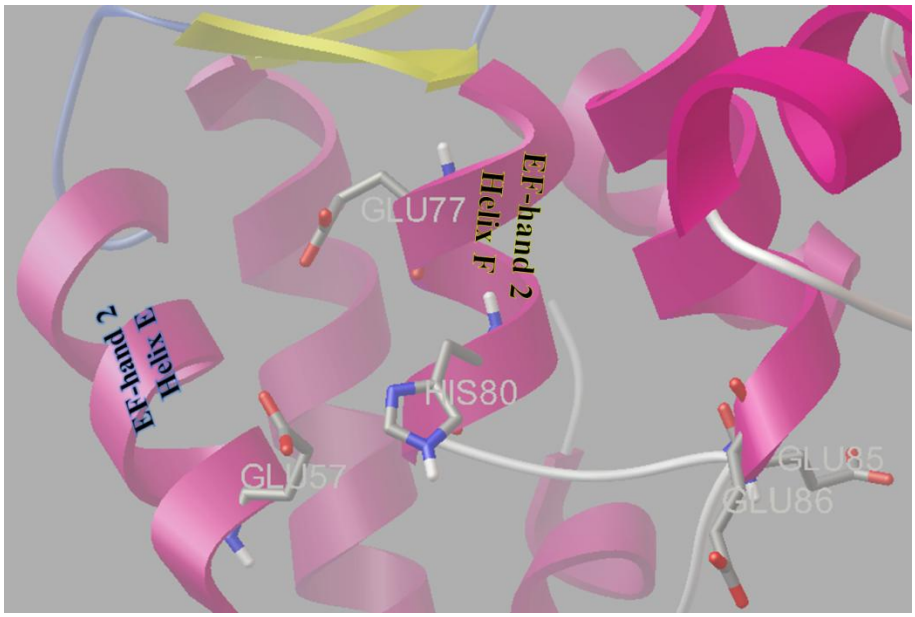


Figure 5.8: A snapshot of relaxed CB-D_{28k} showing His80 as well as the glutamic and Aspartic acid residues close to it, namely Glu57, Glu77, Glu85 and Glu86.

5.2.4 Zn^{2+} Docking into Apo-CB-D_{28k} Results

So far, the 2G9B model of CB-D_{28k} was relaxed in a 50 ns MD simulation by NAMD. Four representative structures were obtained from the last 2.5 ns of the simulation using PTRAJ. Experimental evidence has shown that His80 binds Zn^{2+} in CB-D_{28k}. An investigation of the relaxed, apo form of CB-D_{28k} has revealed that Glu57 and Glu77 are likely the other coordinating residues to Zn^{2+} in CB-D_{28k}. Docking calculations of zinc ions were therefore performed at the region containing these residues. For this

calculation, the suspected binding residues Glu57, Glu77, and His80 were set to be flexible. In addition, the amino acids Leu78, Ala79 and Leu81 were also set to be flexible, to allow for more spatial freedom of residues Glu77 and His80. As part of the protein preparation (discussed in detail in Section 2.1.1), histidine residues were set to be deprotonated, since this is their form in zinc binding proteins. The docking simulation was set to run 100 times on RepZ1-RepZ4, each with a maximum of 50,000,000 energy evaluations. The energy evaluations were set to run on 28,000 generations each with 300 individuals. The results of the docking calculations on all four representative structures are shown in this subsection. The choice of the docked conformations was based on interaction distance criteria, which will be explained in detail in the discussion, section 5.2.5.

The 100 predicted Zn^{2+} locations in RepZ1 are shown in Figure 5.9 (A). In most of these docked conformations, Zn^{2+} interacted with the backbone carbonyl of His80 as well as the nitrogen on its side-chain. The other two interactions are formed between Zn^{2+} and the oxygen atoms of the carboxyl side chain of Glu57, as shown in Figure 5.9 (A) and for the chosen Zn^{2+}-bound structure, in Figure 5.9 (B) respectively.

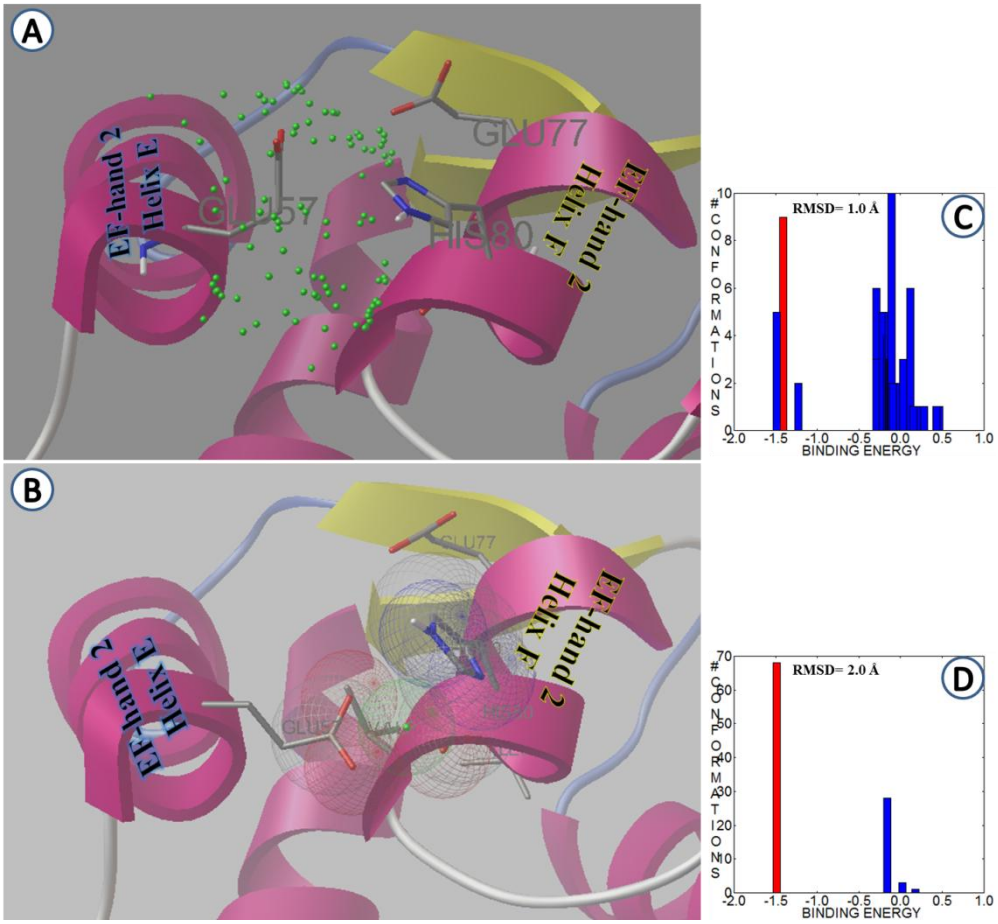


Figure 5.9: Results for Zn^{2+} docking in RepZ1: (A) Results from the 100 runs are represented by the green spheres. (B) The chosen conformation for bound Zn^{2+} in RepZ1, in which Zn^{2+} is bound to His80 backbone carbonyl as well as the nitrogen on its ring. In addition, Zn^{2+} is interacting with the oxygen atoms in the carboxyl of Glu57. (C and D) Two histograms of the different conformations were used with RMSD values of 2.0 Å and 1.0 Å, respectively. The chosen conformation is represented by the red bar on the histogram.

The results for docking Zn^{2+} in RepZ2 are shown in Figure 5.10. Although some conformations showed interactions with Glu57, Zn^{2+} in most docked conformations interacted with the carboxyl oxygen atoms of Glu57 and Glu77 as well as the side-chain nitrogen of His80.

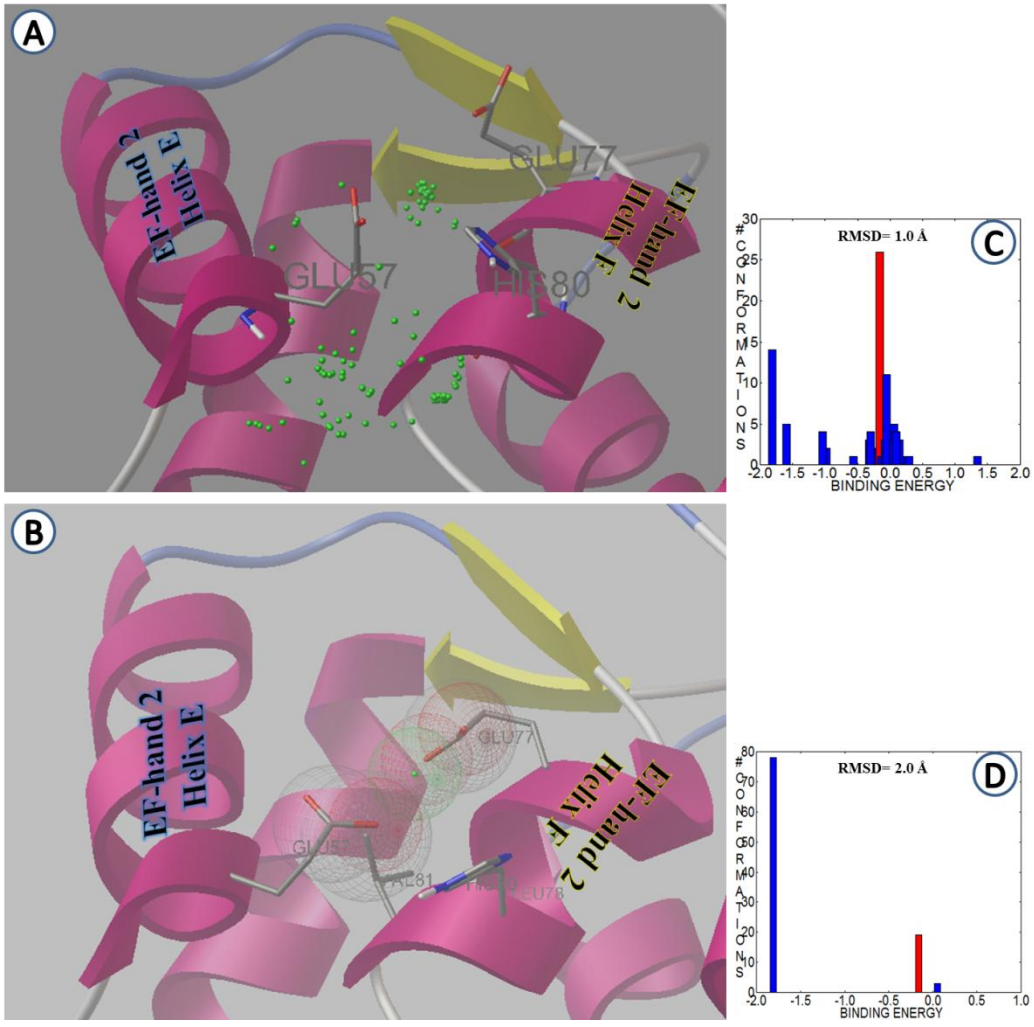


Figure 5.10: Results for Zn^{2+} docking in RepZ2: (A) Results from the 100 runs, zinc ions are represented by the green spheres. (B) The chosen location of Zn^{2+} in RepZ2. In this conformation, Zn^{2+} interacts with the side-chain N of His80 as well as the carboxyl side-chain oxygen atoms of Glu57 and Glu77. (C and D) Two histograms of the docked conformations were used with RMSD values of 2.0 Å and 1.0 Å, respectively. The chosen conformation belongs to the red bars in the histograms.

Figure 5.11 shows the results of Zn^{2+} docking into RepZ3. Some docked conformations involved an interaction with the carboxyl groups of Glu57 and Glu77 but most involved the interaction with the carboxyl of Glu57, the backbone carbonyl of His80 and its side-chain nitrogen, as shown in Figure 5.11 (B).

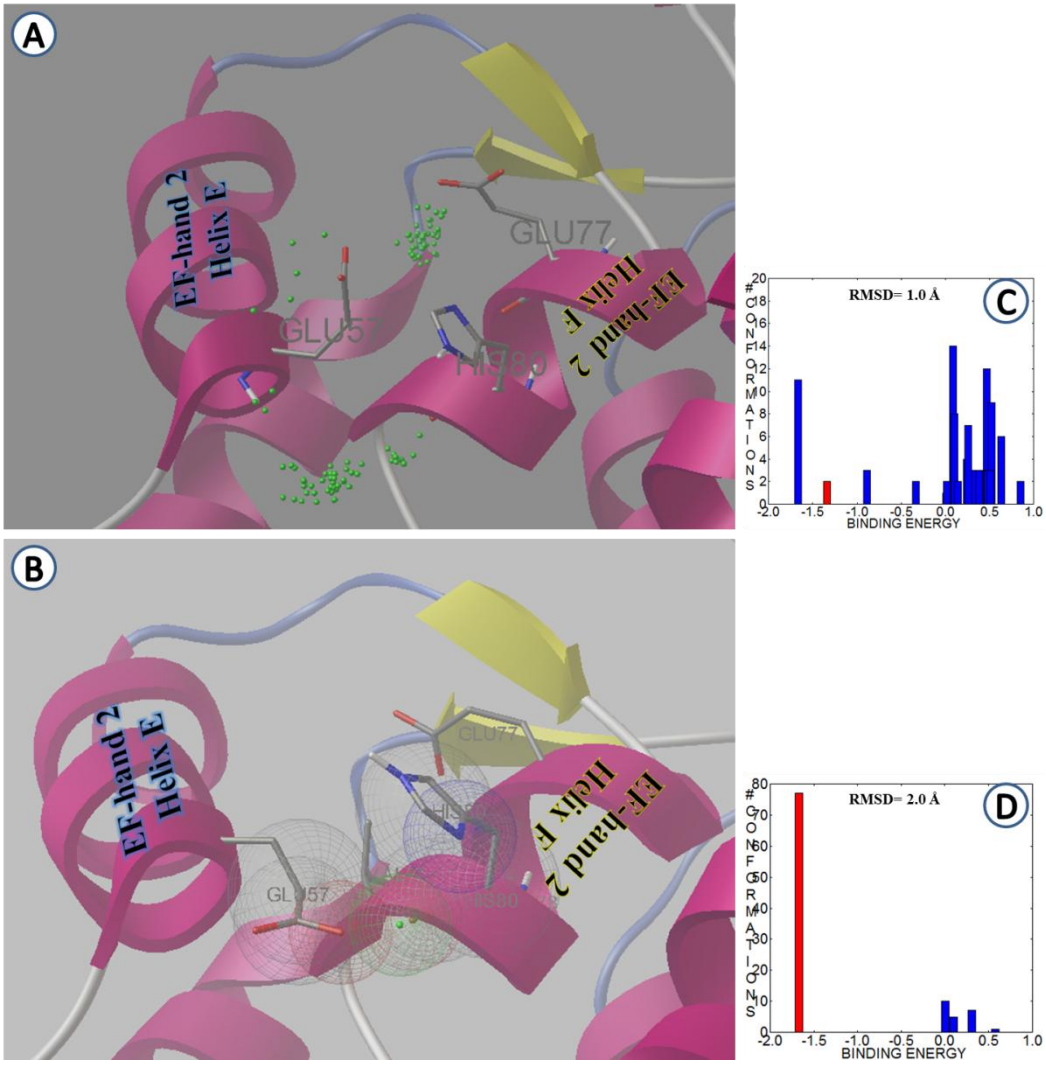


Figure 5.11: (A) Green spheres represent the 100 docked conformations predicted by Autodock 4.2. (B) The chosen Zn^{2+} position in RepZ3. In this conformation, Zn^{2+} interacts with the side-chain N of His80 and its backbone carbonyl group as well as the carboxyl side-chain oxygen atom of Glu57. (C and D) Two histograms of the docked conformations were used with RMSD values of 2.0 Å and 1.0 Å, respectively. The chosen conformation belongs to the red bars in the histograms.

Lastly, the results for the docking calculations performed on RepZ4, using Zn^{2+} as a ligand are shown in Figure 5.12. The majority of the predicted conformations were at the region in which Zn^{2+} interacted with the carboxyl side-chain of Glu77 and the side-chain nitrogen of His80. However, a few other conformations showed Zn^{2+} interacting with the two oxygen atoms of the carboxyl side-chain of Glu57, the side-chain nitrogen of His80, and its backbone carbonyl group. Since both types of conformations fulfilled the interaction distance criteria, they were both carried forward for the next series of calculations. These structures, shown in Figure 5.13, will be referred to as Zn-RepZ4a and Zn-RepZ4b respectively.

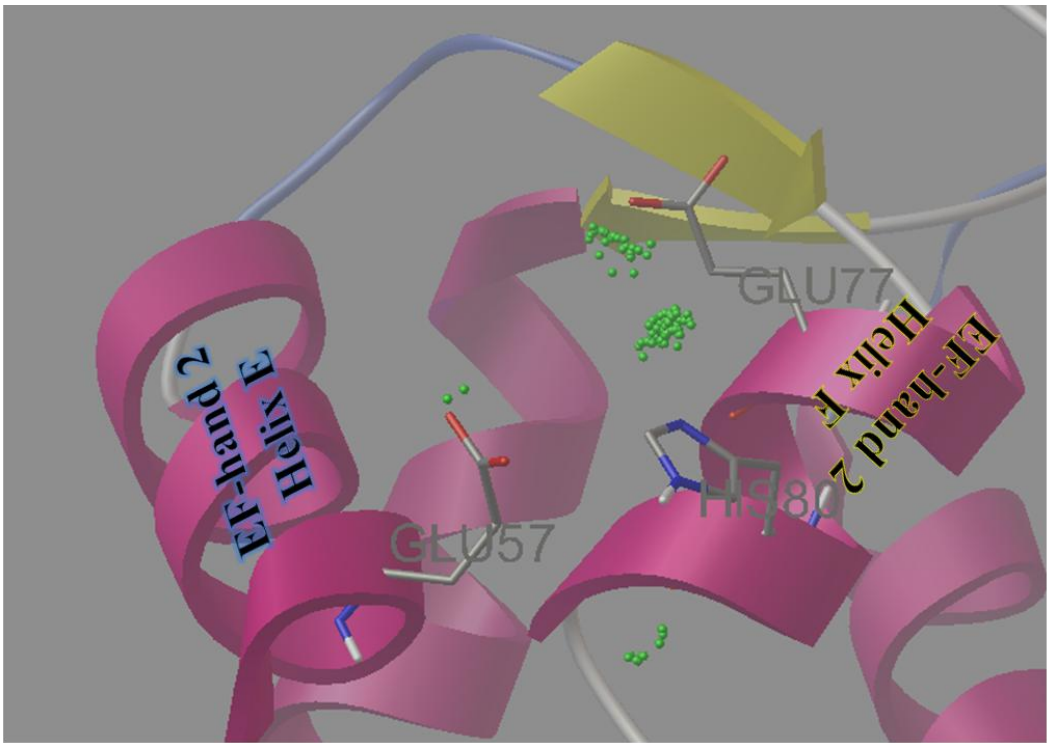


Figure 5.12: A snapshot of the 100 predicted Zn^{2+} positions in CB-D_{28k} near His80. While most conformations allow Zn^{2+} to interact with Glu77 and His80, some conformations in which Zn^{2+} interacts with Glu57 and His80 are also observed.

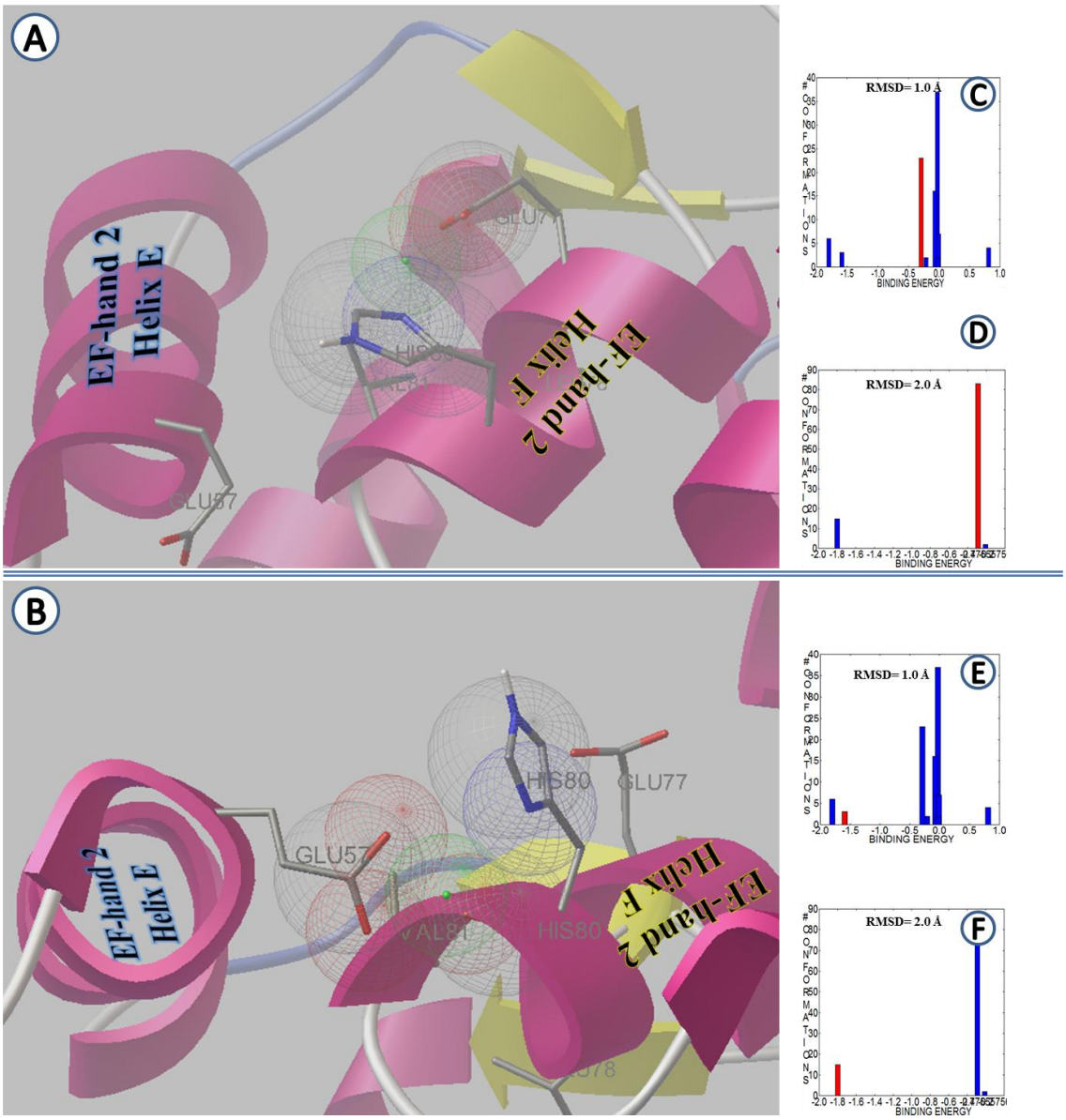


Figure 5.13: (A) The first chosen conformation from Zn^{2+} docking calculations in Rep4Z, showing Zn^{2+} interacting with Glu77 and His80. (D) The second chosen conformation of Zn^{2+} in RepZ4, showing Zn^{2+} interacting with Glu57 and His80. The docked conformations were clustered using RMSD values of 1.0 Å and 2.0 Å respectively. The first chosen conformation belongs to the red histograms in (C-D) while the second chosen conformation belongs to the less populated histograms shown in red in (E-F).

Table 5.3: A list of the distances between Zn^{2+} and the atoms in the residues with which it interacted in the conformations chosen above. Autodock reports distances to 4 significant digits.

	Zn-RepZ1	Zn-RepZ2	Zn-RepZ3	Zn-RepZ4a	Zn-RepZ4b
Residues Interacting with Zn^{2+}	**Glu57:** O^-= 2.016 O^-= 1.951	**Glu57:** O^-= 2.216	**Glu57:** O^-= 1.877	**Glu57:** O^-= 1.945 O^-= 1.984	
		Glu77: O^-= 2.023			**Glu77:** O^-= 1.944 O^-= 1.969
	His80: O= 1.860 N= 2.616	**His80:** *N= 5.705*	**His80:** O= 1.757 N= 2.914	**His80:** O= 1.738 N= 2.850	**His80:** N= 2.210

5.2.5 Discussion of Zn^{2+} Docking into Apo-CB-D_{28k} Results

The docking calculations above were the last step in Stage I, as outlined in the flowchart, Figure 5.5. The purpose of these calculations was to introduce Zn^{2+} into the apo form of CB-D_{28k}. The choice of the best-docked conformation of Zn^{2+} in CB-D_{28k} is very difficult since there is no evidence or guidance for its exact binding residues, except for His80. However, a statistical study that involved the crystal structures of 994 zinc-complexes reported the ranges for the interaction distances between zinc and its different coordinating residues in metalloproteins (76). According to this study, the most common Zn-Glu (O^-) interaction distances lie within the range of 1.94 Å to 3.36 Å. The average

distance between Zn^{2+} and the side-chain nitrogen of histidine was found to be 2.09 Å ±0.14, while the average distance between other types of oxygen atoms (for example carbonyls) and zinc ions was 2.19 Å ±0.24. The docked conformations that fulfilled, or nearly fulfilled, these interaction ranges (Table 5.3), were chosen as the poses for the next calculations (in Stage II, Section 5.2.6) and are discussed in more detail below.

RepZ1: The chosen conformation for this docking calculation had an interaction distance between the backbone carbonyl of His80 and Zn^{2+} of 1.860 Å, compared to a minimum accepted value of 1.95 Å. In addition, the side-chain nitrogen of His80 was 0.39 Å off the maximum allowed interaction distance. The interaction distances between Zn^{2+} and the two oxygen atoms of Glu57, 2.016 Å and 1.951 Å, were both within the acceptable Zn-O^- distance range of 1.94 Å to 3.36 Å.

RepZ2: None of the docked conformations formed interactions with Zn^{2+} within the distance criteria. In fact, the best-docked conformation had only three interacting sites. In addition, His80 was 5.705 Å away from Zn^{2+}; this is almost 3.5 Å off the maximum allowed Zn-N (His) distance. The distances of 2.216 Å and 2.023 Å between the two oxygen atoms of Glu57 and Glu77, respectively, were both within range.

RepZ3: None of the interaction distances in the calculated conformations met the criteria for interaction distances for this representative structure. Given this situation, the closest docked structure was chosen from the docking results. In this conformation, only three interactions were made between Zn^{2+} and the protein residues. The first involved the oxygen of Glu57 carboxyl group, which was less than 0.1 Å off the minimum accepted value. The other two interacting residues were the carbonyl of His80 and its side-chain nitrogen atom. These were about 0.07 Å and 0.68 Å off the accepted range.

Since, the differences of 0.1 Å and 0.07 Å were extremely small, they were considered successful, making this an acceptable candidate.

RepZ4: As shown in the results section, two conformations were chosen from the docking calculations on that representative structure. These were especially interesting since they involved different interactions for the same structure. In RepZ4a, Glu57 carboxyl interaction distances were both within the accepted range, 1.945 Å and 1.984 Å versus an acceptable range of 1.94 Å to 3.36 Å. The two interactions with His80, however, were slightly off the accepted range; the carbonyl oxygen was 0.21 Å off and the side-chain nitrogen was 0.62 Å off. RepZ4b had interaction distances that were all within range, but formed only three interactions instead of four, one with His80 and the other two with the carboxyl of Glu77 rather than Glu57.

Note that it is expected that none of the conformations would meet all the criteria, since the docking calculations were performed on the pseudo-version of apo-CB-D_{28k} and not the true zinc-bound form of the protein. In fact, all these docked structures show some promise as the starting point for getting the zinc-bound form of the protein. At this stage, five structures of apo-CB-D_{28k} with Zn^{2+} bound to it had been obtained. These are referred to as Zn-RepZ1, Zn-RepZ2, Zn-RepZ3, Zn-RepZ4a and Zn-RepZ4b for the rest of the thesis.

5.2.6 Stage II MD Simulations and Analysis

The five structures obtained from the docking calculations in Stage I were solvated and neutralised in a water box using VMD. NAMD was used to relax them for 50 ns. The change in the protein positions for all the structures was calculated over time using the

PTRAJ utility. These are shown on the next page in Figure 5.14. It can be seen from the plots that the protein structure fluctuated significantly during the simulation. However, over the last 2500 frames, all the structures showed deviation of less than 1.0 Å, with the exception of Zn-RepZ3. To keep the number of structures manageable, the representative structure for the last 2.5 ns of each of the five simulations was obtained by PTRAJ. These five structures were used for Stage III calculations.

Relaxation of the protein structure with Zn^{2+} bound to it would change the conformation of the protein. However, the structure obtained would not be of the fully zinc-bound form of the protein, since it was reported to bind at least three ions, while the structures used at this stage involve only one Zn^{2+} in the most likely first site. Nevertheless, it would be useful to test whether the structure of relaxed CB-D_{28k} with one Zn^{2+} bound to it would still bind Ca^{2+}.

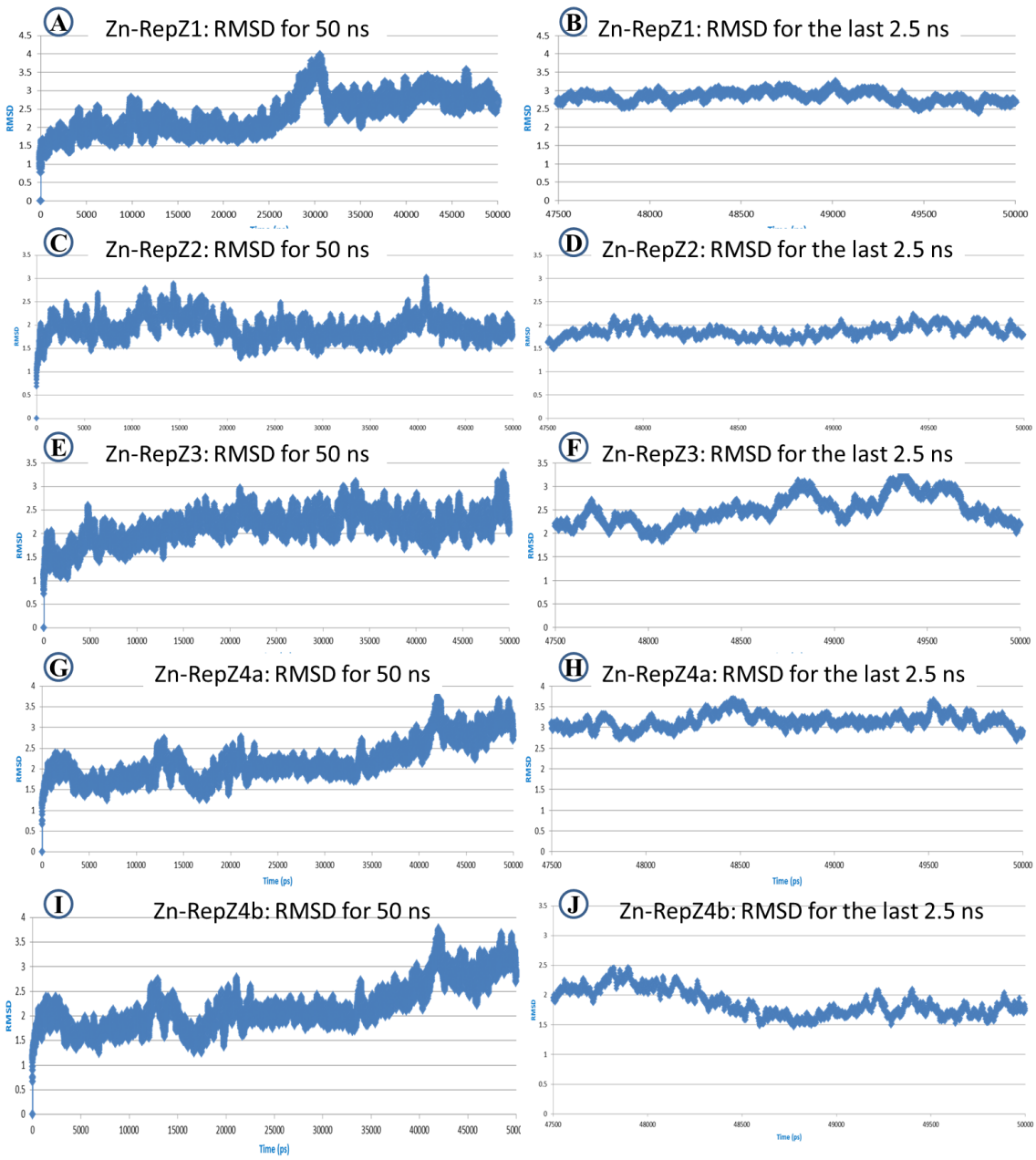


Figure 5.14: The plots of RMSD versus time for the five structures obtained from docking. Plots A, C, E, G and I show the RMSD for the protein over the while simulation period. Plots B, D, F, H and J show the RMSD of the protein over the last 2.5 ns (47,500 ps to 50,000 ps).

5.3 The Zinc- Calcium Conundrum

Due to the absence of any experimental structures for Zn^{2+}-CB-D_{28k}, the computational methods employed in this project were used to create a pseudo version for the next calculation. The relaxed protein structures from the MD simulation in Section 5.2.6 will be referred to a 1Zn-RepZ# complex (# =1, 2, 3, 4a, 4b). Since EF-hand 1 is the first to bind Ca^{2+}, docking calculations were performed on this hand to test whether its calcium binding ability had been affected by the change in structure due to binding of Zn^{2+} to the protein. This constitutes the third and last stage to the second goal of the project as outlined in Figure 5.15.

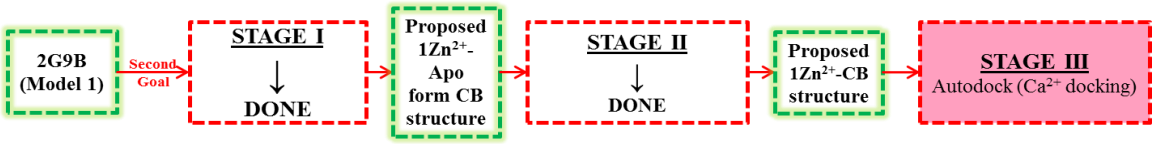


Figure 5.15: The workflow for the second goal of the project. Stages I and II have been accomplished in Section 5.2. This section involves calculations at Stage III.

Autodock 4.2 was used to predict the possible Ca^{2+} binding positions in EF-hand. The docking calculations were set to run 100 times each in which 50,000,000 energy evaluations were performed on 28,000 generations each containing 300 individuals. The docking results are shown and discussed in the following sub-sections.

5.3.1 Ca^{2+} Docking in 1Zn-CB-D_{28k} Complex Results

The docking results on 1Zn-RepZ1 complex are shown in Figure 5.16. Most of the docked conformations were localised at the first half of the loop, as shown in (A). At this region of the loop, Ca^{2+} interacts with the carboxyl groups of Asp24 and Asp26 as shown in (B). The red histograms in (C) and (D) show that this was the most frequently predicted Ca^{2+} position in the hand. In a few other conformations, Ca^{2+} interacted with the carboxyl groups of Asp24 and Glu35 as well as the Phe23 backbone carbonyl group.

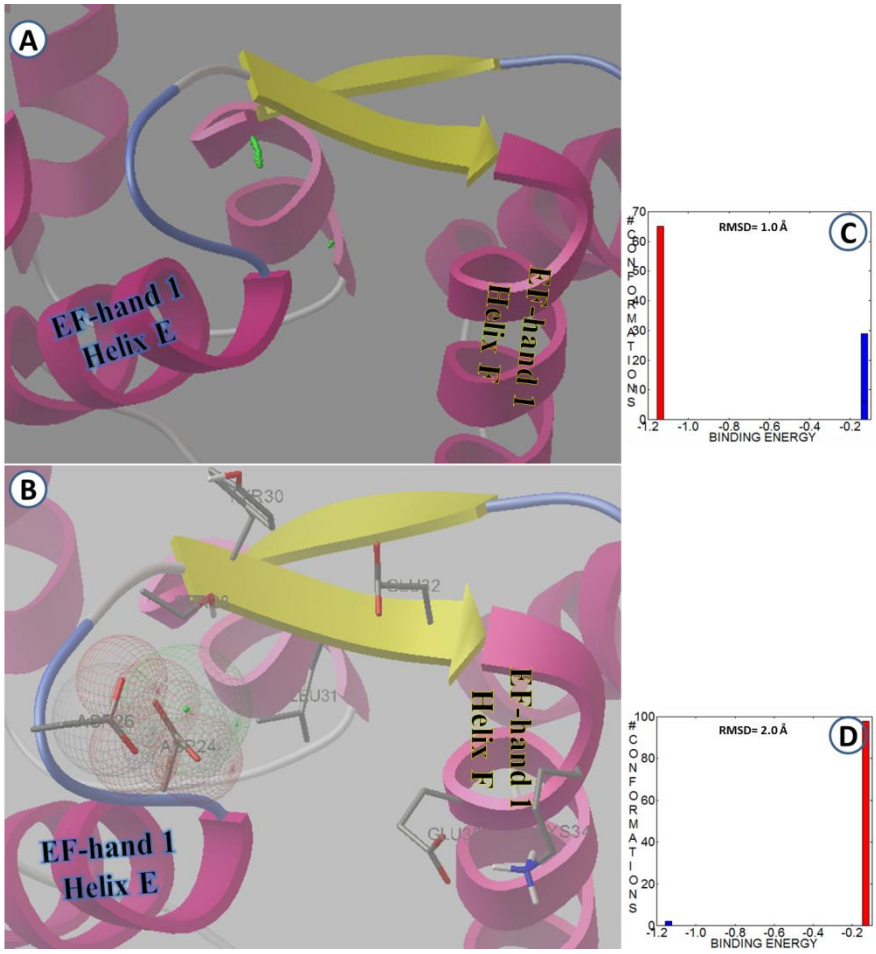


Figure 5.16: (A) The predicted locations of Ca^{2+} (represented by the green spheres) in the EF-hand 1 of 1Zn-RepZ1 complex. **(B)** The most common docked conformation. **(C-D)** Histograms of docked conformations clustered at RMSD values of 1.0 Å and 2.0 Å respectively.

Figure 5.17 shows the docked conformations of Ca^{2+} into EF-hand 1 of 1Zn-RepZ2 complex. The docked conformations are more widely distributed in this structure, and Ca^{2+} interacts with only two residues in each location. The most common conformations involved Ca^{2+} interaction with the carboxyl groups of Asp26 and Glu32, shown in (B). In the second most common conformation, Ca^{2+} interacted with the carboxyl groups of Asp24 and Glu35.

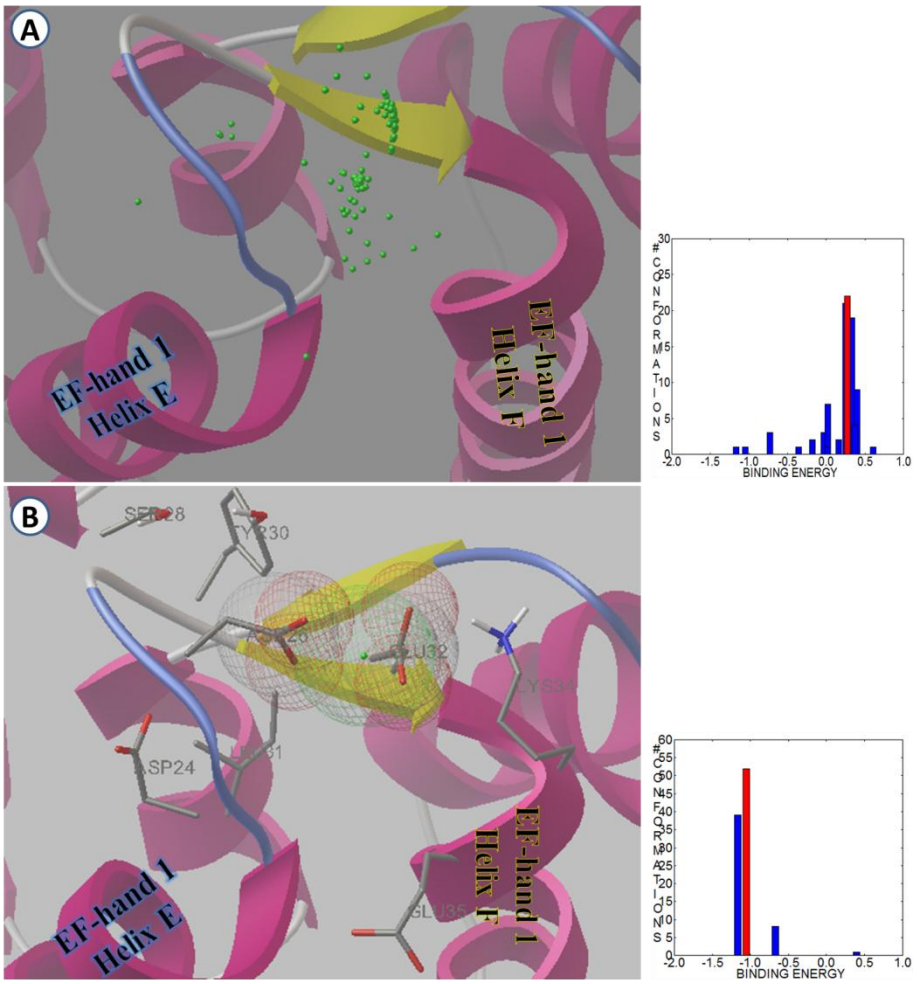


Figure 5.17: (A) Predicted locations of Ca^{2+} (represented by the green spheres) in the EF-hand 1 of 1Zn-RepZ2 complex. **(B)** The most common docked conformation. **(C-D)** Histograms of docked conformations clustered at RMSD values of 1.0 Å and 2.0 Å respectively.

The Ca^{2+} positions in the EF-hand of the third Zn^{2+}-bound model, 1Zn-RepZ3 complex, are shown in Figure 5.18. The image (A) shows that Ca^{2+} was mostly centered in the loop, but in-depth analysis of the docked conformations revealed that Ca^{2+} interacted with the carboxyl groups of Asp26, Glu32 and Glu35 only, as shown in (B). In one conformation Ca^{2+} was interacting with the carboxyl of Asp24 and Asp26 as well as the backbone carbonyl of Tyr30, rather than the former set.

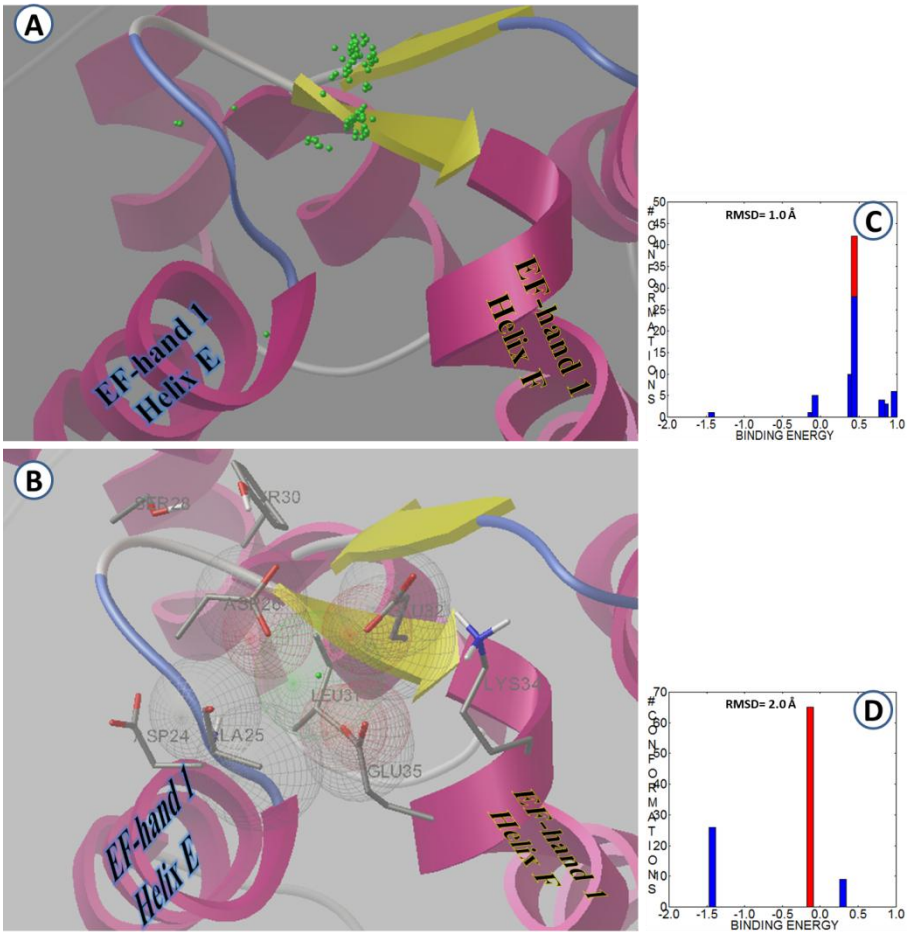


Figure 5.18: (A) 1Zn-RepZ3 complex with the predicted Ca^{2+} positions (represented by the green spheres) in the EF-hand 1. **(B)** The most common docked conformation. **(C-D)** Histograms of docked conformations clustered at RMSD values of 1.0 Å and 2.0 Å respectively.

The docking results for EF-hand 1 of the 1Zn-RepZ4a complex are shown in Figure 5.19. While a few conformations showed that Ca^{2+} interacted with the carboxyl groups of Asp24 and Glu35, most conformations involved the interaction between Ca^{2+} and the carboxyl groups of Asp24 and Asp26 as well as the hydroxyl of Ser28. This conformation is shown in Figure 5.19 (B).

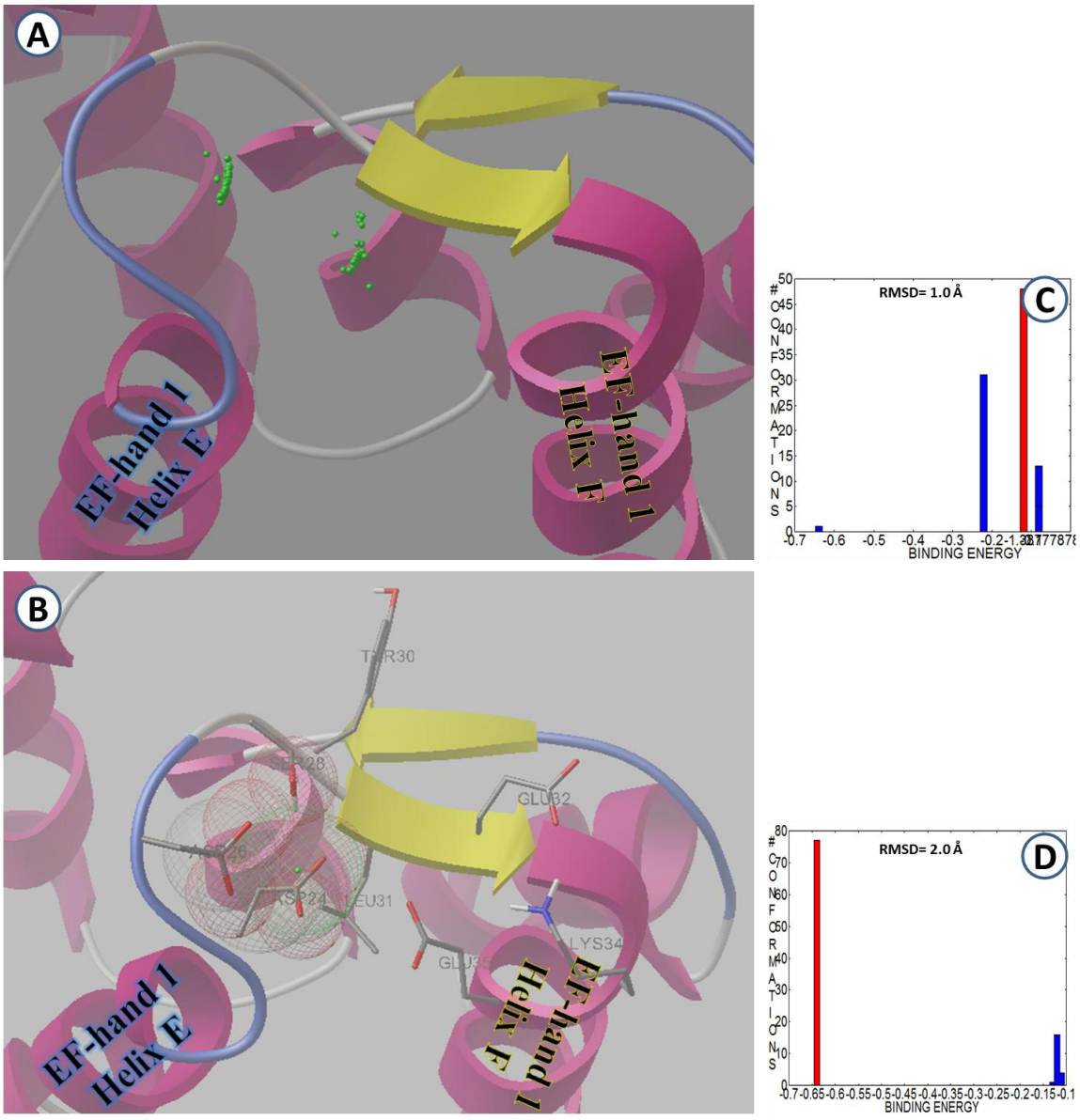


Figure 5.19: Results for Ca^{2+} docking in EF-hand 1 of 1Zn-RepZ4a complex. See previous caption description for full details.

The results for the last docking calculations, which were performed on 1Zn-RepZ4b complex, are shown in Figure 5.20 below. The most frequent Ca^{2+} location was near the Asp24 end of the loop. At that position, Ca^{2+} interacted with the carboxyl groups of Asp24 and Asp26. In only one conformation, an additional interaction with the hydroxyl group of Ser28 was also formed.

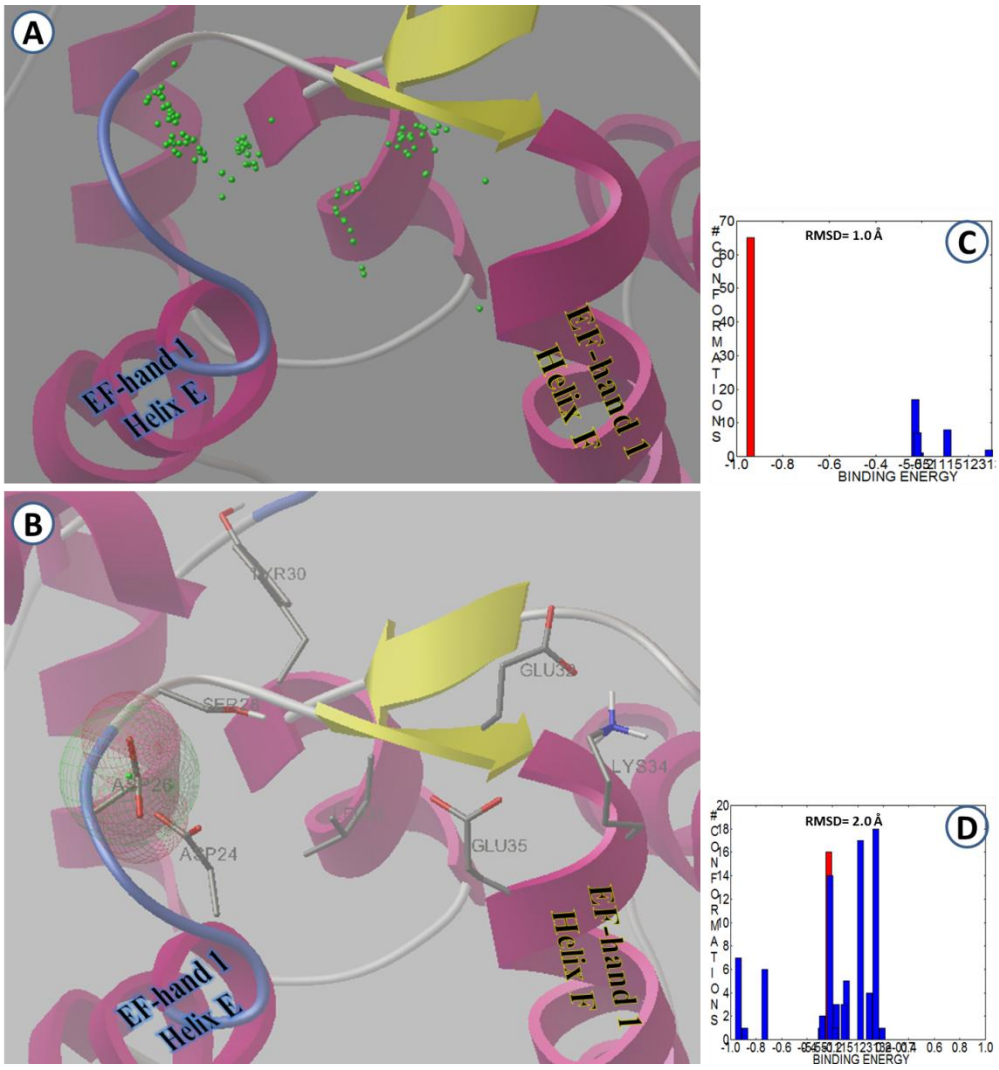


Figure 5.20: Results of Ca^{2+} docking calculations in EF-hand 1 of 1Zn-RepZ4b complex. See Figure 5.19 caption description for full details.

5.3.2 Discussion of Ca^{2+} Docking in 1Zn-CB Complex Results

The structures of CB-D_{28k}, containing one bound Zn^{2+}, were tested for their Ca^{2+} binding ability using Autodock 4.2. Since CB-D_{28k} normally binds the first Ca^{2+} at EF-hand 1, docking calculations were performed on this hand with the results as shown in Section 5.3.1, and discussed here.

1Zn-RepZ1 complex: Most of the docked conformations had Ca^{2+} interacting with only two canonical residues of the loop, Asp24 and Asp26. A third interaction site was detected in only a few conformations, at Ser28. Other conformations with three interacting residues involved Asp24, Glu35 and the backbone carbonyl of Phe23, which is an incorrect interaction, that is, one that is not part of the canonical binding residue set.

1Zn-RepZ2 complex: Most of the predicted conformations for this structure involved only two of the six interacting residues, those being either Asp26 and Glu32 or Asp24 and Glu35, in the typical canonical binding. In addition, a few conformations showed the docked Ca^{2+} located entirely outside the loop of the EF-hand (Figure 5.16).

1Zn-RepZ3 complex: The docking results for most of the predicted conformations showed that Ca^{2+} interacted with three (Asp26, Glu32 and Glu35) of the six canonical residues. The conformation displayed in Figure 5.18 (B) shows a Ca^{2+} that is the most centered of docked conformations for this representative structure. However, Asp24 is no longer one of the interacting residues (position 1 in the loop). Note that this is the one position in a canonical EF-hand protein that has always been conserved as an aspartate residue (30).

1Zn-Rep4a complex: The docked locations of Ca^{2+} for this structure were divided between two main regions. The region closer to the first half of the loop was more frequently populated; here, Ca^{2+} interacted with Asp24, Asp26 and Ser28. The more centralised Ca^{2+} showed interactions only with residues Asp24 and Glu35. Thus, none of the guessed conformations involved more than three correct interacting residues.

1Zn-Rep4b complex: The most common interacting residue was Asp24 only. Some conformations, however, showed an additional interaction with Asp26 and one showed an interaction with Ser28.

It is evident from the results above that the binding of one zinc ion to this approximate structure of CB-D_{28k} has decreased the Ca^{2+} binding affinity to EF-hand 1 significantly. This is supported by the fact that fewer canonical residues interact with Ca^{2+} in the new protein form.

5.4 Comparisons and Conclusions

The results of the Ca^{2+} dockings above show that the binding of Zn^{2+} to CB-D_{28k} renders the protein incapable of proper canonical binding to Ca^{2+}. It can be deduced from those calculations that both the canonical amino acid sequence as well as the loop geometry are important for proper canonical binding of Ca^{2+} to CB-D_{28k}. As discussed above, the docking results show poor Ca^{2+} interactions with the loop residues compared to the typical canonical binding, especially in the 1Zn-RepZ3 complex. In fact, even the 1Zn-RepZ4a complex, which had the best-docked conformation, showed only three interactions with the loop residues instead of six.

A comparison was made between the docked Ca^{2+} conformation in the holo form of CB-D_{28k} (Section 4.2.3) and in the 1Zn-RepZ4a complex as shown in Figure 5.21 (A and B) and (C and D), respectively. This comparison illustrates the best objective reasoning for the decreased capacity of the protein to bind Ca^{2+} when Zn^{2+} is already bound to it.

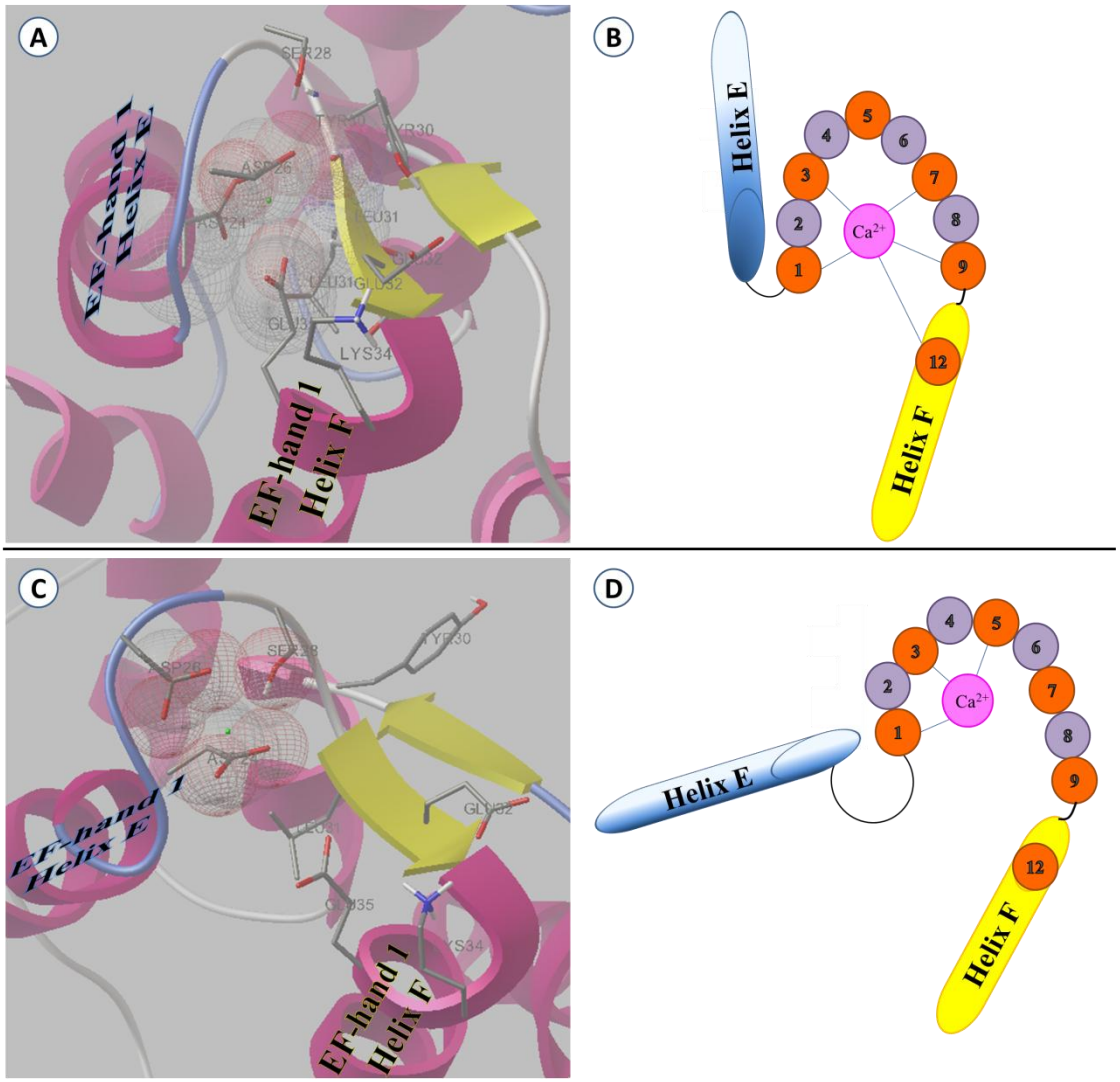


Figure 5.21: (A) The best binding pose for Ca^{2+} in RepC5 obtained in 4.2.3. **(B)** A schematic of the docked conformation in figure A. **(C)** The best binding pose for Ca^{2+} in 1Zn-RepZ4a complex obtained in 5.3.1. **(D)** A schematic of the docked conformation in figure (D). Note: for (B) and (D), positions in orange are the correct canonical sites and the blue sticks represent interactions between Ca^{2+} and the respective residue.

The most obvious difference is in the number of correct interactions in each docked pose. In Figure 5.21 (A), Ca^{2+} interacts with all the canonical loop residues except at position 5, while (C) shows that Ca^{2+} interacts with only the residues at positions 1, 3 and 5. On the examination of the hands above, it can also be seen that the canonical residues are far away from each other in Figure 5.21 (C). Specifically, residues Ser28 and Glu35 are spatially distant in the loop that even if fully extended, these residues would not be able to both interact with Ca^{2+} simultaneously.

The structures of the loops are also quite different as expressed in Figure 5.21 (B) and (D) above; in (D), the loop forms a wider U-shape compared to the loop in holo CB-D_{28k} shown in (B). In fact, the structure of the hand is distorted as helices lose their somewhat perpendicular angularity. In addition, the loop of EF-hand 1 in 1Zn-RepZ4a complex starts at residue His22 and ends at Glu32, instead of starting at Asp26 and ending at Glu32. These structural variations all provide support to the hypothesis that the binding of Zn^{2+} to CB-D_{28k} interferes with its Ca^{2+} binding capability. These findings imply that increased Zn^{2+} levels in cells could decrease the calcium buffering capability of the protein. A decrease in Ca^{2+} buffering inside the cells could cause an increase in the intracellular levels of Ca^{2+}, which would contribute to the altered Ca^{2+} homeostasis in AD. Thus, it could play a role in the calcium hypothesis for the development of the pathology of AD (1, 10).

Chapter 6

Future Work

The main goals of this project have been achieved as explained in Chapters 4 and 5. As with any research, several factors could be tested further to provide more confidence in the results obtained. In this chapter, future work specific to the calculations in this project is presented as well as other possible work on the protein.

6.1 Project Specific Future Work

A possible enhancement to the first goal results could be to run MD calculations on the holo protein obtained in section 4.2.3 for a second round. Since the quality of computational modeling results generally depend on the starting structures, a second round of calculations could provide an even better structure of the protein. In addition, docking calculations could be run on software that is better parameterised for electrostatic interactions so that comparisons in binding energies could be made. This would provide insight to the binding strength of both cations to the protein.

An important variable in the calculations for the second goal were the durations of the MD simulations. Longer relaxation calculations on the Zn-CB_{28k} structure, obtained in section 5.2.4, could be run and the structures obtained would be compared to the ones in this project; structural similarity would provide support that the protein was relaxed long enough. In addition, the relaxed 1Zn-CB-D_{28k} structures could be examined for the other possible Zn^{2+} binding sites. Other software that supports blind docking (in which the region of binding is unknown), could also be used to aid in the search for the other Zn^{2+} binding sites. Another, more difficult, enhancement would be the re-parameterisation of Autodock 4.2 for CDM.

6.2 General Future Work

The results from the Zn^{2+}-Ca^{2+} conundrum show important findings with respect to the calcium binding capability of the protein when bound to Zn^{2+}. These findings should motivate an experimental group to test the effect of Zn^{2+} on the Ca^{2+} binding capability of CB-D_{28k}. Further computational modeling of this possible competition between the cations would provide more structural insight into the nature of this competition. Evidence from this modeling could inspire the design of pharmaceutical interventions to stop or reduce the Zn^{2+} interference with Ca^{2+} binding in CB-D_{28k}, thus reducing the possible consequence of elevated Zn^{2+} levels on the Ca^{2+} buffering property of CB-D_{28k}. Note that this would be a challenging task since the balance of Ca^{2+} and Zn^{2+} must be delicately maintained within their physiological limits.

A very important aspect of the protein is its function as a Ca^{2+} sensor protein, in specific, its anti-apoptotic ability through its binding to the apoptotic protein caspase-3. A model of the interaction mechanism between CB-D_{28k} and caspase-3 has recently been proposed (77). Success in modeling this interaction would be a very important step towards understanding the paradox of the role of CB-D_{28k} in AD (neuronal loss) and cancer (uncontrolled cell division) (77).

The function of CB-D_{28k} as a sensor when Zn^{2+} is bound to it should also be modelled. If Zn^{2+} does bind to the protein in AD and inhibits it from interacting with caspase-3, then the damage caused by Zn^{2+} in AD would be two fold; it would contribute to altered Ca^{2+} homeostasis as well as promote neuronal apoptosis.

Bibliography

References

1. Berridge MJ. Calcium hypothesis of alzheimer's disease. Signalling and cell physiology. 2009(459):441-9.

2. Prince M, Bryce R, Ferri C. The benefits of early diagnosis and intervention. Institute of Psychiatry, King's College London: Alzheimer's Disease International, World Alzheimer; 2011.

3. Brookmeyer R, Johnson E, Ziegler-Graham K, Arrighi HM. Forecasting the global burden of alzheimer's disease. Alzheimers & Dementia. 2007 JUL 2007;3(3).

4. Gonzalez-Dominguez R, Garcia-Barrera T, Gomez-Ariza J. Metabolomic approach to alzheimer's disease diagnosis based on mass spectrometry. Chemical Papers. 2012 SEP 2012;66(9).

5. Alzheimers Assoc. 2012 alzheimer's disease facts and figures. Alzheimers & Dementia. 2012 MAR 2012;8(2).

6. Odero GL, Oikawa K, Glazner KAC, Schapansky J, Grossman D, Thiessen JD, et al. Evidence for the involvement of calbindin D28k in the presenilin 1 model of alzheimer's disease. Neuroscience. 2010 8/11;169(1):532-43.

7. Wang Jian-Zhi, Tian Qing. Molecular mechanisms underlie alzheimer-like tau hyperphosphorylation and neurodegeneration. Progress in Biochemistry and Biophysics. 2012 AUG 2012;39(8).

8. Colletier J, Laganowsky A, Landau M, Zhao M, Soriaga AB, Goldschmidt L, et al. Molecular basis for amyloid-beta polymorphism. Proc Natl Acad Sci U S A. 2011 OCT 11 2011;108(41).

9. Permi͡akov, E. A. (Evgeniĭ Anatol′evich). Calcium binding proteins. 2011.

10. Berridge MJ. Calcium signalling and alzheimer's disease. Neurochem Res. 2011;36(7):1149-56.

11. Supnet C, Bezprozvanny I. The dysregulation of intracellular calcium in alzheimer disease. Cell Calcium. 2010 FEB 2010;47(2).

12. Religa D, Strozyk D, Cherny RA, Volitakis I, Haroutunian V, Winblad B, et al. Elevated cortical zinc in alzheimer disease. Neurology. 2006 JUL 11 2006;67(1).

13. Palop JJ, Jones B, Kekonius L, Chin J, Yu GQ, Raber J, et al. Neuronal depletion of calcium-dependent proteins in the dentate gyrus is tightly linked to alzheimer's disease-related cognitive deficits. Proc Natl Acad Sci U S A. 2003 AUG 5 2003;100(16).

14. Palop JJ, Mucke L, Roberson ED. Quantifying biomarkers of cognitive dysfunction and neuronal network hyperexcitability in mouse models of alzheimer's disease: Depletion of calcium-dependent proteins and inhibitory hippocampal remodeling. Alzheimer's Disease and Frontotemporal Dementia: Methods and Protocols. 2011 2011;670.

15. Greene JRT, Radenahmad N, Wilcock GK, Neal JW, Pearson RCA. Accumulation of calbindin in cortical pyramidal cells with ageing; a putative protective mechanism which fails in alzheimer's disease. Neuropathol Appl Neurobiol. 2001 OCT 2001;27(5).

16. Iacopino AM, Christakos S. Specific reduction of calcium-binding protein (28-kilodalton calbindin-D) gene-expression in aging and neurodegenerative diseases. Proc Natl Acad Sci U S A. 1990 JUN 1990;87(11).

17. Riascos D, de Leon D, Baker-Nigh A, Nicholas A, Yukhananov R, Bu J, et al. Age-related loss of calcium buffering and selective neuronal vulnerability in alzheimer's disease. Acta Neuropathol. 2011;122(5):565-76.

18. Bauer MC, Nilsson H, Thulin E, Frohm B, Malm J, Linse S. Zn2+ binding to human calbindin D28k and the role of histidine residues. Protein Sci. 2008;17(4):760-7.

19. Lutz W, Frank EM, Craig TA, Thompson R, Venters RA, Kojetin D, et al. Calbindin D28K interacts with ran-binding protein M: Identification of interacting domains by NMR spectroscopy. Biochem Biophys Res Commun. 2003 April 18, 2003;303(4):1186-92.

20. Wassermann RH, Corradino RA, Taylor AN. Vitamin D-dependent calcium-binding protein - purification and some properties. J Biol Chem. 1968 1968;243(14).

21. Kojetin DJ, Venters RA, Kordys DR, Thompson RJ, Kumar R, Cavanagh J. Structure, binding interface and hydrophobic transitions of Ca2+-loaded calbindin-D-28K. Nature Structural & Molecular Biology. 2006 JUL 2006;13(7):641-7.

22. Kordys DR, Bobay BG, Thompson RJ, Venters RA, Cavanagh J. Peptide binding proclivities of calcium loaded calbindin-D28k. FEBS Lett. 2007 OCT 2 2007;581(24):4778-82.

23. Jande SS, Maler L, Lawson DEM. Immunohistochemical mapping of vitamin-D-dependent calcium-binding protein in brain. Nature. 1981 1981;294(5843).

24. Baimbridge KG, Miller JJ, Parkes CO. Calcium-binding protein distribution in the rat-brain. Brain Res. 1982 1982;239(2).

25. German DC, Ng MC, Liang CL, McMahon A, Iacopino AM. Calbindin-D-28k in nerve cell nuclei. Neuroscience. 1997 DEC 1997;81(3).

26. Gross M, Kumar R. Physiology and biochemistry of vitamin D-dependent calcium-binding proteins. Am J Physiol. 1990 AUG 1990;259(2).

27. Bu J, Sathyendra V, Nagykery N, Geula C. Age-related changes in calbindin-D-28k, calretinin, and parvalbumin-immunoreactive neurons in the human cerebral cortex. Exp Neurol. 2003 JUL 2003;182(1).

28. Schwaller B. Cytosolic Ca2+ buffers. Cold Spring Harbor perspectives in biology. 2010 2010-Nov;2(11).

29. Kretsing.Rh, Nockolds CE. Carp muscle calcium-binding protein .2. structure determination and general description. J Biol Chem. 1973 1973;248(9).

30. Gifford JL, Walsh MP, Vogel HJ. Structures and metal-ion-binding properties of the Ca2+-binding helix-loop-helix EF-hand motifs. Biochem J. 2007 JUL 15 2007;405.

31. Hobbs CA, Deterding LJ, Perera L, Bobay BG, Thompson RJ, Darden TA, et al. Structural characterization of the conformational change in calbindin-D28k upon calcium binding using differential surface modification analyzed by mass spectrometry. Biochemistry. 2009;48(36):8603-14.

32. Laskowski RA. Structural quality assurance. Methods Biochem Anal. 2003 2003;44.

33. Venters RA, Benson LM, Craig TA, Bagu J, Paul KH, Kordys DR, et al. The effects of Ca2+ binding on the conformation of calbindin D28K: A nuclear magnetic resonance and microelectrospray mass spectrometry study. Anal Biochem. 2003 November 15, 2003;322(2):292.

34. Ikura M. Calcium binding and conformational response in EF-hand proteins. Trends Biochem Sci. 1996 JAN 1996;21(1).

35. Hehre WJ. A guide to molecular mechanics and quantum chemical calculations. Wavefunction; 2003.

36. Levine IN. Quantum chemistry. Sixth ed. Pearson Prentice Hall; 2009.

37. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. Journal of Computational Chemistry. 2009 DEC 30 2009;30(16).

38. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, et al. Scalable molecular dynamics with NAMD. Journal of Computational Chemistry. 2005;26:1781-802.

39. Www.ks.uiuc.edu/research/namd/ [Internet].

40. Case DA, Darden TA, Cheatham I,T.E., Simmerling CL, Wang J, Duke RE, et al. AMBER 12. University of California, San Francisco: ; 2012.

41. Sousa SF, Fernandes PA, Ramos MJ. Protein-ligand docking: Current status and future challenges. Proteins-Structure Function and Bioinformatics. 2006 OCT 1 2006;65(1).

42. Goodsell DS, Morris GM, Olson AJ. Automated docking of flexible ligands: Applications of AutoDock. Journal of Molecular Recognition. 1996 JAN-FEB 1996;9(1).

43. Morris GM, Goodsell DS, Pique M, E., Lindstrom W, Huey R, Forli S, et al. AutoDock version 4.2: Automated docking of flexible ligands to flexible receptors; 2010.

44. Sanner MF. Python: A programming language for software integration and development. Journal of Molecular Graphics and Modelling. 1999;17(February):57-61.

45. Gasteiger J, Marsili M. Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. Tetrahedron. 1980;36(22):3219-28.

46. Http://Www.rcsb.org/pdb/home/home.do [Internet].

47. Morris GM, Goodsell DS, Huey R, Olson AJ. Distributed automated docking of flexible ligands to proteins: Parallel applications of AutoDock 2.4. J Comput Aided Mol Des. 1996 AUG 1996;10(4).

48. Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, Alagona G, et al. A new force-field for molecular mechanical simulation of nucleic-acids and proteins. J Am Chem Soc. 1984 1984;106(3).

49. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, et al. Automated docking using a lamarckian genetic algorithm and an empirical binding free energy function. Journal of Computational Chemistry. 1998 NOV 15 1998;19(14).

50. Solis FJ, Wets RJB. Minimization by random search techniques. Mathematics of Operations Research. 1981 1981;6(1).

51. Huey R, Morris GM, Olson AJ, Goodsell DS. A semiempirical free energy force field with charge-based desolvation. Journal of Computational Chemistry. 2007;28(6):1145-52.

52. Mehler EL, Solmajer T. Electrostatic effects in proteins: Comparison of dielectric and charge models. Protein Engineering. 1991 December 01;4(8):903-10.

53. Huey R, Goodsell DS, Morris GM, Olson AJ. Grid-based hydrogen bond potentials with improved directionality. Lett Drug Des Discovery. 2004;1(2):178-83.

54. Stouten PFW, Frommel C, Nakamura H, Sander C. An effective solvation term based on atomic occupancies for use in protein simulations. Molecular Simulation. 1993 1993;10(2-6).

55. Humphrey W, Dalke A, Schulten K. VMD - visual molecular dynamics. Journal of Molecular Graphics. 1996;14:33-8.

56. Www.ks.uiuc.edu/research/vmd/ [Internet].

57. Http://Www.ks.uiuc.edu/training/tutorials/#namd [Internet].

58. Http://Www.ks.uiuc.edu/training/tutorials/#vmd [Internet].

59. Shao J, Tanner SW, Thompson N, Cheatham TE,III. Clustering molecular dynamics trajectories: 1. characterizing the performance of different clustering algorithms. Journal of Chemical Theory and Computation. 2007 NOV-DEC 2007;3(6).

60. Lepsik M, Field MJ. Binding of calcium and other metal ions to the EF-hand loops of calmodulin studied by quantum chemical calculations and molecular dynamics simulations. J Phys Chem B. 2007 AUG 23 2007;111(33).

61. Ren Y, Gao J, Xu J, Ge W, Li J. Key factors in chaperonin-assisted protein folding. Particuology. 2012 FEB 2012;10(1).

62. Dudev T, Lim C. - Principles governing mg, ca, and zn binding and selectivity in proteins. - Chem Rev. 2003(- 3):- 773.

63. Www.westgrid.ca [Internet].

64. 10.2210/pdb1b1g/pdb [Internet].

65. Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD. Improved protein-ligand docking using GOLD. Proteins-Structure Function and Genetics. 2003 SEP 1 2003;52(4).

66. Bohari MH, Sastry GN. FDA approved drugs complexed to their targets: Evaluating pose prediction accuracy of docking protocols. Journal of Molecular Modeling. 2012 SEP 2012;18(9).

67. Chang MW, Ayeni C, Breuer S, Torbett BE. Virtual screening for HIV protease inhibitors: A comparison of AutoDock 4 and vina. PLoS ONE. 2010 08/04;5(8):e11955.

68. Schwaller B. The continuing disappearance of "pure" Ca2+ buffers. Cellular and Molecular Life Sciences. 2009 JAN 2009;66(2).

69. Svensson LA, Thulin E, Forsen S. Proline cis-trans isomers in calbindin D9k observed by X-ray crystallography. J Mol Biol. 1992 FEB 5 1992;223(3).

70. Pang YP. Successful molecular dynamics simulation of two zinc complexes bridged by a hydroxide in phosphotriesterase using the cationic dummy atom method. Proteins-Structure Function and Genetics. 2001 NOV 15 2001;45(3).

71. Pang YP. Novel zinc protein molecular dynamics simulations: Steps toward antiangiogenesis for cancer treatment. Journal of Molecular Modeling. 1999 1999;5(10).

72. Roe RR, Pang YP. Zinc's exclusive tetrahedral coordination governed by its electronic structure. Journal of Molecular Modeling. 1999 1999;5(7-8).

73. 10.2210/pdb3d10/pdb [Internet].

74. Ostendorp T, Diez J, Heizmann CW, Fritz G. The crystal structures of human S100B in the zinc- and calcium-loaded state at three pH values reveal zinc ligand swapping. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research. 2011 5;1813(5):1083-91.

75. 10.2210/pdb2g9b/pdb [Internet].

76. Tamames B, Sousa SF, Tamames J, Fernandes PA, Ramos MJ. Analysis of zinc-ligand bond lengths in metalloproteins: Trends and patterns. Proteins-Structure Function and Bioinformatics. 2007 NOV 15 2007;69(3).

77. Bobay BG, Stewart AL, Tucker AT, Thompson RJ, Varney KM, Cavanagh J. Structural insights into the calcium-dependent interaction between calbindin-D28K and caspase-3. FEBS Lett. 2012 OCT 19 2012;586(20).