THE ATOMIC BASIS FOR METAL-DNA SIGNALLING IN CUER

By

CHRISTOPHER IAN COLLINS

UNIVERSITY OF
BIRMINGHAM

A thesis submitted to

The University of Birmingham

For the degree of

DOCTOR OF PHILOSOPHY

School of Chemistry
The University of Birmingham
August 2008

ABSTRACT

E. coli and other Gram-negative bacteria control the concentrations of metal ions *via* a variety of routes including through the MerR family of transcriptional activators. Several proteins in this family have been characterised, including MerR, which removes toxic Hg(II) ions from the bacterial cells, CueR, which removes toxic Cu(I) ions and ZntR, which removes Zn(II) which becomes toxic at high concentrations but is an essential nutrient at lower concentrations.

MerR family proteins work through an unusual route, for which there is little chemical data in the literature. Crystal structures are available for activator forms of CueR and ZntR, but the change to these structures upon binding metal ions was not previously characterised. This thesis studies mechanisms of binding, information transfer and activation of transcription.

CueR binds exceptionally strongly to Cu(I), and this thesis shows that this is due to the binding site being very well set up. An essential hydrogen bonding chain which transfers data about the metal binding event to the DNA has been characterised. ZntR works through a similar mechanism with the hydrogen bonding chain replaced by a direct ionic bond to the metal ion.

Mutants are considered which disrupt the hydrogen bonding chain. CueR with and without Cu(I) is studied using Molecular Dynamics along with mutants and a resting state for CueR is proposed with a larger dihedral angle than the activator form.

Alternative binding sites are proposed for a variety of metal ions to produce new sensors and sequesters.

CONTENTS

1. INTRODUCTION

1.1 Overview

This thesis studies toxic metals and approaches to reducing the effect that they have on organisms. Certain bacteria have developed a method of removing toxic metals from their cells by using a family of proteins called the MerR family. These proteins bind very strongly to any of the metals that enter the cell and then turn on cell machinery that then removes all of the metal from the cell. The way in which these proteins work is very unusual and the strength and specificity of binding is extremely high. The way that information on the metal binding event produces the signal is also very interesting, and knowledge of this process may be used to develop a variety of highly selective and sensitive metal ion detectors and sequesters.

1.2 Toxic metals

Many metal ions are essential to life, however all metal ions are toxic in high enough concentrations. The toxicity limit varies depending on the ion, the organism, the environment and various other factors. It can be very important to remove excess amounts of metal ions from cells, especially those that are toxic at low concentrations.

Different ionic forms of each metal will have different toxicities; for example, Cu(I) is highly toxic at all concentrations whereas Cu(II) is an essential nutrient for proteins such as blue copper proteins.[1-3] Even ions that are essential to the correct function of the organism will become toxic when too much is present, such as Zn(II).[1] Other ions may be toxic in all forms, or may easily change between forms (for example arsenic).[2] Thus it can be very important to regulate the amount of metal ions in cells.

Mercury is known to be poisonous but has been used over thousands of years by many different cultures. It has been used by felt hat makers, dentists, physicians and scientists and may enter the environment through the burning of contaminated crops, spillages from thermometers, breakdown of amalgam fillings and mercury-based pesticides.[3] It is toxic in both its elemental form and as organic compounds.[4] In humans it causes deterioration of the nervous system, impairs hearing, speech and vision, causes involuntary muscle spasms, corrodes skin and mucous membranes and causes various other health effects. It has no known beneficial use to any organism. It can also bioaccumulate so that small concentrations in water may become much larger concentrations higher up the food chain.[5] Thus even small concentrations must be removed from organisms.

As mentioned above, copper is an essential nutrient in the form of Cu(II); however, it is highly toxic in the form of Cu(I). Metallic copper has been used for between seven and ten thousand years for making bronze, brass and other alloys and poses no health threat.[6,7] Thus all Cu(I) must be removed from organisms, but other forms of copper must not be affected.

Zinc is known to be an essential nutrient for life; however, in excess amounts it can become toxic.[1] It is an essential nutrient for the formation of certain proteins and zinc fingers.[8-10] In humans it is relatively non-toxic and animals show tolerance to high intakes of zinc.[11] However, some research has shown that it may replace copper and iron ions leading to nausea and vomiting, anaemia, leucopoenia and neutropenia.[12,13] In bacteria different effects will be produced and excess concentrations of Zn(II) may be more toxic. Thus intracellular concentrations must be carefully monitored and controlled.

Lead is known to be toxic to humans and has historically been used in a wide variety of applications, from bullets to church roofs. Lead has found its way into the environment in

dangerous forms particularly through the burning of leaded petrol.[14] It can cause long term health problems by replacing calcium ions in bones and being stored in the body from where it may be released and thus cause problems later. It interferes with enzymes in the central nervous system and affects kidney function, blood cells and the metabolism of vitamin D and calcium.[15] It has no known positive use in organisms, thus all must be removed from the intracellular medium.

Cadmium is a rare metal, but one which has severe health implications. It is present in the environment due to humans using it in battery production and plating. Like lead, it replaces calcium and can bioaccumulate due to non-bacterial organisms being unable to remove it from their cells.

Many other metal ions are also toxic, and whilst some may be regulated by some organisms, there are many that are not. Understanding how some organisms can cope with high concentrations of ions that are toxic to humans may be very useful.

1.3 Methods of removing toxic metal ions from cells

Some organisms have evolved methods of removing toxic metal ions from their cells. There are several different routes through which the concentration of ions may be reduced: reduction of metal uptake, enhanced metal exportation and metal sequestration mechanisms.[16] Each of these may proceed through different methods.

Reduction of metal uptake may proceed through reducing the membrane concentration of the receptor for membrane transport, making the extracellular metal unavailable for uptake

through chelation with polysaccharides, mineralization of the metals by inorganic ions such as sulfide or various other methods.[17-19]

Enhanced metal exportation is not fully understood, but various ATPases are employed. Some of the methods involve reduction of the metal ion, and mercury may even be reduced to volatile Hg(0), which diffuses out of the cell.[20] The ion may also be moved into an organelle within the cell from where it cannot re-enter the bulk cell and various chemical processes may be performed on it.

Metal sequestration may be performed by binding the metal ion to various peptides or proteins such as Metallothioneins. Metallothioneins are widely used throughout nature.[16] These are proteins that contain a large number of cysteine residues and, as such, are very good for binding to certain metal ions, such as Cu(I) and Cd(II). They are generally non-specific and will bind to any group 11 or 12 metal ions present. They bind the metal ions in four coordinate tetrahedral (in the case of Zn(II) and Cd(II)) and three coordinate trigonal planar (Cu(I)) sites.[21]

The methods through which the above mechanisms are employed must be carefully regulated. Regulation can occur at the transcriptional, translational or enzymatic level. In mammals metallothionein production is regulated by the presence of copper, cadmium and zinc, and occurs through the metal regulated transcription factor ACE1.[22,23]

1.3.1 MerR family regulators

Escherichia coli and other Gram-negative bacteria have developed a family of transcriptional activators that respond to the presence of toxic metal ions and trigger

production of proteins that can remove the metal ions from the cell.[24] There are a variety of these regulator proteins for different metal ions, including CueR, MerR and ZntR. Once the metal ion has been removed from the cell it may be oxidised so that is will no longer pose a threat to the organism.[25]

MerR is the archetype of a family of the metal-dependent transcriptional activators. In the presence of Hg(II), MerR is an activator of *mer* genes, which code for mercury resistance and regulate synthesis of MerR.[26-30] It appears that both transcriptional activation and repression occur with MerR bound to the promoter sequence of the gene. DNA transcription is initiated by MerR twisting the DNA and presenting a face to which RNA polymerase may bind and produce messenger RNA.

Typically DNA transcription is promoted through the -35 and -10 regions, which appear upstream (towards the 5' end) of the protein coding gene.[9] These are usually on the same side of β-DNA and so the RNA polymerase can bind directly. The promoter sequence of MerR is unusual, with the -35 and -10 sequences being 19 base pairs apart, rather than the usual 16-18.[27,31-36] This change causes the base pairs to be approximately 72 degrees further round the helix than would usually be the case. Thus transcription cannot usually happen directly and MerR must be employed to twist the DNA into the correct shape for transcription. The amount of twist is therefore very important and transcription will only occur when MerR is in the correct conformation. The presence of Hg(II) causes MerR to adopt the correct conformation for DNA transcription; any other conformation will repress transcription.[37]

MerR has been shown to activate DNA transcription at very low concentrations of Mg(II) with high selectivity against other metal ions.[38,39] CueR is a similar protein which causes

DNA transcription in the presence of Cu(I) and to a lesser extent in the presence of Ag(I) and Au(I).[40-44] Other members of the family include ZntR, which is activated in the presence of Zn(II) and partial response to Cd(II) and Pb(II), PbrR, which is activated in the presence of Pb(II), and BmrR which is activated in the presence of certain organic molecules.[45-51]

Members of the MerR family have been shown to have high similarity with sequence alignments showing CueR to have 52% similarity and 28% identity to MerR despite the differences in selectivity.[25] CueR is also 37% identical to ZntR. Comparison of various strains of MerR from different Gram-negative bacteria show around 90% sequence similarity.[24]

In the case of mercury, the DNA is transcribed to produce a variety of proteins including more MerR, MerP, MerT, MerA and MerD.[36] Each of these performs its own part in mercury resistance. MerP binds to Hg(II) in the periplasm, from where the metal ion is transferred to MerT in the inner membrane. It is then transferred to MerA in the cytoplasm which reduces it to Hg(0), which is volatile and allows it to diffuse out of the cell as mercury vapour. Once all of the Hg(II) has been removed from the cell, MerD binds to MerR and destabilises its interaction with the DNA, stopping transcription.

MerR and related proteins form a cycle of binding, transcription and removal that is self-regulating. A few copies of MerR will be present in the cell to start with, which will bind to the correct part of the DNA. As MerR is a mild repressor of transcription in the absence of DNA, production of any of the proteins will be repressed until the concentration of Hg(II) is high enough. Hg(II) ions then bind to MerR which changes shape, allowing DNA transcription to occur and produce all of the required proteins. These newly produced

proteins then work together until all of the Hg(II) has been removed from the cell, whereupon MerD binds to MerR.[52] MerR then dissociates from the DNA and the cycle begins again. Similar cycles and proteins are likely to be found for other members of the MerR family.[40,53]

The above information has been gathered from various experimental procedures and X-ray crystal structures. The X-ray structures currently available include activator forms of CueR and ZntR, BmrR with DNA bound, C-terminal and N-terminal regions of various members of the MerR family and a few other members of the family.[51,54,55] There is currently no structure for a metal binding repressor motif, nor is there a structure of the whole of MerR.

1.4 Chemical interest

The MerR family of transcriptional activators are highly sensitive and selective.[24] CueR has the highest binding affinity, somewhere in the zeptomolar (10^{-21}M) range for binding to Cu(I).[7,42,56] Upon binding, DNA transcription is activated, which can be seen as an identifiable signal (the production of the proteins coded for by the DNA sequence to which the protein is bound). Thus these proteins are often described as highly sensitive and selective sensors for the metal ions they bind to. The mechanisms for the sensitivity, selectivity and signal production are not currently fully understood. However, the aim of this thesis is to investigate local changes in the metal binding site in response to metal binding and to explore the way in which these changes propagate through the regulator to affect DNA conformation.

There is currently little chemical data on any members of the MerR family in the literature, and the biochemical data is also somewhat limited. There have also been no computational

studies into the exact nature of the binding site. As such, there is no experimental chemical data in the literature with which the data in this thesis may be compared.

1.4.1 Metal binding site

Understanding the binding site is the key to understanding the sensitivity and selectivity of these proteins. In MerR, mutagenesis experiments have shown that three cysteine residues bind directly to Hg(II) and that these are essential to the correct function of the protein; these are known as cys82, cys117 and cys126.[57-59] An X-ray crystal structure has not yet been able to be defined for MerR, although some are available for other MerR family members.[51,54]

ZntR binds to Zn(II) with an affinity of around 10^{-14} M.[45-47] The binding site is relatively similar to that of MerR, with cys79, cys114 and cys124 binding to Zn(II). The change in positions of these residues causes a slight change in the overall shape of the protein. This change may account for the change in binding affinity. X-ray crystal structures have been determined for ZntR and show a tetrahedrally coordinated Zn(II) ion with a phosphate ion taking up the fourth site. There also appears to be a second zinc ion bound in the crystal structure; this ion may or may not be essential to the correct function of the protein. It may only appear in the x-ray crystal structure because of the conditions used during the crystallisation of the protein.[24] Figure 1.1 shows how the binding site appears with the inclusion of the second zinc ion.

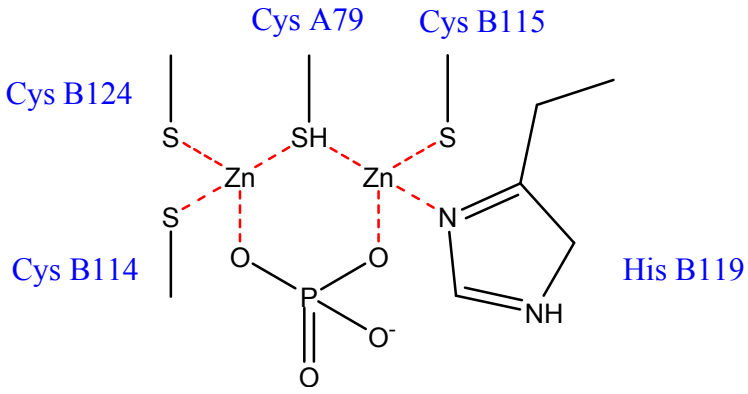


Figure 1.1 - ZntR crystal structure metal binding site

Due to the similarity of ZntR to MerR and other members of the family and the amount of data available on ZntR, it will be used for comparison to CueR and MerR in this thesis.

CueR binds to Cu(I) with an affinity of around 10^{-21} M.[41-43,54] The binding site has changed somewhat from the MerR structure with only cys112 and cys120 binding to the metal ion. The third cysteine found in MerR and ZntR (cys82 and cys79 respectively) has been mutated to a serine residue and therefore does not bind to the metal ion. X-ray crystal structures have been determined for CueR and show a completely linear (>175^{o}) two coordinate binding site, which has been confirmed by EXAFS.[60] Thus the binding site of CueR may be described as shown in figure 1.2.

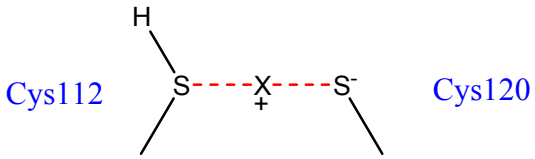


Figure 1.2 - CueR crystal structure metal binding site

The above unusual binding site and high affinity make this protein very interesting to study from a chemical perspective.

1.4.2 Signal-response communication

There is an identifiable signal produced when the correct metal ion has bound to the proteins of the MerR family. In normal bacterial conditions the signal is seen as the production of the proteins coded for by the gene to which the MerR family protein is bound. The protein produced may be a sensor protein, a metal removal protein or an oxidation protein.

The metal binding site is several Angstroms away from the DNA-binding domain, yet information on the metal binding event is passed through to the DNA-binding domain. In MerR and ZntR one of the metal binding residues (cys82/cys79) is in the DNA binding domain and thus information transfer is easily understandable.[61]

ZntR is a dimeric protein with two domains. Each monomer binds 1 zinc ion and one end of the DNA. There are two α-helices joining the domains around which they may flex relative to one another. Several crystal structures have been obtained for ZntR but they all contain Zn(II).[54] Thus the resting state of the protein is not known, though any non-activator form will act as a repressor.

In CueR the equivalent of cys82/cys79 has been mutated into a serine so the method of information transfer seen for ZntR cannot occur. Looking again at the binding site, several hydrogen bonding interactions have been identified.[54] One of these is from the thiol hydrogen of cys112 to the backbone carbonyl oxygen of ser77. The length of this hydrogen bond is likely to change with addition of the metal ion and thus the binding information can be transferred to the DNA binding domain. The exact nature of the change will be discussed in chapter 3.

The binding event is likely to have both local and whole protein effects. The residues around the binding site will change positions and this can cause the entire protein to change shape and twist the DNA into the correct conformation for transcription.

1.4.3 Sensitivity

Each of the MerR family proteins has a different sensitivity to their metal ions. MerR promotes transcription at around 10^{-8} M, whereas CueR has a binding affinity of 10^{-21}M.[7,38,39,42] ZntR has a binding affinity of around 10^{-14} M that will allow free Zn(II) to be removed from the cell, but Zn(II) bound to other molecules will not cause DNA transcription.[45]

Measuring the exact binding affinity of CueR is not possible as extremely low concentrations are required and cannot be measured very accurately. However, Changela *et al* performed thermodynamic calibrations with CN^- at pH 8.0 to show that half maximal CueR induction is reached at $2x10^{-21}$ M ± $1x10^{-21}$.[54] This binding affinity suggests that a single Cu(I) ion in an *E. coli* cell has a greater than 50% chance of being bound to CueR.

1.4.4 Selectivity

Each of the MerR family proteins are highly specific to the ion they detect. CueR produces a large signal when Cu(I) is bound and a much smaller one when Ag(I) and Au(I) are bound.[36,42] Some of the experiments appeared to show that Cu(II) also activated the protein, but it was soon realised that the signal was produced *via* intracellular reduction of Cu(II) to Cu(I).[42]

It appears that binding other metal ions does not cause a signal to be produced. Li(I) has a similar size and charge to Cu(I) but does not produce a signal when bound to CueR. It is currently unknown whether other metal ions do not bind or if they bind but twist the DNA the wrong amount for the binding of RNA polymerase.

1.5 Aims for this thesis

This thesis aims to give a greater understanding of how CueR works at an atomic level. The methods used are described in Chapter 2. The binding site will be studied in great detail and different metal ions will be compared to see how it achieves its selectivity and sensitivity (Chapter 3). These results will be compared to another member of the MerR family, ZntR, for which a crystal structure and other data is available (Chapter 4). The method of information transfer from the binding event will be studied both by considering the local effects on the binding site and the global effects on the protein as a whole (Chapters 3 and 5). The resting state of the protein will be considered and calculations performed in an attempt to ascertain what happens when Cu(I) is not bound (Chapters 3 and 5). Finally applications of the protein will be considered along with studying whether or not other metal binding proteins could be designed (Chapter 6).

2. COMPUTATIONAL METHODOLOGY

2.1 Introduction to computational techniques

There are several different levels of theory that can be used to study biological systems. Two basic levels are used in this project: *ab initio* Quantum Mechanics (QM) and Molecular Mechanics (MM). There are large differences in the way that these calculations work, the length of time that they take to run and thus the size and type of systems which may be studied using them.

2.2 Quantum Mechanics

Quantum Mechanics is divided into two main categories: *ab initio* and semi empirical calculations. Only the former of these was used and thus will be discussed in detail.

Ab initio calculations start with the Schrödinger equation which is used to describe the wavefunction (ψ) of electrons.[62] Many useful properties of a molecule can be calculated from the wavefunction of the atoms. The probability density of finding an electron in a given region of space can be obtained from the square of the wavefunction, which can then be used to study the classical concept of "bonds". The time independent version of the Schrödinger equation takes the basic form as shown in equation 2.1.

$$\hat{H}(\psi) = E\psi$$

Equation 2.1 – The Schrödinger equation

There are various forms of the Hamiltonian Operator ($\hat{H}$) including time dependent and time independent versions. The Hamiltonian operator can be written as a sum of all of the electronic interactions and kinetic motions. For a system with M nuclei (A and B) and N electrons (i and j) the Hamiltonian can be written in the form shown in equation 2.2.[63]

$$\hat{H} = -\sum_{i=1}^{N} \frac{1}{2} \nabla_i^2 - \sum_{A=1}^{M} \frac{1}{2M_A} \nabla_A^2 - \sum_{i=1}^{N} \sum_{A=1}^{M} \frac{Z_A}{r_{iA}} + \sum_{i=1}^{N} \sum_{j>i}^{N} \frac{1}{r_{ij}} + \sum_{A=1}^{M} \sum_{B>A}^{M} \frac{Z_A Z_B}{R_{AB}}$$

Equation 2.2 – The Hamiltonian Operator

Qualitatively this equation means that the Hamiltonian is the sum of the total electron motion, the nuclear motion, the interaction between each of the nuclei and the electrons, the interaction between the electrons and the interaction between the nuclei.

Unfortunately the Schrödinger equation is analytically insoluble for systems containing more than two particles, so numerical methods must be used. The Born-Oppenheimer Approximation may be used to simplify the calculations and thus study larger systems.[64-66] The Born-Oppenheimer approximation states that as nuclei and electrons move at very different speeds, these two motions can be separated out to produce a different form as shown in equation 2.3.

$$\psi_{total} = \psi_{electronic} + \psi_{nuclear}$$

Equation 2.3 – The Born-Oppenheimer Approximation

Essentially the electrons move around stationary nuclei and the nuclei move in an averaged field of electron density. Thus some of the terms in the above equation are removed, producing a much more simple form for the electronic Hamiltonian as shown in equation 2.4.

$$\hat{H}_{elec} = -\sum_{i=1}^{N} \frac{1}{2} \nabla_i^2 - \sum_{i=1}^{N} \sum_{A=1}^{M} \frac{Z_A}{r_{iA}} + \sum_{i=1}^{N} \sum_{j>i}^{N} \frac{1}{r_{ij}}$$

Equation 2.4 – The electronic Hamiltonian

This form does not contain any information on nuclear motion or the interaction between the nuclei. Such motions and interactions must still be calculated but can be performed separately.

The electrons can thus be placed in orbitals, each with a unique spin and spatial distribution combination. These are referred to as Molecular Orbitals. These molecular orbitals can be conveniently described using a Linear Combination of Atomic Orbitals (LCAO) approach. These molecular orbitals are not necessarily localised onto one atom as seen for Atomic Orbitals but will be affected by the surrounding nuclei. Each new Molecular Orbital will have a contribution from the various Atomic Orbitals making it up as shown in equation 2.5.

$$\psi_i = \sum_{\mu=1}^{K} c_{\mu i} \varphi_\mu$$

Equation 2.5 – Linear Combination of Atomic Orbitals

Where:

ψ_i is the molecular orbital

φ is the atomic orbital contributing to the molecular orbital

$c_{\mu i}$ is the coefficient which determines the amount that the specified atomic orbital is contributing to the molecular orbital

K is the number of atomic orbitals contributing to the molecular orbital

The Atomic Orbitals used in the above equation are referred to as Basis Sets. Basis Sets can take a wide variety of different forms depending on the circumstances (see section 2.3.2).

The problem with the Molecular Orbital approach is that the electron-electron interaction is not calculated in the scheme described so far. Various approximations are available to simulate the effects of electron-electron interactions, such as the Hartree-Fock approximation. The Hartree-Fock approximation does not simultaneously calculate the energies of each electron, but rather allows each electron to move in an average field of the rest of the electrons. In this case the Fock operator is used to describe electron *i* as shown in equation 2.6.

$$f(i) = -\frac{1}{2}\nabla_i^2 - \sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \upsilon^{HF}(i)$$

Equation 2.6 – The Fock operator

Where $\upsilon^{HF}(i)$ is the average potential experienced by the *i*th electron due to the other electrons. Solutions to the Schrödinger equation can be more accurately determined by the Self-Consistent-Field (SCF) method. In this case the atomic orbital coefficients for each molecular orbital are initially guessed. The resulting electron density determines the field of the surrounding electrons, which is in turn used to calculate the energy and distribution of each electron. A new field will then be calculated by the individually changed electrons and this field can then be reused to perform the calculation for each electron and so on. This process continues until the field no longer changes and self-consistency is reached.

The above models still do not take account of all of the electron-electron interactions. The electrons are only affected by the general field of the other electrons, not the instantaneous position of any of them. However, electron motions may be correlated (*i.e.* move in a non-independent fashion) and must be accounted for. Electron correlation can be calculated through a variety of methods including Configuration Interaction techniques (such as

CASSCF), Møller-Plesset Perturbation Theory (such as MP2) and Density Functional Theory (DFT) techniques.[67-73]

DFT is a different approach from Hartree-Fock theory. Whilst Hartree-Fock theory uses a combination of the individual electron wavefunctions to calculate the total electron wavefunction, DFT only attempts to calculate the total electronic energy and the overall electronic density distribution. This approach is useful as the ground state energy and other properties of a system are uniquely defined by the electron density.[74] Thus the total energy, E, is a unique functional of $\rho(r)$. The energy functional can be written as shown in equation 2.7.[75]

$$E[\rho(r)] = \int V_{ext}(r)\rho(r)dr + F[\rho(r)]$$

Equation 2.7 – DFT energy functional

In this equation, the external potential, $V_{ext}(r)$, is caused by the Coulomb interaction between the nuclei. This potential is added to the total kinetic electron energy and interelectronic interactions, $F[\rho(r)]$. There are many different ways of calculating these functions, and as such there are a large number of different DFT functionals available.

The first practical suggestion for calculating the kinetic electron energy and interelectronic interactions was suggested by Kohn and Sham.[76] Their method involves splitting the function $F[\rho(r)]$ into contributions from kinetic energy (E_{KE}), electron-electron Coulombic energy (E_H) and exchange and correlation (E_{XC}) as shown in equation 2.8.

$$F[\rho(r)] = E_{KE}[\rho(r)] + E_H[\rho(r)] + E_{XC}[\rho(r)]$$

Equation 2.8 – Kohn and Sham energy functional

The kinetic energy term involves non-interacting electrons with the same density as the real system as shown in equation 2.9.

$$E_{KE}[\rho(r)] = \sum_{i=1}^{N} \int \psi_i(r) \left(-\frac{\nabla^2}{2} \right) \psi_i(r) dr$$

Equation 2.9 – K-S kinetic energy

The electron-electron Coulombic energy (also known as the Hartree electrostatic energy) calculates the interaction between two charge densities as shown in equation 2.10.

$$E_H[\rho(r)] = \frac{1}{2} \iint \frac{\rho(r_1)\rho(r_2)}{|r_1 - r_2|} dr_1 dr_2$$

Equation 2.10 – K-S Coulombic energy

There are also various methods for calculating the exchange and correlation, such as the local density approximation (LDA). LDA is based on the uniform electron gas model, which assumes the electron density is constant throughout all space. The exchange-correlation energy can then be calculated by equation 2.11.

$$E_{XC}[\rho(r)] = \int \rho(r) \varepsilon_{XC}(\rho(r)) dx$$

Equation 2.11 – LDA Exchange-Correlation energy

Where $\varepsilon_{XC}(\rho(r))$ is the exchange-correlation energy per electron as a function of the density in the uniform electron gas.

As mentioned previously, DFT can also be used in conjunction with Hartree-Fock methods to calculate electron correlation. Many of the calculations in this thesis were performed using combined DFT/HF methods. Such methods involve calculating all of the required parameters apart from correlation effects *via* Hartree-Fock, followed by adding correlation effects *via* a DFT calculation such as LDA. Becke has suggested that the exchange-correlation energy should take the form shown in figure 2.12.[77]

$$E_{XC} = \int_0^1 U_{XC}^{\lambda} d\lambda$$

Equation 2.12 – Becke Exchange-Correlation energy

This equation involves a coupling parameter, λ, which takes values from 0 to 1 and describes various possible states of the system with different amount of correlation. A λ value of 0 implies the system with no correlation, whereas a λ value of 1 implies full correlation. The real system will be somewhere between these two states and a wide variety of functional forms may be employed to study these states.

2.3 Programs and parameters used for Quantum Mechanics

The main program used for the quantum mechanics calculations was Gaussian 98.[78] It can be programmed with a variety of parameters for different systems. The main parameters used are discussed in more detail in the following sections.

2.3.1 Hartree-Fock

Several different forms of the Hartree-Fock approximation were used for the calculations: Unrestricted Hartree-Fock (UHF), Restricted Hartree-Fock (RHF) and Restricted Open-shell Hartree-Fock (ROHF).

UHF is the normal Hartree-Fock calculation and should work for all systems. RHF is a faster calculation that assumes that all of the electrons will be paired (normally true for singlet ground state molecules) and thus the number of orbitals that need to be calculated is reduced. RHF and UHF should always produce the same results overall in these cases. However, as RHF assumes paired electrons, if the molecule contains one unpaired electron (*i.e.* a doublet state) then RHF cannot be used. RHF is also unable to model systems where

electrons become unpaired during the calculation, such as through dissociation of bonds to form two radicals (*e.g.* the dissociation of the H_2 bond to form two hydrogen atoms).

ROHF may be used to study systems with one unpaired electron as it pairs all but one of the electrons and then tries to place the pairs and single unpaired electron in the orbitals. ROHF is a much faster way of modelling transition metals with a single unpaired electron compared to UHF.

RHF was used for all of the calculations involving singlet states, whereas ROHF was used for those involving doublet states (*e.g.* Cu^{2+}).

2.3.2 Basis sets

Two basis sets were used: 6-31G** and LANL2DZ.[79-91] These basis sets are explained in more detail on the following pages. 6-31G** was used for the calculations in Chapters 3 and 4, and LANL2DZ for Chapter 6 and reused in Chapters 3 and 4 to confirm that the same results were obtained.

2.3.2.1 6-31G**

Though Slater-Type Orbitals (STOs) reproduce the shapes of hydrogen atomic orbitals quite reliably, they can be computationally demanding to calculate as more electrons are added to the system. One way of avoiding these is to use Gaussian curves to produce Gaussian-Type Orbitals (GTOs).[92-94] These curves are much faster and model the MO curves well apart from when very close to the nucleus. For a 1 dimensional hydrogen-like orbital the Slater-Type Orbital takes the form shown in equation 2.13 whereas the

Gaussian-Type Orbital takes the form shown in equation 2.14. These can be plotted using a zeta value of 1.24 (for hydrogen) and an alpha value of 0.4166 as shown in figure 2.1.

$$STO = \left(\frac{\zeta^3}{\pi}\right)^{0.5} e^{(-\zeta r)}$$

Equation 2.13 – Slater-Type Orbitals

$$GTO = \left(\frac{2\alpha}{\pi}\right)^{0.75} e^{(-\alpha r^2)}$$

Equation 2.14 –Gaussian-Type Orbitals

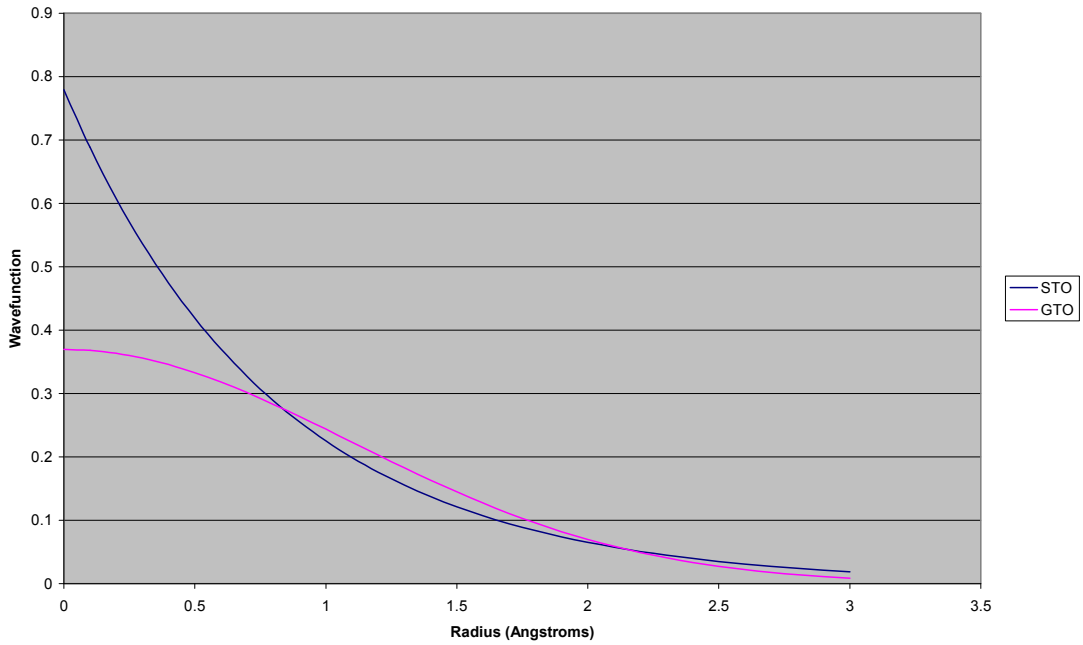


Figure 2.1 – Slater-type vs. Gaussian-type Orbitals

Adding multiple Gaussians together improves the modelling of the curve. The curve produced will have different alpha values and need a contraction coefficient for each Gaussian added. This model is then called a minimal basis set and takes the form shown in equation 2.15. For a hydrogen 1s orbital the parameters shown in table 2.1 are used for this basis set. Comparing these values to the STO and simple GTO values produces figure 2.2.

$$\phi = cc_1\left(\frac{2\alpha_1}{\pi}\right)^{0.75} e^{\left(-\alpha_1 r^2\right)} + cc_2\left(\frac{2\alpha_2}{\pi}\right)^{0.75} e^{\left(-\alpha_2 r^2\right)} + cc_3\left(\frac{2\alpha_3}{\pi}\right)^{0.75} e^{\left(-\alpha_3 r^2\right)}$$

Equation 2.15 – Minimal Basis Gaussian

	Alpha value	Contraction coefficient
GTO1	3.42424091	0.154328970
GTO2	0.623913730	0.535328140
GTO3	0.168855400	0.444634540

Table 2.1 – Alpha values and contraction coefficients used to describe a minimal basis Gaussian for a hydrogen 1s orbital.

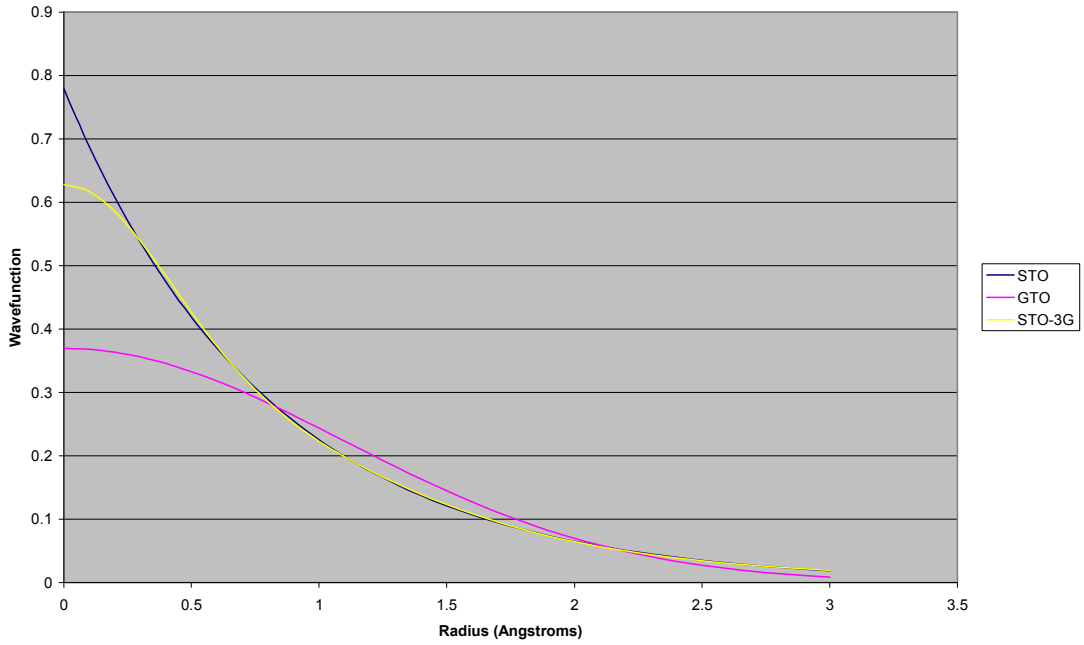


Figure 2.2 – STO-3G vs. STO and GTO

Adding more Gaussians will improve the accuracy further, although 6 is usually the maximum used. To further improve the calculation a Split-Valence Basis Set is often used; this method uses a rough calculation for the core electrons and a more accurate calculation for the valence electrons. The valence electrons are then studied in more detail by doubling the number of STOs, thus requiring the use of two zeta values to produce a double zeta basis set. The first zeta value is for the part of the orbital closer to the nucleus and the second for that further away. 6-31G uses 6 Gaussians for the core, 3 for the inner part of the valence electrons and 1 for the outer part of the valence electrons.

Orbitals are often polarised by the presence of electrons in nearby orbitals, causing them to change shape. This polarization is partially due to the orbital taking on some properties of other low lying orbitals (in the example of a hydrogen 1s orbital the polarization is due to the electron going into the 2p orbital). 6-31G* is an example of a basis set which incorporates polarisation functions on non-hydrogen atoms. The 6-31G** basis set adds polarisation functions on hydrogen atoms as well.

2.3.2.2 LANL2DZ

6-31G** and related basis sets only work for elements that do not contain normally occupied *f*-orbitals, not including the second and third rows of the transition metals. Since mercury, cadmium, silver, gold and lead all bind to CueR-like proteins it would be useful to use a basis set that can handle such elements.

One problem with dealing with f-orbitals is that the electrons are moving within them at relativistic speeds, which causes a breakdown in the Schrödinger equation. Several alternative basis sets are available that can avoid such problems.

Four alternative basis sets were compared (SDD, STO-3G, LANL2MB and LANL2DZ), which can all handle second and third row transition metals.[88-92,95-118] A model similar to CueR's binding site was designed with silver, gold or mercury replacing the copper atom (model 1 in Chapter 3) and the above basis sets used for the calculations. They produce similar geometric results but with different energies due to the methods of calculation.

SDD took much longer to complete than the other basis sets. STO-3G is a minimal basis set and therefore not as accurate as the other models. LANL2MB and LANL2DZ produced

similar results in a similar time, though LANL2MB uses a single zeta value whereas LANL2DZ uses double zeta calculations making the results far more accurate.

LANL2DZ was therefore used for many of the calculations in this thesis. This basis set uses an approximation to deal with the core electrons of larger elements; only the valence electrons may take part in interactions with other atoms. The core of each atom is given an effective nuclear charge to account for electron shielding. This approximation drastically reduces computation time with minimal effects on accuracy.

2.3.3 Correlation functions and DFT

MP2 was initially used to calculate electron correlation. However, the calculations were relatively slow and frequently produced files larger than the 2 GB maximum file size allowed on many Linux systems. Using the MaxDisk keyword the maximum was increased to 8GB, but doing so sometimes made the calculations even slower. In order to avoid these problems DFT was used.

Two different DFT functionals were implemented: B3LYP and MPW1PW91.[77,119-129] B3LYP was used first as it is a commonly used functional for a wide variety of applications. However, several of the calculations did not converge as both structure and energy oscillated around the minimum. To avoid oscillation, several other functionals were compared and MPW1PW91 seemed to give good results in much less time than any of the alternative methods.

2.3.4 Optimisation

Once the electron calculation has completed the nuclear motion calculation must also be performed. As mentioned previously, optimisation is a separate calculation where the nuclei move in an average field of the electrons. The interactions are calculated and second derivatives tell the atoms in which direction to move and how far. However, there are still several ways in which the atom optimisation can be performed; in the calculations in this thesis Z-matrix and Cartesian optimisations were employed.

Z-matrix calculations rely on the positions of atoms relative to other atoms. When one atom moves, all of the atoms specified relative to it move. This movement can be useful when moving whole fragments; however it presents its own problems. Sometimes a small motion (for example a rotation) in one area can cause large motions in other parts of the system, changing interactions there. Another problem is that sometimes angles will move to over 180^{o} (or to less than 180^{o} with a torsion angle change of 180^{o}), causing the calculation to fail. Depending on how the Z-matrix is set up, failure may occur quite frequently.

Cartesian calculations specify each atom relative to the origin. This method of optimisation removes all of the above problems encountered with Z-matrix calculations but has its own problems. As each atom is now free relative to the atoms around it (apart from the usual electronic and steric conditions) whole fragments will not move together. However, the lack of movement of other atoms boned to the atom moving can mean that extra geometry optimisation steps are required.

Both methods should produce the same results eventually, assuming that both complete. Z-matrix calculations were used to start with in conjunction with 6-31G**. Cartesian calculations were used later in conjunction with LANL2DZ, especially for the calculations in Chapter 6 where each calculation was not individually set up so could not be carefully controlled to avoid the problems.

2.4 Molecular Mechanics

Quantum Mechanics is very useful for studying small systems in great detail; however, as soon as the system gets larger than a few atoms, the calculations become unreasonably long. For studying a system as large as a protein simplifications must be used.

Molecular Mechanics is based on classical mechanics and so looks at "bonds" between the atoms as opposed to the quantum mechanical method of studying electron density to work out the behaviour of the system. Such simplifications vastly reduce the number of required calculations per atom and remove the need for the time-consuming SCF process. The energy of the system can still be calculated by comparing various properties to the "equilibrium" values (see section 2.4.1).

Molecular Mechanics can be used for far more than just working out the approximate energy of a system; it can also be used to study how the whole protein is likely to move.

2.4.1 Force fields

Force fields use predefined parameters to calculate various properties of the molecules. These parameters vary from one force field to another. They are based on assigning "atom types" to each atom, which then affects how the atoms interact with one another.

The AMBER force field (Assisted Model Building and Energy Refinement) was implemented in various forms, all of which have a similar form.[130-132] AMBER has parameters for bond lengths, bond angles, torsion angles and non-bonded interactions. It has been specifically designed for use with biomolecules and, as such, is very useful for modelling proteins and DNA. Unfortunately it does not model other molecules (*e.g.* metal ions and other organic molecules) as well, but may still produce reasonably accurate results.[131,132]

Using classical mechanics the bond length can be modelled using the Lennard-Jones potential shown in equation 2.16.[133] The Lennard-Jones potential is a reasonably good fit to bond stretching; however, it is very time consuming to calculate. An alternative would be a Morse potential shown in equation 2.17.

$$V(r) = 4D_e\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$$

Equation 2.16 – Lennard-Jones potential

Where:

D_e is the depth of the potential energy minimum

σ is the finite distance at which the inter particle potential is zero

r is the distance between the particles

$$V(r) = D_e\left(1-\exp\left(-\omega\sqrt{\frac{\mu}{2D_e}}(r-r_0)\right)\right)^2 + V(r_0)$$

Equation 2.17 – Morse potential

Where:

ω is the frequency of the bond vibration and is related to the force constant at the energy minimum, k_0, by $\omega = \sqrt{\frac{k_0}{\mu}}$

μ is the reduced mass.

r is the current bond length

r_0 is the reference bond length

The Morse potential is still reasonably time consuming to calculate and requires three parameters to be specified for each bond. A much simpler alternative would be a simple harmonic oscillator. This model produces a very differently shaped potential in which the atoms are not allowed to dissociate as the energy continues to increase as distance increases. Modelling bonds in this way can be useful for making sure that biomolecules retain their predefined configuration. The modelling is good close to the reference bond length but gets worse as it gets further away. In this model the bond lengths are likely to remain close to their reference values. The reference value has been worked out by studying various biomolecules and measuring their bond lengths. Such reference values may not be the lowest energy length for a given bond depending on its environment but gives a good measure of the amount of energy needed to stretch/compress the bond. The simple harmonic potential (as used by AMBER) takes the form shown in equation 2.18. The bond angle is also measured relative to a reference value and uses a similar harmonic potential of the form shown in equation 2.19.

$$V(r) = \frac{k}{2}(r - r_0)^2$$

Equation 2.18 – Simple harmonic potential (AMBER bond length)

$$V(\vartheta) = \frac{k}{2}(\vartheta - \theta_0)^2$$

Equation 2.19 – AMBER bond angle

The bond length and angle are unlikely to vary from the reference values by a large amount, whereas the other parameters (such as torsion angles and non-bonded interactions) can vary much more easily. Torsion angles are calculated from a cosine series expansion of the form shown in figure 2.20. The non-bonded (van der Waals and electrostatics) interactions are calculated from equation 2.21.

$$V(\omega) = \sum_{n=0}^{N} \frac{V_n}{2}(1 + \cos(n\omega - \gamma))$$

Equation 2.20 – AMBER torsion angle

$$V(nb) = \sum_{j=1}^{N-1} \sum_{i=j+1}^{N} \left(4D_{e(i,j)} \left(\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right) + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right)$$

Equation 2.21 – AMBER non-bonded interactions

The first term in the above equation is caused by the van der Waals interaction and is modelled by a Lennard-Jones potential. Unfortunately there is no way of simplifying this term for a non-bonded interaction. AMBER, unlike some other force fields, does not contain an explicit hydrogen bonding term.

As non-bonded interactions are time consuming to calculate it may be useful to restrict the number of them that need to be calculated. The number of atoms in a system may be quite large but longer range interactions will affect the atoms less than shorter range ones. As

such, calculations are usually performed with a maximum distance cut-off for non-bonded interactions.

2.4.2 Minimisation

The energy of a system can be calculated from a combination of equations 2.18, 2.19, 2.20 and 2.21. Each of the parameters can change the overall energy of the system, but many of them are linked. A given system will have a multi dimensional energy surface which will have many different energy minima. There are many different methods through which the energy surface may be explored and energy minima found. Most minimisation techniques, including those used in this project, are only able to find the local energy minimum (nearest to the starting point) and not the global energy minimum as shown in figure 2.3.

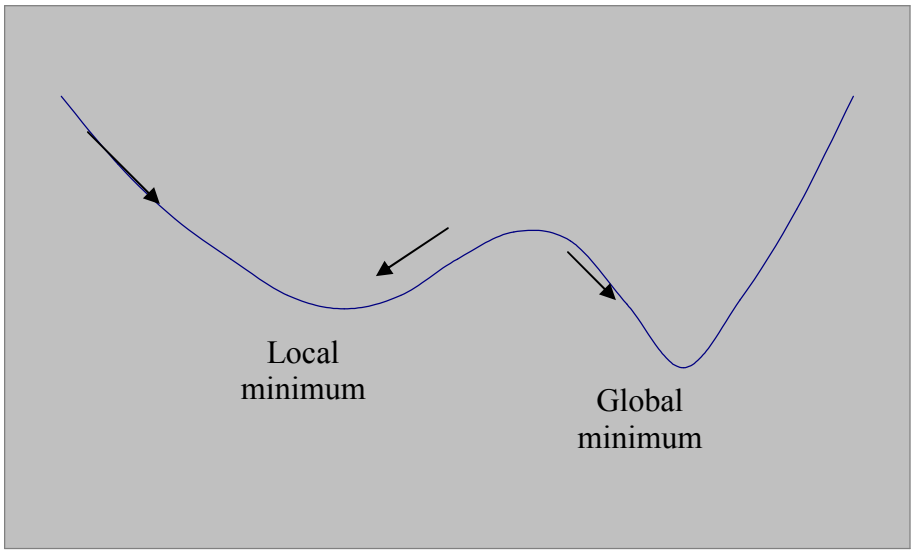


Figure 2.3 – Minimisation to local and global minima

One of the simplest forms of minimisation is "steepest descent". Steepest descent involves measuring which direction each parameter is being pushed in by the combination of interactions (by measuring the gradient of the graph) and taking a small step in that direction for every parameter in the system. Depending on the size of step taken, such

methods may either overshoot the minimum or require a large number of steps to get to the minimum. After each step all of the parameters must be recalculated.

A much more efficient method of finding the minimum is to use a second derivative method such as the Newton-Raphson Method. This method calculates the rate of change of the gradient (*i.e.* the second derivative of the curve) and uses it to predict where the energy minimum will be. In order to calculate the second derivate, a second data point is needed, so at least the first step of a Newton-Raphson minimisation is always steepest descent. Unfortunately due to the multi dimensional nature of the energy surface the minimum along a given axis might have changed due to another variable, so many steps will still be required. Newton-Raphson minimisation is employed by AMBER to estimate the local energy minimum.

The minimisation of a protein will be affected by the environment. In cells proteins are often in aqueous solution and may have metal ions, cofactors and many other molecules surrounding them. As such, the protein will usually be modelled in a box of water with the other molecules. To stop any of the molecules from getting too far away from the others, periodicity is applied. Essentially periodicity means that when a molecule leaves one side of the periodic box it reappears on the opposite side and the interactions must also be calculated across the periodic boundaries. Problems can be found when calculating large molecules; it may be possible for a molecule to interact with itself across the periodic boundary. To avoid a molecule interacting with itself, large molecules are surrounded with a box of water molecules that is at least twice as thick as the maximum non-bonded interaction cut off.

2.4.3 Molecular Dynamics

Molecular Dynamics (MD) is a process by which the movements of the whole protein may be studied. It requires a minimised protein in a water box for the starting coordinates.

There are several steps involved in running a complete MD run. First the protein must be prepared: water and other molecules must be added, periodic boundary conditions must be set and the protein minimised. Next an equilibration run must be performed to allow the temperature to distribute evenly throughout the protein. Then a full MD data gathering run can be performed. Finally the data must be analysed.

Equilibration is a process by which energy is added to the system in the form of kinetic energy to simulate the effect of being at around body temperature. Every atom in the system is given a random velocity until the required temperature is reached (usually 300 K). The energy may then transfer from one atom to another through collisions until the more mobile sections of the system have more velocity and the less mobile ones have less. This process often takes up to around 100 ps and so the equilibration run was set to 100 ps for all of the calculations in this project. The temperature is kept constant by removing excess energy every 100 fs by using the weak-coupling algorithm.[134]

The full MD data gathering run takes the output from the equilibration (including velocities and coordinates) and allows the motions to continue. The coordinates, energies and velocities for every atom are output to files every 1000 fs for later analysis.

2.5 Programs and parameters used for Molecular Mechanics

AMBER 8 was used for all of the Molecular Mechanics calculations. The preparation of the proteins was performed using the version of LEaP provided with AMBER 8.[135] All parameters were set to the defaults apart from those specified below.

The force fields used were AMBER ff03 and ff02EP. ff03 contains the latest parameters for all atoms at the time of production of AMBER 8. ff02EP uses slightly older parameters (from 2002 and previously) but attempts to make hydrogen bonds more effective by adding "extra points". These extra points are dummy atoms held a short distance away from any atoms with lone pairs (*e.g.* oxygen, nitrogen and sulfur) with a small negative charge. The core atom then has a more positive charge to produce the correct overall charge.

Minimisation was performed with a maximum of 10,000,000 cycles. However, the run sometimes exited prematurely stating that the convergence criteria had been met. Minimisation was performed three times for each calculation to ensure that a low energy starting point was found.

MD was performed using the SHAKE algorithm.[136,137] SHAKE removes bond stretching from the calculations and reduces the number of parameters to be calculated, allowing a larger time step to be used. The calculation was run for 1 ns after the equilibration and a time step of 2 fs was used. See chapter 5 for detail on protein setup.

3. MODEL SYSTEMS (QUANTUM MECHANICS)

3.1 Studying model systems

CueR is a fairly large protein with over 4000 atoms,[54] thus it can be challenging to model the whole protein to a high degree of accuracy. As discussed in chapter 2, molecular dynamics (MD) can be used to study the whole protein, with or without DNA attached. Unfortunately MD is unable to give any information about the way in which the copper binds, what local changes may occur to eventually be transferred through the protein or what the repressor conformation may be. Therefore various parts of the protein can be studied in greater detail using quantum mechanics.

The most chemically important part of the protein is the metal binding domain.[56] To aid understanding of the mode of action, key amino acids were identified and the interaction between them and the metal ion have been studied.[60]

All calculations in this chapter were performed using Gaussian 98, the basis set 6-31G**, z-matrix optimisation and the DFT functional B3LYP for electron correlation calculations.

3.2 Protein-ion interactions

CueR uses two amino acids bind to Cu(I): Cys112 and Cys120 (see chapter 1).[56] It is generally accepted that Cu(I) prefers a two-coordinate linear arrangement of ligands and also that it will bind to "soft" ligands such as sulfur-based ligands. As such, it is feasible to assume that this binding site is already well set up to accept Cu(I). However, the strength of binding may also be affected by the charge on the ligands. In this case the cysteine residues may be thiols or thiolates. Thiols are softer ligands than thiolates due to their

ability to donate a lone pair rather than a whole charge, thus they might be expected to bind better to soft Cu(I).

It is impossible to find out whether hydrogen atoms are present through X-ray diffraction.[54] However, there are several pointers that suggest that at least one of the cysteine residues must be a thiol.

The sulfur of Cys112 is quite close to the oxygen of the backbone carbonyl group of Ser77. The proximity of these partial negative charges could cause destabilisation in the structure in this area. The distance is similar to that usually found for a hydrogen bond to be present between these atoms and the angle between the copper and where the hydrogen bond would need to be is approximately 109.5°; also similar to that of a hydrogen bond from an sp^3 hybridised atom. This hydrogen bond could be important for the function of the protein, as discussed later.

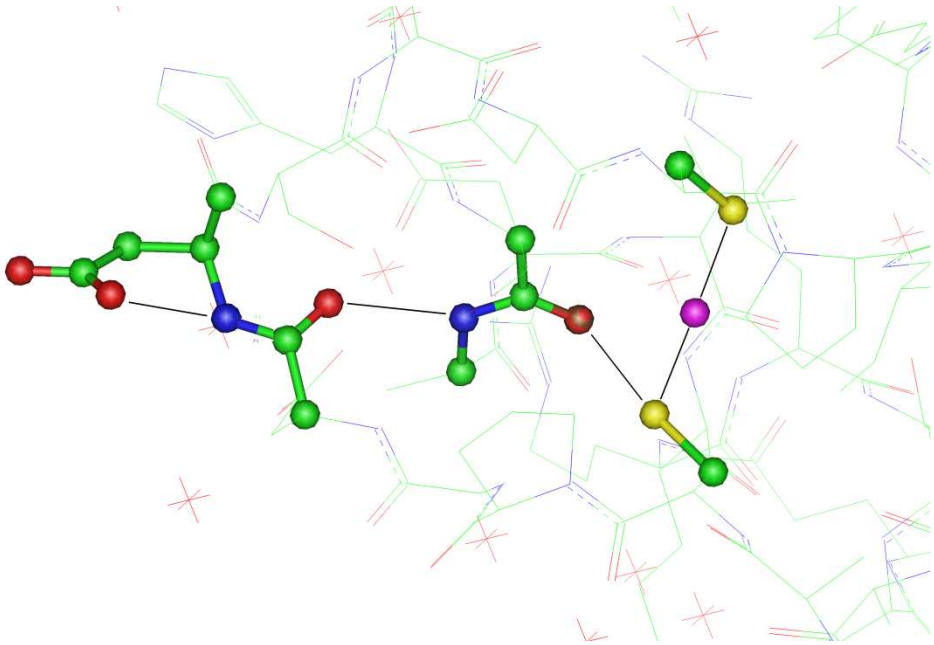


Figure 3.1 – XRD structure of CueR metal binding site.[54] The residues discussed later are shown in ball and stick representation; the proposed hydrogen bonds are shown in white.

3.2.1 Protein-ion communication in other MerR family members

All members of the MerR family show that binding the metal ion causes the entire protein to twist and thus change the shape of the DNA.[24] A schematic diagram of the protein is shown in figure 3.2 and a view of the crystal structure in figure 3.3.

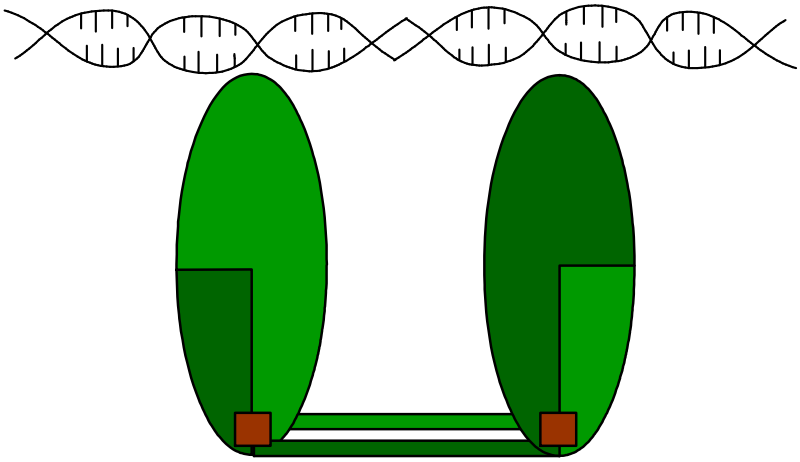


Figure 3.2 – Schematic diagram of MerR family bound to DNA. Each monomer is shown in an individual shade of green, the metal ion in brown and the DNA in black.

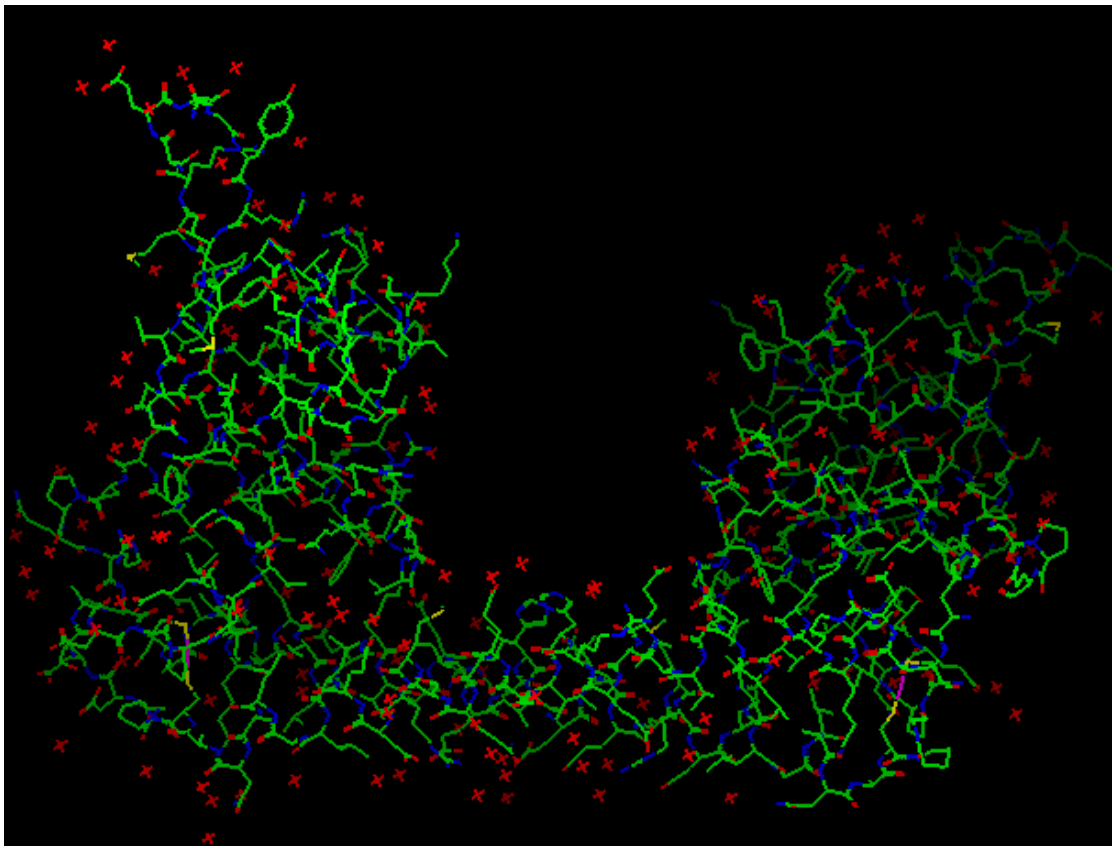


Figure 3.3 – XRD structure of CueR (1Q05)[54]

Each monomer has residues on both the left and right of the diagrams shown in figures 3.2 and 3.3. The DNA binds to the protein in two places: once in each monomer near the top of the diagram shown. In contrast, the metal ion binds near the bottom of figure 3.3. Two alpha helices run between the domains, and the ends of these helices act as pivots around which the protein may twist.[40] Therefore the exact conformations of these areas are extremely important.

In ZntR three cysteine residues bind to the metal ion: Cys79, Cys114 and Cys124.[46] Cys79 and Cys114 lie at opposite ends of the alpha helix so it is very important that there is an interaction between these residues to cause the global conformational change. If this interaction is directly through the metal ion it is easy to see how metal binding can be communicated through to the DNA.[47]

3.2.2 Protein-ion communication in CueR

As already mentioned, CueR has a different binding site to other MerR family members. In particular, only Cys112 and Cys120 (analogous to Cys114 and Cys124 in ZntR) interact directly with the metal. Such changes mean that it is much more difficult to see how the metal binding can affect the alpha helix pivot and thus the DNA transcription.

CueR is unable to bind to Cu(I) through the equivalent of ZntR's Cys79 as it has been mutated to Ser77. The oxygen of the serine residue is less likely to be deprotonated and is not a good ligand in its protonated form. In the crystal structure, this oxygen atom appears to be approximately 4.41Å away from Cu (I) and pointing in the opposite direction. Therefore Ser77 cannot be directly involved in metal-DNA communication. However, the hydrogen bond mentioned earlier could go some way to helping to explain this interaction.

The metal binding to the sulfur of the thiol will act as a Lewis acid and withdraw electrons from the sulfur-hydrogen bond and indirectly change the distance between Cys112 and Ser77.[36]

A simple hydrogen bond from a thiol to the oxygen of a carbonyl group may not be very strong as the oxygen will not be a good electron donor. However Ser77 can pass the hydrogen bond on through the amide hydrogen onto the backbone of Asp115, which can finally neutralise onto the side chain of Asp116 on the outside of the protein.[36] Thus the full hydrogen bonding scheme can be seen as shown in figure 3.4.

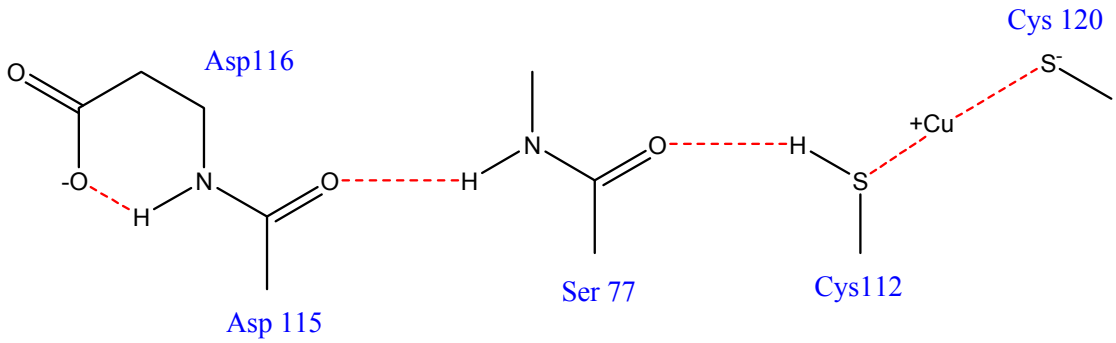


Figure 3.4 – The proposed hydrogen bonding chain linking the metal binding domain (Cys112) to the DNA binding domain (Ser77)

Figure 3.4 implies that hydrogen transfer could occur, though full transfer is unlikely to be the case due to unfavourable pK_as. However, the chain shown in figure 3.4 could help to strengthen the interaction between Cys112 and Ser77. This system was built and studied using Gaussian 98 and the results compared to those from XRD and EXAFS.[54,60] Also truncated versions of the hydrogen bonding network were used and these were compared as discussed on the following pages.

The simplest model studied (model 1) was that of the two thiols bound directly to the metal ion as shown in figure 3.5. The results of model 1 showed reasonable similarity to the XRD structure with a sulfur-copper distance of 2.06/2.05Å, compared to 2.14Å from

experiments.[54] The S-Cu-S bond angle was also similar to the experimental data, which predicted >175° and found 171.5° for model 1.[60] These results suggest that the model was reasonably good.

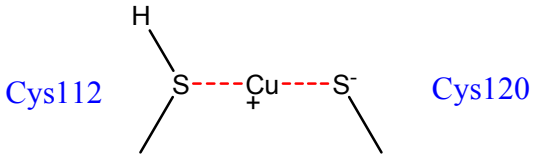


Figure 3.5 – A diagram of CueR model 1

Model 1 was then extended (to produce model 2) to include the hydrogen bond to the backbone of Ser77 as shown in figure 3.6. The geometry of model 2 was even closer to that of the crystal structure, with a sulfur-copper distance of 2.10/2.15Å and a sulfur-copper-sulfur bond angle of 171.62°. These results imply that the addition of this hydrogen bond brought the calculated values even closer to the experimental values.[54,60] The lengthening of these bonds may be due to the formations of the O-H hydrogen bond which will increase the negative charge on the sulfur atom. This will cause repulsion between the sulfur atoms, and thus the distances between each of the sulfur atoms and the copper ion. It should also be noted at this point that the δ- charged cysteine residue is closer, and therefore more strongly bound, to the copper ion than the fully negatively charged cysteine residue. This suggests that the exact nature and strength of these bonds is highly dependent on the local conditions of the binding residues.

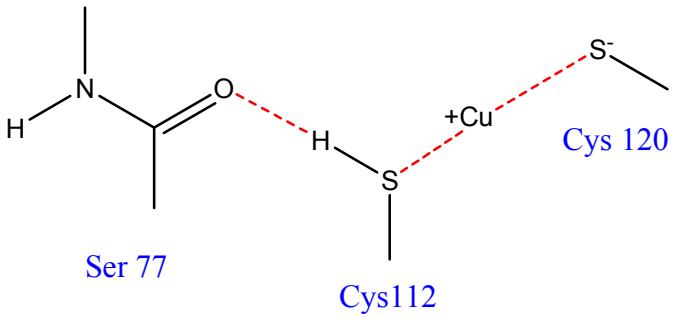


Figure 3.6 – A diagram of CueR model 2

The next model (model 3) took the above system one step further by adding another hydrogen bond acceptor onto the far side of the amide bond. It is designed to model the effect of the backbone of Asp115 as shown in figure 3.7. The results of model 3 could easily be superimposed upon the crystal structure with very little deviation in the structures as shown in figure 3.8. The sulfur-copper distances were 2.10/2.13Å and the angle was 172.17°.

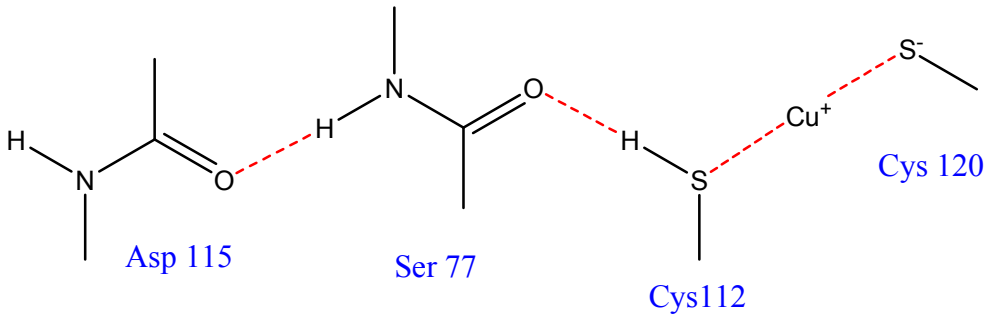


Figure 3.7 – A diagram of CueR model 3

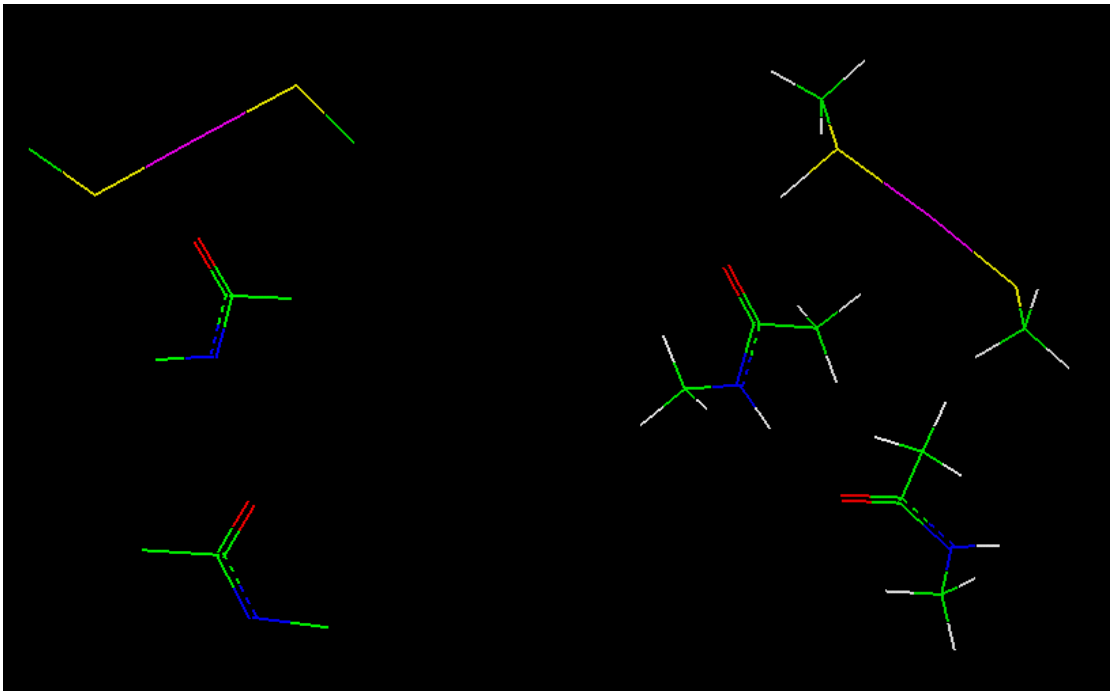


Figure 3.8 – A comparison of the structures of the calculated and the experimental binding site of CueR. The structure on the right is the result of calculations using model 3 (B3LYP/6-31G**) and the structure on the left is the XRD structure

The final model (model 4) for this section was the full hydrogen bonding chain as shown in figure 3.9. The calculation of model 4 took a very long time to run. Unfortunately it could not converge using any of the basis sets or DFT functionals tested, so the final structure could not be calculated. However, the lowest energy conformation produced during the calculation using B3LYP/6-31G** was even closer to the experimental data.[54,60] At this point the bond lengths were 2.10/2.12Å with a bond angle of 170.56° and the force had converged (max 0.000250 vs. threshold 0.000450, RMS 0.000043 vs. threshold 0.000300) but the displacement had not (max 0.016519 vs. threshold 0.001800, RMS 0.002877 vs. threshold 0.001200) showing that the potential energy surface is very flat.

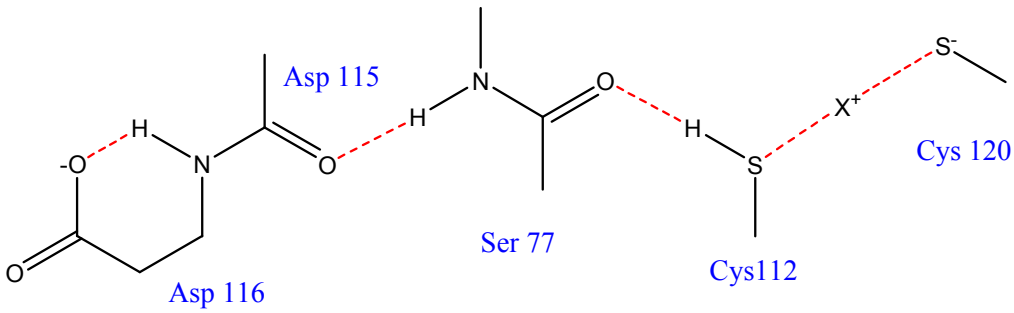


Figure 3.9 – A diagram of CueR model 4

All of the calculations on the model systems discussed so far were carried out as free gas phase calculations. In the protein the residues would be held in place by the backbone, so the minima proposed by these calculations could be geometrically impossible to achieve. The fact that all of the above calculations produced such similar results shows how perfectly the protein is set up to accept Cu(I).

These results give credibility to the idea that there is a hydrogen bond between Cys112 and Ser77 and raise the possibility that the proposed hydrogen bond may do the same as a direct cysteine-Zn bond would do in ZntR.[46,47]

3.3 Resting state

The XRD structure currently studied for CueR is the activator form of the protein. Since Cu(I) binds so tightly it is very difficult to crystallise a sample without Cu(I) being present.[54] As such the repressor structure is not known.

When the protein is in the activator form it twists the DNA exactly the right amount to cause transcription.[29] Any other form of the protein would twist the protein too much or too little, thus leading to a range of repressor forms.[28] In order to improve understanding of these repressor forms, it would be useful to find out what the local structure of the protein is when Cu(I) is not present.

3.3.1 Simple system calculations

Using model 1, various potential resting states were calculated and compared. When Cu(I) is removed it may be replaced by nothing (which may cause the thiol hydrogen to move around, removing the hydrogen bond) or various different ions found in biological systems could take its place. A simple alternative would be to protonate the thiolate to produce two thiols and form a new hydrogen bond between them. Also a water molecule could enter and form hydrogen bonds. All of these structures were tested for model 1.

To calculate the "affinity" of the system for a particular ion, a reference state was needed. These calculations were based on an "empty" system, which involved no positively charged ion present between the two sulfur atoms. Upon minimisation, the hydrogen atom (which would be involved in a hydrogen bond to Ser77 in the protein) swung around to be form a hydrogen bond to the thiolate, thus removing the repulsive effect of the sulfur

atoms. All of the calculations were referenced to this "empty" system, shown in figure 3.10.

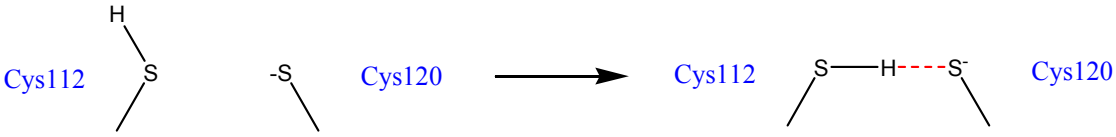


Figure 3.10 – A diagram of the change to the reference state used for the simple system calculations

Other effects are important to these calculations, most notably solvent effects. These can be taken account of by subtracting the solvation energy for the ion from the energy of binding.[138] As all of the calculations are based on the combination of the nuclei and electrons, it is also important to calculate the energy of each of the species to be combined individually. Therefore, the energy of binding can be calculated from equation 3.1. The energies and geometries of various calculations using model 1 are shown in table 3.1.

$$\Delta H_{total} = \Delta H_{f,\,bound} - \Delta H_{f,\,empty} - \Delta H_{f,\,ion} - \Delta H_{hyd,\,ion}$$

Equation 3.1 – Calculation of binding energy from calculable elements

Where:

ΔH_{total} is the affinity of the system for the ion, compared to the ion free in solution

$\Delta H_{f,\,bound}$ is the enthalpy of formation of the complete system

$\Delta H_{f,\,empty}$ is the enthalpy of formation of the reference structure (the protein with nothing bound)

$\Delta H_{f,\,ion}$ is the enthalpy of formation of the free ion

$\Delta H_{hyd,\,ion}$ is the enthalpy of hydration of the free ion[138]

Ion	Affinity (kJ mol^{-1})	HS-X^{+} (Å)	$^{-}$S-X^{+} (Å)	HS-S^{-} (Å)	HS-X^{+}-S^{-} (^{o})
Cu^{+}	-398	2.06	2.05	4.10	171.75
"H^{+}"	-411	2.79	1.33	4.04	155.15
H_2O	-11	3.29	2.17	4.77	95.19
"H_3O^{+}"	-394	2.21	2.48	5.27	102.17
Cu^{2+}	+461	2.12	2.04	4.15	170.60
Zn^{2+}	+152	2.28	2.10	4.36	170.43
Mg^{2+}	+290	2.50	2.25	4.72	166.89
Ca^{2+}	+357	2.93	2.59	4.08	94.97
Na^{+}	-164	2.86	2.54	4.26	104.06
K^{+}	-143	3.29	2.92	4.21	84.97
Empty	0(defined)	1.47	1.88	3.33	168.16
XRD	-	2.24	2.21	4.45 (4.27)	175.32

Table 3.1 – Formation energies and selected geometric parameters for minimised structures of CueR model 1 (Figure 3.5), incorporating various ions (calculated using B3LYP/6-31G**)

The data in table 3.1 suggests that doubly charged ions (such as Zn^{2+}, Mg^{2+} *etc*) will not bind due to a positive binding energy; conversely, any singly charged ion (*e.g.* Na^{+}, K^{+}) will bind, although these may not bind as tightly as Cu^{+}. However, this is assuming that the reference state is the resting state of the protein, which is unlikely to be the case. As such, the actual energies shown in table 3.1 are irrelevant; the relative energies must be considered.

The data in table 3.1 suggests that the tightest binding ion is "H^{+}" (it has the highest binding affinity). Looking at the geometry of this result, a protonated thiol is produced; this arrangement is likely to be the resting state. Cu(I) has a binding energy which is only slightly lower than H^{+}, although these results imply that it would be less strongly bound than the resting state. The only ions which will bind, assuming that the protonated thiol is the resting state, are Cu^{+}, H^{+} and H_3O^{+}; other ions with apparently favourable affinities will not bind, as those energies are at least 200 kJ mol^{-1} lower.

In gas phase calculations, other ions (mainly +2 charged ions, especially Cu (II)) appear to bind more strongly than Cu(I) or H^+, but the energy required to remove their solvent shells is too great in comparison to the calculated binding affinities.

The geometry of binding is also important as it may affect the global conformation of the protein; some conformations may not be possible (see section 3.4). The most important interaction is the one between Cys112 and Ser77; however this interaction cannot be studied using this model, as Ser77 is not present.

3.3.2 Other ions in more complex models

To see what affect the hydrogen bonding chain had on the affinities of these ions, the same process was used for model 2 as for model 1. For model 2 the energies were again referenced to a system with no ion. In this case the system rearranged itself to remove the S^--$S^{\delta-}$ interaction and thus the reference state is structurally and energetically different from the structure found in the protein. The results of these calculations are shown in table 3.2.

Ion	E (kJ mol^{-1})	HS-S$^-$ (Å)	SH-O=C (Å)
Cu^+	-3693	4.24	1.83
H^+	-351	4.05	2.05
Cu^{2+}	-3303	4.15	1.02
Zn^{2+}	-3383	4.32	1.02
Mg^{2+}	-889	4.63	1.02
Na^+	-1254	5.32	3.20
K^+	-2070	6.13	2.27
Empty	0	8.03	1.92
XRD	-	4.27	approx. 2.2

Table 3.2 – Formation energies and selected geometric parameters for minimised structures of CueR model 2 (Figure 3.6), incorporating various ions (calculated using B3LYP/6-31G**)

As was seen for model 1, the reference structure is not the resting state of the protein, so the energies shown in table 3.2 are correct relative to one another but do not yield absolute binding energies. Therefore the binding energy as shown in table 3.2 would need to be

strongly negative for an ion to bind (possibly at least 3500 kJ mol^{-1}). Assuming that energies over 3500 kJ mol^{-1} are the favoured forms, these results again suggest that Cu(I) should bind and other ions would show unfavourable binding. For model 2 the protonated thiol ("H^+") does not appear to be a favourable resting state; therefore the assumption that the resting state involves a protonated thiol may be incorrect. However, such a dramatic change in binding affinity suggests that the structure calculated in model 2 may be incorrect for the protonated thiol.

The geometries of these results show that the distance between Cys112 and Ser77 can vary a fair amount upon binding. The stronger Lewis Acids (Cu^{2+}, Zn^{2+}, Mg^{2+}) promote transfer of the proton onto the carbonyl of Ser77 as shown in figure 3.11. Proton transfer may occur as electron density is pulled away from the S-H bond and into the S-X^{2+} bond. This shift in electron density could transfer along the hydrogen bonding chain changing the interaction with Asp115 and thus the distance between Cys112 and Ser77, finally resulting in a change in the geometry of the entire protein. The nature of the proton transfer is discussed in section 3.5.

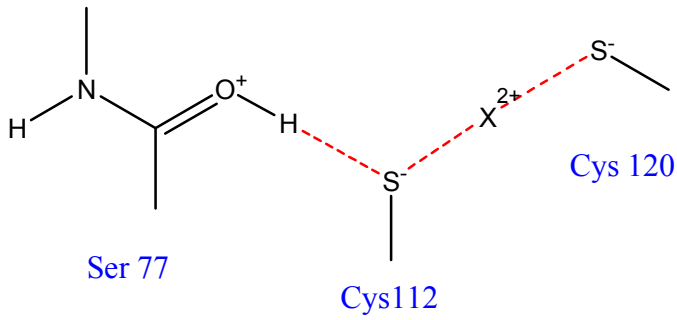


Figure 3.11 – A diagram of the result of proton transfer in model 2 (Figure 3.6)

3.3.3 Resting state conclusion

It appears that the resting state of CueR when Cu(I) is not present contains two protonated thiols. The calculations show Cys112 and Cys120 forming a hydrogen bond, bringing them

closer together. In this hydrogen bonded structure the distance between Cys112 and Ser77 increases to be larger than the crystal structure.[54]

3.4 Stationary models

As mentioned earlier, these calculations are based on "free" gas phase calculations where the cysteine residues have free movement. In the activator form of the protein these residues are likely to be more rigid, being held in place by the backbone and steric/electrostatic interactions with other parts of the protein.[54] The positioning and mobility of these residues may affect which ions can bind and the proposed hydrogen bonding chain. It is not obvious what the effect of repressor forms would have on the flexibility of the protein. The following models are all based on model 1 using a variety of different stationary parameters.

3.4.1 Stationary carbon model

If the backbone of the residues of interest is held in place by the rest of the backbone then the sulfur atoms could move and the carbons would be held in place. In the crystal structure the carbons are 6.03Å apart, so the next calculations were performed with this distance set.[54] Since there were no constraints on rotation around the carbon centres, the stationary carbon model still allowed the binding sulfur atoms to move and accommodate different sized ions. This motion meant that the geometries and thus the energies became very close to those for the unconstrained system and alternatives had to be studied.

3.4.2 Stationary sulfur model 1: 4.45Å

In the protein it may be more that the sulfur atoms are constrained than the carbon atoms due to other hydrogen bonds or the backbone being unable to flex and allow movement of

the β carbon. Thus the next few systems were based on constraining the distances between the sulfur atoms (the space in which the metal binds).

The first model held the sulfur atoms 4.45Å apart, as found in the better defined monomer of the crystal structure.[54] Again, various different metals were compared and the results are shown in table 3.3.

Ion	Affinity (kJ mol^{-1})	HS-X^{+} (Å)	$^{-}$S-X^{+} (Å)	HS-S^{-} (Å)	HS-X^{+}-S^{-} (^{o})
Cu^{+}	-1235	2.30	2.15	4.45	177.74
"H^{+}"	-1276	3.15	1.33	4.45	166.05
Zn^{2+}	-713	2.33	2.12	4.45	173.70
Mg^{2+}	-574	2.48	2.24	4.45	141.61
Ca^{2+}	-509	2.96	2.60	4.45	106.20
Na^{+}	-1030	2.88	2.54	4.45	110.10
Empty	0			4.45	
Cu^{+}(-H^{+})	-1152	2.23	2.22	4.45	179.01

Table 3.3 – Formation energies and selected geometric parameters for minimised structures of stationary sulfur model 1, incorporating various ions (calculated using B3LYP/6-31G**). The last line refers to a version of model 1 with the thiol being changed to a thiolate and Cu(I) being bound.

The results shown in table 3.3 are similar to those for the unconstrained system; as seen for the unconstrained models the protonated thiol is lowest in energy with Cu(I) bound next lowest. However, all of the energies are overall lower than those for the unconstrained system due to the distances being too great to form strong enough hydrogen bonds (in the empty system in particular). The calculations on the other ions did not run with the basis set and DFT functional used.

3.4.3 Stationary sulfur model 2: 4.27Å

The distance used for the first stationary sulfur system may not be a good model for the enzyme; in the less well defined monomer of the crystal structure the sulfur atoms are 4.27Å apart. The above procedure was repeated and afforded the results in table 3.4.

Ion	Affinity (kJ mol^{-1})	HS-X$^+$ (Å)	$^-$S-X$^+$ (Å)	HS-S$^-$ (Å)	HS-X$^+$-S$^-$ (^{o})
Cu^+	-1252	2.16	2.11	4.27	176.13
"H^+"	-1273	3.00	1.33	4.27	158.40
Cu^{2+}	-764	2.18	2.10	4.27	175.44
Zn^{2+}	-709	2.24	2.08	4.27	163.85
Mg^{2+}	-567	2.48	2.24	4.27	129.90
Ca^{2+}	-505	2.95	2.60	4.27	100.58
Na^+	-1026	2.87	2.53	4.27	104.26
K^+	-1005	2.24	2.08	4.27	163.85
Empty	0			4.27	
$Cu^+(-H^+)$	-1144	2.14	2.13	4.27	177.98

Table 3.4 – Formation energies and selected geometric parameters for minimised structures of stationary sulfur model 2, incorporating various ions (calculated using B3LYP/6-31G**)

Again, all of these structures are less favourable than the unconstrained system and H^+ appears to bind more strongly than Cu(I).

3.4.4 Stationary sulfur model 3: 4.10Å

The system may wish to remain at the distance seen for Cu(I) depending on the flexibility of that region of the protein. If this DFT functional/basis set (B3LYP/6-31G**) predicts that model 1 will minimise to a distance of 4.10Å then setting this value might result in a better model. Thus the results in table 3.5 were obtained *via* modelling stationary sulfur model 3. Once again, all of these structures are less favourable than the unconstrained system. H^+ appears to bind more strongly than Cu(I).

Ion	Affinity (kJ mol^{-1})	HS-X^{+} (Å)	$^{-}$S-X^{+} (Å)	HS-S^{-} (Å)	HS-X^{+}-S^{-} (^{o})
Cu^{+}	-1492	2.06	2.05	4.10	171.73
"H^{+}"	-1506	2.84	1.33	4.10	156.85
Cu^{2+}	-999	2.10	2.03	4.10	166.45
Zn^{2+}	-930	2.22	2.07	4.10	146.23
Mg^{2+}	-796	2.62	5.74	4.10	39.98
Ca^{2+}	-737	2.93	2.59	4.10	95.61
Na^{+}	-1258	2.86	2.53	4.10	98.81
K^{+}	-1238	3.30	2.89	4.10	82.76
Empty	0			4.10	
$Cu^{+}(-H^{+})$	-1143	2.05	2.05	4.10	174.81

Table 3.5 – Formation energies and selected geometric parameters for minimised structures of stationary sulfur model 3, incorporating various ions (calculated using B3LYP/6-31G**)

3.4.5 Stationary sulfur model 4: 5.27Å

An alternative resting state of the protein may be with "H_3O^{+}" bound. The minimised structure for this system involves a protonated thiol (as seen for H^{+}) with a water molecule bridging the gap and forming hydrogen bonds between the cysteine residues. This change in geometry forces these residues apart and could cause a change in the shape of the protein. Model 4 has the sulfur atoms 5.27Å apart. To test the energies of various other ions in this potential resting state the same procedure as before was used to obtain the results in table 3.6. These results show that Cu(I) still binds extremely strongly. However, other singly charged ions also appear to bind fairly well.

Ion	Affinity (kJ mol^{-1})	HS-X^{+} (Å)	$^{-}$S-X^{+} (Å)	HS-S^{-} (Å)	HS-X^{+}-S^{-} (^{o})
Cu^{+}	-1356	3.22	2.05	5.27	176.94
Cu^{2+}	-856	3.09	2.18	5.27	178.54
Mg^{2+}	-764	2.92	2.35	5.27	178.85
Ca^{2+}	-724	2.99	2.61	5.27	140.37
Na^{+}	-1242	2.90	2.53	5.27	151.72
K^{+}	-1221	3.46	2.91	5.27	111.26
Empty	0			5.27	
$Cu^{+}(-H^{+})$	-1223	2.64	2.63	5.27	179.08

Table 3.6 – Formation energies and selected geometric parameters for minimised structures of stationary sulfur model 2, incorporating various ions (calculated using B3LYP/6-31G**)

3.4.6 Stationary model conclusion

The above results show similar trends for all of the systems. It appears that the resting state of the protein is likely to be two protonated thiols (whatever this may do to the conformation of nearby residues and long range interactions). The proposed resting state will be able to form good hydrogen bonds from sulfur to sulfur, and the change of Cu(I) to H^{+} changes the strength of the hydrogen bond to Ser77 (as seen by the change in hydrogen bond length) and thus the global conformation.

3.5 Hydrogen transfer

The previous experiments showed that it may be possible for the proposed thiol hydrogen to transfer onto the backbone of Ser77 in the presence of strong enough Lewis Acids. The largest model (with the side chain of Asp116) also shows that the hydrogen may transfer with this many hydrogen bonds as shown in figure 3.12.

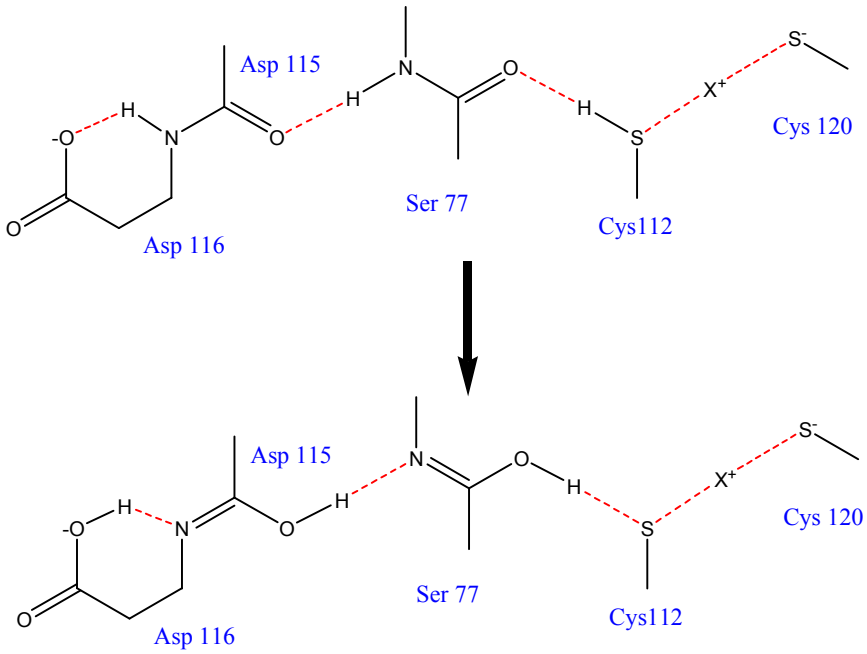


Figure 3.12 – Hydrogen transfer in the proposed hydrogen bonding chain

It seems unlikely that such a transfer would occur according to the normal pK_a values; overall the hydrogen is being transferred from a group with a pK_a of around 10 (thiol) to around 4 (carboxylic acid).[9] However, such a transfer occurs in serine protease enzymes.[139] These enzymes use hydrogen bonds to weaken a serine O-H bond to make the oxygen a better nucleophile. Their hydrogen bonding chain goes from the serine OH onto the deprotonated nitrogen of a histidine residue and on from the histidine's opposite NH to neutralise onto the carboxylic side chain of an aspartate/glutamate residue as shown in figure 3.13. This mechanism of partial hydrogen transfer is relatively similar to the system proposed for CueR that starts from a more acidic SH and ends up on a carboxylic acid side chain.

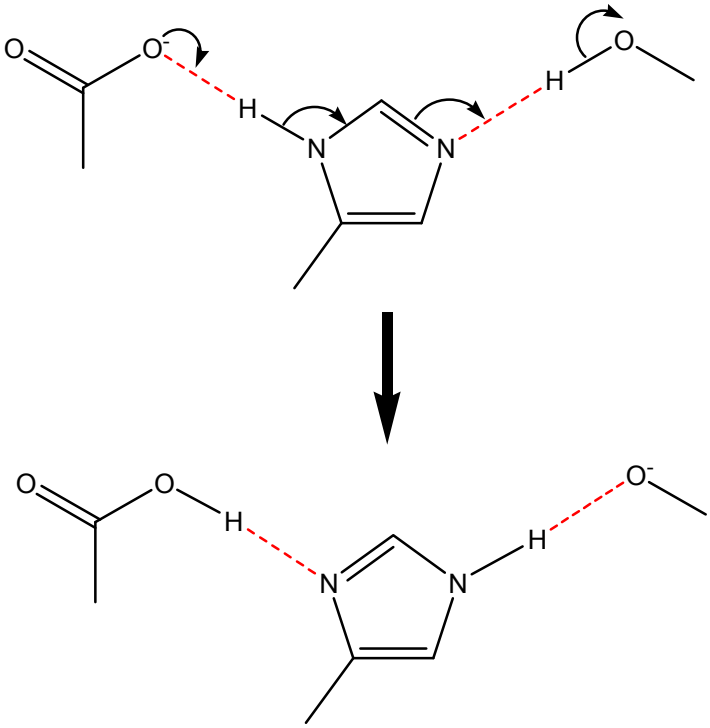


Figure 3.13 – A diagram of electron transfer in the serine protease mechanism

The proposed system for CueR is different from the serine protease mechanism in two main ways; the initial hydrogen donor is stronger (SH *vs.* OH) and the intermediate hydrogen acceptor is weaker (CO: *vs.* N:). The catalytic triad of serine proteases and related enzymes can vary and may include backbone amides and various other groups.[140,141] Neither serine proteases nor the proposed CueR system would be likely to result in full hydrogen transfer, but the hydrogen donor will become more negatively charged and the distances between the groups will be affected.

Proton transfer could be a way of communicating the binding of Cu(I) to the DNA-binding domain. Even if the transfer is not complete the S-H bond length is likely to stretch and change both the relative positions of the chains and hydrogen bond strength. Various different systems were designed to test this theory.

3.5.1 Proton affinities

The simplest way of measuring the energy of hydrogen transfer is through the proton affinity of the system. Each metal ion will pull on the electrons differently and the energies of the system with and without hydrogen (which has no enthalpy of formation for H^+) will vary according to the length and strength of these bonds as shown in figure 3.14. The ions used for the previous section were again tested and yielded the results in table 3.7.

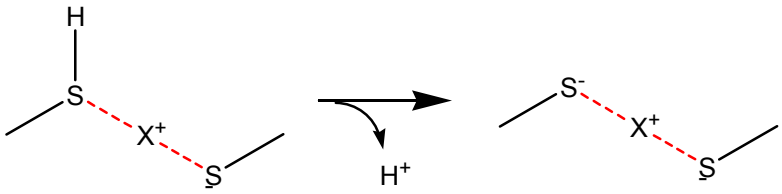


Figure 3.14 – A diagram of the system used for the proton affinity calculations

Ion	Proton Affinity (kJ mol^{-1})
Cu^+	-1369
"H^+"	-1503
"H_3O^+"	-1493
Zn^{2+}	-861
Mg^{2+}	-878
Ca^{2+}	-959
Na^+	-1381
K^+	-1413

Table 3.7 – Proton affinities for various ions for the system shown in figure 3.14

The data in table 3.7 shows a general trend of systems involving stronger Lewis Acids having a lower affinity for the hydrogen. These ions would most likely be the cause of the largest change in the shape of the protein.

There is a difference between the proton affinities for the Cu(I) bound system and the protonated thiol of approximately 150 kJ mol^{-1}. Such a change in proton affinity would cause the hydrogen to be pulled closer to the sulfur in the case of the protonated thiol and thus reduce the strength (and increase the length) of the hydrogen bond to Ser77. With

Cu(I) bound, the hydrogen bonding chain would be stronger and the two inter-domain alpha helices would be held together more.

3.5.2 Sulfur-hydrogen energy plot 1

Proton affinities can be useful for measuring how easy it is to remove the hydrogen from the system to an infinite distance. Such calculations should show the order of the affinities for a system where the proton acceptor is part of the same molecule (as is the case for the above system) but will not measure length or strength of the hydrogen bond quantitatively. To calculate this change a model of the acceptor can be used as shown in figure 3.15.

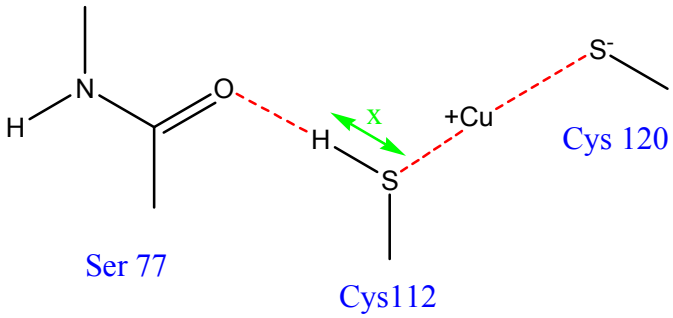


Figure 3.15 – A diagram of the system used to generate the results for sulfur-hydrogen energy plot 1

The sulfur-hydrogen distance (x) was fixed at various values whilst the rest of the system was allowed to minimise. If there is hydrogen transfer then a double-well energy plot will be observed. This potential energy surface would correspond to the hydrogen either being located on the sulfur (thiol) or the carbonyl oxygen as shown in figure 3.16. The sulfur-hydrogen distance was varied to produce the results shown in figure 3.17.

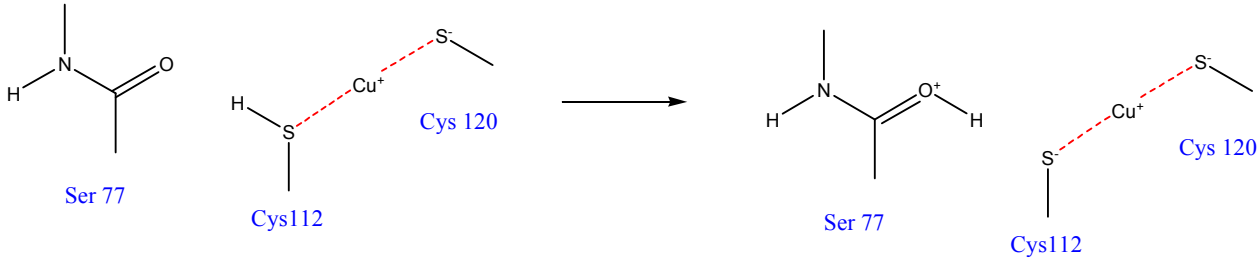


Figure 3.16 – A diagram of hydrogen transfer in S-H plot 1

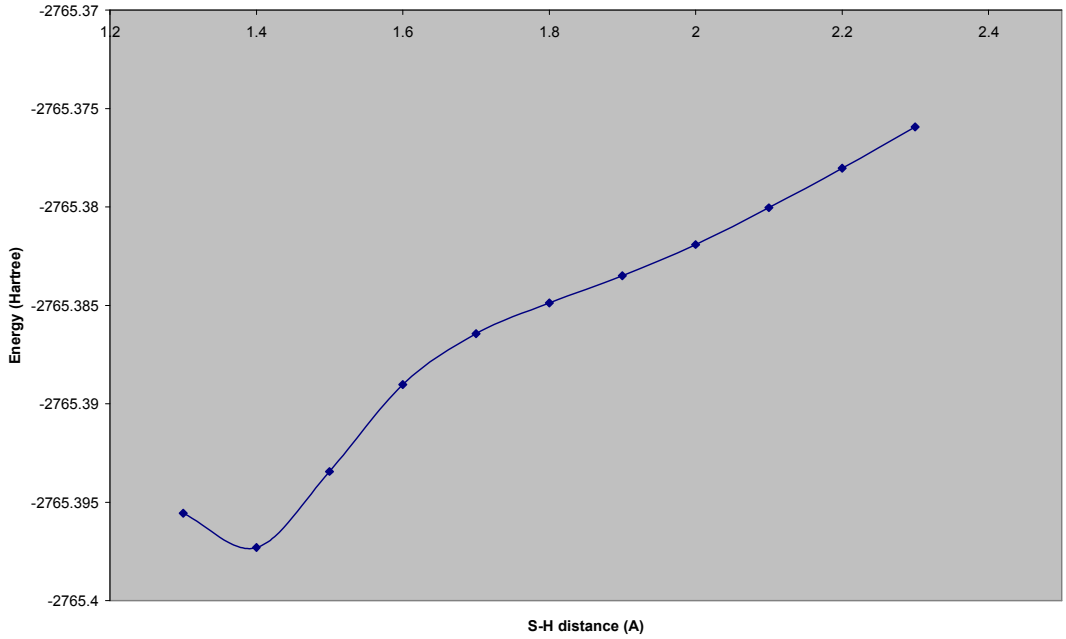


Figure 3.17 – Energy as a function of S-H bond length in sulfur-hydrogen plot 1 model (calculated using B3LYP/6-31G**)

Figure 3.17 shows an energy minimum at around 1.4 Å that corresponds to the equilibrium bond length. Compressing the bond (to less than 1.4Å) causes the nuclei to repel one another, whereas stretching it removes the favourable interaction between the sulfur and the hydrogen. The raw data for this graph may be found in Appendix I, table I.1.

The suggested double-well hydrogen bond did not appear as the acceptor used in this case is not electronegative enough to allow the formation of $C{=}OH^+$. Though not constrained, the distance between H^+ and C=O remained approximately constant. As such, the structure used in plot 1 may still not be a good model for the system.

3.5.3 Sulfur-hydrogen energy plot 2

The next energy plot modelled the effect of removing the copper from the system, and had the benefit of speeding up calculation times, thus producing the system shown in figure 3.18.

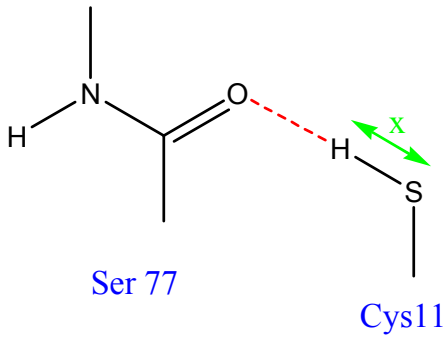


Figure 3.18 – A diagram of the system used to generate the results for sulfur-hydrogen energy plot 2

Calculations on the structure shown in figure 3.18 should show a different equilibrium bond length and is less likely to show a double-well hydrogen bond (or a higher energy well for the longer bond length). The results are shown in figure 3.19.

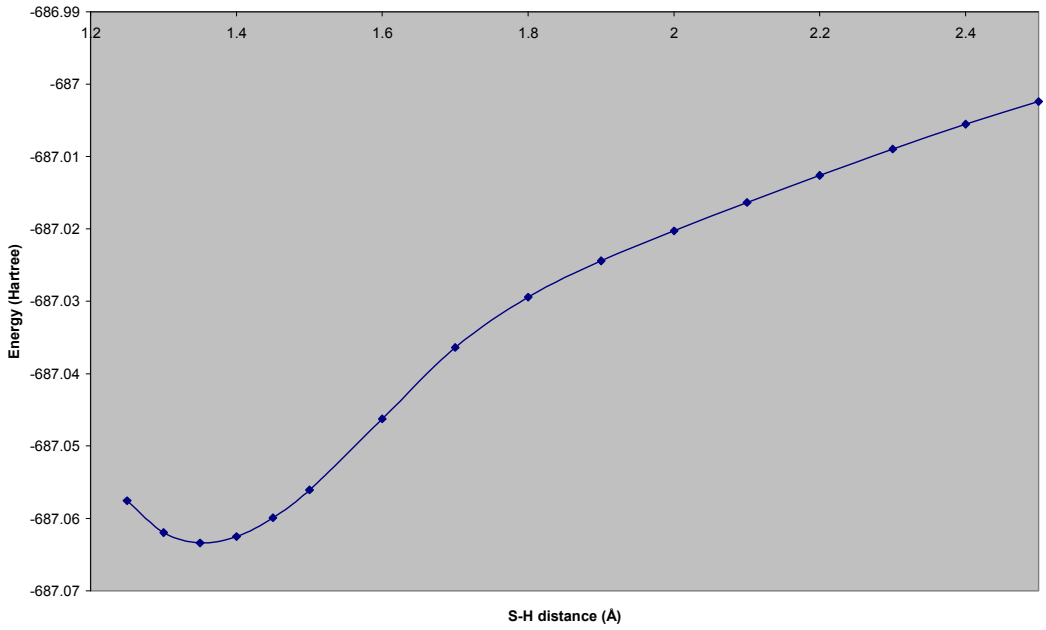


Figure 3.19 – Energy as a function of S-H bond length in sulfur-hydrogen plot 2 model (calculated using B3LYP/6-31G**)

Figure 3.19 shows an equilibrium bond length of approximately 1.35Å (vs. 1.4Å for the previous system) and a very similar energy profile. This distance suggests that the proton acceptor may not be strong enough to show an effect on the binding of copper. This lack of strength is not unusual as a carbonyl group is a very weak Lewis Base unless activated by

another group (such as a deprotonated carboxylic acid). The raw data for this graph may be found in Appendix I, table I.2.

3.5.4 Sulfur-hydrogen energy plot 3

The previous experiment highlights the need to have a better acceptor such as that found in the protein. To model a better acceptor the full hydrogen bonding chain could be used (model 4) as shown in figure 3.20.

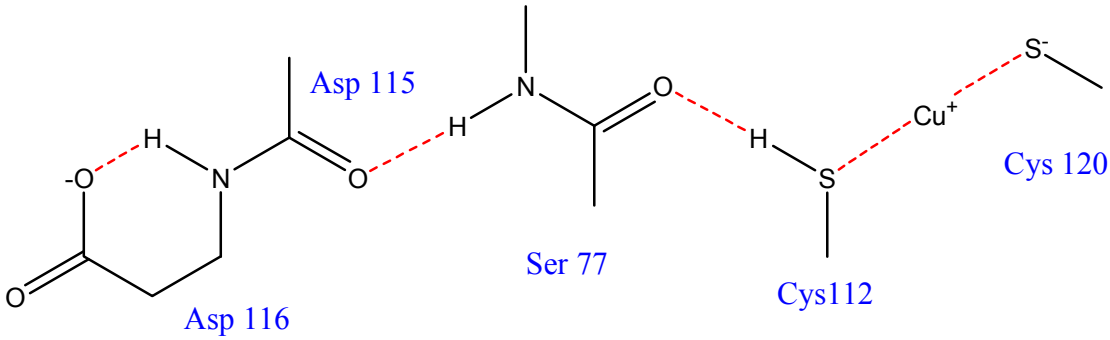


Figure 3.20 – A diagram of CueR model 4

Unfortunately calculations on model 4 did not complete previously (see section 3.2.2) so it could not be used directly. A smaller model with similar properties and fewer fragments was designed to speed up the calculation as shown in figure 3.21.

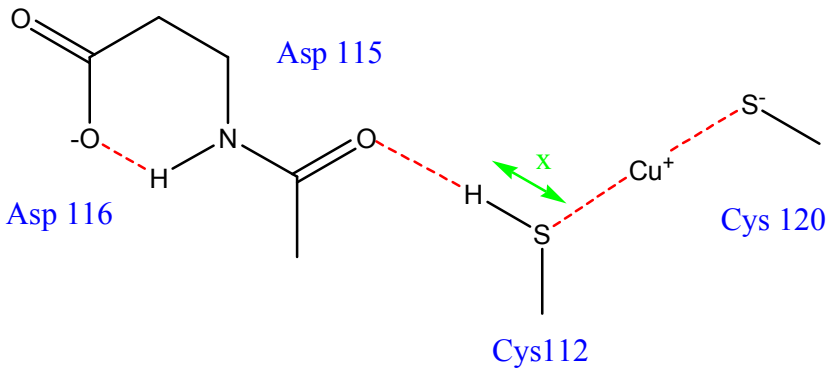


Figure 3.21 – A diagram of the system used to generate the results for sulfur-hydrogen energy plot 3

The structure shown in figure 3.21 shortens the distance between the carboxylic acid proton acceptor and the thiol proton donor. If the amide group is acting as a way of passing the proton and the binding information then this model should behave in a similar manner

to the real system (see chapter 4 for more information on the global conformation change of this mutation). The results are shown in figure 3.22.

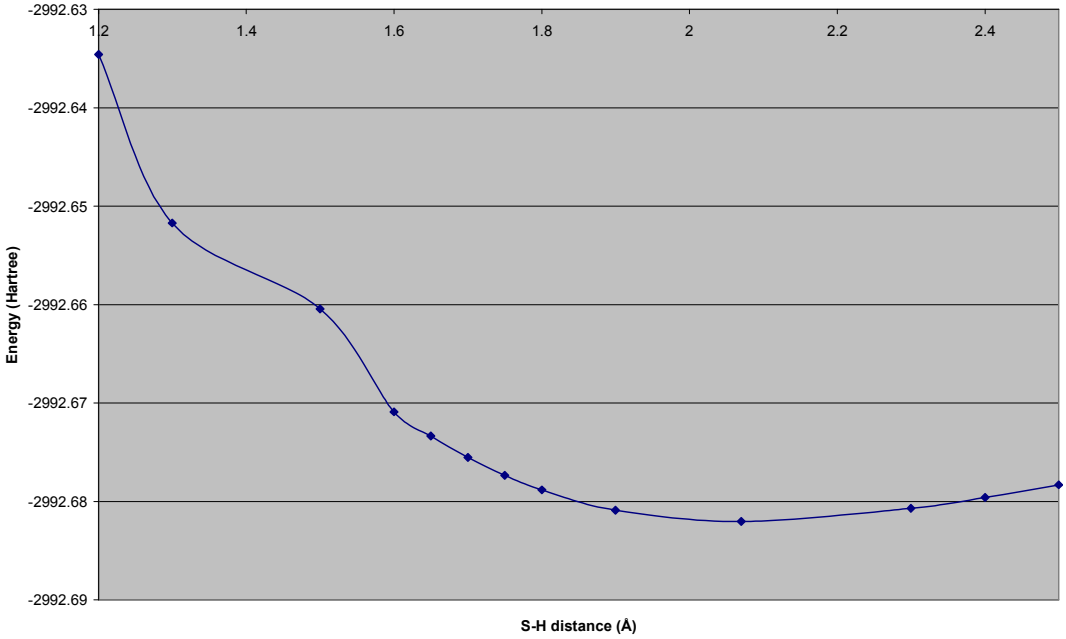


Figure 3.22 – Energy as a function of S-H bond length in sulfur-hydrogen plot 3 model (calculated using B3LYP/6-31G**)

Again there is no evidence of a second well for this system. However, the equilibrium bond length has become much longer (2.07Å) and it appears that the hydrogen has transferred fully onto the carbonyl oxygen. The bump on the left hand side of the graph is caused by this shift. Attempts were made to fill in the points between those shown but the calculations would not converge. The raw data for this graph may be found in Appendix I, table I.3.

3.5.5 Sulfur-hydrogen energy plot 4

As with the first energy plot it would be useful to compare the above truncated version to one without the copper ion being present. Thus the system shown in figure 3.23 was used, with the results shown in figure 3.24.

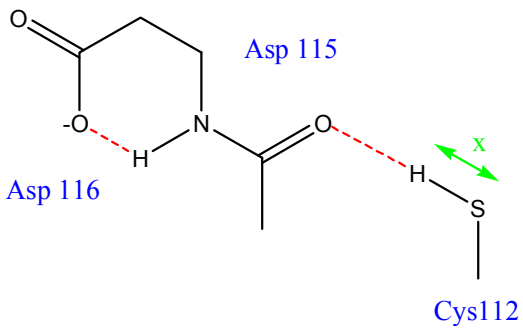


Figure 3.23 – A diagram of the system used to generate the results for sulfur-hydrogen energy plot 4

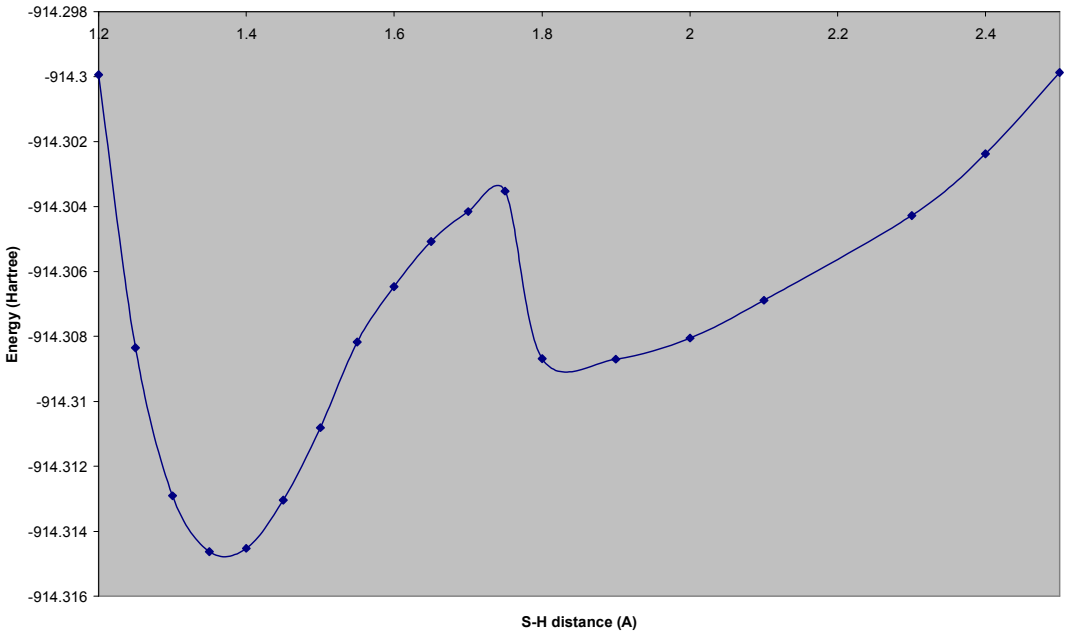


Figure 3.24 – Energy as a function of S-H bond length in sulfur-hydrogen plot 4 model (calculated using B3LYP/6-31G**)

Figure 3.24 shows the hydrogen bond expected for the other systems. There is a low energy minimum (corresponding to -S-H…O=C<) with a transition state (high energy barrier, -S…H…O=C<) leading to a higher energy minimum (corresponding to -S…H-O=C<). There appears to be a jump from one curve to another: this jump is the change between the hydrogen positions on the amide/thiol. The raw data for this graph may be found in Appendix I, table I.4.

3.5.6 Proton transfer conclusion

The results of the S-H plots may be combined to produce figure 3.25. Figure 3.25 shows plots 1, 2 and 4 to all have energy minima around 1.4Å, close to that seen in the crystal structure and the EXAFS data.[54,60] Conversely, plot 3 has its energy minimum at 2.09Å, showing full hydrogen transfer to the carbonyl group.

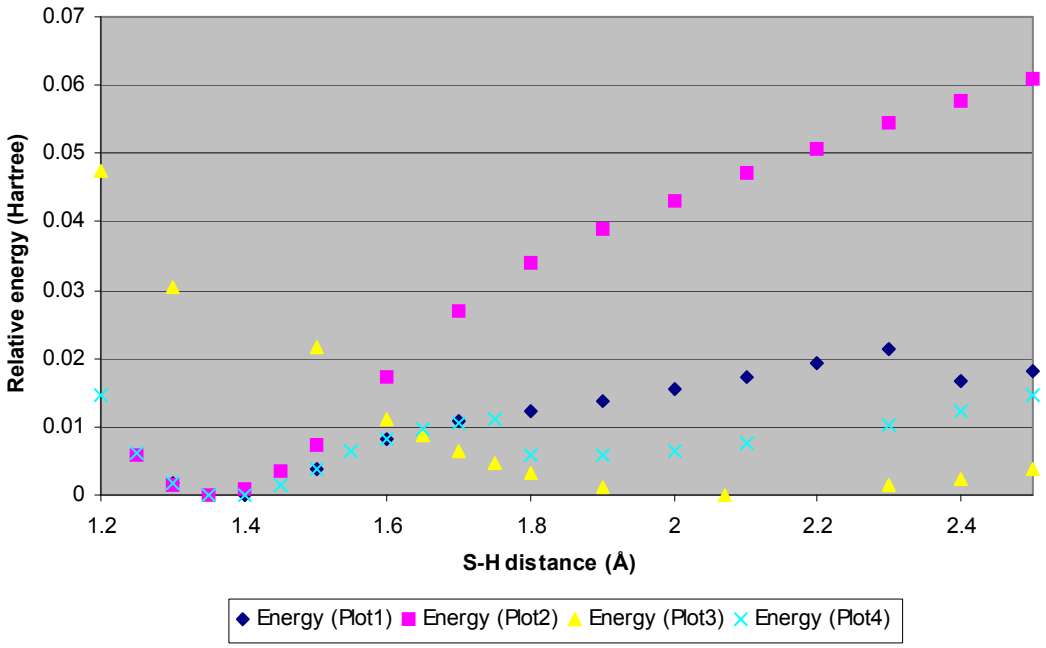


Figure 3.25 – Combined sulfur-hydrogen plots 1, 2, 3 and 4. The energy minimum for each curve is set to zero, but relative energies are unaffected

The above data shows that there is a large difference in the strength of the hydrogen bond depending on whether or not Cu(I) is bound. The proton affinities show that there is 150kJmol^{-1} less energy holding the hydrogen onto the thiol when Cu(I) is bound than when it is not. The first two energy plots suggest that the backbone of Ser79 alone is not a strong enough hydrogen acceptor to change the shape of the binding site and thus the global conformation upon binding of Cu(I). The final two energy plots show that, for a truncated model of this system, the hydrogen will transfer onto the carbonyl group when Cu(I) is present, and that hydrogen transfer is less likely to occur when Cu(I) is absent.

3.6 Model system conclusion

The data presented in this chapter gives more information on the unusual binding site of CueR. Biological data shows two cysteine residues are essential for the correct function of CueR, whilst XRD and EXAFS data give further chemical detail.[41-43,54,60]

The results of the experiments in this chapter identify the resting state of the binding site, which contains two protonated thiols. Cu(I) is the only metal ion able to bind; other metal ions will not bind due to unfavourable binding enthalpies.

A hydrogen bonding chain essential to the function of the protein is identified by the results in this chapter. This chain links together the metal binding and DNA binding domains. The four models discussed in section 3.3 explained the need for a hydrogen bonding chain comprising of residues Cys120-Cu-Cys112-Ser77/Ala78-Asp115/Asp116. As more of this hydrogen bonding network was modelled, the calculated structure became closer to the crystal structure. The important interaction in this chain is the hydrogen bond between Cys112 and Ser79, the analogues of which affect the function of MerR and ZntR.[43,47,59,142] The presence of Cu(I) changes the distance between these residues, as confirmed by the sulfur-hydrogen energy plots shown in section 3.5.

4. ZNTR

4.1 Introduction

As mentioned in section 3.2.1, there are several members of the MerR family.[16] Of these, the mechanism of ZntR has been studied thoroughly and several X-ray crystal structures produced.[54]

The basic structure of ZntR appears to be similar to that of CueR including the DNA binding domain.[24] However the metal binding site seems to be very different, causing the switch in selectivity and sensitivity. Whilst CueR exhibits 10^{-21} M affinity for Cu(I) and selectivity for only group XI metals, ZntR exhibits only 10^{-14} M affinity for Zn(II). This switch is important as Zn(II) is an essential nutrient so any bound to other proteins (such as zinc fingers) must remain so, whilst excess must be removed, whereas Cu(I) has no use in cells, so all Cu(I) must be removed.

4.2 ZntR experimental binding site

It would be useful to find what the differences are between the binding modes of ZntR and CueR and how these affect the switch in selectivity and sensitivity.[24] The binding site of ZntR is very similar to that for CueR with a few key differences. In the crystal structure there appear to be two Zn(II) ions binding per monomer. These may not both be required for activation as the concentration of Zn(II) used in crystallisation is very high and any sites that could be occupied by a positively charged ion are more likely to be occupied than *in vivo*. The other major difference is that the binding site that is retained in other members of the MerR family has changed Ser77 (in CueR) to Cys79 in ZntR. Finally, an additional phosphate ion has bound directly to both of the

adjacent zinc ions. It is not clear either whether this phosphate is an essential part of the protein, whether it could be replaced by a water molecule/hydroxide group or whether it is merely a product of crystallisation from a phosphate-based buffer solution. Thus, overall the binding site may appear as shown in figure 4.1 in which the left-hand Zn(2) occupies a position analogous to that of Cu(I) in CueR.[54]

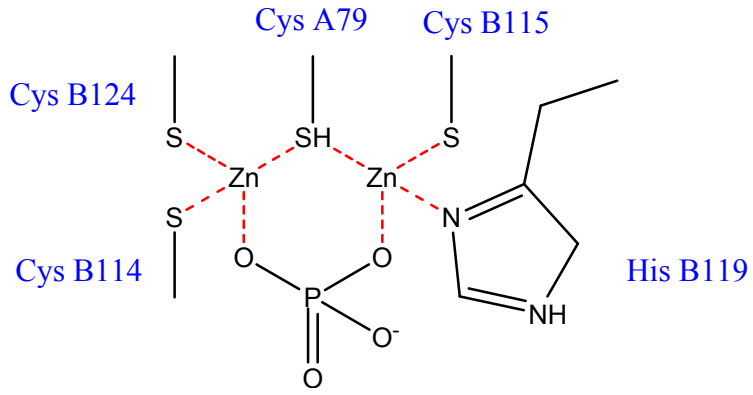


Figure 4.1 – A diagram of the metal binding site of ZntR as seen in the crystal structure[54]

4.3 ZntR initial model

The binding site above is too large to model using QM methods, especially due to the number of fragments. As mentioned previously, the right hand zinc ion may not be required, thus removing the need to model it or the ligands attached to it. The phosphate can also be approximated to a hydroxide ion as hydroxide has a similar charge on the oxygen, without the need for a computationally demanding phosphorous atom.[143]

All calculations in this chapter were performed using Gaussian 98, the basis set 6-31G**, z-matrix optimisation and the DFT functional B3LYP for electron correlation calculations.

Figure 4.2 shows the initial model used for the active site of ZntR (model 1). As with the CueR model, a variety of ions were tested in model 1 to see how they would affect

the binding affinity. Many of the results could not be calculated due to SCF convergence problems. The DFT functional was changed to both mpw1pw91 and b1lyp, the starting geometry was changed and various other techniques were used, all with no positive results.

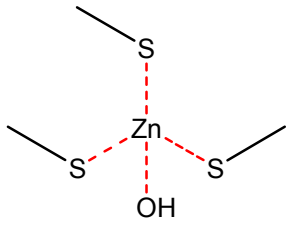


Figure 4.2 – A diagram of ZntR model 1

The singly charged ions often failed due to the overall highly negative charged nature of the system (3-). They frequently formed metal hydroxide complexes with the thiolate ions being repelled by the negative charge. The production of such complexes during the minimisation shows that the metal-hydroxide bond is probably more stable than the metal-thiolate bond, and that thiolate is more stable alone than hydroxide. Energies could not be derived for these systems as they could not converge. It seems unlikely that all of the cysteine residues would be deprotonated as the pK_a of isolated hydrated cysteine is around 10.

For Cu (I) the calculation again did not complete. However, the calculated structure had become a linear hydroxide-copper-thiolate complex with two separate thiolate ions. 1^+ charged ions in group 11 often form 2-coordinate linear structures due to the similarity in energy of the outer *s*, *p* and *d* orbitals, permitting the formation of collinear *spd* hybrid orbitals.[144] The fact that the minimisation produces this structure shows that the model is working correctly.

The 2+ charged ions did complete their calculations and all formed tetrahedral complexes. The calculated energies are shown in table 4.1. These results take into account hydration energies of the metal ions and are all compared to the Zn(II) bound structure. These results show that Zn(II) is the strongest binding 2+ charged ion. These results need comparing to those for singly charged ions (which appeared to bind most strongly for CueR).

Ion	Affinity (kJ mol^{-1})
Zn^{2+}	0 (Reference)
Mg^{2+}	4382
Ca^{2+}	3420

Table 4.1 – Relative formation energies for minimised structures of ZntR model 1 (Figure 4.2), incorporating various ions (calculated using B3LYP/6-31G**)

4.4 ZntR singly protonated model

The binding site model used above may not be accurate for ZntR as XRD cannot show the positions of hydrogen atoms. Whilst it is possible for the protein to retain its shape with three negatively charged residues in close proximity without the residues moving apart as in the model, this structure would be extremely energetically unfavourable, so at least one of the residues is likely to be protonated. Protonation should hopefully also reduce the repulsion of the ligands enough to compare the effects of a singly charged metal ion. Thus model 2 was used as shown in figure 4.3. As predicted, model 2 converged for almost all of the ions used although some produced different structures to those predicted as shown in table 4.2.

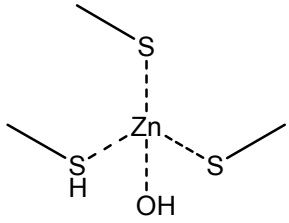


Figure 4.3 – A diagram of ZntR model 2

Ion	Charge	ΔH_a (kJ mol^{-1})	Geometry
Zn^{2+}	-1	0 (Ref)	Trigonal $[Zn(OH)(SMe)_2]^-$ + HSMe
Cu^+	-2	2208	Linear Cu(OH)(SMe), H-bonded to HSMe + MeS^-
Cu^{2+}	-1	-	-
H^+	-2	-	-
Mg^{2+}	-1	4419	Trigonal $[Mg(OH)(SMe)_2]^-$ + HSMe
Ca^{2+}	-1	3508	Trigonal $[Ca(OH)(SMe)_2]^-$ + HSMe
Na^+	-2	6399	
K^+	-2	5228	Trigonal $[K(H_2O)(SMe)_2]$ H-bonded to MeS^-

Table 4.2 – Relative formation energies for minimised structures of ZntR model 2 (Figure 4.3), incorporating various ions (calculated using B3LYP/6-31G**)

These results show that Zn (II) is still the strongest binder, and that Cu (I) has a high affinity for sulfur-based ligands.

Several of the calculations did not complete as neutral molecules moved away from the rest of the system (HSMe). Once again, the d^{10} ion, Cu (I) formed a linear species with additional hydrogen bonding, showing that the methods used work well for d^{10} ions.

4.5 ZntR doubly protonated model

The protonation state of the cysteine residues in ZntR's binding site could allow an even smaller negative charge. Calculations on such a charged site could potentially change the order of the binding affinities. Thus model 3 was used as shown in figure 4.4., with the results shown in table 4.3.

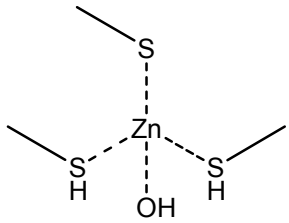


Figure 4.4 – A diagram of ZntR model 3

Ion	Charge	ΔH_a (kJ mol^{-1})	Geometry
Zn^{2+}	0	0 (Ref)	Distorted tetrahedron
Cu^{+}	-1	1934	Linear Cu(OH)(SMe), H-bonded to 2 x HSMe
Cu^{2+}	0	-	-
H^{+}	-1	6471	Tetrahedral HSMe around S- (inc. 1 Me)
Mg^{2+}	0	4418	Distorted tetrahedron
Ca^{2+}	0	3499	Distorted tetrahedron
Na^{+}	-1	6050	
K^{+}	-1	4897	Octahedron with 2 missing ligands.

Table 4.3 – Relative formation energies for minimised structures of ZntR model 3 (Figure 4.4), incorporating various ions (calculated using B3LYP/6-31G**)

These results show the 2+ charged ligands producing distorted tetrahedra and the 1+ charged ligands in other shapes and generally much higher in energy. Zn(II) is the strongest binding ion with Cu(I) also binding reasonably well. This binding affinity can be expected as both bind fairly well to sulfur and the geometries are again those expected for these ions.

4.6 ZntR conclusion

As with CueR, the binding site of ZntR appears to be well set up for binding selectively to Zn(II) and the unconstrained minimised structure is similar to the crystal structure.[54] However, the sensitivity is less than that for CueR allowing Zn (II) to be bound in other parts of the cell. The binding of Zn(II) will cause a direct interaction between cysteine A79 and B114 (the ends of the inter domain alpha helix and the proposed hinge region) explaining how the binding information can be transferred through to the DNA.[37] It would seem sensible to suggest that this interaction must also be important in CueR *via* the proposed hydrogen bonding chain. This would provide an atomic basis for the initial stages of information transfer from the metal-binding event to changes in DNA-binding for both CueR and ZntR.

5. WHOLE PROTEIN STUDIES (MOLECULAR MECHANICS)

5.1 Introduction

Chapter 3 studied the binding site of CueR and explained the selectivity and specificity for Cu(I).[42] To fully understand how CueR works it is important to study the global effects of binding Cu(I) and how the binding information may be transferred to the DNA. To study models the size of the whole of CueR, Quantum Mechanics methods would take far too long so Molecular Mechanics based calculations must be used. Molecular Mechanics will not allow hydrogen transfer or similar changes but should show some long range interactions.[135]

To study the effect of binding, various models must be compared. Unfortunately there is currently only one X-ray crystal structure known for the activator complex with Cu(I) bound (ref 1Q05) and no MD results are apparent in the literature.[54] To fully understand the mode of action of CueR, the effects of Cu(I) binding, DNA binding and removal of the hydrogen bond *via* mutation must be studied. Due to the inability of Molecular Mechanics to study electron density and the effect of partial hydrogen transfer, the effect of the hydrogen bonding interaction may not be well modelled by this technique.

Molecular Dynamics (MD) is used to see what motions the protein may make and which areas are mobile. Unfortunately with current limits on computer power only a relatively short amount of time may be simulated – from a few hundred picoseconds to a few nanoseconds. In this time frame we are unlikely to see a substantial change in the overall conformation of the protein due to the Cu(I) unbinding event. However, we may see some

destabilisation of the structure around the binding site when Cu(I) is not bound and some movement of the whole protein may start to take place.

5.2 Setting up the structure

The X-ray crystal structure for CueR found in the pdb archives with Cu(I) bound has reference 1Q05.[54,55] This structure is the activator form of CueR with Cu(I) bound, but no DNA. This structure required tidying up and hydrogen atoms to be added using both InsightII and AMBER8's LEaP functionality as described on the following pages.[135,145]

One monomer of the crystal structure is better defined than the other; the less well defined monomer has residues missing between cys112 and cys120. To ensure that the calculation was performed correctly, the missing residues were built using InsightII's biopolymer module.[145] Both monomers were missing the C-terminal residues from around 122-135. It is likely that these residues are too mobile to be seen using X-ray diffraction. However, these residues may be important in ensuring that the residues near the binding site are held in place and so they must be added into the structure used for MD. Once these residues had been added, a secondary structure was required for them. However, there is no literature on the nature of the structure of these residues. Upon studying the crystal structure further a surface groove was observed along the side of the DNA-binding domain. This groove is the correct charge, size and shape for an α-helix, and so the remaining residues were added in this manner.

Hydrogen atoms were added to the structure using LEaP. For hydrogen atoms to be added at the correct positions the residues had to be named correctly. For example, cysteine can be modelled as a thiol or a thiolate by using either "CYS" or "CYM" respectively. The

ability to change between these forms is very useful for making different models of CueR, depending on the setup of the hydrogen bond proposed in Chapter 3. Some models were designed with both cysteine residues being a thiol, some with only cys112 as a thiol and cys120 as a thiolate and some had both residues as a thiolate. See section 5.5 for detail on the models used.

The structure of the DNA was based on a BmrR crystal structure with DNA bound (ref 1EXI).[51] Whilst the ligand binding C-terminal domain of BmrR is quite different from CueR, the DNA binding N-terminal domain is similar to other members of the MerR family including CueR.[24,49] The DNA and DNA-binding region of BmrR was overlaid on the structure of CueR and the DNA copied into the CueR structure. The DNA sequence was modified to match that of the PcopA target sequence, keeping the sugar-phosphate backbone fixed, and then minimised to relax clashes with CueR (CueR coordinates fixed).

5.2.1 Parameters used in AMBER8

For the majority of the calculations, the forcefield used was AMBER's ff03, the latest forcefield available in AMBER8.[146] ff03 is a modified version of the ff99 forcefield which has parameters derived from quantum calculations to give more accurate charges and adjust the φ and ψ backbone dihedral angles.[147]

In an attempt to explore the effect of the hydrogen bond another forcefield was used for some of the calculations: AMBER's ff02EP.[148-150] ff02EP is a polarizable variant of ff99 with "extra points". These are dummy atoms added to electron donating atoms during the implementation of LEaP which are designed to mimic the effect of lone pairs. They are held a short distance from the rest of the atom and have a partial negative charge (with the

atom having a more positive charge to attain the correct charge). Addition of lone pair modelling should improve directionality and allow hydrogen bonds to be more readily formed from a Molecular Mechanics system that would not normally model them very well.

Each calculation was performed three times to avoid any errors caused by the randomisation of velocities at the beginning of the equilibration run. The minimisation was performed once for each protein, then the equilibration started using random number seeds of 71277 (default), 71276 and 71278 respectively.

Neither of the force fields used contained all of the parameterisation required for all of the calculations. In particular Cu(I) is missing from both the library and the forcefield. As such the parameters were added using antechamber.[135] This method produced a single point charge Cu(I) atom with a mass of 65.36 a.u, a charge of +1 and non-bonded parameters taken from those for Li(I) which has a similar radius and charge.[151]

5.3 Comparisons of structures

To understand the mechanism of activation of CueR, various models must be built to model different forms of the protein. The protein as found in the crystal structure may have either protonated or deprotonated cysteine residues (or one of each).[54] For each of these, extra points may be added to model the effect of lone pairs by using the ff02EP force field. As mentioned before, removing the metal ion may cause a change in the conformation of the protein (although maybe not in the timescale of an MD run) so calculations must be performed on systems without Cu(I).[61] Various other mutants may change the way CueR

works as described below. Finally, any of the above models may have DNA added, leading to a huge number of possible combinations. These are discussed in further detail below.

5.3.1 Removing the hydrogen bond

The Quantum Mechanics calculations show that the hydrogen bond should be essential for the correct function of the protein. Thus if this interaction could be removed, the protein should behave as a repressor irrespective of whether or not Cu(I) is present. The interaction between Cys112 and Ser77 is from the hydrogen of the thiol SH to the backbone carbonyl oxygen of Ser77. Unfortunately there is no way of directly removing this interaction by simple site directed mutagenesis except replacing Cys112, which would be expected to completely destroy the function of the protein. However, it may be possible to remove the extended hydrogen bonding chain and thus weaken this interaction. One way of removing the interaction would be by removing the NH hydrogen bond donor on Ala78 (next to Ser79) by changing it to a proline as shown in figure 5.1.

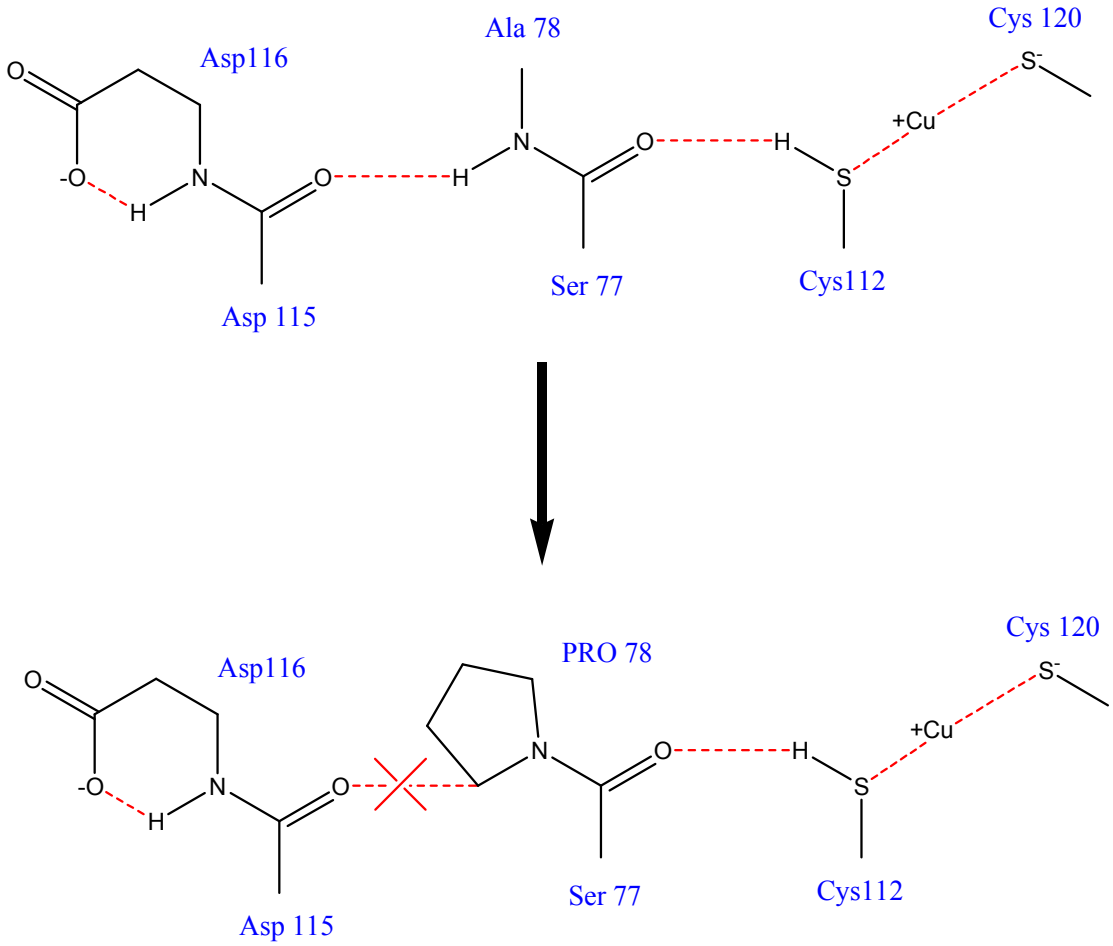


Figure 5.1 – A diagram of a mutant of CueR which would disrupt the hydrogen bond chain (A78P)

There are problems associated with the mutation proposed in figure 5.12; though both Ser77 and Ala78 have similar φ and ψ angles to proline, the additional bulk of the proline side chain may disrupt the local protein fold.[135] Thus it would be useful to replace Ala78 with another residue which would have the same effect on the protein's ability to fold as proline but which could restore the activity of the protein. One way of producing this model would be to truncate the hydrogen bonding chain (as suggested in chapter 3) by mutating residue 78 to an aspartate as shown in fig. 5.2.

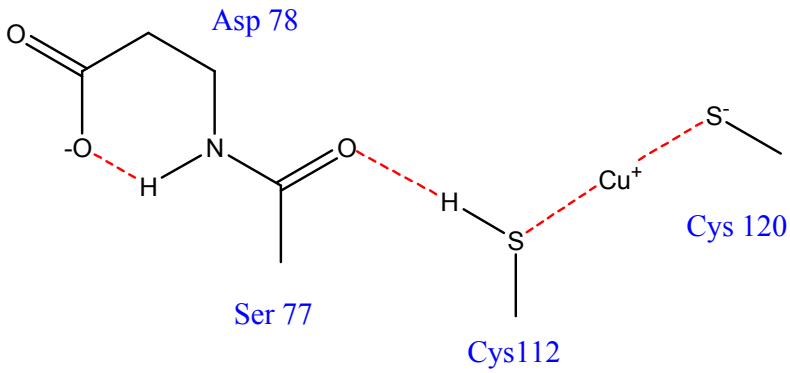


Figure 5.2 – A diagram of a mutant of CueR which would truncate the hydrogen bond chain (A78D) and could restore activity lost in the production of A78P

If the model is working as we expect it to, we should see a loss of activity for the first mutant, A78P, and restoration of activity for the second, A78D. Alternatively, if the increased bulk of the proline side-chain is an issue, then both mutants would be expected to be equally inactive.

Experimental data for these mutants has been gathered by colleagues in the biosciences department at the University of Birmingham and is shown in figure 5.3.[152] The first pair of columns shows the effect of addition of copper and unmodified plasmids to a wild type bacterial cell which contains CueR. The second pair of columns shows the effect of addition of copper and unmodified plasmids to a bacterial cell which contains no CueR. The third pair of columns shows the effect of addition of copper, unmodified plasmids and CueR to a bacterial cell which cannot produce CueR. The fourth, fifth and sixth pair of columns are similar to the third but each with addition of a different mutant of the CueR coded for on the plasmids. Essentially the first three sets of columns are controls to check that CueR (and mutants thereof) may be added and that only CueR will cause DNA transcription.

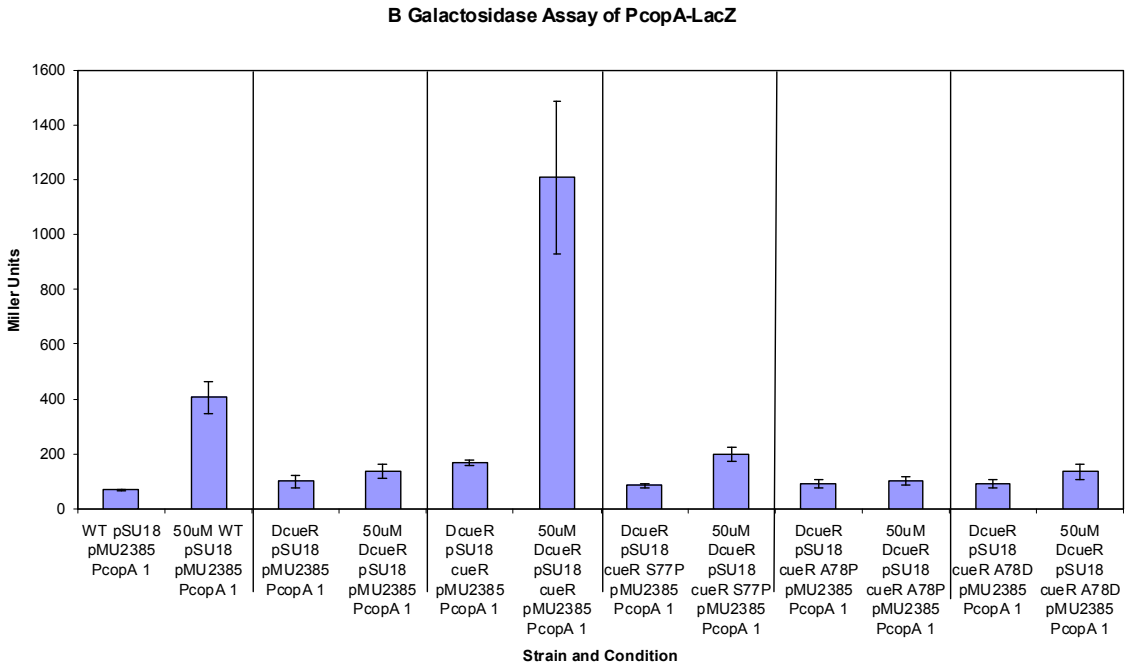


Figure 5.3 – Activity of bacterial CueR and related mutants. The activity studies were performed as β-Galactosidase Assays with levels of induction determined following the method of Stoyanov & Brown.[43]

In figure 5.3:

WT is the wild-type (not mutated) *E. coli* bacterium

pSU18 and pMU2385 are plasmids which may be modified to produce mutant CueR

PcopA is the promoter region to which DNA binds (usually promoting transcription of CopA, a Cu(I) export protein but in this case attached to DNA coding for β-galactosidase)

50μM indicates addition of a 50μM solution of Cu(II). This solution produces some *in situ* reduction to Cu(I). As Cu(II) was added rather than Cu(I), these results are not quantitative, and each column will only be relative to the one next to it.

DcueR is a deletion mutant of the *E. coli* bacterium (one which cannot naturally produce CueR)

cueR indicates addition of CueR

S77P has a gene for production of CueR mutant S77P on the plasmid

A78P has a gene for production of CueR mutant A78P on the plasmid

A78D has a gene for production of CueR mutant A78D on the plasmid

Miller units are a measure of the amount of DNA transcription produced, measured via β-Galactosidase assays

Figure 3.5 shows that mutant S77P is much less active than the wild type, and as such any changes to this region are likely to produce a less active strain of CueR than the wild type. The mutant A78P shows no response to Cu(I), as the transcription of the DNA occurs at the same rate regardless of the addition of Cu(I). Mutant A78D shows some restoration of activity as shown by the higher transcription of DNA in the presence of Cu(I). Due to the nature of the biochemical experiment, it is not known how active A78D will be and whether the binding event or the transcription event is affected. Studying mutations through MD will hopefully elucidate the mode of activation of these strains of CueR.

5.4 Data analysis

The easiest way of visualising the output from MD runs is to load the "mdcrd" files into a program which can read them, such as VMD.[153] VMD can then show a movie of the complete MD run. However, it is not possible to convert such movies directly to a format which can be shown on paper, and they would yield only qualitative results. Several different properties of the protein and surrounding molecules were calculated and are discussed on the following pages. The files required for viewing the movies are contained in the attached DVD.

The root-mean-squared deviations (RMSd) of the atomic positions of all non-solvent atoms were calculated vs. time, relative to the input coordinates. RMSd gives an overall idea of how mobile the whole protein is and how much it has moved. The motion of side chains may be random and not correlate with the mobility of the entire protein. As such, it may be more useful to study the RMSd of the alpha carbons. The RMSd of individual areas of the protein may also be calculated, so the binding site can be studied *via* this method.

The twist of the whole protein can be measured *vs.* time to see if there are any changes during the timescale studied. The twist of the whole protein is very important as it will affect the bending of the DNA and thus the activation of transcription.[61] In order to do study the twist, the dihedral angle of four points must be used. The first and last points used were a group of residues at the top of the DNA binding domain: residues 31-37 in each monomer. The middle two residues were the two ser77 residues which are at the end of the inter-domain α-helix. This dihedral angle may be drawn as shown in figure 5.4.

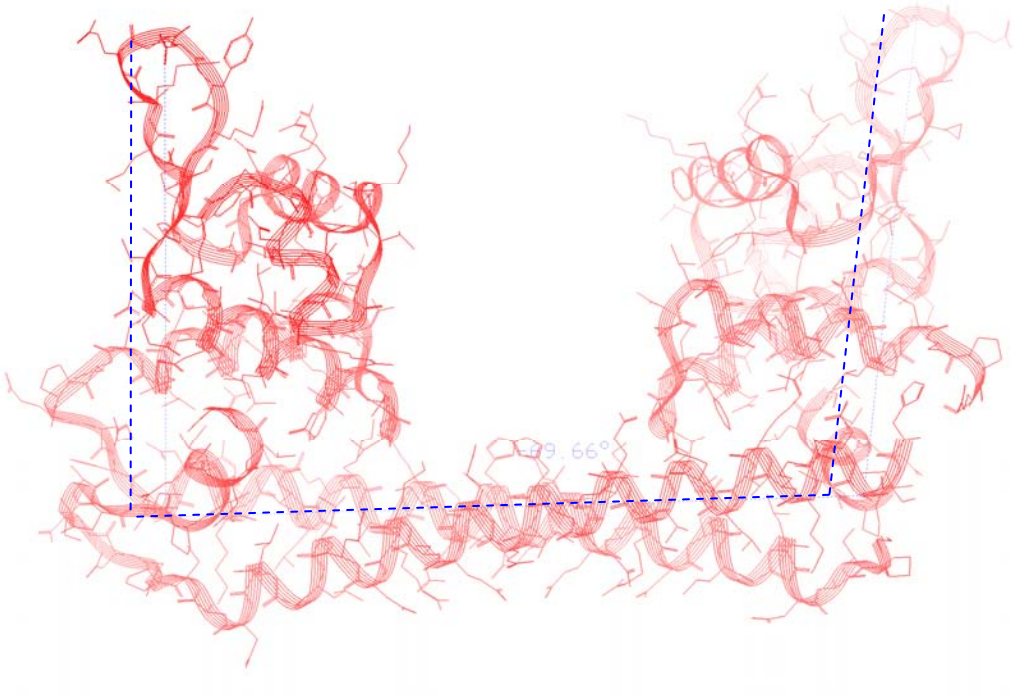


Figure 5.4 – A diagram of how the dihedral angle was measured in this chapter

Changes in the degree of protein twist will modify the extent to which bound DNA is bent, causing it to either adopt the conformation required for transcription, resulting in an activator form, or moving the conformation away from that required for transcription, resulting in a repressor form. In the CueR crystal structure, the DNA-bend angle is around 115^{o} but the calculations involving DNA may allow the DNA to bend. Three points must be specified for an angle. Due to the way AMBER8's "ptraj" program works, measuring the angle of duplex DNA can be tricky. A single atom was used for each of the points: one nearest to the middle of the strand and one near the middle of each end. The DNA used is 21 residues long, so the angle was measured from the 6th residue to the 11th and ending on the 16th. Measuring from this far from the ends of the DNA should avoid potential problems with the ends being too mobile. The closest atom to the middle of each base pair will be nitrogen 1 (N1) in the purine base. Due to the sequence of the DNA, N1 was guanine 6 from chain 1, guanine 11 from chain 1 and guanine 6 from chain 2 (opposite cytosine 16 in chain 1). The DNA bend may be drawn as shown in figure 5.5.

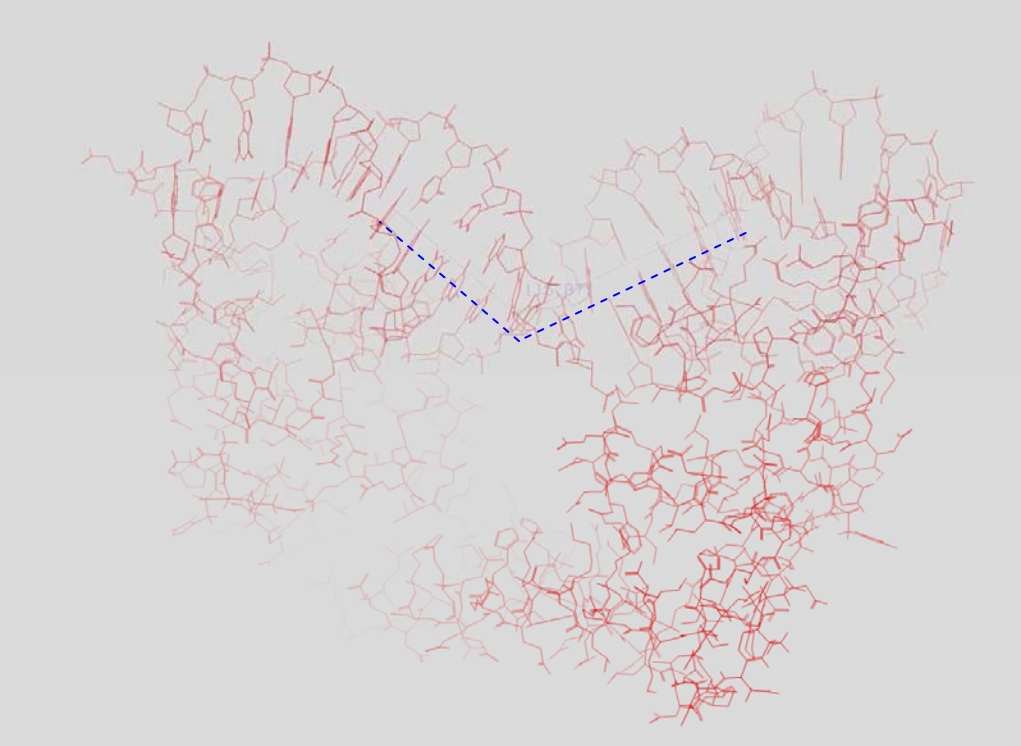


Figure 5.5 – A diagram of how the angle of the DNA was measured in this chapter

Finally the length of the hydrogen bond chain may be an important indicator of the way the protein changes. In particular, the binding site may disorder upon removal of Cu(I) and this disordering may be transferred through the protein. The best bond to study is that between the side chain thiol sulfur of cysteine 112 and the backbone carbonyl oxygen of serine 77. These residues are at opposite ends of the inter-domain alpha helices and so will be affected directly by removal of Cu(I) and should affect the protein twist angle. In some of the calculations there was no hydrogen present in this bond (see later for details) and, as such, this distance will be referred to as the C112-S: S77-O distance.

5.5 Setup of specific experiments

All movies referred to in the following sections are provided in the attached DVD along with instructions on viewing them (contained in the file "ViewingMDtraj.htm"). The structures were minimised, then a 100 ps equilibration run was performed. The movies were collected after equilibration.

5.5.1 CueR with protonated cys112, deprotonated cys120 and Cu(I) bound

The simplest calculation is based on the crystal structure; one of the cysteine residues is protonated and the other deprotonated, as suggested in chapter 3 (referred to as CueRH). Cu(I) is bound, but DNA and mutations are not present. The system was initially studied using the ff03 force field.[146]

Looking at the resulting movies it appears that Cu(I) does not remain bound due to insufficient charge on the thiol sulfur atom. Chapter 3 shows that Cu(I) would remain bound but that the parameters used by the molecular mechanics force field allow the Cu(I) to escape the binding site. Thus these results may not show exactly how CueR works. The

movies also show the region between cys112 and cys120 to be fairly mobile. The mobility observed would explain why one monomer of the crystal structure does not show this loop.[54]

5.5.1.1 Addition of extra points to model lone pairs

As mentioned previously, there is an AMBER force field which models the effect of lone pairs on electron donating atoms: ff02EP.[148,149] Using a model of lone pairs should help with the directionality of the copper binding and may help to prevent loss of Cu(I) for this model (referred to as CueRHep).

Looking at the movies generated by using this force field, it appears that Cu(I) still does not bind. Looking more closely it seems that the extra points are not added to the negatively charged thiolate ligands and so these are treated as point charges as before. The thiols do have extra points added but they may not have a large enough negative charge to hold the Cu(I) in place. Due to the lack of polarizability of this force field, any hydrogen bond formed would not increase negative charges on the thiol sulfur in this model.

5.5.2 CueR with deprotonated cys112 and cys120, Cu(I) bound

Since CueRH did not keep Cu(I) bound, the larger negative charge in the binding site of the doubly deprotonated version (referred to as CueR-) was employed. This model is the protonation state suggested by Changela *et al*.[54] Calculations involving this protonation state will not contain the hydrogen bonding chain.

The movies generated from this system show that Cu(I) does remain bound in the binding site for most of the calculations. In one calculation one of the Cu(I) ions was still able to

move away from the binding site during the course of the calculation. The lack of Cu(I) remaining bound suggests that the molecular mechanics force field used is not a perfect model for this system.

The calculations in Chapter 3 show that Cu(I) binds very strongly to CueR whether the residues are protonated or not. The fact that Cu(I) is not always remaining bound for the parameters used, irrespective of the protonation state of the cysteine residues, shows that Molecular Mechanical calculations are insufficient to ascertain the correct protonation state of the binding site. As such, the requirement of the hydrogen bond for the correct function of the protein (and the lack of reliable evidence to the contrary from Molecular Mechanical calculations) gives further credence to the idea that one of the cysteine residues may be protonated, as in CueRH.

5.5.2.1 Addition of extra points to model lone pairs

As before, the force field ff02EP was implemented to model the lone pairs (referred to as CueR-ep). In this case there are no extra points on either of the copper-binding ligands and no directionality for interacting with other residues.

The movies generated looked very similar to those generated with the ff03 forcefield (CueR-), with one of the runs allowing the copper ion escape the binding site again.

5.5.3 CueR without Cu(I) bound (protonated cys112&120)

The next calculation that was run involved no copper present, modelling the resting state of the protein (referred to as CueRnc). This model may produce the same effect as CueRH when the Cu(I) ion has left the protein if it is the binding of Cu(I) that changes the

dynamics of the protein. This model has two protonated cysteine residues as predicted for the resting state in Chapter 3. These calculations were only for 1ns.

The movies generated appear to show the cysteine residues in close proximity and forming hydrogen bonds to each other. It also appears that the dihedral angle is larger than with the earlier models.

5.5.4 CueR mutant A78P

The effects of the proposed mutants were studied through MD (see section 5.3). The model of A78P was built with Cu(I) present and two deprotonated thiols to ensure that the Cu(I) was not able to escape the binding site.

Looking at the movie generated it appears that both the N-terminal (DNA-binding) and C-terminal (metal binding) ends of the protein are more mobile than in the previous calculations. The binding site appears to remain approximately structurally similar to other models despite this increase in mobility; it is only the alpha helix that was added on and not seen in the crystal structure. The inter-domain helix appeared to remain stable throughout this increased mobility.

5.5.5 CueR mutant A78D

The other mutant suggested was A78D, which was modelled with Cu(I) present and two thiolates. In theory this mutation should reactivate the protein, but as cys112 is a thiolate, there will still be no hydrogen bonding chain present. If it were present, the truncation of the hydrogen bonding chain may still cause the cys112-ser77 hydrogen bond to be too strong and either stop activation of the protein or force it to be an activator whether or not

Cu(I) is present. Even so, it would be useful to model the effects of putting a residue the size of aspartic acid into position 78.

The movie generated shows high mobility in the N-terminal (DNA-binding) domain as was seen for A78P. However, the C-terminal (metal binding) domain is much less mobile and appears to be similar to the wild type. The inter-domain helix is still stable.

5.5.6 Addition of DNA to CueR

The model of CueR with DNA attached (taken from BmrR) was studied using two thiolates for the copper-binding cysteine residues with Cu(I) bound (referred to as CueR.DNA). The movie showed CueR to be less mobile with respect to the dihedral angle due to being held in place by the DNA.

5.5.7 Removal of Cu(I) from CueR with DNA bound

The effect of removing Cu(I) from CueR, and that of adding DNA, have now been studied individually. It would be very useful to understand how the removal of Cu(I) affects the DNA when it is bound to CueR. The model used was the same as CueR.DNA but without Cu(I) (referred to as CueRncDNA). The movie generated looks to be quite similar to that for CueR.DNA. The dihedral angles obtained from the simulations without Cu(I) were all similar to each other, but somewhat smaller than that seen in the presence of Cu(I).

5.5.8 CueR mutant A78P with DNA bound

The effects of mutating some of the residues of CueR have now been studied on the protein alone. How this information is passed onto the DNA can now be studied and compared to the effect of removing Cu(I). The model was built as for A78P but with the DNA added as

CueR.DNA (referred to as A78P.DNA). The movie produced appears to be quite similar to that for CueR.DNA so other effects must be studied.

5.5.9 CueR mutant A78D with DNA bound

CueR mutant A78P shows a loss of activity *in vivo* when DNA is bound; the effect of changing to mutant A78D was studied next. The model was built as before, essentially as CueR.DNA but with residue alanine 78 mutated into an aspartate (referred to as A78D.DNA). The movie generated again looks very similar to that found for CueR.DNA.

5.6 Dihedral angles

The dihedral angle shown in figure 5.4 was studied for each of the models. Due to the nature of DNA binding, this angle will directly affect the shape of the DNA, and, presumably, DNA transcription. Specific graphs may be found in appendix I, part a).

The average dihedral angles were measured for each of the models. The data gathered is summarised in table 5.1.

Species	Average Dihedral	Standard deviation	Long term effects on dihedral angle
CueRH	-64.5870	7.660481	No obvious change
CueRHep	-63.6425	4.381135	No obvious change
CueR-	-70.8670	7.057756	No obvious change
CueR-ep	-62.7364	6.757647	Increases with time
CueRnc	-75.5307	6.303169	Increases with time
A78P	-60.2065	4.768684	No obvious change
A78D	-70.0386	4.113808	No obvious change
CueR.DNA	-71.2589	4.277364	No obvious change
CueRncDNA	-60.0963	3.023914	No obvious change
A78P.DNA	-73.4960	5.393307	Increases with time
A78D.DNA	-67.9199	3.899579	No obvious change

Table 5.1 – Summary of the MD calculation dihedral angles (as described in figure 5.4) for various forms and mutants of CueR

Several models of CueR (such as CueR-ep, CueRnc and A78P.DNA) show a gradual increase in dihedral angle over time. This change suggests that the geometry being produced may be more stable than the input geometry. Some of the models have stable geometries that are larger than others; for these models the geometry may have changed during the minimisation calculations.

The data in table 5.1 shows that there is an obvious change in dihedral angle for the forms of the protein expected to be repressors. The activator form (shown here as CueR- and CueR.DNA) has a dihedral angle of around 70^{o} ($\pm 7^{o}/4^{o}$), which is transferred to the DNA to produce a bend of around 110^{o}. There are various repressor forms produced; some have a dihedral angle as small as 60^{o} ($\pm 3^{o}$) in the case of CueRncDNA whereas others have a larger dihedral angle, such as CueRnc, with a dihedral angle of around 75^{o} ($\pm 6^{o}$).

5.7 C112-S: S77-O distance

The C112-S: S77-O distance will be directly affected by the presence of Cu(I) if the hydrogen bonding chain is important. This distance may also affect the twist of the protein. A simultaneous change may occur between this distance and the dihedral angle. As such, it was studied for each model to produce the graphs found in Appendix I b) and the averages are shown in table 5.2. The nature of concerted changes in described in section 5.10.

Species	Average S-O distance	Standard deviation	Long term effects on S-O distance
CueRH	4.483067	0.594158	No obvious change
CueRHep	5.077424	0.791282	No obvious change
CueR-	4.819452	0.624726	No obvious change
CueR-ep	6.366492	0.633347	Increases slightly with time
CueRnc	7.311148	2.038414	Increases with time
A78P	6.804145	0.310746	No obvious change
A78D	9.583841	1.145798	No obvious change
CueR.DNA	5.149360	0.546404	No obvious change
CueRncDNA	5.724376	0.615464	No obvious change
A78P.DNA	6.010018	0.472296	Increases with time
A78D.DNA	5.169874	0.516987	No obvious change

Table 5.2 – Summary of MD calculation C112-S: S77-O distances for various forms and mutants of CueR

Some of the models of CueR have larger C112-S: S77-O distances than others. The distance observed for CueRH (with hydrogen bonding present) is much lower than for the models which are unable to form a hydrogen bond between C112-S and S77-O. All of the calculations have relatively large standard deviations and so it may be difficult to be certain whether any changes are statistically significant.

The models with DNA bound produce C112-S: S77-O distances which are more similar to one another than the models without DNA bound do. However, there is a shift from around 5Å for the proposed activator structures (CueR.DNA and A78D.DNA) to around 6Å for the repressors (CueRncDNA and A78P.DNA).

5.8 RMSd

The Root-Mean-Squared deviation (RMSd) gives a measure of how mobile the protein is over time. Three ways of measuring the RMSd (using different atom sets) were initially employed: all atoms, alpha carbons and binding site atoms. Measuring the RMSd of all of the atoms is the most obvious and should show how the entire system changes during the course of the MD run. Measuring the RMSd of the alpha carbons will remove the effects of

side chain motion and DNA bending. Measuring the RMSd of the binding site atoms should show how the binding site disorders on the removal of Cu(I).

Each of the ways of measuring the RMSds were compared. The RMSd for all atoms and for the alpha carbons produced very similar graphs for the systems without DNA; systems involving DNA had constantly increasing RMSd for all atoms due to the flexibility of the ends of the DNA and so the RMSd for alpha carbons was far easier to interpret. The RMSd for the binding site was too highly dependent on the C112-S: S77-O distance to show anything else.

The RMSd initially rises sharply for each calculation, so the first 200 ps were ignored when calculating the average values. The RMSd was studied for all atoms to produce table 5.3 and the graphs found in Appendix I c).

Species	Average RMSd	Standard deviation	Long term effects on RMSd
CueRH	3.400027	0.793383	Increases with time
CueRHep	3.523015	0.462517	No obvious change
CueR-	3.086274	0.487085	Increases slightly with time
CueR-ep	3.386243	0.375365	No obvious change
CueRnc	2.972211	0.375187	No obvious change
A78P	2.655601	0.355083	Decreases slightly with time
A78D	2.744750	0.400693	Increases with time
CueR.DNA	3.536265	0.330328	Increases with time
CueRncDNA	3.699930	0.280600	Increases with time
A78P.DNA	3.492292	0.449777	Increases with time
A78D.DNA	3.397012	0.545887	Increases slightly with time

Table 5.3 – Summary of the MD calculation RMSd for all atoms for various forms and mutants of CueR

The long term effects on the RMSd for all atoms show interesting results; many of the calculations have increasing RMSd, showing that they are slowly disordering.

CueRH shows increasing RMSd with a large standard deviation, showing that the loss of Cu(I) during the calculation leads to a change of shape of the protein.

The models involving DNA show increasing RMSd (due to the flexibility of the ends of the DNA) yet CueR.DNA and CueRncDNA have a relatively small standard deviation. This suggests that the structures are overall more rigid.

A78P shows a slight decrease in RMSd over time, suggesting that the structure is gradually getting more stable. Looking at the other parameters for A78P (dihedral angle and C112-S: S77-O distance) the structure appears to be a repressor with a small standard deviation but these parameters are not changing with time.

The RMSd was also studied for alpha carbons for the systems incorporating DNA, thus ignoring the effects of the mobility of the DNA. The results may still be compared to those for models without DNA. This data is summarised in table 5.4 with the graphs shown in Appendix I d). As was seen for RMSd for all atoms, RMSd for alpha carbons is lower for the structures incorporating DNA.

Species	Average RMSd	Standard deviation	Long term effects on RMSd
CueRH	2.782817	0.766742	Increases with time
CueR.DNA	2.117219	0.150674	No obvious change
CueRncDNA	1.776463	0.292270	Increases with time
A78P.DNA	2.139045	0.225075	Decreases slightly with time
A78D.DNA	1.925036	0.303746	Increases slightly with time

Table 5.4 – Summary of the MD calculation RMSd for alpha carbons for various forms and mutants of CueR

5.9 DNA bend

The bend of the DNA directly affects transcription and will be directly affected by dihedral angle. The DNA bend was measured for each model to produce the average values in table 5.5 and the graphs in Appendix I e).

Species	Average DNA bend	Standard deviation	Long term effects on DNA bend
CueR.DNA	108.5263	5.356007	No obvious change
CueRncDNA	136.5420	4.981913	No obvious change
A78P.DNA	118.2329	8.162705	Increases with time
A78D.DNA	109.0558	4.929734	No obvious change

Table 5.5 – Summary of the MD calculation DNA bends for various forms and mutants of CueR

5.10 Concerted motions

The parameters described in 5.6 to 5.9 may be reliant on one another. As such, it is important to compare the changes which occur within similar time frames. For example, the C112-S: S77-O distance may become longer, changing the shape of the protein such that the dihedral angle increases, the DNA angle increases and the overall mobility of the protein (measured *via* RMSd) increases.

CueRH shows a concerted motion between the dihedral angle and the C112-S: S77-O distance; whenever there is a rise in C112-S: S77-O distance it is followed a few nanoseconds later by an increase in dihedral angle and *vice versa*. This suggests that the dihedral angle is affected by C112-S: S77-O distance. The nature of the concerted motion for CueRH is illustrated in figure 5.6. CueR- shows a concerted motion between dihedral angle and RMSd for all atoms as shown in figure 5.7. CueR-ep shows a concerted motion between dihedral angle and C112-S: S77-O distance, similar to that seen for CueRH. This is illustrated in figure 5.8. The form of CueR modelled without Cu(I), CueRnc, shows

increasing dihedral angle and C112-S: S77-O distance over time, as shown in figure 5.9. A78P shows similar (although less pronounced) changes to CueRnc as shown in fig 5.10.

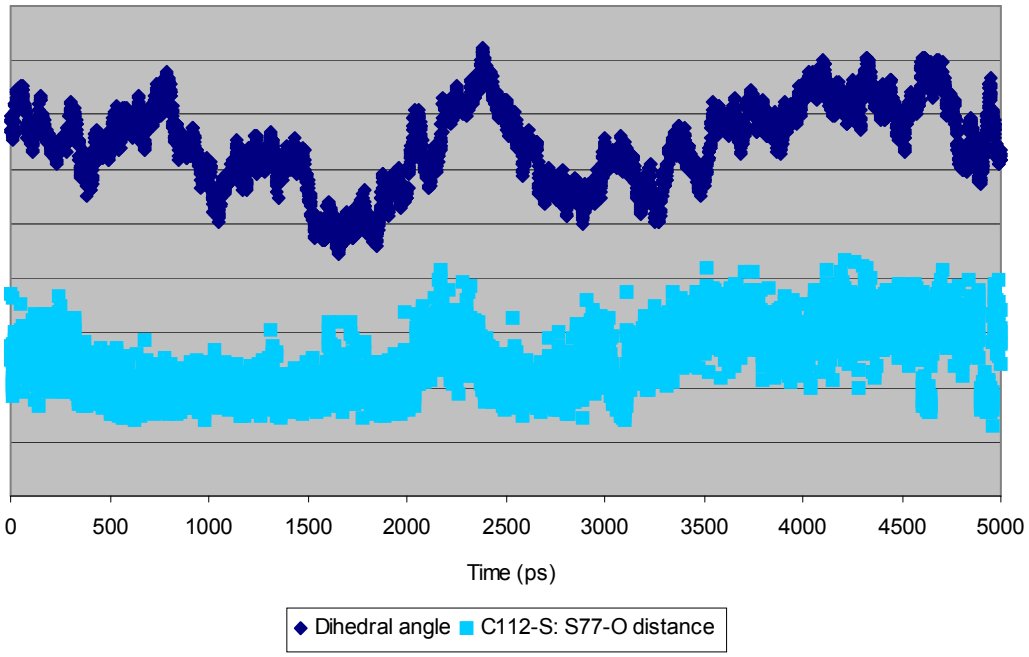


Figure 5.6 – Concerted motion of dihedral angle and C112-S: S77-O distance *vs* time for CueRH

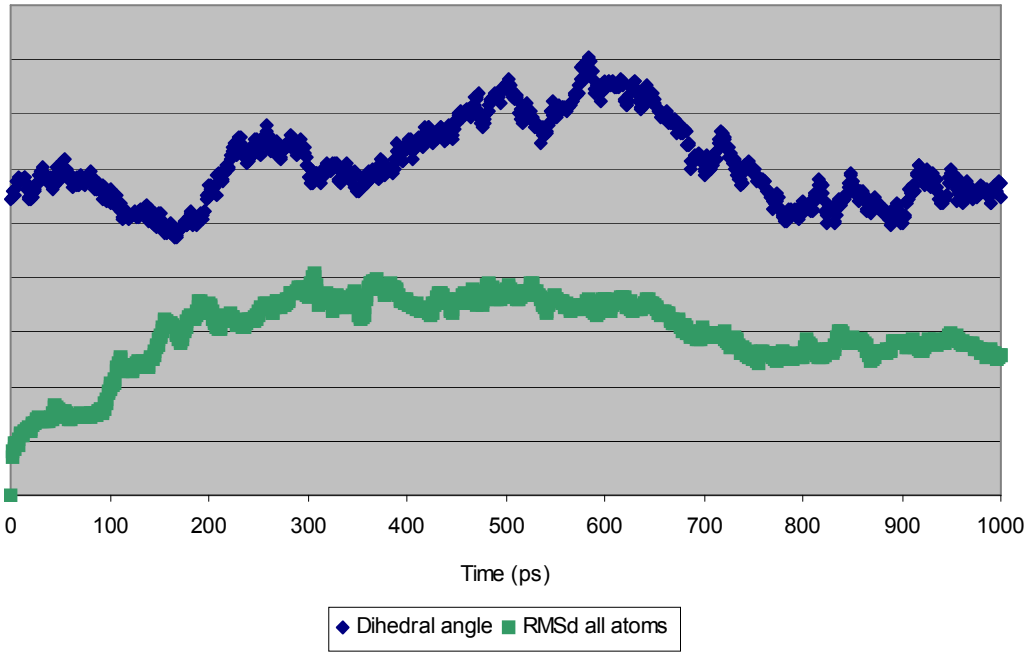


Figure 5.7 – Concerted motion of dihedral angle and RMSd (all atoms) *vs* time for CueR-

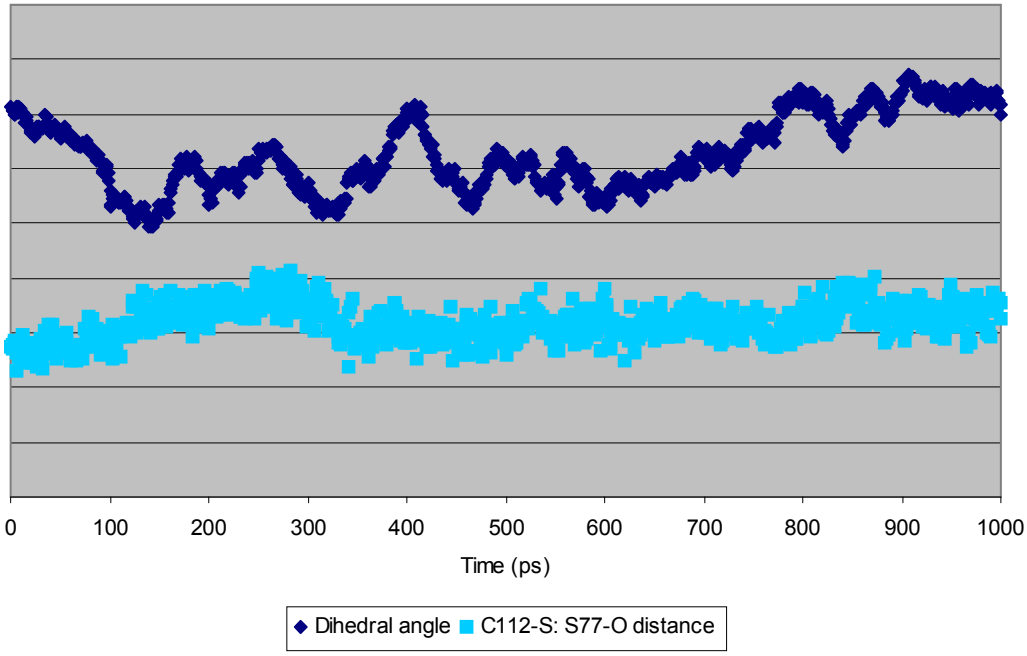


Figure 5.8 – Concerted motion of dihedral angle and C112-S: S77-O distance *vs* time for CueR-ep

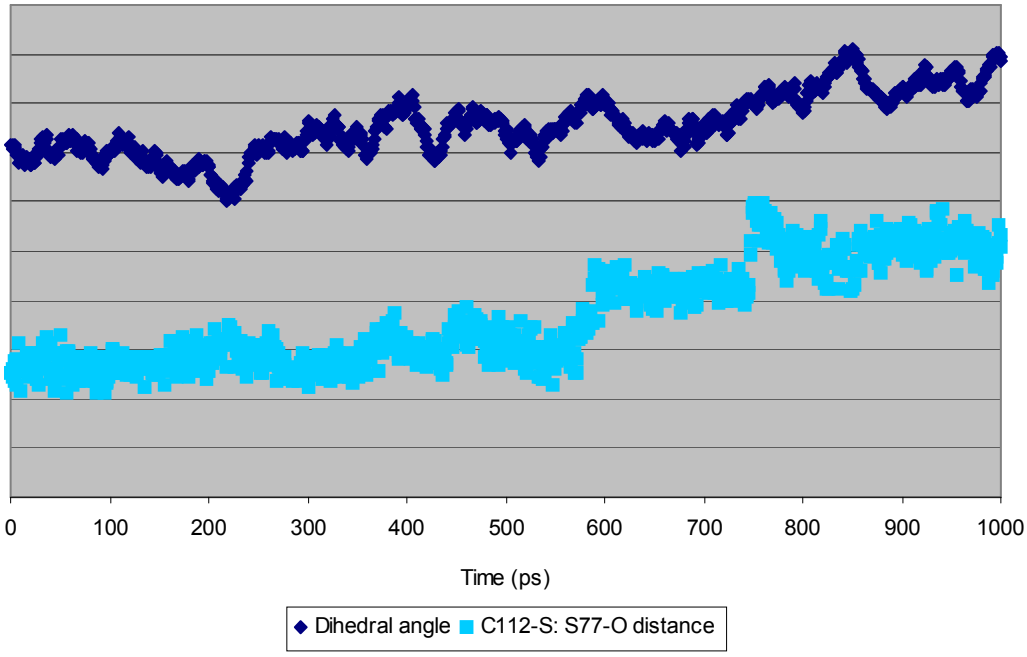


Figure 5.9 – Concerted motion of dihedral angle and C112-S: S77-O distance *vs* time for CueRnc

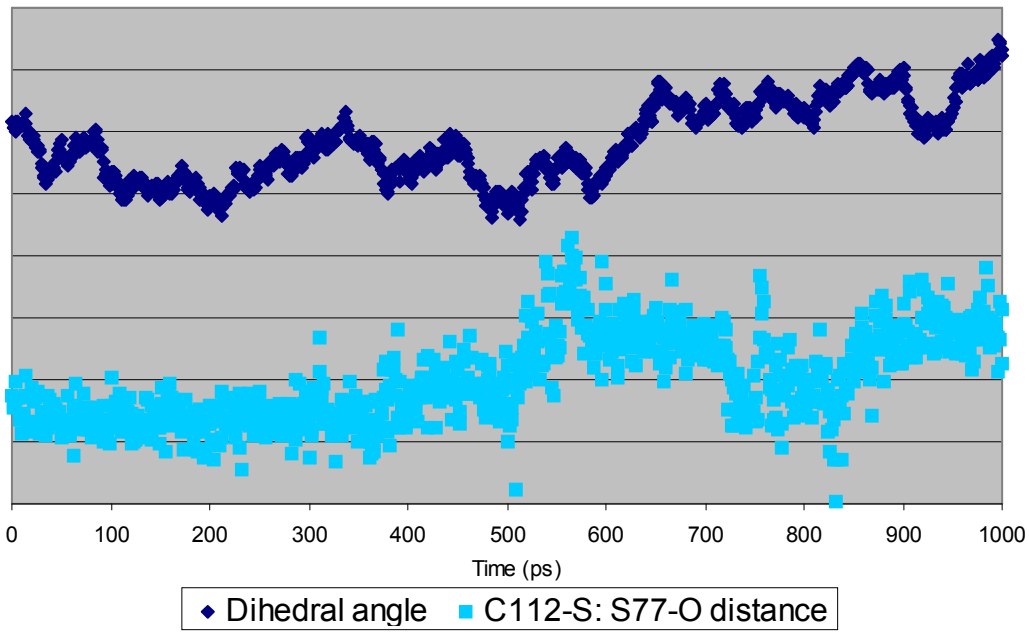


Figure 5.10 – Concerted motion of dihedral angle and C112-S: S77-O distance *vs* time *vs* time for A78P

The models incorporating DNA show that dihedral angle is strongly correlated to DNA angle. Figures 5.6, 5.8, 5.9 and 5.10 show that dihedral angle is correlated to C112-S: S77-O distance, therefore changes in this distance will also be correlated to DNA angle, as shown in figure 5.11. A78P.DNA shows the same correlation as CueR.DNA, with an overall slight increase in all three parameters as shown in figure 5.12.

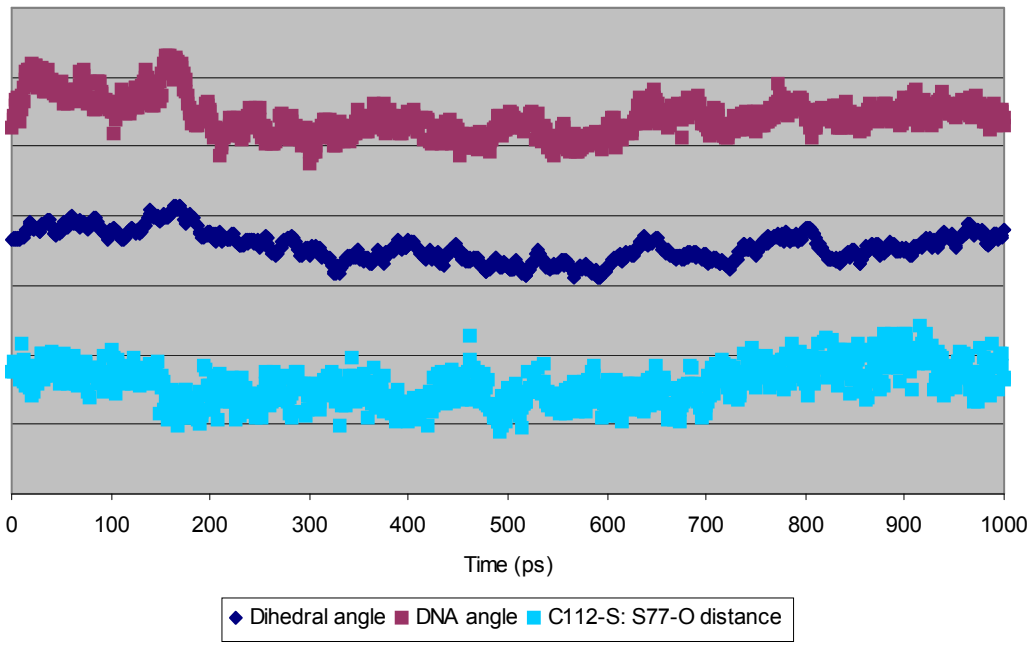


Figure 5.11 – Concerted motion of dihedral angle, DNA angle and C112-S: S77-O distance *vs* time for CueR.DNA

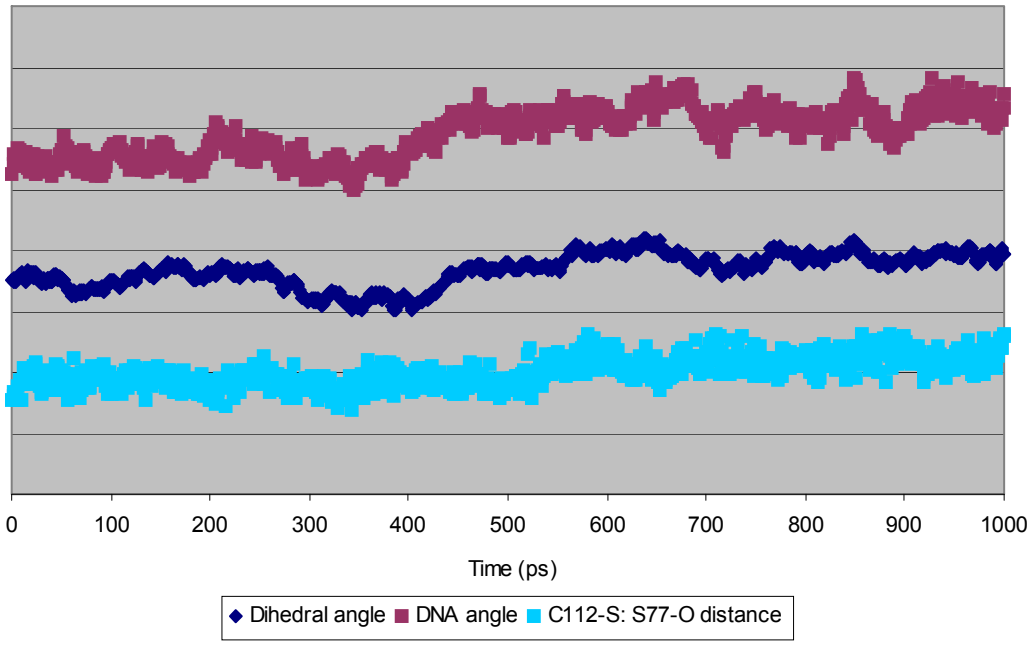


Figure 5.12 – Concerted motion of dihedral angle, DNA angle and C112-S: S77-O distance *vs* time for A78P.DNA

5.11 Summary and Conclusions

Several different models of CueR have been designed, each explaining more about the way in which information is passed to DNA and what is required for it to work. This data may be summarised to produce table 5.6.

Species	Average Dihedral	Standard deviation	Average DNA Bend	Standard deviation	Average S-O distance	Standard deviation
CueRH	-64.587	7.660481	-	-	4.483067	0.594158
CueRHep	-63.6425	4.381135	-	-	5.077424	0.791282
CueR-	-70.867	7.057756	-	-	4.819452	0.624726
CueR-ep	-62.7364	6.757647	-	-	6.366492	0.633347
CueRnc	-75.5307	6.303169	-	-	7.311148	2.038414
A78P	-60.2065	4.768684	-	-	6.804145	0.310746
A78D	-70.0386	4.113808	-	-	9.583841	1.145798
CueR.DNA	-71.2589	4.277364	108.5263	5.356007	5.14936	0.546404
CueRncDNA	-60.0963	3.023914	136.542	4.981913	5.724376	0.615464
A78P.DNA	-73.496	5.393307	118.2329	8.162705	6.010018	0.472296
A78D.DNA	-67.9199	3.899579	109.0558	4.929734	5.169874	0.516987

Table 5.6 – Summary of structural data for all MD runs

The initial structures (CueRH and CueR-) were designed to find out what happens when hydrogen is added or removed from the proposed hydrogen bonding chain. It was discovered that the force field does not model the interaction between Cu(I) and a thiol (protonated cysteine) correctly, which allowed Cu(I) to leave the active site in CueRH. For this reason all other models had two thiolates in the active site. This change removes the hydrogen bonding chain for all other models and so the full effect of it cannot be known from these experiments. Adding extra points (lone pairs) to the structures using the ff02EP force field has little effect on the way that the proteins behave overall.

Including DNA in the complex lead much reduced variation in both protein dihedral and DNA bend angles during the course of simulations as indicated by the reduced std. deviation values.

Unfortunately, deficiencies in the forcefield lead to the charge neutral Thiol-Thiolate complex being unable to bind Cu(I) throughout the whole MD simulation, so we were unable to investigate the role of the hydrogen bond in communicating between Cu(I) binding and DNA-bending events. This deficiency in the forcefield is highlighted by a number of thiolate-thiolate complexes that were also unable to bind Cu(I) for the duration of the MD simulation, despite the very tight experimentally observed Cu(I)-binding by CueR.

A number of A78 mutants were also studied to investigate their effect on protein behaviour. As seen in the experimental data, changing this residue to a proline removed activity, whilst changing it to an aspartate restored limited activity.[152] However, our proposed hydrogen bond could not be present in either structure due to the need to model the system using cysteine thiolates. This suggests that the change in activity is not caused by the presence or absence of the hydrogen bond but by some change in the structure of this region of the protein. The mutations modelled were at sites in the hinge region of the protein; thus the change in structure observed for A78P is not a relic of disrupting the hydrogen bond, but rather due to a change in the nature of this hinge.[37]

The data shows that when Cu(I) is removed from CueR there is a change in conformation in the protein. The Cu(I) bound form is similar to the crystal structure, whereas the resting state has an average dihedral angle around 5^{o} larger.[54] Such a change in dihedral angle could change the twist in the DNA and thus transcription would no longer be able to occur.[37] With DNA bound, the dihedral angle changed by around 10^{o} upon binding of Cu(I), leading to change in DNA twist of around 30^{o}. This should easily be enough to allow RNA polymerase to effect transcription of the DNA.[24]

5.12 Further work

The calculations in this chapter were unable to correctly model the nature of the hydrogen bond due to Cu(I) leaving the binding site in the calculations involving cys112 as a thiol. Chapter 3 shows that Cu(I) binds very strongly in this binding site and so the general force field used is an incorrect model for this site. It should be pointed out that force fields such as ff03 and ff02ep have been developed primarily to simulate protein structure rather than being concerned with the specific interactions of thiols and thiolates with Cu(I) ions. It is possible to modify the force field in several different ways which could improve the modelling of this binding site.

Chapter 3 shows that the hydrogen bonding chain pushes electron density into the Cu(I)-SH bond by partially transferring the hydrogen onto the backbone oxygen of serine 77. This means that the thiol sulfur has a larger negative charge than would usually be seen for a cysteine residue. The exact charges of all of the atoms involved will also be affected by this chain. A new residue could be added to the forcefield (as commonly occurs with AMBER8) with increased polarisation of the cys112 thiol.

Alternatively, the S-Cu(I) bond could be modelled directly using a modified residue which contains the protonated thiol bound to Cu(I). This “cys-Cu” residue would then not allow the Cu(I) ions to escape the binding site but would allow the Cu(I) ion to interact with the other thiolate (cys120). In order for this to work, parameters would need to be derived for the S-Cu bond length and strength. These parameters could be derived from QM calculations or using AMBER8’s antechamber program. Unfortunately, as each of the MD simulations takes approximately 3 months to run on this protein, use of modified force fields or residues falls outside the scope of this thesis.

6. APPLICATIONS AND EVOLUTIONS

6.1 Alternative binding sites

CueR is part of the MerR family of metal sensing transcriptional activators. There are several proteins in the MerR family which are found in different organisms and are specific for different metal ions.[24] These include MerR which senses Hg(II), ZntR which senses Zn(II), Cd(II) and to a lesser extent Pb(II), PbrR which senses Pb(II), CadR which senses Cd(II) and, of course, CueR which senses Cu(I) and to a lesser extent Ag(I) and Au(I). [26,42,43,154-156]

Whilst the above proteins are used to trigger the removal of the relevant toxic metal ions from cells, there are plenty of other toxic metal ions which do not have a similar inducible removal mechanism. These metal ions include Arsenic and Chromium which are toxic at low doses, along with lithium and magnesium at higher doses. It could be very useful to be able to detect low concentrations of these ions and potentially remove them. Thus, the metal-binding site of one of the known MerR-family regulators could be modified to bind strongly to the required metal ion. The new metal binding site would need to fulfil certain criteria to be successful:

1. To bind to the metal strongly
2. To be based on binding elements which may readily be found in proteins (*i.e.* amino acid side chains)
3. To have a similar geometry to the binding site of one of the already known MerR family members such that they may be incorporated into the structure
4. To have a different geometry when the metal is not bound (although any structure which is not an activator will be a repressor)

The extent to which these criteria must be met will depend upon the use to which the protein will be put. MerR family proteins are transcriptional activators, binding to the metal to produce a signal and removing that metal ion from the cell. The proteins designed in this chapter may require different elements of the above scheme for their mechanism. There are three basic types of protein which could be produced: transcriptional activators, simple signalling devices and sequestration proteins. These will all have slightly different requirements, and some may be easier to design than others.

Transcriptional activators bind to the metal and cause the correct conformational change in the DNA. This process occurs with the natural proteins (MerR, CueR etc) but may be difficult to achieve with a mutated version. As shown throughout this thesis, the metal ion is required to generate the correct conformational and other metal ions may cause a change in the binding site that would be transferred through the rest of the protein and cause a different conformational change in the DNA. It is possible that such proteins may be designed here, especially if the overlay of the binding site on a known structure is very close. However, the other protein types may be easier to produce.

Simple signalling proteins could work in a variety of ways, such as being attached to fluorescent DNA. In this case, various different changes in the overall conformation of the protein may produce a signal so long as it binds well to the metal ion. However, depending on the nature of the signal produced, the exact nature of the conformational change may be important.

Sequestration proteins should be the easiest to design; their only major requirement is to bind to the metal ion strongly. The overlay on the binding site of a known protein will still be important to maximise the effectiveness of the protein produced.

All of the above types of protein can, in principle, be developed from MerR proteins as they already bind the metal ions extremely tightly, transfer information on the binding site to the DNA and cause the correct conformational change.

6.2 Uses of high affinity proteins

Protein biosensors have already been used to detect low concentrations of Hg(II) and Cu (I) in soil samples, so all members of the MerR family could potentially be used on-site to detect the concentrations of various toxic metal ions.[157,158] A MerR-family promoter is used to control expression of a luminescent protein such as green fluorescent protein so that when the toxic metal is present the protein is produced and the sample would fluoresce.[159] In order to detect different metals, a gene for a suitably modified MerR-family regulator could be included on the plasmid, though not under the control of the MerR-family promoter itself. It is likely that changes to the metal-binding site of a MerR-family regulator may cause it to bind metal ions less strongly than in the wild type, especially for CueR and MerR, which bind so strongly already. However, a slightly lower detection limit may actually be useful for some applications (*e.g.* biosensors) as it may be preferable to detect metal ions at a range of different concentrations or at levels somewhat higher than that usually detected by CueR and MerR.

As the metal binding affinity of the MerR family regulators is generally so high (*e.g.* CueR $K_d \approx 10^{-21}$ M) an alternative use of these proteins lies in clearing up toxic metals. Theoretically an organism could be genetically modified to overexpress the protein and be placed in an area contaminated with the toxic metal. The organism could then soak up the surrounding metal and be harvested and removed from the area.

6.3 Studying alternative toxic metal binding sites

The binding site of CueR seems to be extremely well set up to accept Cu(I) ions. Chapter 3 showed that the unconstrained minimum for the binding site residues was very similar to that found in CueR in terms of Cu-S bond lengths and S-Cu-S bond angles. Chapter 4 showed that ZntR is designed for a different purpose in which it only needs to bind with $K_d \sim 10^{-14}$ M through a 4-coordinate binding site. Thus it may be possible for an adapted version of ZntR to bind to Zn(II) more or less strongly.

All calculations in this chapter were performed using Gaussian 98, the basis set LANL2DZ, Cartesian optimisation and the DFT functional MPW1PW91 for electron correlation calculations.

6.3.1 Metals studied

Most, if not all, metal ions are toxic in large concentrations, with each ion having its own toxicity limit. The reasons for the ions studied in this project are discussed on the following pages.

Since CueR has such a good binding site, Cu(I) was tested as a comparison. This binding site should produce a good control to check that the methodology used works.

If the methodology is effective at finding a good binding site for Cu(I) then it seems likely that the suggested structure should be similar to that for CueR. As discussed in Chapter 3, Cu(I) is extremely thiophillic and normally binds to two ligands. The test for this methodology is that Cu(I) should bind to two sulfur-based ligands.

MerR binds to Hg(II) and thus serves as another comparison.[26] However, it may also be possible to design a stronger binding site and thus increase the sensitivity of this sensor. Mercury is known for being extremely toxic. It can bioaccumulate to reach toxic levels from sub toxic amounts.[5,160,161] Thus even a few parts per billion can be extremely important. Mercury is known to contaminate various water supplies and the ground throughout the world. One example of such a place is Battery Park in Birmingham, UK, where mercury has leaked into the soil from the old battery factory and thus the ground can no longer be used for housing. In these places an on-site testing kits based on MerR could detect extremely low concentrations of Hg(II).

Mercury is currently removed from contaminated soil by completely removing the soil and either placing it in landfill sites or chemically cleaning it.[162] This is expensive and inefficient, thus other methods are being sought.[163] One such method is phytoremediation: using plants to absorb the metal ions.[164] Phytoremediation is still in development and increased binding affinity would improve the efficiency of the technique. Phytoremediation may incorporate MerR by producing a genetically modified grass which overexpresses MerR and thus absorb Hg(II). The genetically modified grass could then be sown on contaminated ground and harvested when full of Hg(II). It would then be easy to deal with the waste MerR in the harvested plants elsewhere (by burning or similar).[162]

ZntR has also been studied (in chapter 4) and seems to be carefully tailored to remove the correct amount of Zn(II) from bacterial cells. Zinc is an essential element for human life. However, in extremely large doses it may become toxic; as such ZntR has evolved in bacteria to help remove excess amounts of Zn(II).[1]

Ag(I) and Au(I) are similar to Cu(I) and also bind to CueR.[42,43] They are likely to show a preference for a binding site similar to CueR, although the geometry may be different.

Cd(II) is similar in size, orbital occupation and charge to Hg(II) and appears to show some activation of MerR. However, there is also a specific Cd(II)-dependent regulator, CadR, which does not yet have a published crystal structure.[156] Cd(II) is extremely toxic and is also often found on old battery factory sites where it can contaminate the land for many years. Thus, a method by which Cd(II) could be removed from soil would be extremely useful.

Pb(II) binds to PbrR and is another toxic metal.[154] Again, it would be very useful to be able to remove Pb(II) from contaminated soil or detect low concentrations of it. It may be possible to design a stronger binding site.

Arsenic is toxic in both acute and chronic doses. The effect of arsenic as a poison in acute doses has been well documented throughout history.

There are serious problems with Arsenic poisoning in the groundwater in West Bengal and Bangladesh as well as other areas in the world.[2,165-168] Groundwater poisoning has only recently become a problem. These areas were once frequently

suffering from drought. It was then discovered that there was a large supply of groundwater under the entire area that could easily be tapped though the use of tube wells. When the wells started to be used they were tapping into arsenic free supplies. However as the well continued to be used it drew the water through arsenic rich rocks which gradually leached arsenic into the water. As the wells continue to be used the water is drawn through more and more arsenic rich rock and thus the water has a much higher concentration of arsenic.

A possible solution to removing the arsenic from water would be to genetically modify a reed to overexpress an arsenic responsive version of CueR/MerR (*i.e.* AsrR). The water from the tube wells could then be passed through a reed bed which could in theory absorb free arsenic to produce clean water. These reed beds could then be harvested every so often to remove the arsenic-laden foliage. The "filter" would be automatically regenerating (*i.e.* the plants would grow back) thus removing the need for constantly replacing the filters. Any type of filter which needs replacing regularly may not be replaced in poorer areas of the world; thus this method of producing a simple-to-grow reed bed sounds initially attractive. Whether or not this method is feasible is not known and will not be looked into in this project.

Arsenic is commonly found in two different ionic forms in nature: As^{3+} and As^{5+}. Both are toxic and are readily interchangeable through REDOX processes; thus any method of removing one should indirectly be able to remove both. It is thought that part of the reason for the toxicity of arsenic is caused by arsenates replacing phosphates in such molecules as DNA and thus changing the shape of them. If a protein could be designed to remove either form the above process could work; thus both forms will be tested.

Several other common metals were tested to show that the methodology used is selective. It may be useful to detect concentrations of these ions using the proposed approach.

6.3.2 Setting up the binding sites

A variety of different metal ion-ligand complexes were minimised to identify both the optimum number of ligands and optimum geometry for a range of metal-ion binding sites. These were compared against the optimised Cu(I) binding site to gauge their effectiveness in binding metal ions.

The binding energy for each metal ion can be measured in a variety of potential binding sites. Metal binding sites in proteins are based around amino acids containing nitrogen, oxygen or sulfur atoms. Depending on the environment of each residue these atoms may have different protonation states to those found in bulk water. They can bind to the metal through lone pairs or negative charges. Thus models of the following amino acids were used: aspartic acid/glutamic acid, cysteine, histidine, lysine, serine and tyrosine. Not all of these would usually be available in bulk solution but may be viable ligands within proteins as discussed later. The ligand models used are shown in table 6.1.

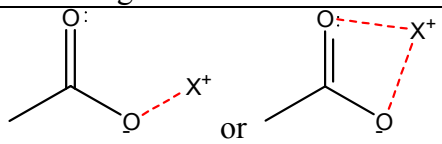
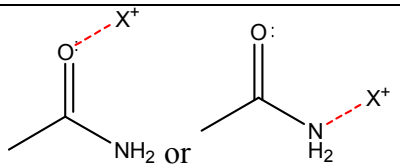
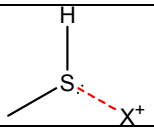
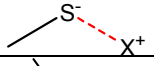
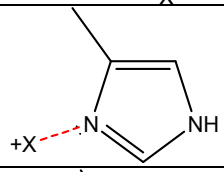
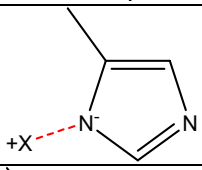
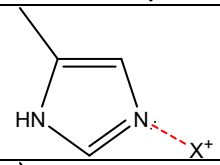
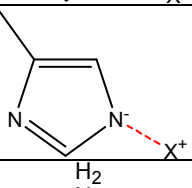
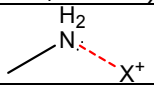
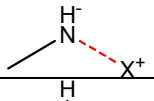
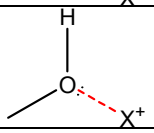
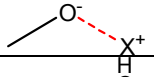
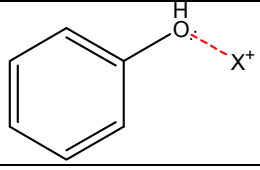
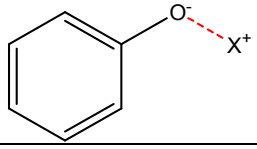
Amino acid	Binding atom	Charge	Diagram of model used
Aspartic Acid (or Glutamic Acid)	O (or 2xO, overall -1 charged)	-1	
Asparagine (or Glutamine)	O (lone pair) or N (lone pair)	0	
Cysteine	S (lone pair)	0	
	S (negative charge)	-1	
Histidine (Delta)	N (lone pair)	0	
	N (negative charge)	-1	
Histidine (Epsilon)	N (lone pair)	0	
	N (negative charge)	-1	
Lysine	N (lone pair)	0	
	N (negative charge)	-1	
Serine	O (lone pair)	0	
	O (negative charge)	-1	
Tyrosine	O (lone pair)	0	
	O (negative charge)	-1	

Table 6.1 – Ligand models used for the calculations in this chapter

Aspartic acid has a pK_a of around 4.1 meaning that it is likely to be deprotonated in the protein (depending on its environment); it is only in its deprotonated form that it will have the necessary charge to bind to the metal ion, thus only the deprotonated form was considered.[9]

Asparagine will be unable to bond through the lone pair on the nitrogen due to hyperconjugation with the carbonyl double bond. The oxygen lone pair did not appear to be strongly enough charged to bind to the metal ions and thus was discarded from all further calculations.

Cysteine has a pK_a of around 8.3 for the terminal S-H group, but it may easily be changed with the addition of a hydrogen bond acceptor. It may be possible to bind through either the lone pair or the negative charge depending on the protonation state of the sulfur (as seen in CueR)

Histidine has two nitrogen groups through which it may bind. The pK_a of the cation is around 6.0 and highly variable depending on environment so the two neutral, singly protonated, forms are easily accessible. There is very little difference in energy for deprotonation of either delta or epsilon N-H so both were considered in the metal-ligand complexes. Whilst the anionic form of histidine is not usually found, once the metal has bound it may be possible for the second proton to dissociate, for example if the metal is a strong Lewis Acid. Producing such a ligand may require going through a higher energy intermediate with both metal and proton bound, but as this project is comparing thermodynamics and not kinetics this will not be a problem; thus for histidine a total of 4 different binding modes are possible.

Lysine is usually found with a $-NH_3^+$ group in solution. The pK_a for one of these to dissociate is around 10.8 but the dissociated form contains a lone pair through which it may bond. As with histidine it may be possible for another proton to dissociate if the metal ion is a strong enough Lewis Acid, thus both the anionic and neutral forms were included in the complexes.

Serine, like cysteine, may be found in either the neutral form or as an anion. The pK_a of the terminal proton is not often quoted, but the serine protease mechanism relies on a series of hydrogen bond acceptors to make the alcohol more acidic so it is certainly possible. The pK_a of such a group could be in the region of 12-15, and the environment should lower it further.[139,169]

Tyrosine has a pK_a of around 10.9, thus it may be deprotonated if necessary to bind through either the lone pair (for the charge neutral ligand) or a negative charge. One of the lone pairs on the oxygen atom may become involved in conjugation with the aromatic ring (producing an sp^2 hybridised oxygen atom) but the other lone pair will still be free for binding in the charge neutral species.

These ligands have a wide variety of full/partial charges, delocalisation of charges, and hardness/softness of ligands. This variety should be useful for comparing different metal ions, each with their own binding preferences.

The number of ligands is likely to be important as well as the type. Many metal ions (excluding lanthanides and actinides) usually form up to 6-coordinate species. Thus systems with 1-6 ligands were tested to find out which bound the most strongly. As there are 14 ligands and up to 6 ligands, 84 calculations are required per metal

studied, plus comparison calculations. It would have been useful to run calculations with a combination of different numbers of each ligand but doing so would have required at least 38760 calculations per metal ion studied which was not practical.

In most of the previous calculations the basis set used was 6-31G**. Unfortunately this basis set only works for up to the fifth row of the periodic table *i.e.* the first row transition metals. Thus any elements involving f orbitals, such as silver or gold, may not be modelled with this basis set. The DFT functional and basis set used for the calculations in this chapter was MPW1PW91/LANL2DZ as discussed in chapter 2.

As the metal-ligand complexes are built up there are two modes that a new ligand can bind to the rest of the complex: either directly to the metal or to another ligand thus beginning a second coordination sphere. Which mode is most energetically favourable will depend upon the number of ligands already bound, the size and charge of the metal ion and various other factors. It would be expected that the first ligand to bind would bind directly to the metal ion. However, for each new ligand bound to the metal the effective charge of the metal will decrease and the repulsion by the currently bound ions will increase. The ligand-ligand binding is likely to be weaker than the initial metal-ligand interaction but should remain fairly constant as more ligands are added (depending on the nature of the ligands). Thus as more ligands are added to the complex, the relative energies of ligand-metal and ligand-ligand interactions will converge to a crossover point as shown in figure 6.1.

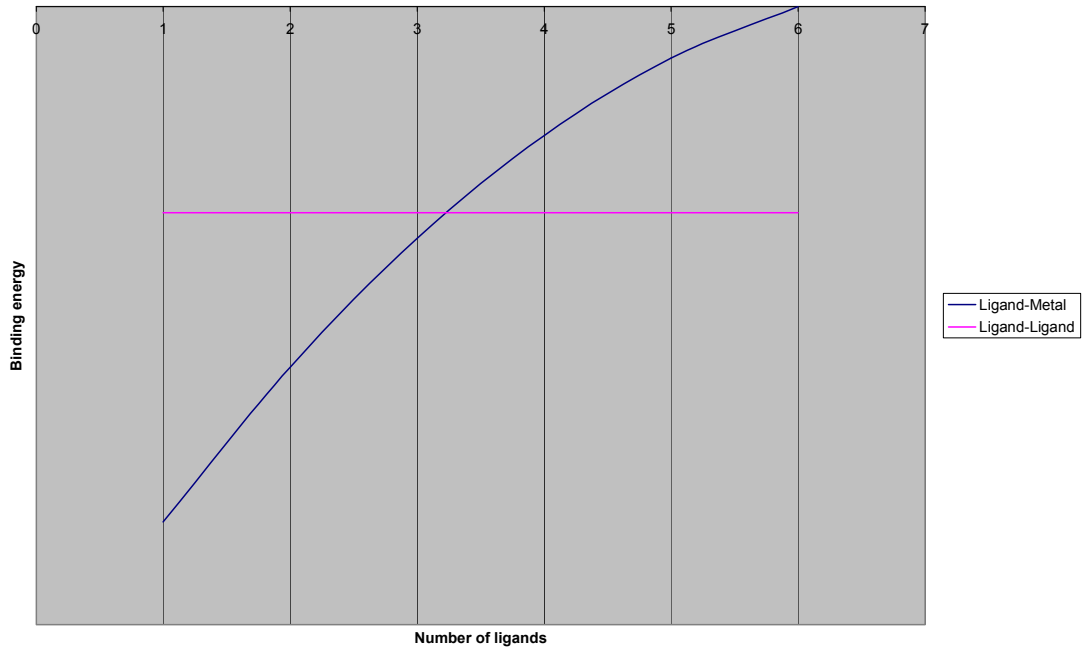


Figure 6.1 – Relative energies of an additional ligand binding to a complex, either directly to the metal ion or to another ligand

The lowest energy setup will always be preferred, so for the metal in the above diagram three ligands would bind to the metal but the fourth would prefer to bind to the other ligands. Since the calculations should produce the lowest energy conformation, both of the lines would not be seen, only the lowest energy conformation producing a hybrid line as shown in figure 6.2.

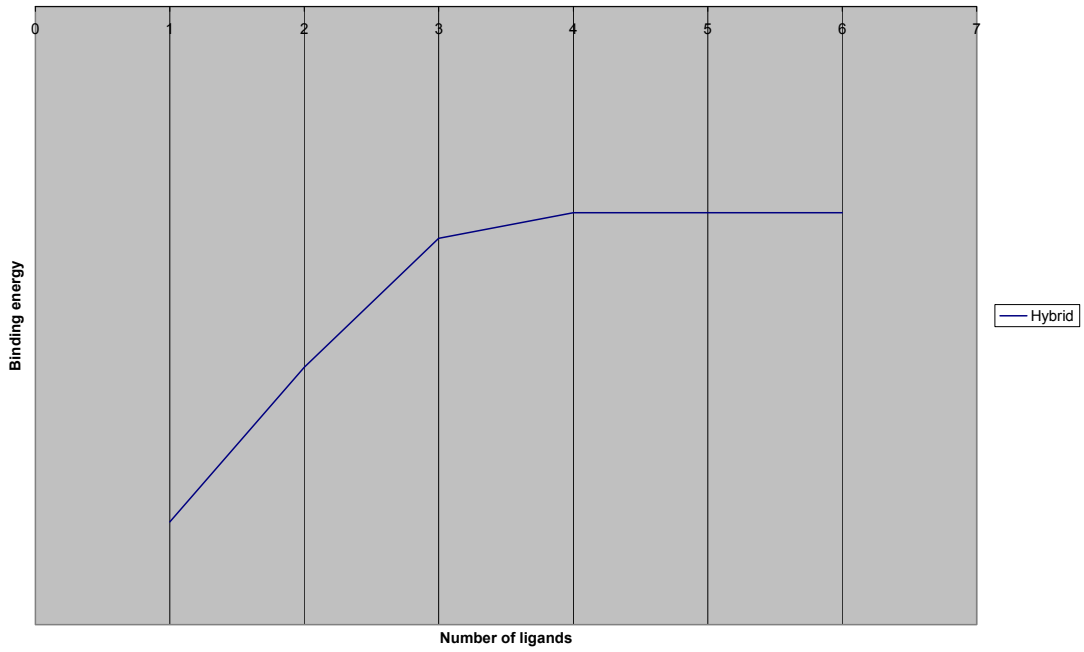


Figure 6.2 – Relative energy of binding an additional ligand to a complex as seen in experiments (*cf* figure 6.2)

In the above diagram the last ligand binds to the metal ion just before the change in gradient of the graph. Such a change should usually be observed for any metal ion. However, as only integral numbers of ligands can be studied, we do not obtain smooth lines as seen in figure 6.1, rather the gradient can appear to change both before and after the cross-over (as in figure 6.2) making identification of the last ligand to bind to the metal problematic. In the example above, there is a question of whether or not the third ligand would bind to the metal and thus whether it would be energetically favourable to design a binding site which would involve the third ligand.

Rather than comparing simple heats of formation, it is more meaningful to compare binding energies, as suggested by Cox *et al.*[170] The heat of formation can be determined using equation 6.1.

$$\Delta H_{(binding)} = \Delta H_{(f,\ total)} - \Delta H_{(f,\ metal)} - n(\Delta H_{(f,\ ligand)})$$

Equation 6.1 – The equation used to determine total heat of formation for a complex

In equation 6.1:

$\Delta H_{(binding)}$ is the total binding energy of the complex

$\Delta H_{(f,\ total)}$ is the energy of formation of each system (the values in the above table)

$\Delta H_{(f,\ metal)}$ is the calculated energy of formation of the metal ion, (-195.6180467 Hartree)

$\Delta H_{(f,\ ligand)}$ is the calculated energy of formation of one ligand

n is the number of ligands

Equation 6.1 gives the net binding energy for all ligands in the complex; it may be more useful to compare either average binding energy per ligand using equation 6.2. Alternatively, incremental binding energy may be useful, calculated through equation 6.3 or 6.4.

$$\Delta H_{ave} = \Delta H_{(binding)} / n$$

Equation 6.2 – The equation used to determine average binding energy for ligands in a complex

$$\Delta H_{inc} = \Delta H_{(binding,n)} - \Delta H_{(binding,n-1)}$$

Equation 6.3 – An equation used to determine incremental binding energy for ligands in a complex

$$\Delta H_{inc} = \Delta H_{(f,total,n)} - \Delta H_{(f,total,n-1)} - \Delta H_{(f,ligand)}$$

Equation 6.4 – The equation used to determine total incremental binding energy for ligands in a complex

Ligand repulsion is included in $\Delta H_{(f,total)}$. The desolvation energy of Cu(I) can be added to the calculation, but it will have the same effect on all of the values, and as such is not important for comparing them to one another.

6.3.3 Hydration effects

When metal ions are not bound, the binding site is likely to be at least partially hydrated. Each ligand is likely to have a different hydration energy, and may lead to a change in the order of ligand binding.

To simulate the effects of partial desolvation of the ligand on binding the metal ion, the energy of each ligand alone was compared with that bound to a single water molecule. Thus the approximate desolvation enthalpies shown in table 6.2 were calculated.

Ligand	Asp	Lys	Cys	His-δ	His-ε	Ser	Tyr
Negative	-0.039431 (-103.53)			-0.036578 (-96.04)	-0.037008 (-97.16)	-0.057358 (-150.59)	-0.032090 (-84.25)
Neutral		-0.018688 (-49.07)	-0.008208 (-21.55)	-0.015963 (-41.91)	-0.016211 (-42.56)	-0.014567 (-38.25)	-0.010057 (-26.40)

Table 6.2 – Calculated hydration enthalpies in Hartree (kJ mol^{-1} in brackets)

The negatively charged ligands can change binding energy to the metal ion by up to 0.057 Hartree, potentially leading to the negatively charged ligand graphs being reordered. However, the neutral ligands are only affected by up to 0.019 Hartree which would not cause a major reordering of binding affinities.

6.4 Results and discussion

6.4.1 Copper (I)

The results were gathered in two batches: those with negatively charged ligands and those with neutral ligands.

The results for the negatively charged ligands are shown in table 6.3. The results of the neutral ligands are shown in table 6.4. Energies correspond to the heat of formation for the metal complex including the specified number of ligands. All values are given in Hartree.

#	aspartate	cysteine	histidine δ	histidine ε	serine	tyrosine
1	-424.563217	-246.1840143	-460.9715384	-460.9718666	-311.2197787	-502.9046153
2	-653.0995822	-296.3278789	-725.9517495	-725.9519526	-426.4019068	-809.8097343
3	-881.449478	-346.2715021		-990.7316063	-541.3325113	-1116.523696
4		-396.0900619				
5						
6						
Ligand alone	-228.574288	-50.1831596	-265.008019	-265.0086215	-115.259328	-306.9347629

Table 6.3 – Enthalpies of formation for negatively charged ligands to Cu(I)

#	cysteine	histidine δ	histidine ε	lysine	serine	tyrosine
1	-246.5196596	-461.3953218	-461.3956128	-291.7675249	-311.6028159	-503.27296
2	-297.1846658	-726.9304852	-726.9315703	-387.682839	-427.3595679	-810.6965576
3	-347.8138067	-992.3867664		-483.5277195	-543.0657551	-1118.076711
4	-398.4360277	-1257.8284		-579.3651107	-658.7634017	
5	-449.0497944				-774.4813235	
6						
Ligand alone	-50.5986996	-265.4241616	-265.4235589	-95.8164834	-115.6748679	-307.3503026

Table 6.4 – Enthalpies of formation for neutral ligands to Cu(I)

In a number of calculations, the SCF failed to converge. Generally speaking, these errors were a result of some ligands being repelled from rest of the complex to leave a 2-coordinate arrangement around the Cu^{+} ion.

The enthalpies of formation are quite difficult to interpret; as such, the total, average and incremental binding energies were studied for later calculations. Tables of raw data may be found in Appendix I b).

Plotting the total binding energies ($\Delta H_{binding}$) yields figure 6.3. Figure 6.3 appears to show the first ligand binding most favourably and the following ligands binding less

so. However, the first two ligands bind reasonably similarly, then the others are more obviously repelled by one another. Cysteine appears to bind the most strongly. There is very little difference in the energies of histidine binding through its delta or epsilon positions so when designing a binding site either may be considered depending on the available backbone positions.

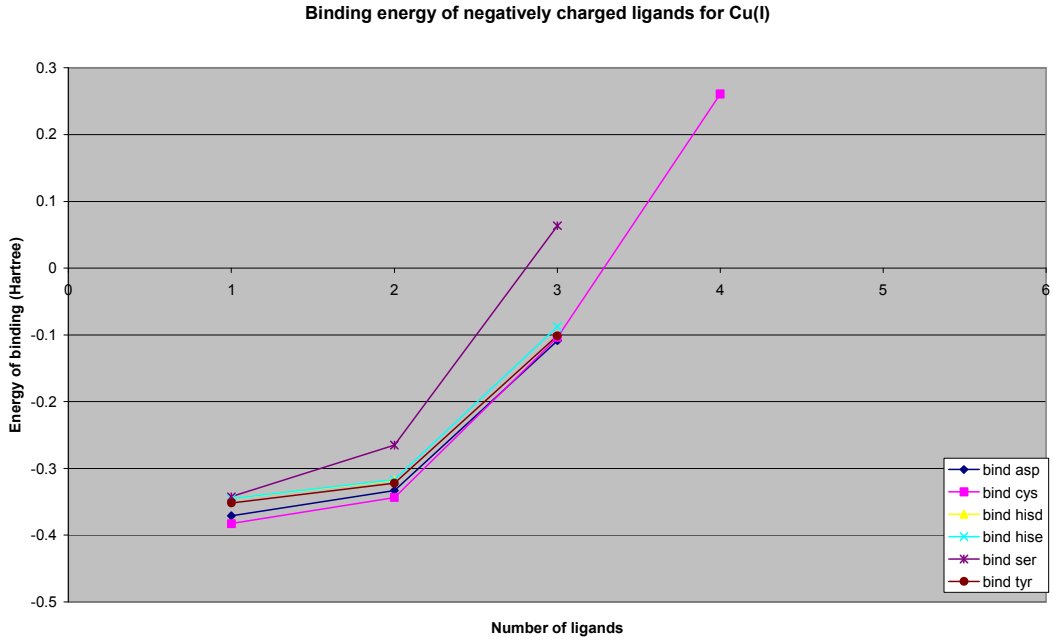


Figure 6.3 – Total binding enthalpies for negatively charged ligands for Cu(I)

ΔH_{ave} can be plotted to produce figure 6.4, whereas ΔH_{inc} can be plotted to produce figure 6.5. For the negatively charged ligands these results show that two ligands may bind. There is less likely to be ligand-ligand binding with negatively charged ligands due to charge repulsion. All of these binding energies appear to be positive and as such binding a second ligand should not be energetically feasible. Part of the reason for the inability to bind a second ligand may be due to ligand-ligand repulsion interactions. If these ligands were in a protein these repulsion effects would not be a problem, but there would be no point in designing a protein with such additional ligands as the change in binding affinity would be negligible. Thus these results imply

that the best binding site for negatively charged ligands binding to Cu(I) would only involve one ligand.

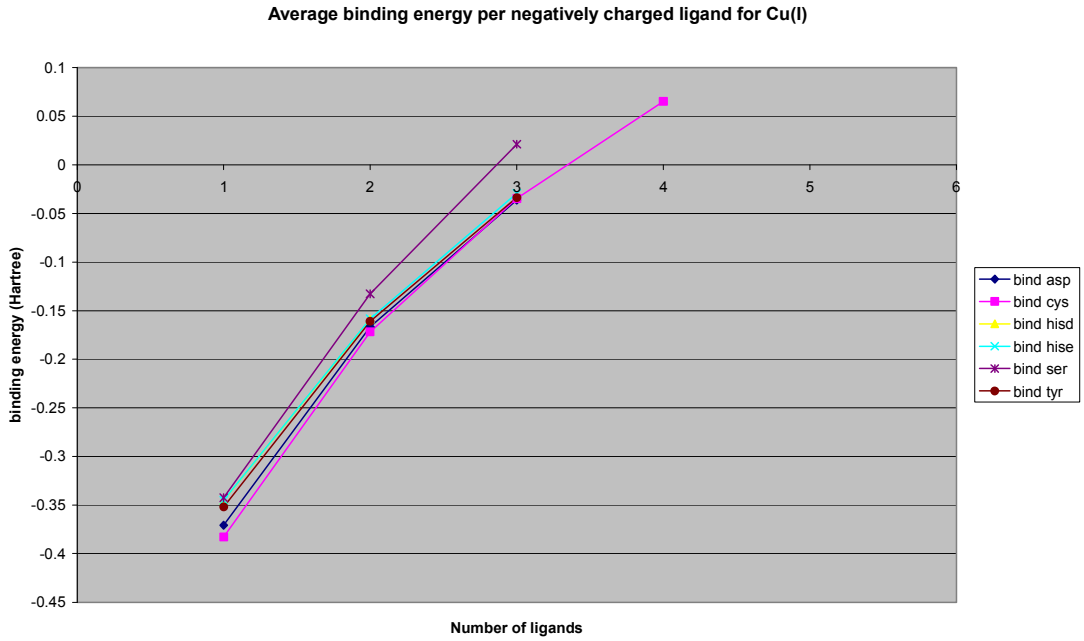


Figure 6.4 – Average binding enthalpies for negatively charged ligands for Cu(I)

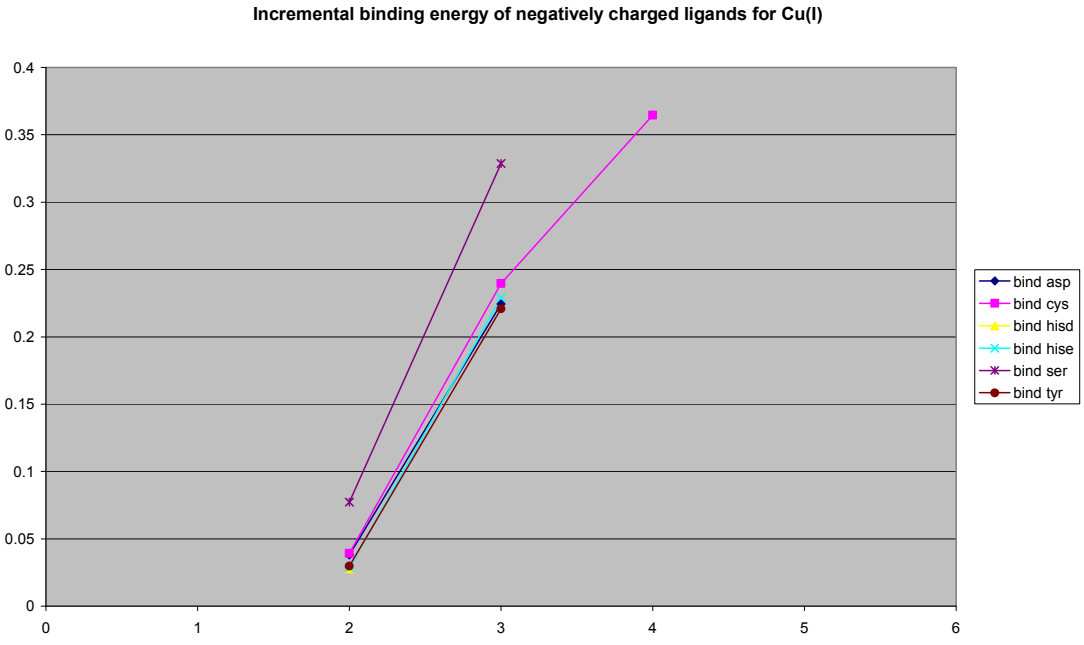


Figure 6.5 – Incremental binding enthalpies for negatively charged ligands for Cu(I)

Plotting the total binding enthalpies for neutral ligands yields figure 6.6. Figure 6.6 does not show any obvious trends. However, the change in binding affinity may be more obvious when ΔH_{ave} and ΔH_{inc} are plotted. ΔH_{ave} is plotted in figure 6.7.

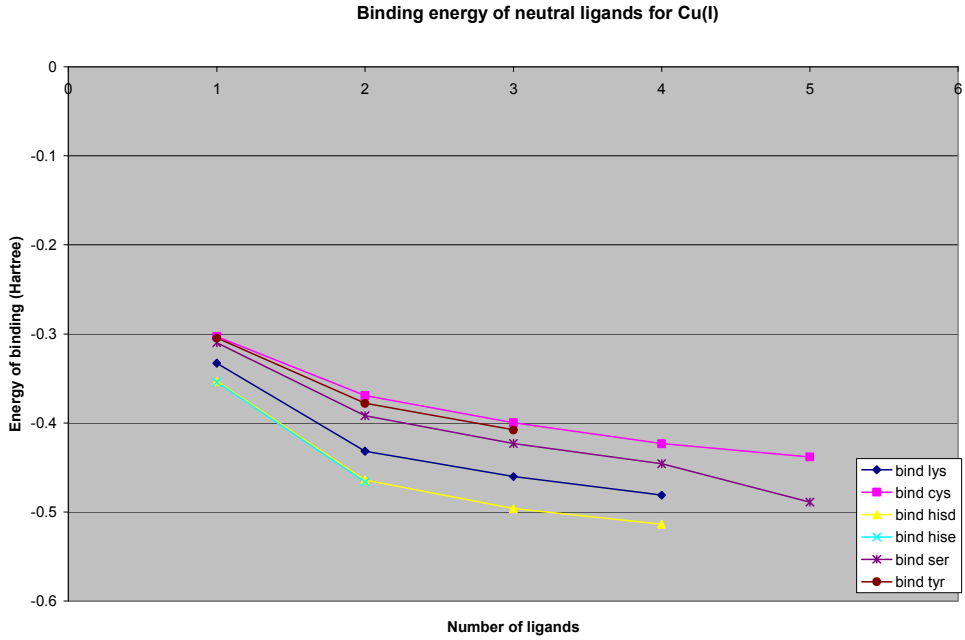


Figure 6.6 – Total binding enthalpies for neutral ligands for Cu(I)

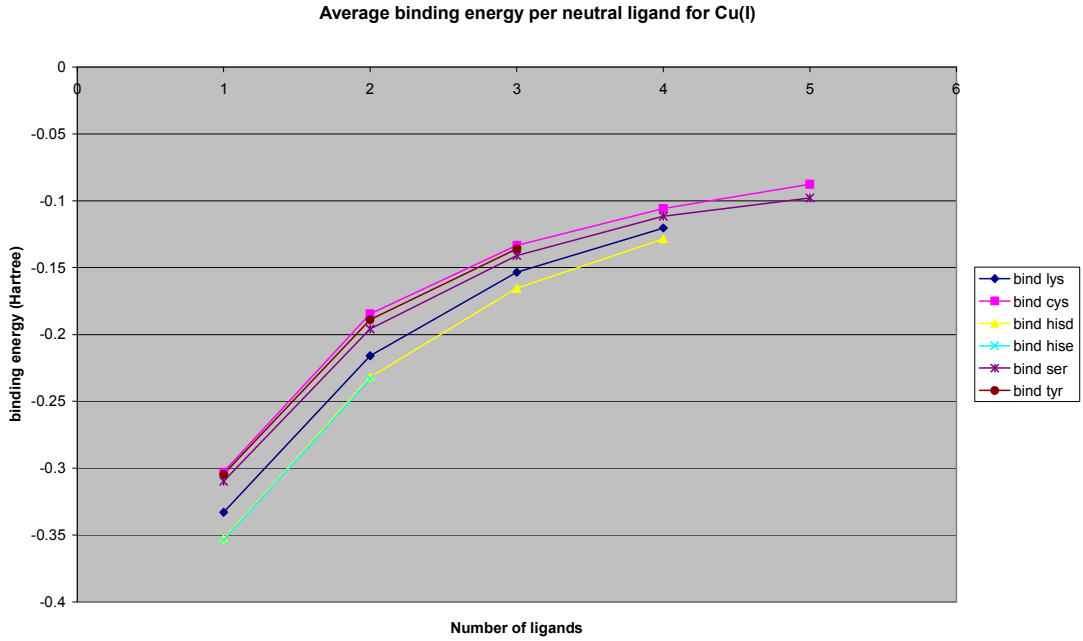


Figure 6.7 – Average binding enthalpies for neutral ligands for Cu(I)

Cox *et al*[170-172] perform incremental binding energy studies on various metal-ligand complexes and ignore the binding energy of the first ligand. Comparing the graphs above for all of the metal ions, the incremental binding energy for one ligand is always much higher than the rest of the ligands. Thus it is almost impossible to see the detail of the rest of the graph. Thus from this point onwards incremental binding energies will be plotted for complexes with two or more ligands. For Cu(I) plotting the incremental binding energy yields figure 6.8.

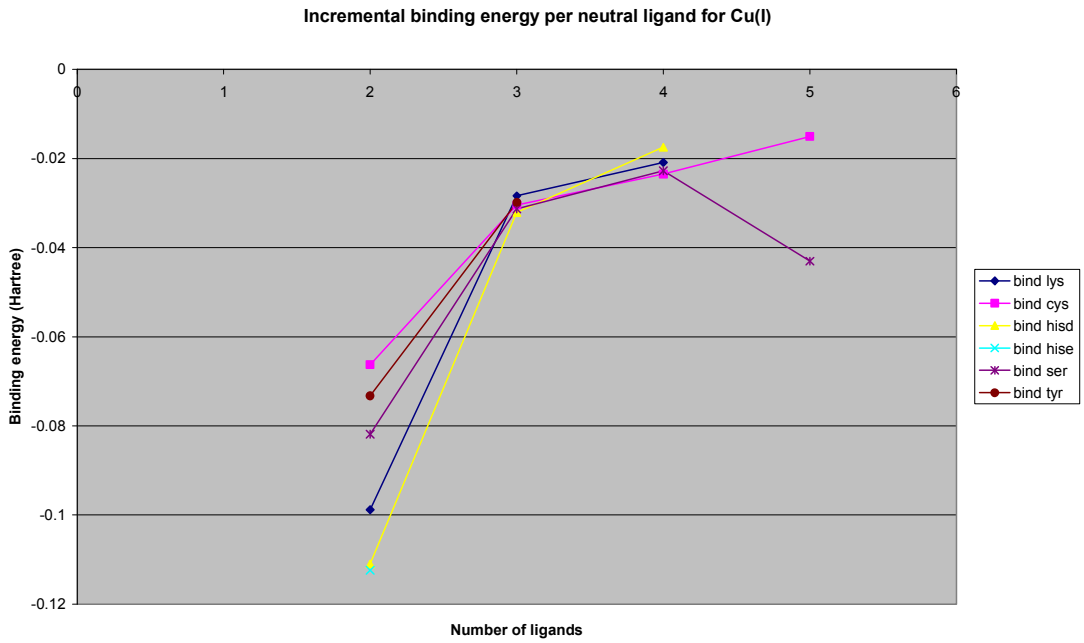


Figure 6.8 – Incremental binding enthalpies for neutral ligands for Cu(I)

Figure 6.8 shows the change in gradient discussed above between two and three ligands, suggesting that the second ligand will bind to the metal ion but the following ligands will bind to one another. There appears to be an outlier on the graph for five serine residues. In this case a distorted tetrahedral complex has been formed around the Cu(I) ion with the extra ligand hydrogen bonding to two of the others, thus increasing the strength of the ligand-ligand interaction and stabilising the structure. The external ligand does not interact directly with the Cu(I) ion as it is too far away.

.

The most stable complex with neutral ligands involves two histidine residues, rather than two cysteine residues as might be expected from the CueR crystal structure. However ΔH_{inc} values for negatively charged ligands show the thiolate form of cysteine to be the preferred first ligand, consistent with our model of CueR in which the metal ion is bound by one thiol and one thiolate. These structures have minimised to the structures found in figure 6.9 and 6.10.

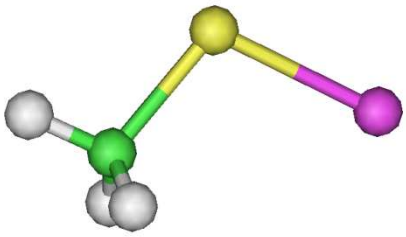


Figure 6.9 – Cu(I) bound to 1 cysteine$^-$ residue (Metal-ligand distance 2.19Å)

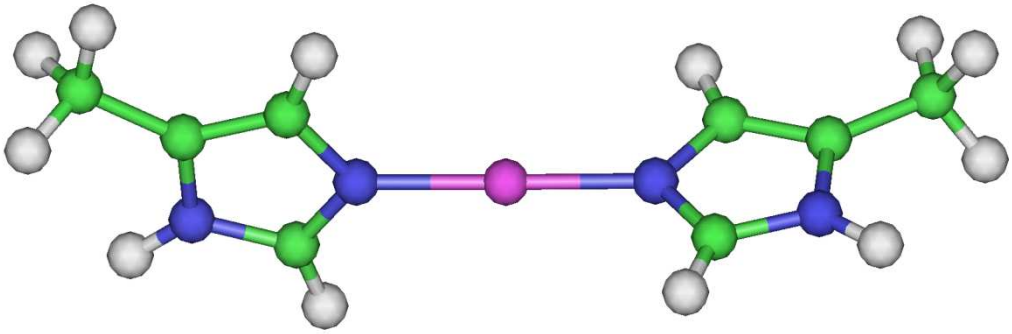


Figure 6.10 – Cu(I) bound to 2 histidine-ε residues (Metal ligand distance 1.89Å)

The structures predicted by these experiments as being the best for Cu(I) are both two coordinate linear structures. Since only one negatively charged ligand binds before the incremental binding energy becomes positive the best binding site will have one negatively charged ligand and one neutral ligand. These calculations show that the negatively charged ligand should be cysteine whilst the neutral ligand should be histidine. The interaction between histidine and cysteine cannot be seen from these

calculations and ideally the relative energies of the second ligand being a cysteine (as seen in CueR) and histidine (as suggested here) would be calculated. Unfortunately insufficient time was available during this project to compare such binding sites.

6.4.2 Mercury (II)

The total binding energy of Hg(II) may be plotted to produce figure 6.11. Very few of the calculations on the negatively charged ligands completed due to convergence failures. There were insufficient results to plot a useful incremental binding energy plot. However, the failure of calculations is a useful result in itself as it shows that no more than three negatively charged ligands will bind to Hg(II) apart from serine. Only one ligand would bind if the energy of binding was positive; the change in gradient approach suggests that the change would be between three and four ligands. Thus these results suggest that three negatively charged cysteine residues would be the most favourable binding site.

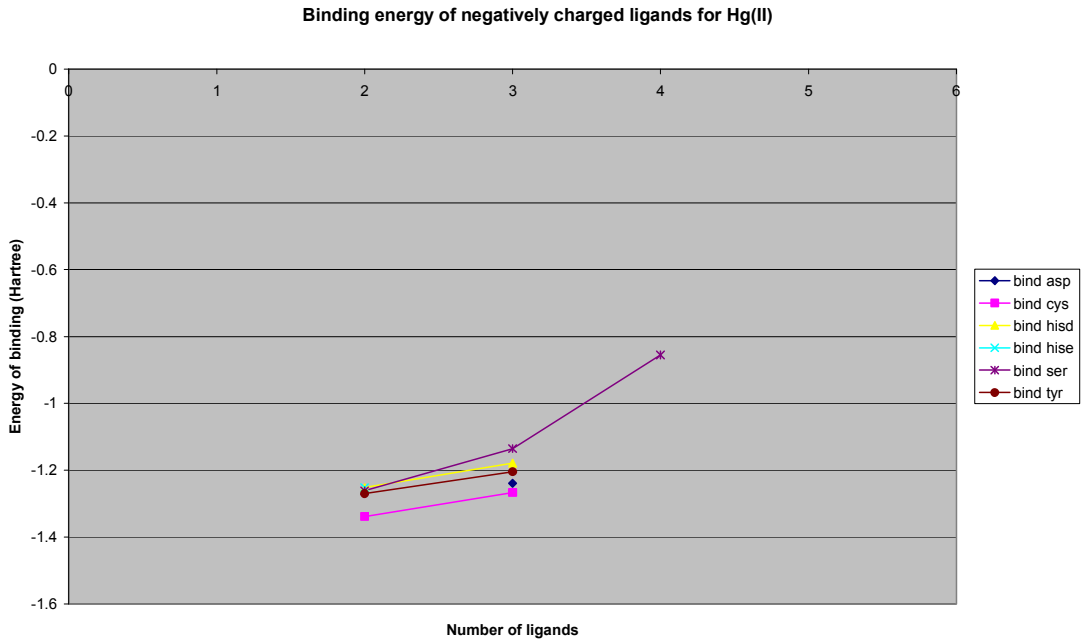


Figure 6.11 – Total binding enthalpies for negatively charged ligands for Hg(II)

For the neutral ligands sufficient data points could be collected to plot a graph of incremental binding energies as shown in figure 6.12. The change in gradient between two and three ligands indicates that two neutral ligands is the preferred arrangement with lysine as the preferred ligand as seen in figure 6.13.

However, as the location of the crossover between ligand-ligand and ligand-metal binding interactions can result in some doubt as to which is the last ligand to bind the metal ion, it is worth checking the structure of the complex visually to identify exactly how many ligands can bind to the metal ion. For Hg(II), visual inspection suggests three ligands can be accommodated around the metal ion. The MerR structure with a metal-binding site formed of three cysteine residues should be a good binding site, but changing some of these to either lysine or histidine may lead to increased binding affinity. The minimised site containing three cysteine residues is shown in figure 6.14.

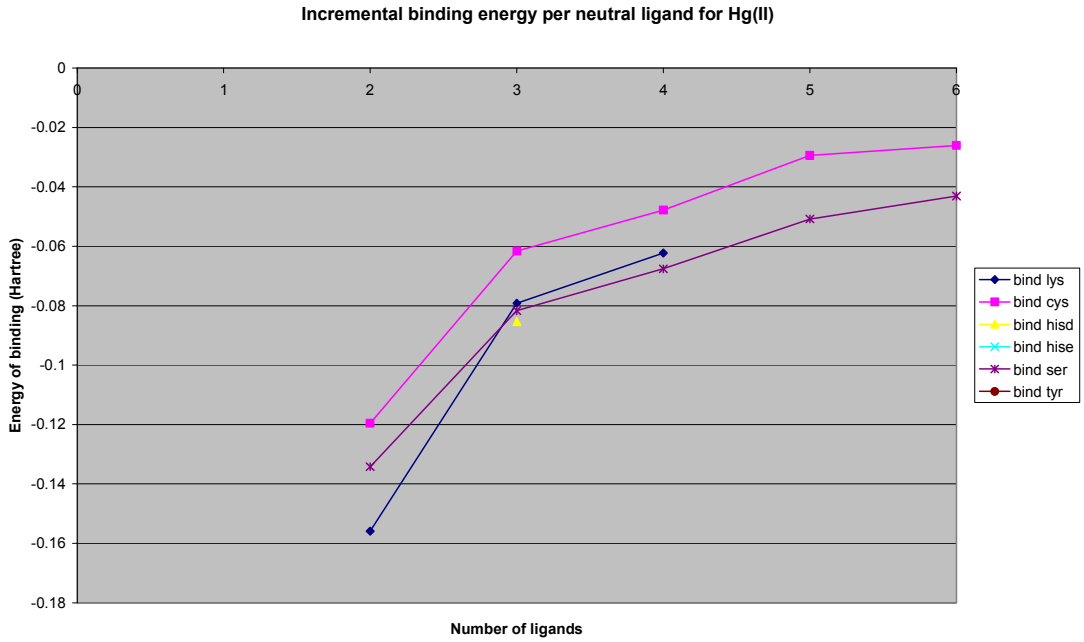


Figure 6.12 – Incremental binding enthalpies for neutral ligands for Hg(II)

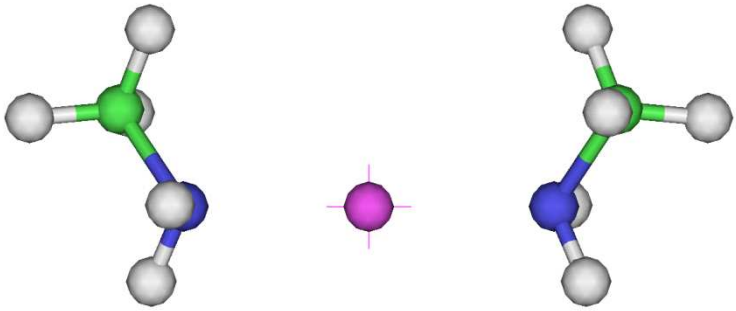


Figure 6.13 – Hg(II) bound to 2 lysine residues (Metal-ligand distance 2.22Å)

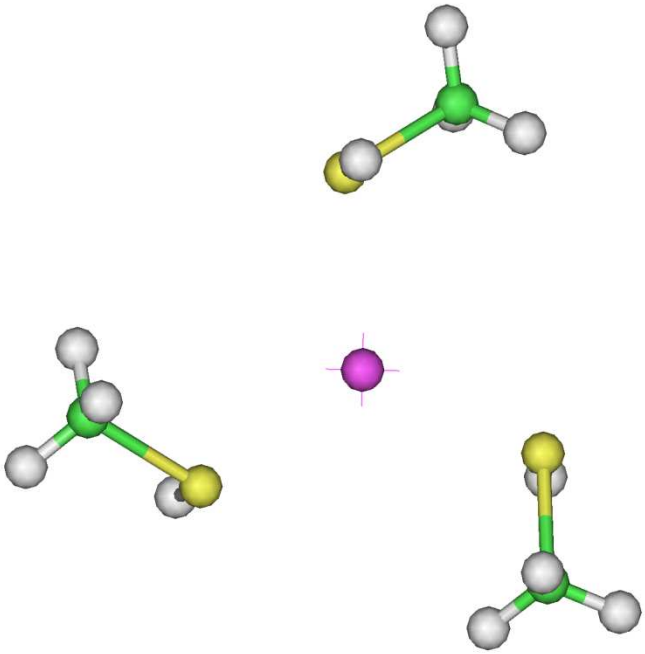


Figure 6.14 – Hg(II) bound to 3 cysteine residues (Metal-ligand distance 2.71Å)

6.4.3 Zinc (II)

The incremental binding energy of Zn(II) may be plotted to produce figures 6.15 and 6.16. ΔH_{inc} for negative ligands shows a pronounced gradient change once four ligands are added to the system, arguing that three negatively charge ligands can be accommodated around the Zn(II) ion, but a positive ΔH_{inc} going beyond two ligands shows that these structures are merely local minima. As in the previous cases, overall

neutral complexes are preferred. When neutral ligands are considered, the gradient change occurs once five ligands have been added. The gradient change suggests that Zn(II) will bind four ligands. The negatively charged ligands show better binding for the "harder" ligand, tyrosine, and worse binding for the "softer" cysteine, whilst the neutral ligands show particularly poor binding for cysteine. These results suggest that oxygen based ligands may bind more strongly to Zn(II). In the crystal structure Zn(II) has been co-crystallised with phosphate ions which should bind strongly according to this data. Overall it appears that Zn(II) prefers four ligands and may bind more strongly to harder ligands such as serine rather than softer ligands such as cysteine, as shown in the structure in figure 6.17.

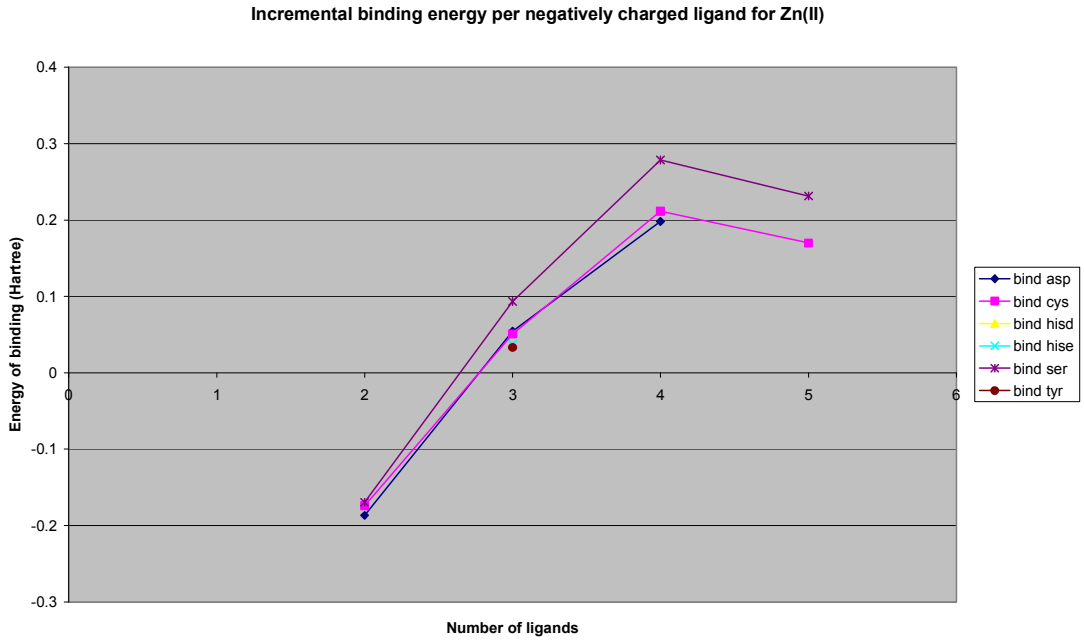


Figure 6.15 – Incremental binding enthalpies of negatively charged ligands for Zn (II)

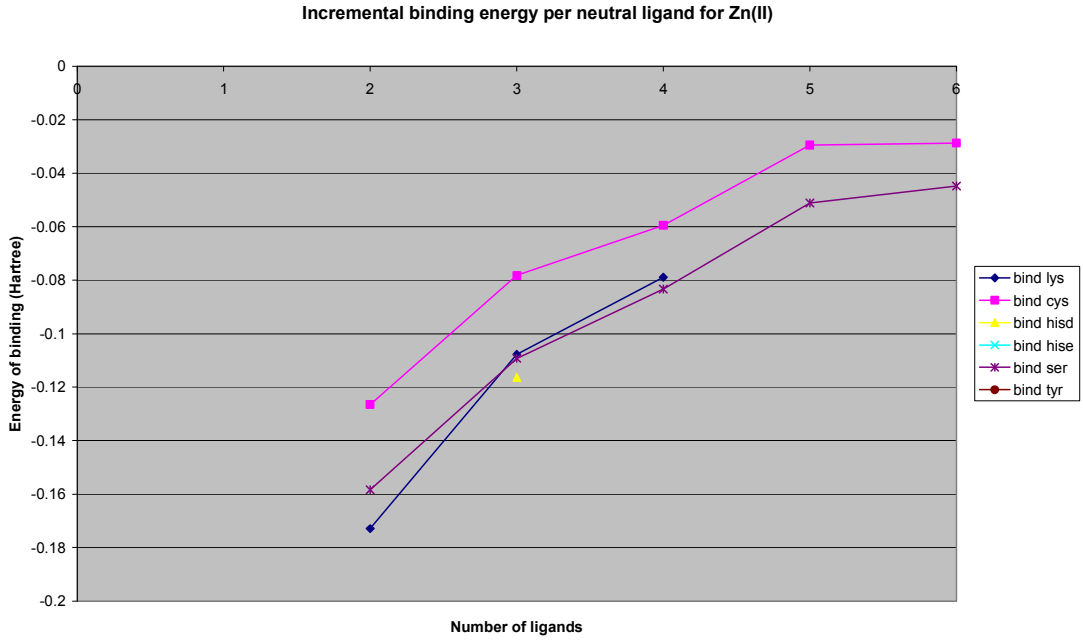


Figure 6.16 – Incremental binding enthalpies of neutral ligands for Zn (II)

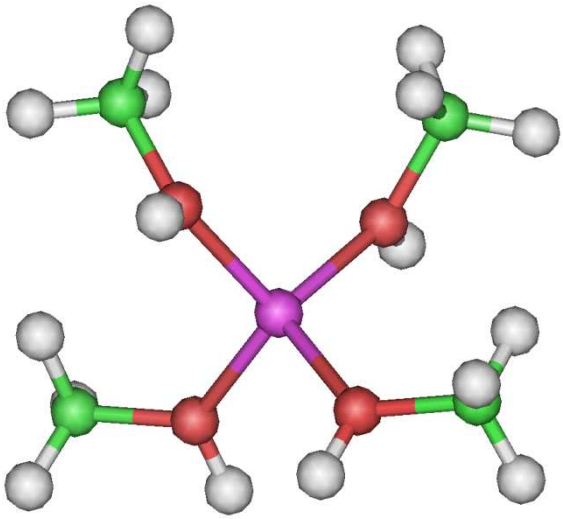


Figure 6.17: Zn(II) bound to 4 serine residues (Metal-ligand distance 2.00Å)

6.4.4 Silver (I)

Ag(I) is expected to bind in a similar way to Cu(I) so both the total binding energy and incremental binding energy were plotted for the negatively charged ligands to

produce figures 6.18 and 6.19. For the neutral ligands only the incremental binding energy need be considered as shown in figure 6.20.

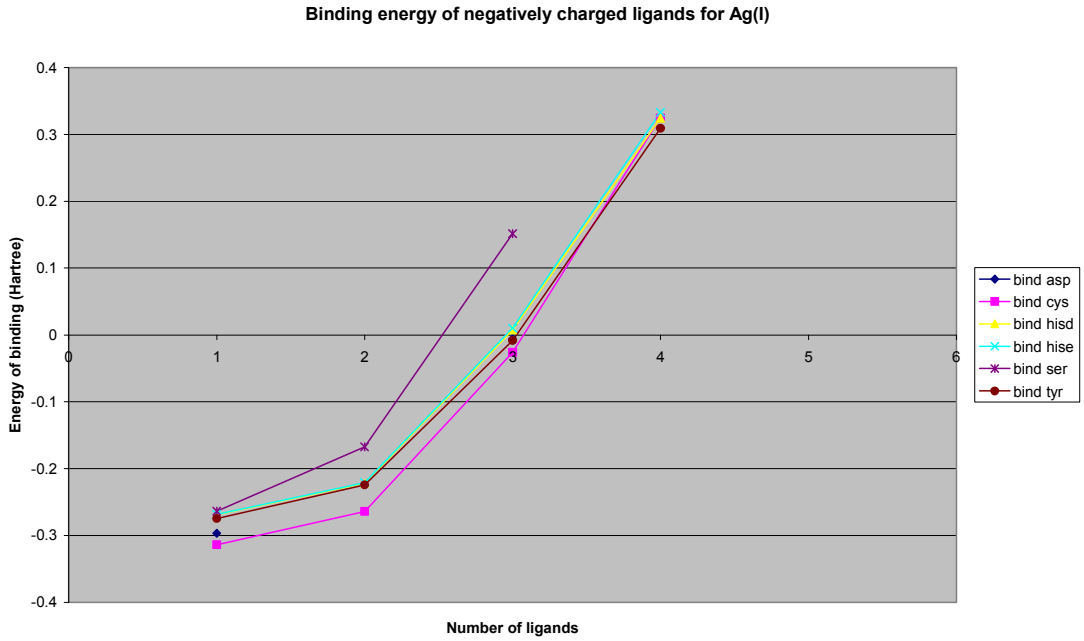


Figure 6.18 – Total binding enthalpies of negatively charged ligands for Ag(I)

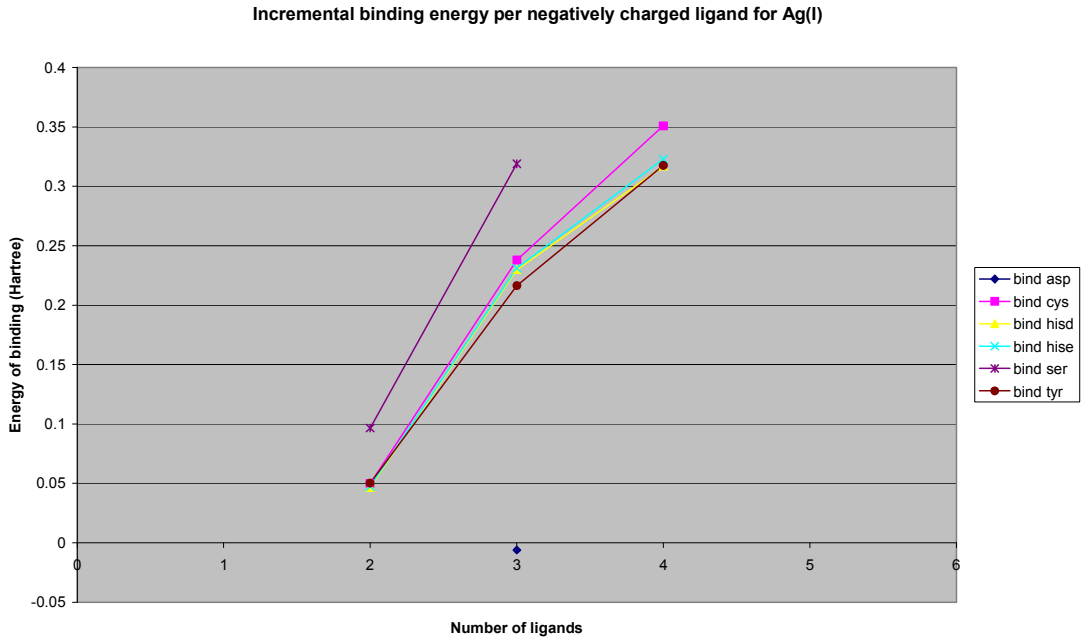


Figure 6.19 – Incremental binding enthalpies of negatively charged ligands for Ag(I)

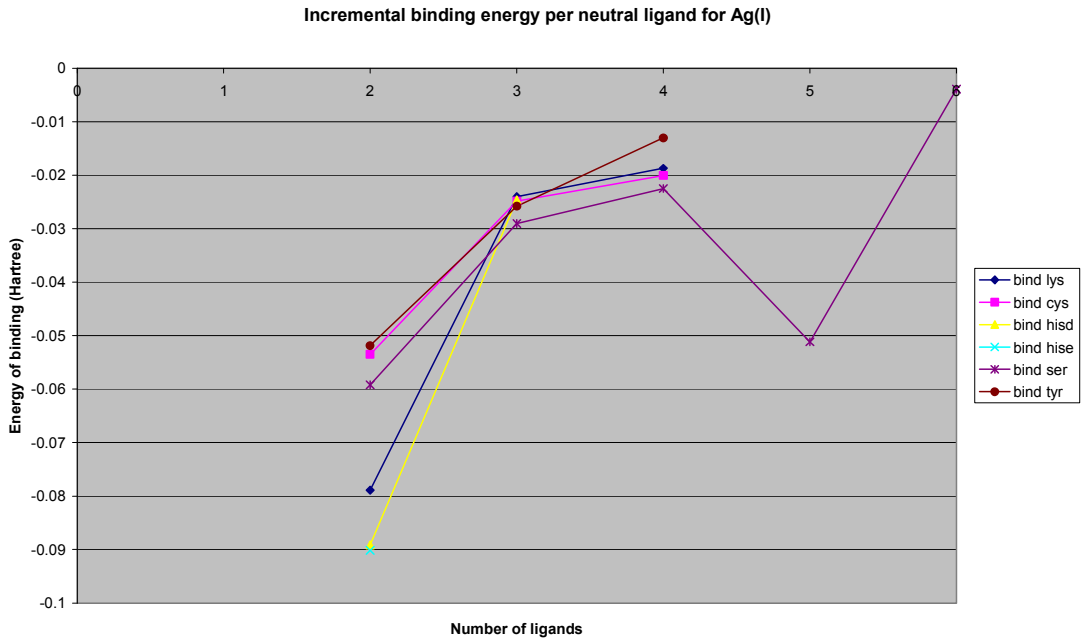


Figure 6.20 – Incremental binding enthalpies of neutral ligands for Ag(I)

These graphs show a change in gradient between two and three neutral ligands. The results are very similar to those for Cu(I). The stable structure involving five neutral cysteine residues was again observed, though, as before, only two of these ligands actually bind the metal ion. It appears that only one negatively charged residue will bind but up to two neutral ligands can be accommodated, as shown in the minimised structures shown in figures 6.21 and 6.22. The similarity to Cu(I) is to be expected, however the exact bond lengths differ (histidine N-M^+ distance is 2.10Å for Ag(I) vs. 1.89Å for Cu(I)).

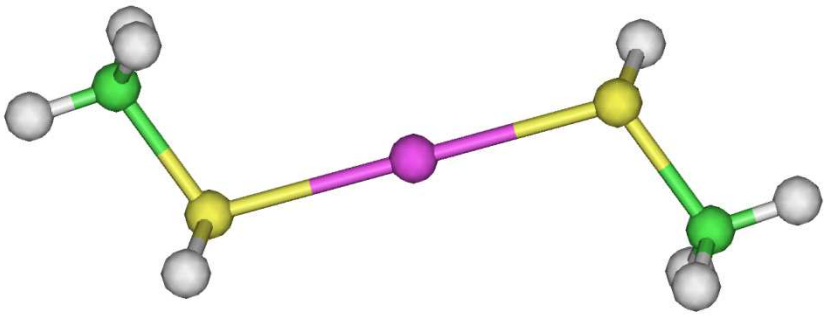


Figure 6.21: Ag(I) bound to 2 cysteine residues (Metal-ligand distance 2.54Å)

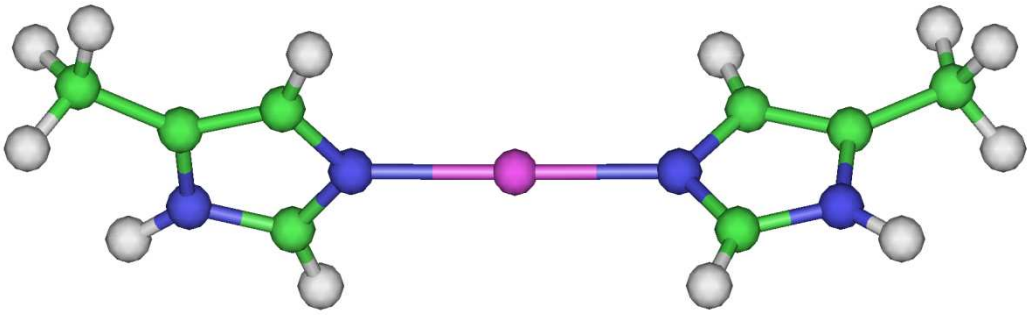


Figure 6.22: Ag(I) bound to 2 histidine-ε residues (Metal-ligand distance 2.10Å)

6.4.5 Gold (I)

Au(I) is similar to both Ag(I) and Cu(I) in size, charge and orbital occupation, therefore the results should be similar. Several of the calculations did not complete so the total binding energy was plotted to produce figure 6.23. The general trend in figure 6.23 seems to be similar to Cu(I) despite the missing data points. The incremental binding energies for the neutral ligands can be plotted to produce figure 6.24.

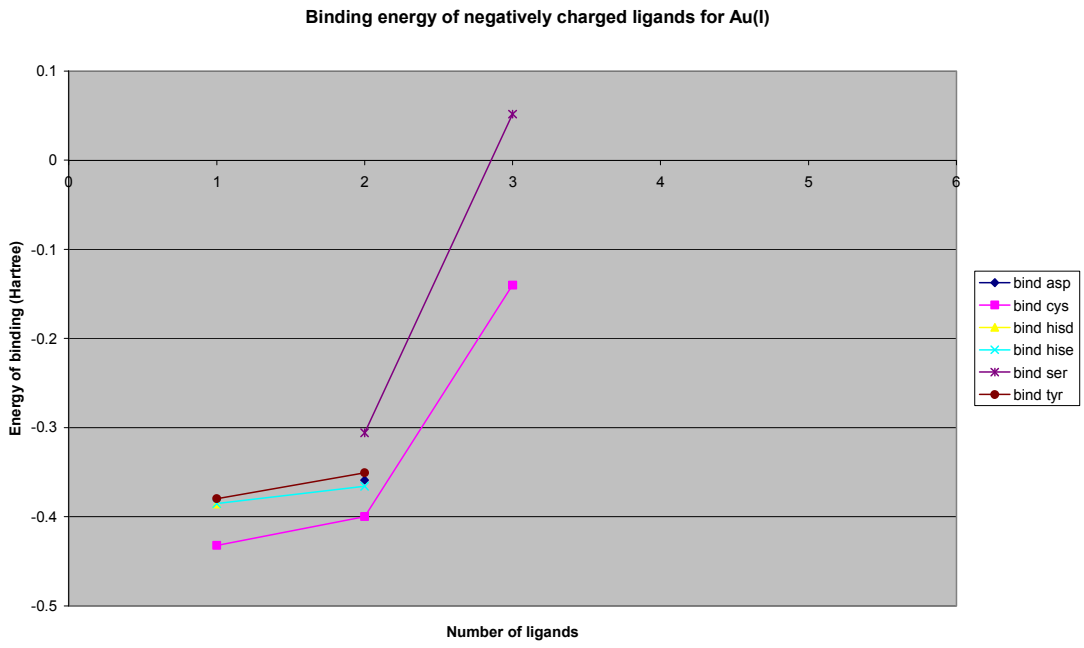


Figure 6.23 – Total binding enthalpies of negatively charged ligands for Au(I)

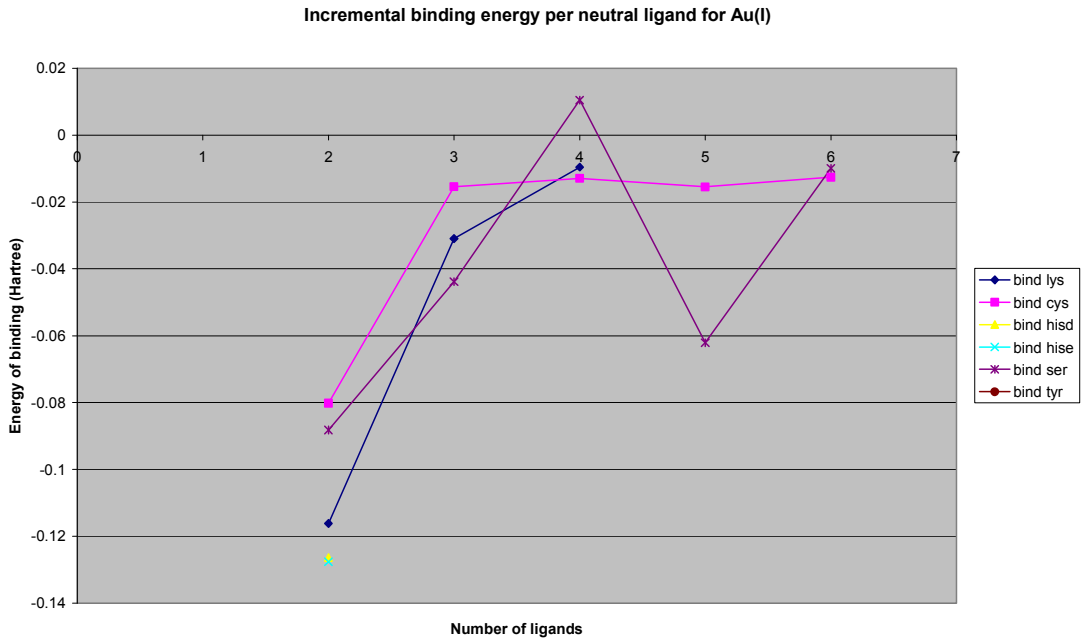


Figure 6.24 – Incremental binding enthalpies of negatively charged ligands for Au(I)

As with both Cu(I) and Ag(I), a two coordinate structure is predicted with either one negative cysteine residue or two neutral histidine residues as shown in figure 6.25. The stable structure with five neutral serine residues was once again observed.

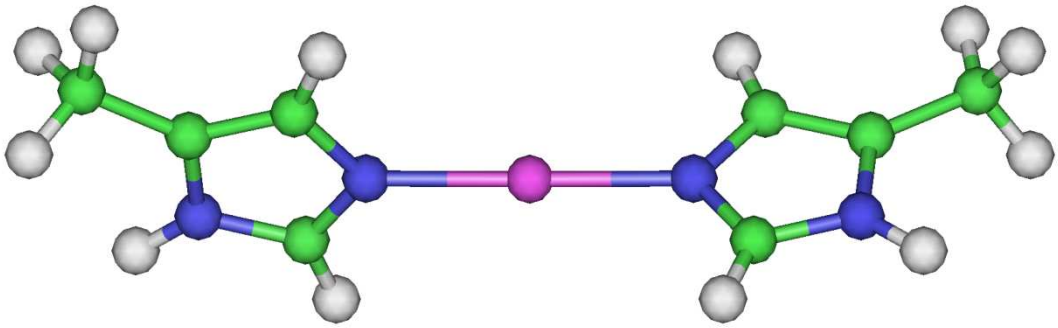


Figure 6.25 – Au(I) bound to 2 histidine-ε residues (Metal-ligand distance 2.01Å)

6.4.6 Cadmium (II)

The incremental binding energy for the negatively charged ligands may be plotted, although the total binding energy graph appears to be more useful as shown in figure

6.26 and 6.27. For the neutral ligands the incremental binding energy is most useful as shown in figure 6.28.

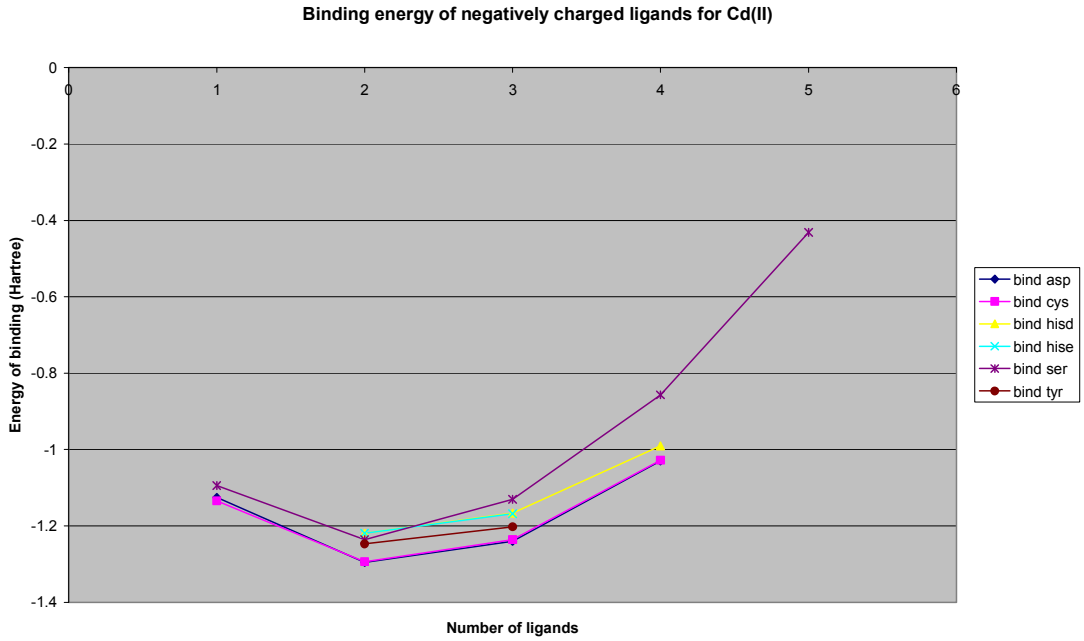


Figure 6.26 – Total binding enthalpies of negatively charged ligands for Cd(II)

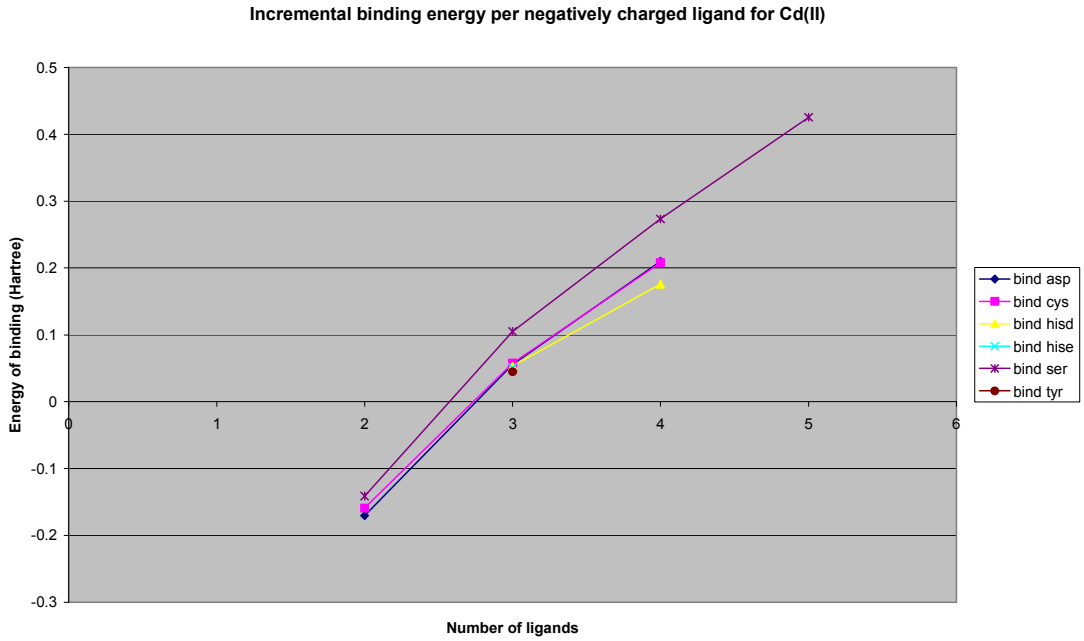


Figure 6.27 – Incremental binding enthalpies of negatively charged ligands for Cd(II)

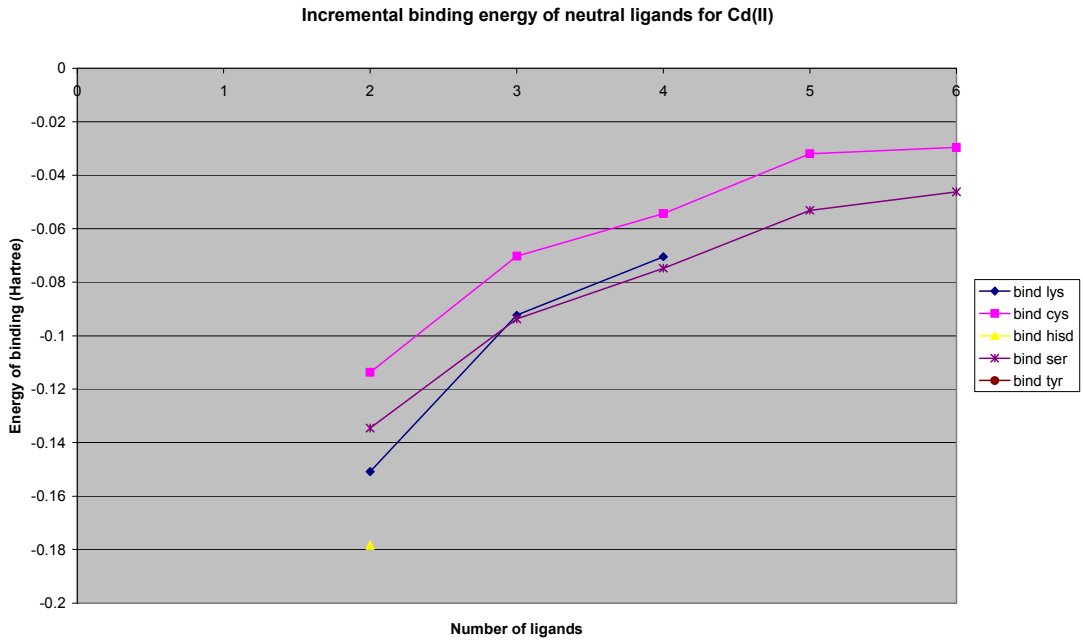


Figure 6.28 – Incremental binding enthalpies of neutral ligands for Cd(II)

For the negatively charged ligands there appears to be a steady decrease in binding energy until around four ligands are bound (a total -2 charge), but as ΔH_{inc} is positive for more than two ligands, these represent local minima rather than useful complexes. The minimised structures with two negatively charged ligands are shown in figure 6.29 and 6.30. Adding more ligands to the 2-coordinate structures becomes difficult due to charge repulsion. For the neutral ligands there is a steady decrease in incremental binding energy until a six-coordinate octahedral structure is produced as shown in figure 6.31.

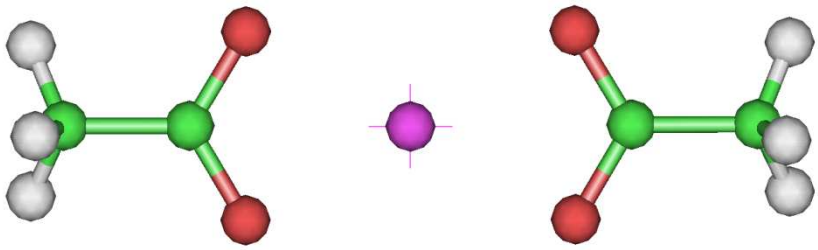


Figure 6.29– Cd(II) bound to 2 aspartate residues (Metal-ligand distance 2.25Å)

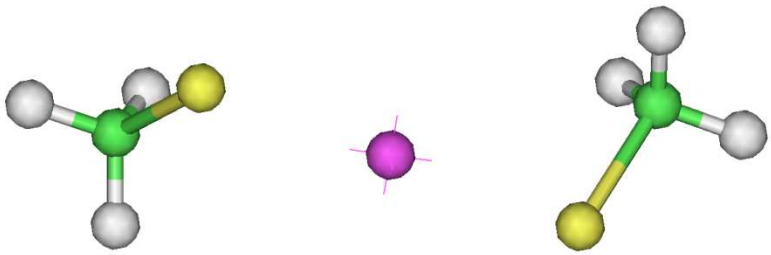


Figure 6.30 – Cd(II) bound to 2 cysteine residues (Metal-ligand distance 2.44Å)

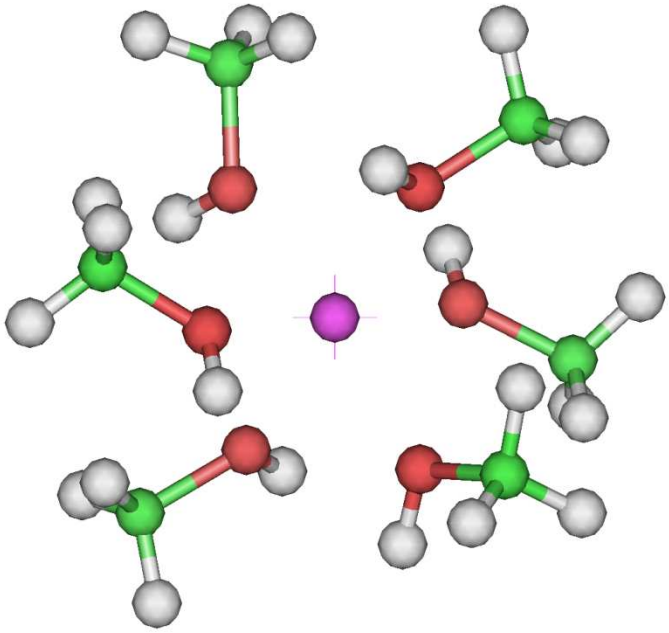


Figure 6.31 – Cd(II) bound to 6 serine residues (Metal-ligand distance 2.29Å)

Only a few of the larger ligands (tyrosine and histidine) were able to bind due to steric effects. There do appear to be slight changes in the gradient of the incremental binding energy for neutral ligands at three and five ligands, but these are due to small rearrangements of the structure rather than a second ligand shell being produced. Serine is by far the strongest binding neutral ligand whereas aspartate and cysteine are the strongest binding negatively charged ligands. The best binding site may therefore contain two cysteine or aspartate residues (to attain charge neutrality) plus four serine

residues to produce an octahedral binding site. However, as aspartate may bind as a bidentate ligand only two serine residues may be required.

6.4.7 Lead (II)

Insufficient data points are available to produce a useful incremental binding energy graph for negatively charged ligands, so the total binding energy must be studied to produce figure 6.32. The incremental binding energies for the neutral ligands may be studied to produce figure 6.33.

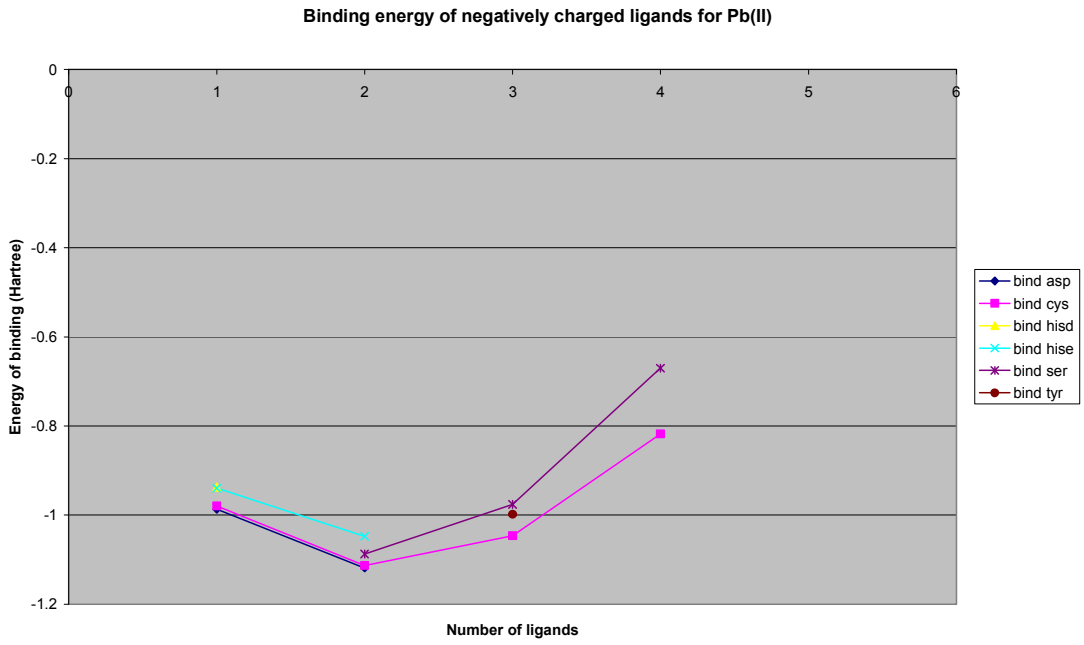


Figure 6.32 – Total binding enthalpies of negatively charged ligands for Pb(II)

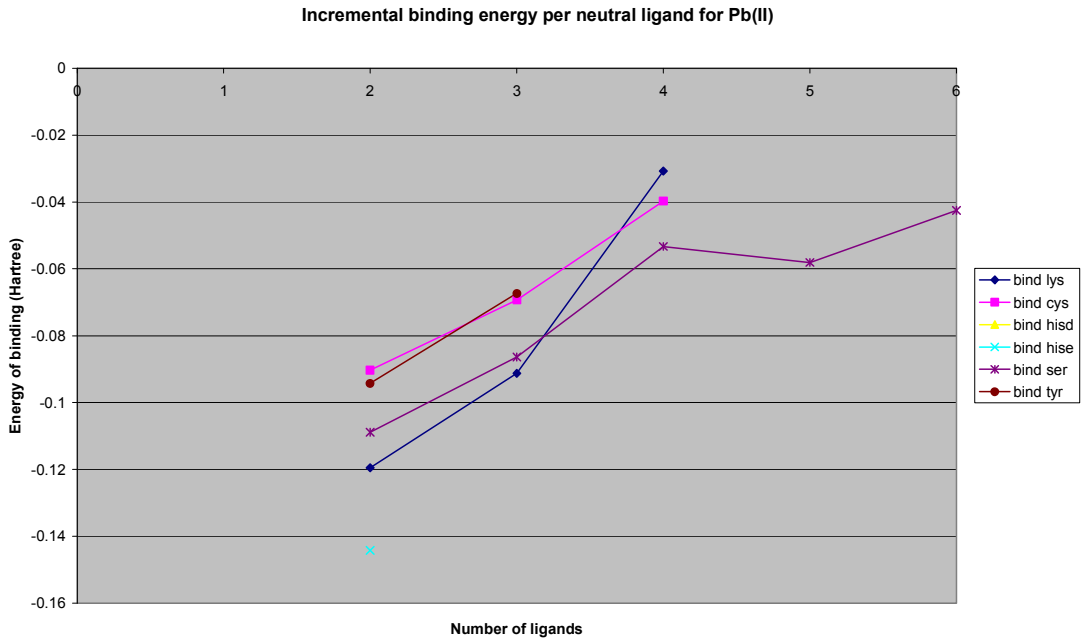


Figure 6.33 – Incremental binding enthalpies of neutral ligands for Cd(II)

Neutral ser shows a significant change in gradient once four ligands have been added. Inspection of structures involving four or more neutral serine residues show an octahedral structure, with two adjacent empty sites and any remaining ligands hydrogen bonding to the others as shown in figures 6.34 and 6.35.

Thus it may not be possible to bind more than four of such ligands. For the other neutral ligands four ligands bind to form a tetrahedral site and others will not bind. A maximum of two negatively charged ligands bind with cysteine or aspartate preferred, while up to four ligands can be accommodated in all. The binding of two negatively charged ligands is shown below in figure 6.36 and 6.37. Thus the strongest binding site is likely to contain two aspartate residues plus two neutral lysine or serine residues.

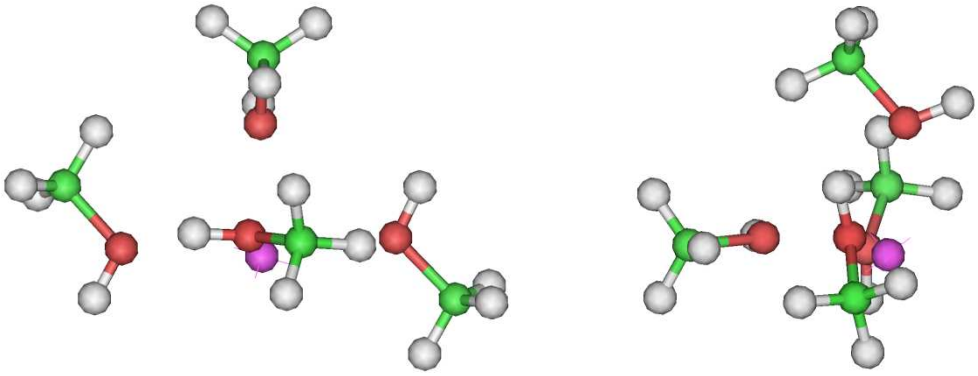


Figure 6.34 (a) and (b) – Pb(II) bound to 4 serine residues, (a) and (b) showing different views (Metal-ligand distance ranging from 2.25-2.53Å)

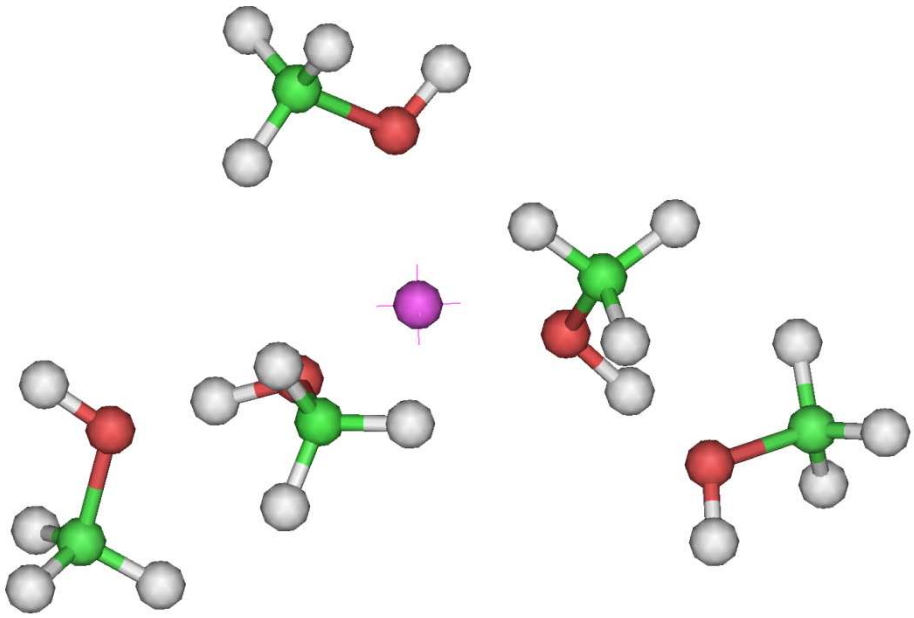


Figure 6.35 – Pb(II) bound to 5 serine residues (Metal-ligand distance 2.19-2.31Å)

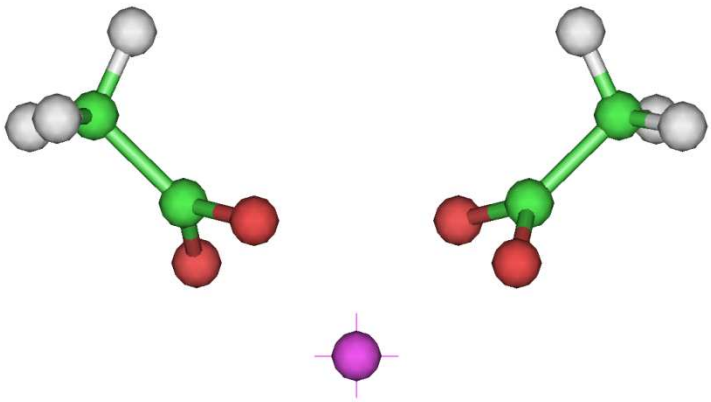


Figure 6.36 – Pb(II) bound to 2 aspartate residues (Metal-ligand distance 2.26Å)

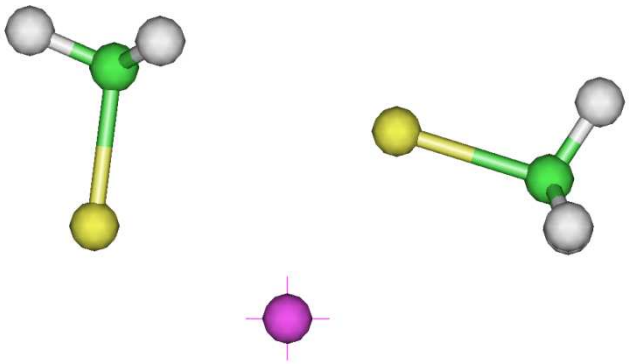


Figure 6.37 – Pb(II) bound to 2 cysteine residues (Metal-ligand distance 2.57Å)

6.4.8 Arsenic (III)

The total binding energy can be plotted for the negatively charged ligands to produce figure 6.38. However, the incremental binding energy clearly shows a maximum of three charged ligands can bind as shown in figure 6.39. Insufficient calculations completed for the neutral ligands to produce an incremental binding energy graph so the total binding energy must be plotted to produce figure 6.40.

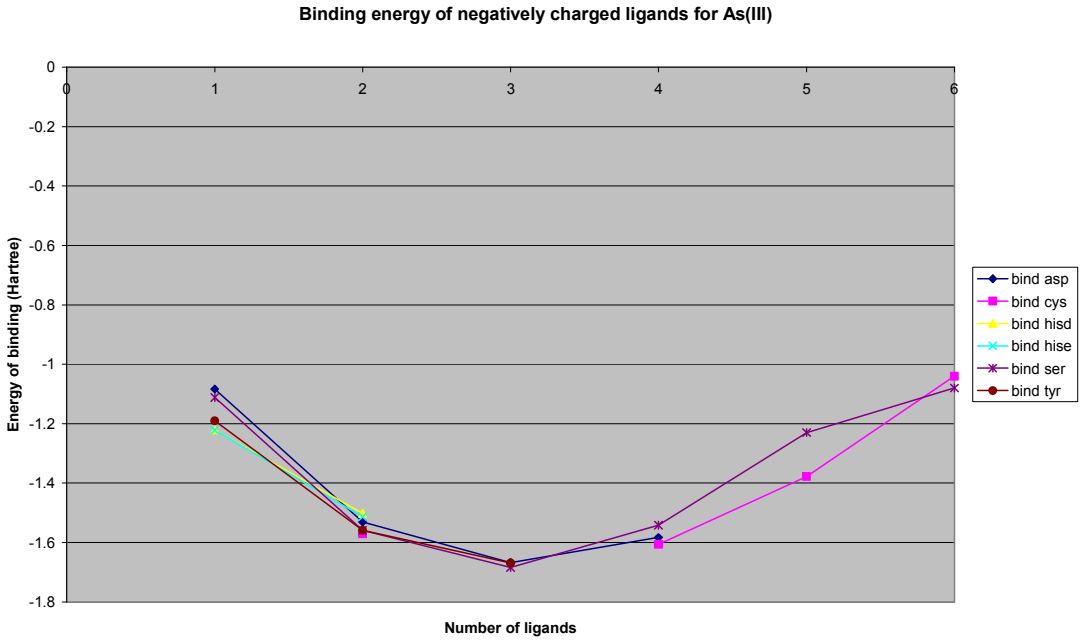


Figure 6.38 – Total binding enthalpies of negatively charged ligands for As(III)

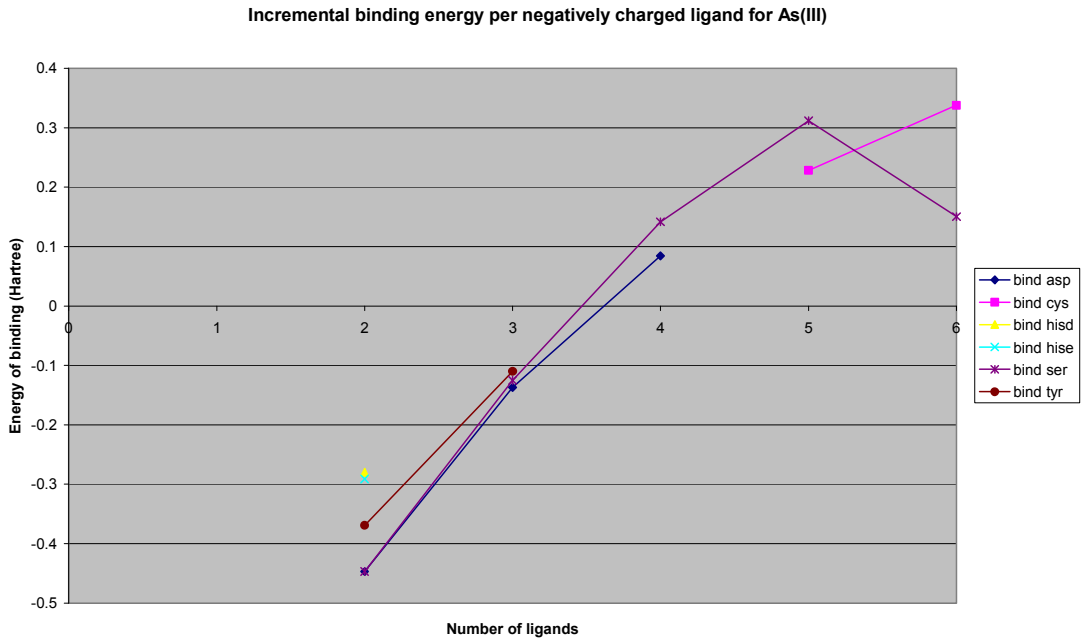


Figure 6.39 – Incremental binding enthalpies of negatively charged ligands for As(III)

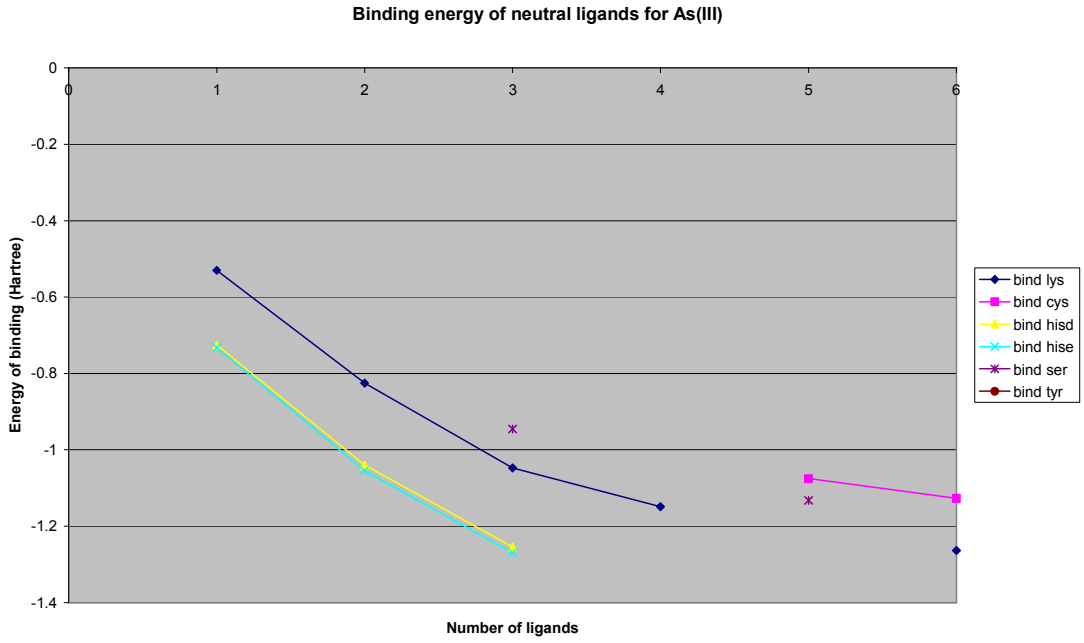


Figure 6.40 – Total binding enthalpies of neutral ligands for As(III)

The negatively charged ligands bind most strongly when charge neutrality is achieved as found in the minimised site shown in figure 6.41. However, binding to the metal may continue to occur for up to six ligands for cysteine (serine with six ligands is a

hydrogen bonded structure). For the neutral ligands histidine binds most strongly when up to three are present; SCF convergence did not occur for more than three ligands as they appeared to drift away from the metal ion. If more than three ligands are bound, neutral lysine is the favoured ligand. The minimised structure involving three neutral lysine residues is shown in figure 6.42.

Combining the two sets of results suggests that the strongest binding would most likely involve three aspartate ligands to attain charge neutrality plus three lysine ligands.

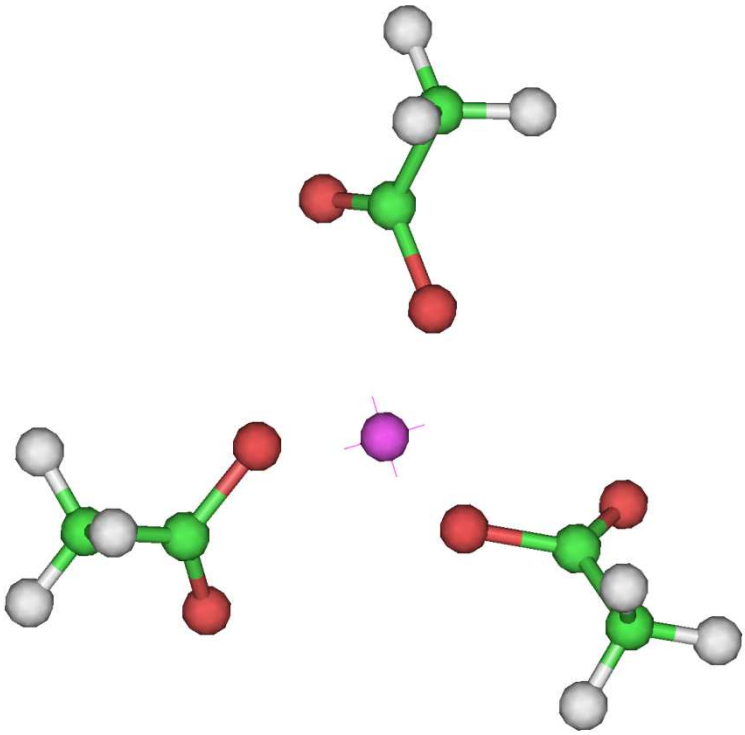


Figure 6.41 – As(III) bound to 3 aspartate residues (Metal-ligand distance 1.86Å)

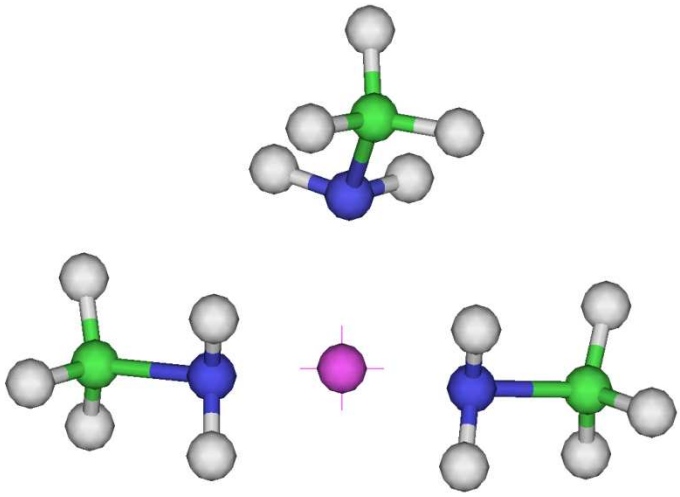


Figure 6.42 – As(III) bound to 3 lysine residues (Metal-ligand distance 2.05Å)

6.4.9 Arsenic (V)

Calculations involving As(V) produced considerable problems with convergence failures and not a single complex involving charge neutral ligands could be obtained. A graph of the total binding energy is most useful for studying the negatively charged ligands due to missing data points as shown in figure 6.43. The incremental binding energy may still be plotted even though very few data points are present as shown in figure 6.44. Figure 6.44 shows that the incremental binding energy becomes positive after five ligands are bound so it may not be energetically favourable to design a protein with six ligands.

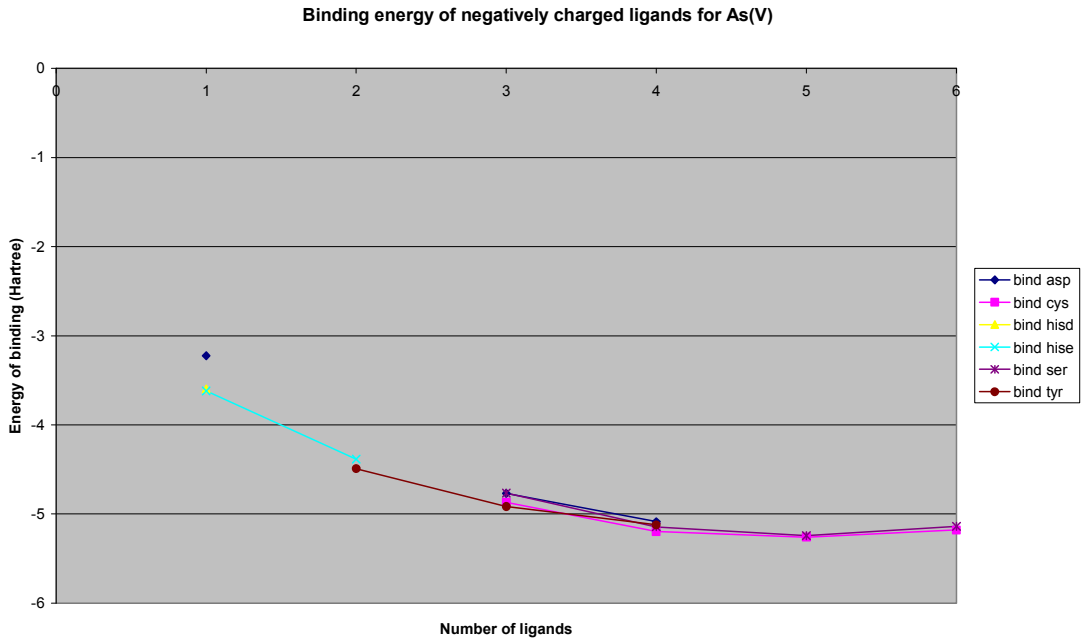


Figure 6.43 – Total binding enthalpies of negatively charged ligands for As(V)

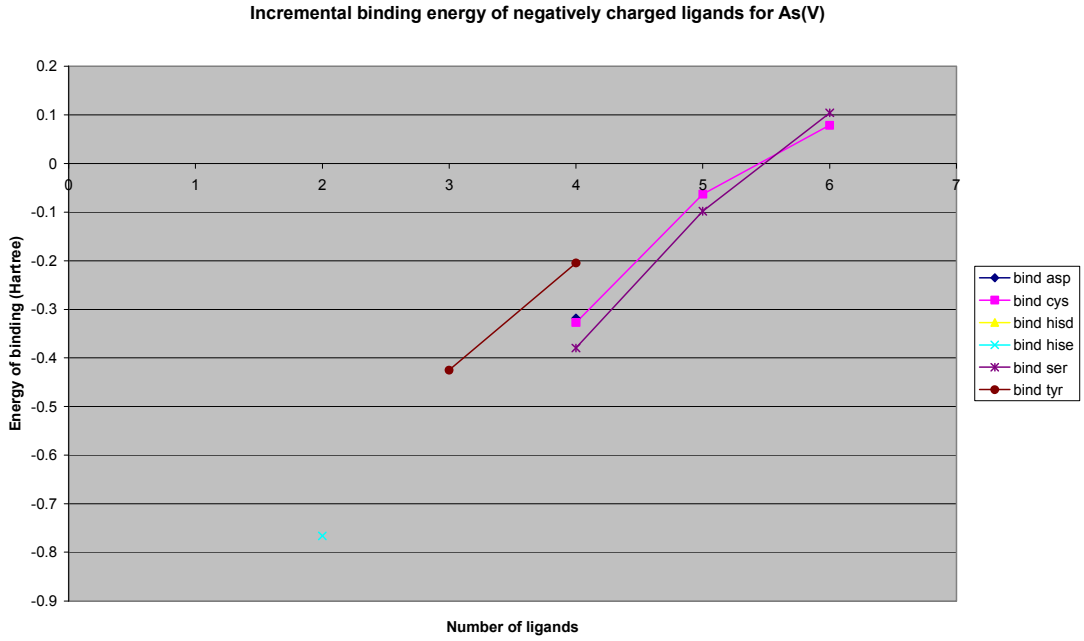


Figure 6.44 – Incremental binding enthalpies of negatively charged ligands for As(V)

The strongest binding again appears to be found for charge neutrality, five ligands, as shown in figure 6.45. Tyrosine binds well up to four ligands where steric effects prevent further ligands from binding as shown in figure 6.46.

Cysteine and serine bind strongly up to a six coordinate octahedral structure. Thus the strongest binding structure may involve up to four tyrosine residues plus enough cysteine or serine residues to make an octahedral binding site and the final ligand being charge neutral.

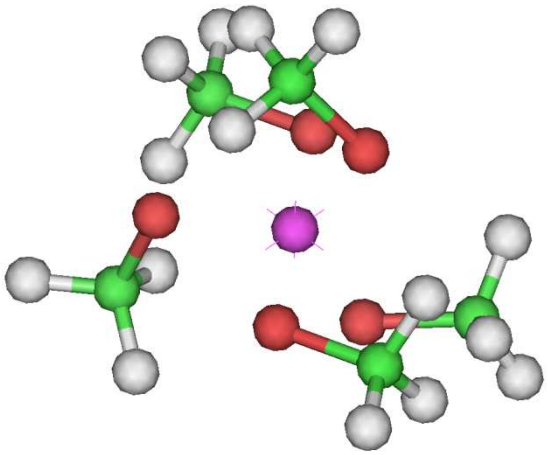


Figure 6.45 – As(V) bound to 5 serine residues (Metal-ligand distance 1.80-1.83Å)

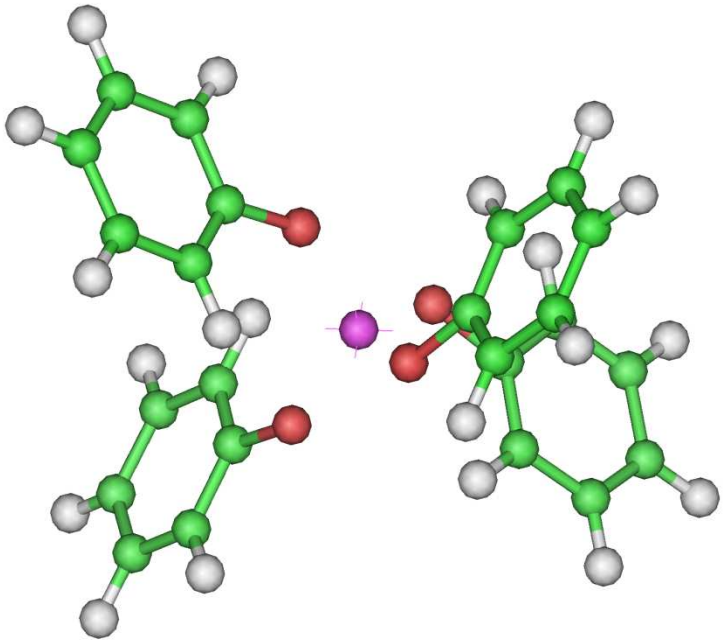


Figure 6.46 – As(V) bound to 4 tyrosine residues (Metal-ligand distance 1.74Å)

6.4.10 Magnesium (II)

The incremental binding energies for both neutral and negatively charged ligands may be plotted to produce the graphs in figure 6.47 and 6.48.

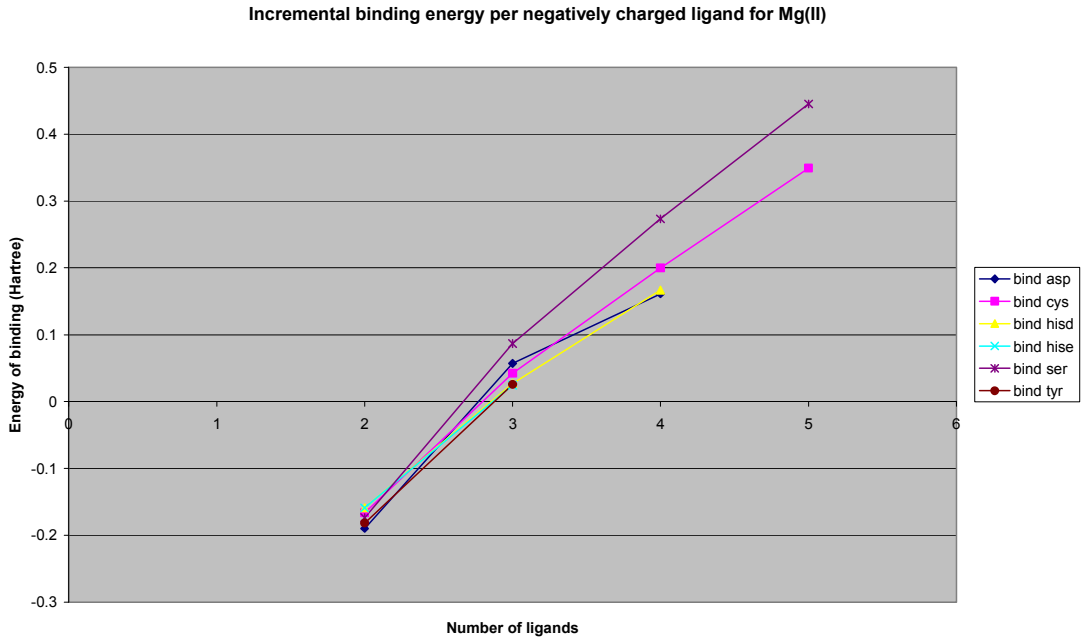


Figure 6.47 – Incremental binding enthalpies of negatively charged ligands for Mg(II)

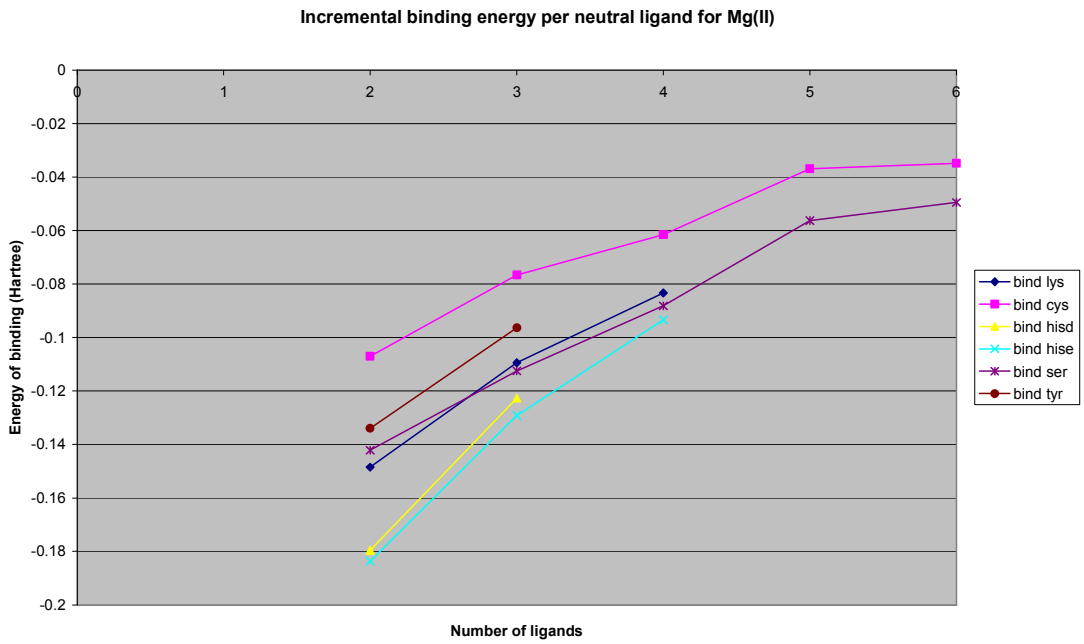


Figure 6.48 – Incremental binding enthalpies of neutral ligands for Mg(II)

For the negatively charged ligands only two ligands were able to bind in a stable fashion, though up to five could be accommodated in structures that were local minima. There is little to choose between the different ligands, but histidine and tyrosine generally bind the most strongly as shown in figure 6.49.

For the neutral ligands up to six ligands were optimised showing a small change in gradient once 5 ligands have been added. Visual inspection confirmed a maximum of four ligands bound directly to the metal with the rest involved in ligand-ligand hydrogen bonding, with the best structure as shown in figure 6.50.

The strongest binding site appears to contain four histidine residues with ser also favoured. Some of these sites may contain partially deprotonated ligands to achieve charge neutrality.

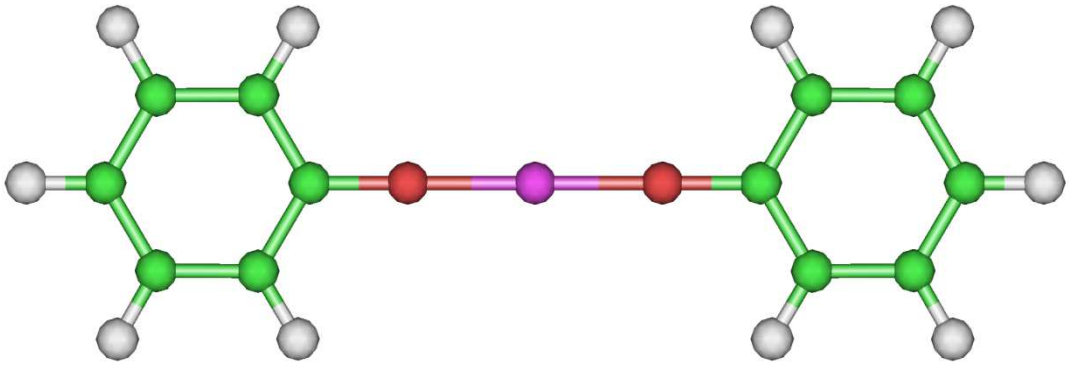


Figure 6.49 – Mg(II) bound to 2 tyrosine residues (Metal-ligand distance 1.76Å)

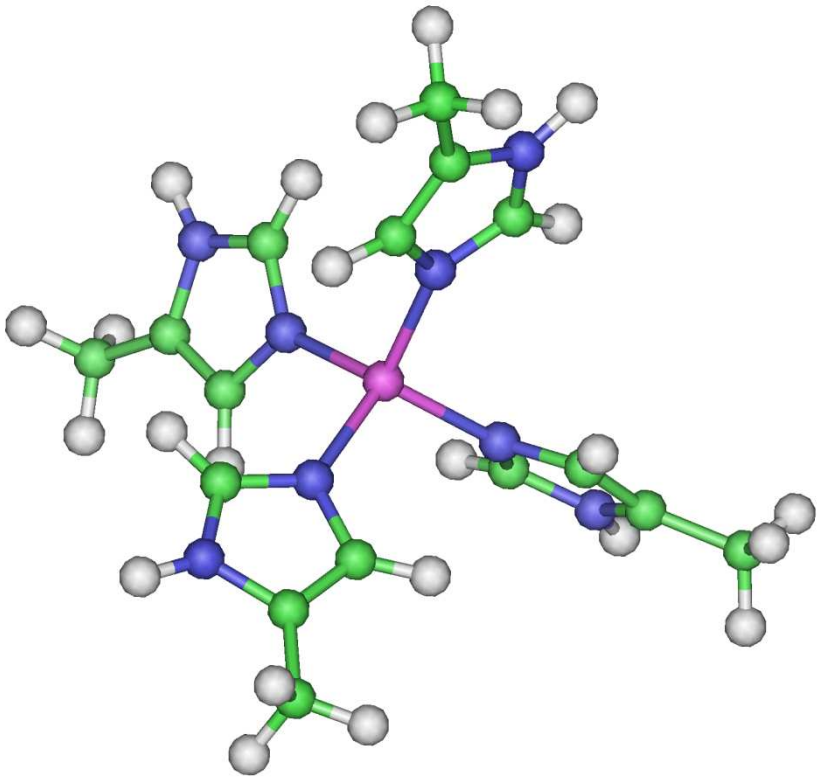


Figure 6.50 – Mg(II) bound to 4 histidine-ε residues (Metal-ligand distance 2.06Å)

6.4.11 Lithium(I)

A graph of the total binding energy is more useful than the incremental binding energy for Li(I) due to some of the calculations failing to converge, as shown in figures 6.51 and 6.52.

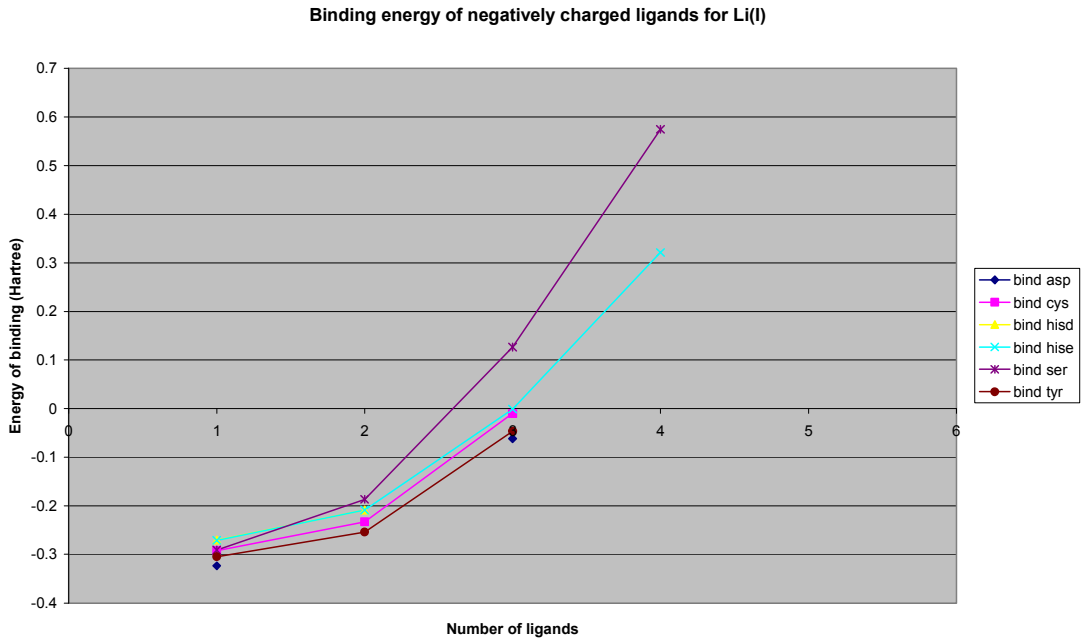


Figure 6.51 – Total binding enthalpies of negatively charged ligands for Li(I)

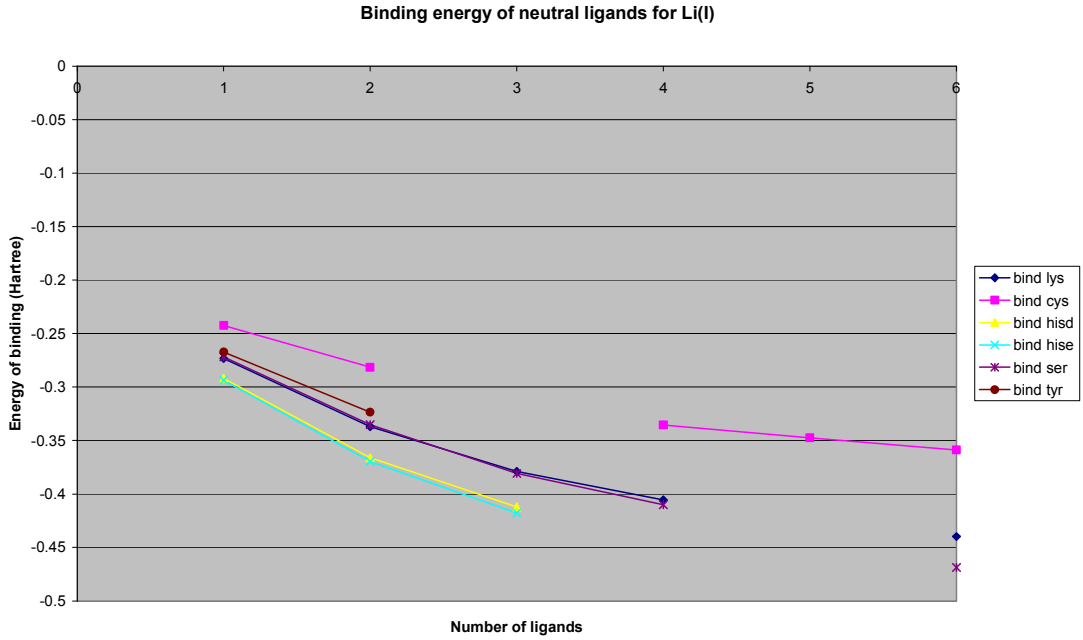


Figure 6.52 – Incremental binding enthalpies of neutral ligands for Li(I)

Only one negatively charged ligand can stably bind Li(I), with aspartate or tyrosine being the preferred ligand as shown in figure 5.53. For the neutral ligands histidine bound most strongly and up to four ligands bound directly to the metal ion with any

additional ligands hydrogen bonding to the inner shell. The most stable minimised site is shown in figure 6.54.

Thus the strongest binding site is likely to contain three histidine residues plus one deprotonated aspartate or tyrosine.

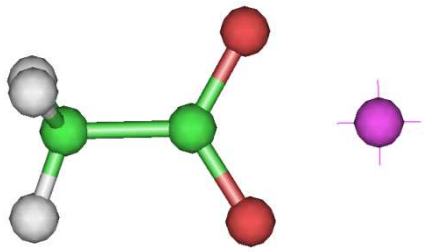


Figure 6.53 – Li(I) bound to 1 aspartate residue (Metal-ligand distance 1.92Å)

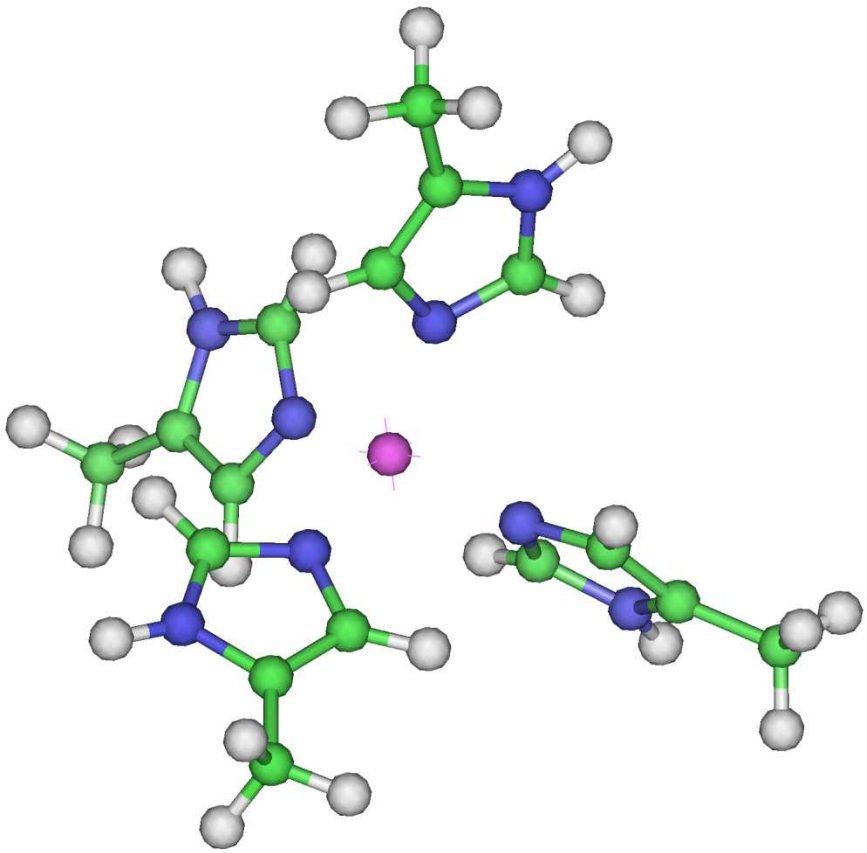


Figure 6.54 – Li(I) bound to 4 histidine-ε residues (Metal-ligand distance 2.06Å)

6.4.12 Calcium(II)

The incremental binding energies may be plotted to produce the graphs shown in figure 6.55 and 6.56.

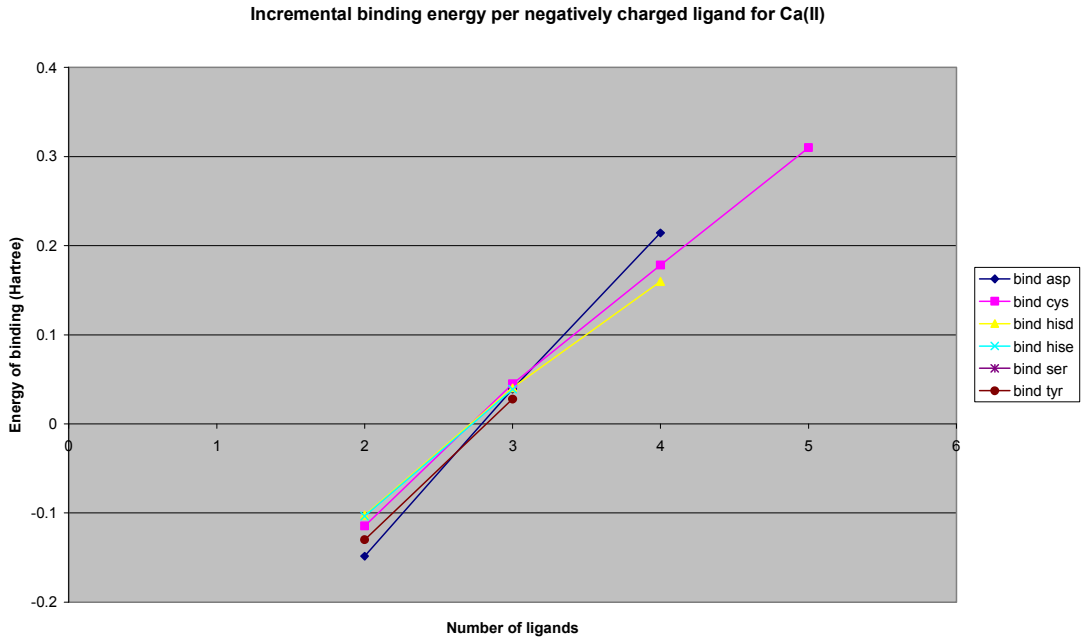


Figure 6.55 – Incremental binding enthalpies of negatively charged ligands for Ca(II)

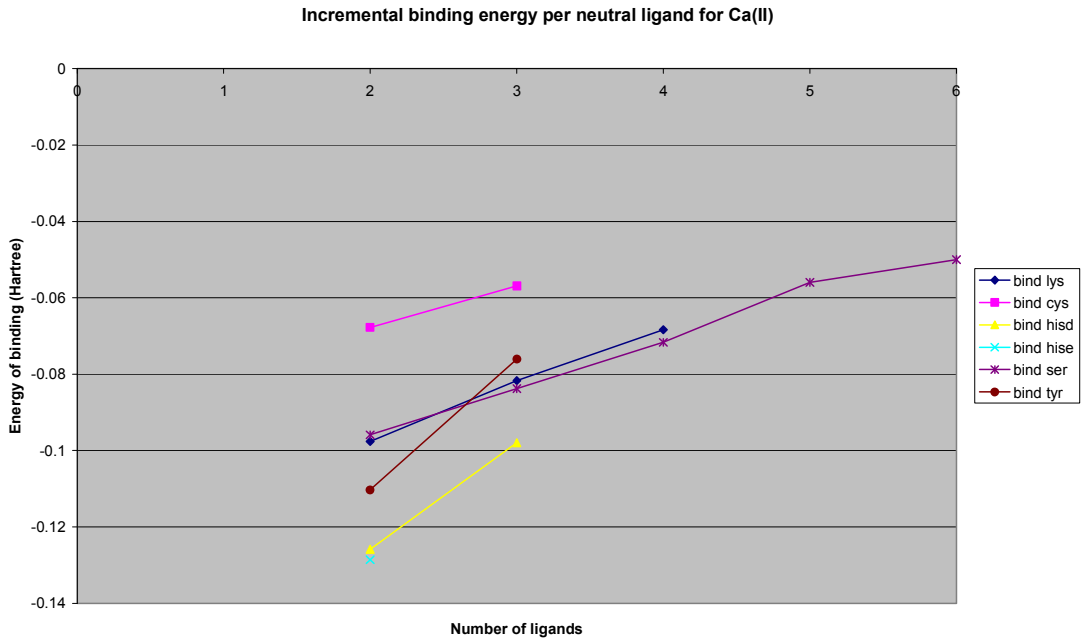


Figure 6.56 – Incremental binding enthalpies of neutral ligands for Ca(II)

For the negatively charged ligands up to five ligands may bind, with tyrosine and histidine binding most strongly though only two are accommodated before ΔH_{inc} becomes positive. The minimised tyrosine structure is shown in figure 5.57.

For the neutral ligands up to six ligands bind in an octahedral geometry, with histidine and serine binding the most strongly as shown in figure 6.58.

As complexes involving more than 6 ligands were not considered, a change ingredient in the plots of ΔH_{inc} is not observed; rather the arrangement of ligands can be identified by visual inspection of the optimised complexes. Numerically, the strongest binding for neutral ligands was observed for 3 histidine residues as shown in figure 6.59.

Thus the strongest binding site should involve two or three histidine ligands plus enough serine residues to complete an octahedral binding site.

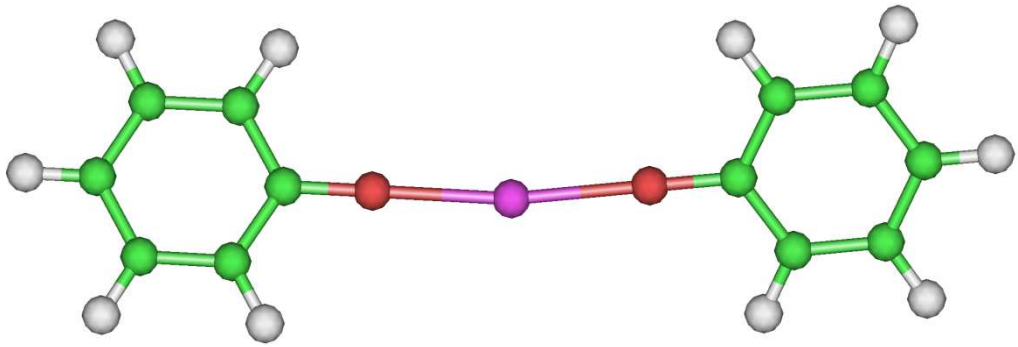


Figure 6.57 – Ca(II) bound to 2 tyrosine residues (Metal-ligand distance 2.11Å)

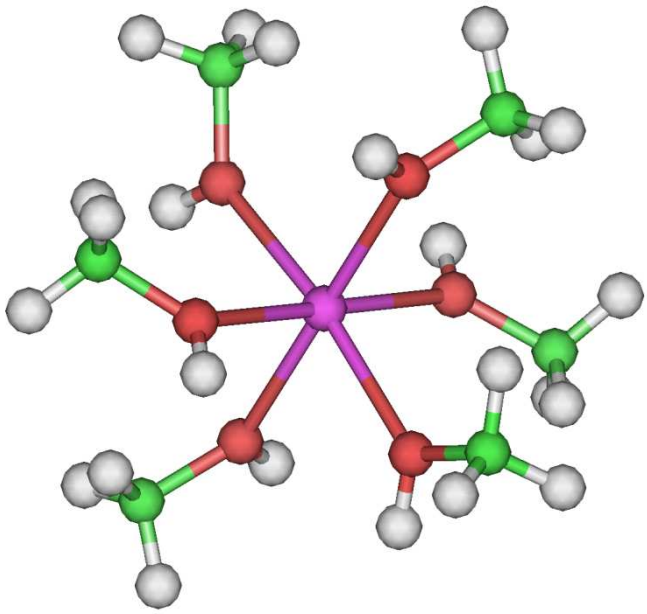


Figure 6.58 – Ca(II) bound to 6 serine residues (Metal-ligand distance 2.41Å)

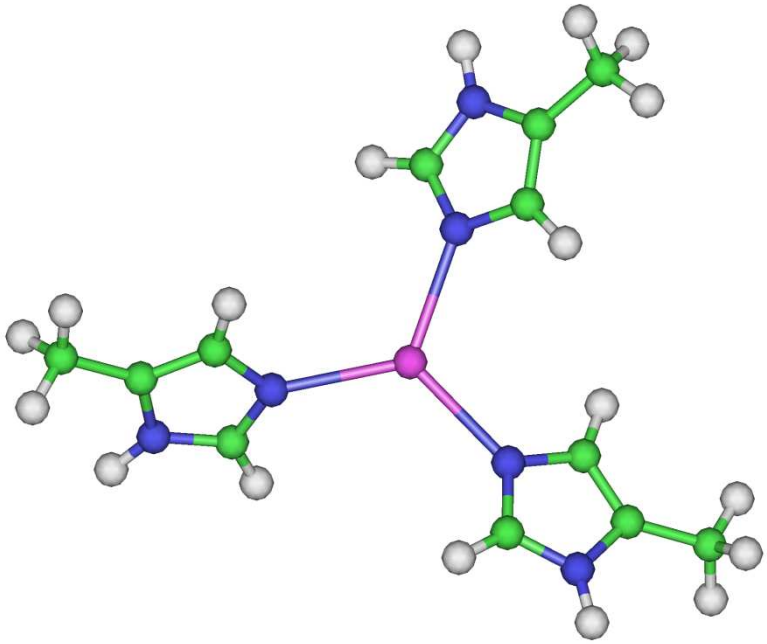


Figure 6.59 – Ca(II) bound to 3 histidine-ε residues (Metal-ligand distance 2.40Å)

6.5 Conclusion

The negatively charged ligands generally appear to show optimum binding for charge neutral sites. The best possible binding site for each metal ion may therefore involve the required number and type of negative ligands shown above with additional charge neutral ligands to complete the shell of ligands.

Cysteine seems to bind well to most of the metal ions as a negatively charged ligand but not as well as a neutral ligand. Our test calculations involving Cu(I) and Zn(II) mimic the X-ray crystal structures of CueR and ZntR.[54] In CueR it has been argued that the neutral ligand has a greater δ- charge due to the hydrogen bonding chain (see Chapter 3). These results show again how well CueR is set up to make such a strong binding site: charge neutrality is maintained whilst the second ligand binds more strongly due to the hydrogen bonding chain. The above binding sites may be improved by including similar environments but that is beyond the scope of this thesis.

The binding sites designed in this chapter may be summarised in table 6.21. The binding sites in the table 6.21 may still not be the strongest possible sites as no account is taken for the size of the ligands/stearic effects. Aspartate may also bind as a bidentate ligand, thus removing the need for some of the additional ligands. If more time were available, calculations would be performed on the proposed sites with mixed ligands.

Metal	Preferred neutral site	Preferred charged site	Overall proposed site
Cu(I)	2 x histidine, 1.89Å	1 x cysteine, 2.19Å	1 x cysteine, 1 x histidine
Hg(II)	2 x lysine, 2.22Å	3 x cysteine, 2.71Å	2 x cysteine, 1 x lysine
Zn(II)	4 x serine, 2.00Å	2 x aspartate	2 x aspartate, 2 x serine
Ag(I)	1 x cysteine	2 x cysteine, 2.54Å or 2 x histidine	1 x cysteine, 1 x histidine
Au(I)	2 x histidine, 2.01Å	1 x cysteine	1 x cysteine, 1 x histidine
Cd(II)	6 x serine, 2.29Å	2 x aspartate, 2.25Å	2 x cysteine, 4 x serine or 2 x aspartate, 2 x serine
Pb(II)	4 x serine, ~2.37Å	2 x aspartate, 2.26Å or 2 x cysteine, 2.57Å	2 x aspartate, 2 x serine
As(III)	3 x lysine, 2.05Å	3 x aspartate, 1.86Å	3 x aspartate
As(V)	None	5 x serine, ~1.81Å	5 x serine
Mg(II)	4 x histidine, 2.06Å	2 x tyrosine, 1.76Å	2 x tyrosine, 2 x histidine
Li(I)	4 x histidine, 2.06Å	1 x aspartate, 1.92Å	3 x histidine, 1 x aspartate
Ca(II)	6 x serine, 2.41Å	2 x tyrosine, 2.11Å	2 x tyrosine, 4 x serine

Table 6.21 – Summary of binding sites. The proposed site is based on the combination of the maximum number of negatively charged ligands with additional neutral ligands until the total maximum number of ligands is reached. For example, Cu(I) will bind to 1 negatively charged cysteine residue, and a maximum of 2 ligands (2 charge neutral histidine residues), thus the proposed site has 2 ligands, one of which is a negatively charged cysteine. The proposed site does not contain charges as biological experiments do not incorporate the charge of the ligands.

Using the RSCB Protein Data Bank and the MESPEUS database the calculated structures and proposed structures may be compared to structures found in the literature.[55,173]

The literature shows some three-coordinate Cu(I) binding proteins involving both cysteine and histidine, such as Human Sco1 which binds through two cysteine and one histidine residues.[174] The yeast metallochaperone Atx1 binds through two cysteine residues and a backbone nitrogen, showing that nitrogen ligands may be used by natural proteins.[175] Other structures have been seen which use two histidine and two cysteine residues, although some of these residues may be forced to bind due to their proximity rather than for energetic reasons.[176] There is also a structure of a trypsin mutant that binds to cysteine through two histidine residues.[177] These structures all show that histidine and cysteine can bind strongly to Cu(I) and

calculations are needed to incorporate both of these residues into one site. Unfortunately such calculations are beyond the scope of this thesis.

MerR binds to Hg(II) through three cysteine residues, which is predicted to be a strong binding site by these calculations.[24] Crystal structures are available for structures with Hg(II) bound to lysine, cysteine and other residues, but none could be found with the predicted structure.[178] Such a structure may still be possible to design, and may bind Hg(II) strongly.

ZntR binds to Zn(II) strongly through 3 cysteine residues, suggesting that the above assumptions may be incorrect.[54] However, there are many other structures (such as alkaline phosphatases) with serine and aspartate bound to Zn(II) showing that such modes of binding are possible.[179-181]

Several Metallothionein structures can be found in the RCSB Protein Data Bank incorporating Ag(I), some of which have structures very similar to those shown above.[55,182] There is also a structure of CueR with Ag(I) bound in place of Cu(I) where Ag(I) is held between two cysteine residues (ref 1QO6) as predicted above.[54]

Crystal structures of proteins which bind Au(I) through two histidine residues are available in the RCSB PDB archives.[55] One such protein is an Oxidoreductase with reference 3ES3.[183] This structure appears to contain one Au(I) ion bound to two histidine residues, and another bound to one cysteine residue. Thus the predicted structures should bind Au(I) correctly.

As seen for some of the other metal ions, Metallothionein structures are available incorporating Cd(II) bound to cysteine residues as predicted above.[184] Structures may also be found for Cd(II) bound to aspartate as predicted.[185,186] However, no structures were found containing six serine residues bound to Cd(II); it may not be possible to place so many residues close enough to the metal ion.

Several crystal structures are available containing Pb(II) bound to aspartate as predicted.[187] However, a search of the literature did not show up any structures with cysteine or serine bound to Pb(II). It may be that the full negative charge of the aspartate residue is required for binding to Pb(II); if so, it may still be possible to bind cysteine or serine to Pb(II) so long as there is a strong enough proton acceptor present.

There are literature examples of arsenic bound to three cysteine residues.[188] The calculations did not complete for this system due to convergence failures. However, cysteine bound more strongly than any other ligand for systems involving at least 4 residues. This suggests that a structure involving three cysteine residues bound to As(III) may be stable. Searching the literature did not show up any other structures bound to As(III).

A search of the literature did not show up any structures involving proteins bound directly to As(V). It may be that neither As(III) nor As(V) exist as free ions and the energy of removing anything bound to them is too great. If they are commonly found as negatively charged arsenate ions, proteins based on negatively charged ligands will not bind them.

There are several structures involving Mg(II) bound to four ligands including histidine and tyrosine.[189,190] The presence of such structures in the literature shows that the proposed mode of binding works.

The majority of structures in the literature which contain Li(I) have it bound to DNA. However, there are structures with Li(I) bound to aspartate and histidine as predicted.[191-194]

Calcium may bind to six ligands as predicted, and there are structures available with both serine and tyrosine bound to a calcium ion.[195,196]

Cu(I), Ag(I) and Au(I) all predicted similar binding sites to one another (1 x cysteine, 1 x histidine) which explains why CueR binds to all three ions.[42,43] This predicted site incorporates a negatively charged cysteine residue and a δ-negatively charged histidine residue. Such a site appears to be remarkably similar to that found for CueR with the hydrogen bonding chain proposed in Chapter 3.

To continue the work in this chapter, the binding sites suggested above should be superimposed upon the known MerR family structures, and then mutants of these proteins could be designed and synthesised.

7. FINAL CONCLUSIONS AND FURTHER WORK

This thesis has studied CueR, ZntR and mutants of CueR. Chapter 1 looked at the role of MerR family regulators in biological systems. It was shown how important removal of toxic metals from biological systems is. Several toxic metals were considered, including copper, zinc, mercury, lead and cadmium, and modes of removal were discussed. There are several members of the MerR family which are designed to regulate amounts of various toxic metals. MerR family proteins have an extremely high sensitivity and selectivity for their respective metal ions. However, not all metals have a MerR family regulator.

The method of information transfer has been carefully studied for MerR; however, in CueR there is no direct bond between the Cu(I) ion and the DNA binding domain. Thus the mechanism by which information about the metal binding event is transferred to the DNA binding domain was one of the aims of this thesis. An X-ray crystal structure is available for CueR (and ZntR) but not for MerR. As such, it may be possible to learn more about the way that these proteins work from this structure. However, the only structures available are for activator complexes and the repressor form is not currently known. There is currently little chemical literature available on MerR family members, especially their mode of binding and information transfer.

The exact conformation of the DNA is extremely important for transcriptional activation. In the absence of the MerR family protein there is little transcription, and if the protein is present but the metal ion is not then a repressor form of the DNA is produced and no transcription may occur.

Two different computational methods were used to probe the activity of CueR: *ab initio* Quantum Mechanics (QM) and Molecular Mechanics (MM, Molecular Dynamics). As discussed in Chapter 2, these techniques show very different amounts of information. The QM calculations show a great deal of information but are computationally demanding and can only be used on small systems; as such they were used to study the binding site. The MM calculations use pre-defined parameters to make them less computationally demanding, but do not allow certain transformations such as electron/hydrogen transfer; MM calculations were used to study the motions of the whole protein.

The binding event was studied using QM calculations as described in Chapter 3. An important hydrogen bonding chain was identified between residues in the metal binding and DNA binding domains. One hydrogen bond in this chain is of particular interest because it is between residues at opposite ends of the inter-domain helices (The C112-S: S77-O bond). Thus any change in the length of this hydrogen bond could easily be transferred to a large scale change in the dihedral angle of the protein and thus the DNA bend. However, it is not possible to prove that such a hydrogen bond exists by studying the X-ray crystal structure as such structures do not contain any hydrogen atoms.

Various models were built of the binding site to varying degrees of success. The models which appear to be most similar to the crystal structure and EXAFS data involve the full hydrogen bonding chain. These results provide evidence that this chain is present and important to the correct function of the protein.

The sensitivity and selectivity of CueR was probed using these models. The binding site seems to be perfectly set up to accept Cu(I). It appears that other metal ions will not bind in the binding site for energetic reasons. However, even if they were able to bind they would

distort the binding site to such an extent that the protein would not produce the correct dihedral angle for transcription. The results of the calculations suggest that the resting state involves two protonated thiols hydrogen bonding to one another.

The length and strength of the hydrogen bond was studied through a variety of models and by considering the proton affinity. The calculations show that the length of the C112-S: S77-O hydrogen bond (and thus the twist of the whole protein through the inter-domain helices) is highly dependent on the presence of Cu(I). Such a change would not occur without the presence of such a hydrogen bond.

ZntR was compared to CueR in Chapter 4. It was shown that the binding site is well set up to accept Zn(II) ions (as was seen for Cu(I) in CueR). Various other metal ions were considered and Zn(II) was shown to bind the most strongly. Zn(II) binds directly to the DNA binding domain through cysteine 79 (analogous to S77 in CueR) and so the mechanism of information transfer is obvious. However, the mechanisms of both CueR and ZntR are extremely similar assuming that the information is passed through the proposed hydrogen bonding chain.

The global effect of binding Cu(I) to CueR was studied using MD in Chapter 5. Various models were used to study the effect of removal of the hydrogen bond, the removal of copper and the addition of DNA. Due to problems with the force field the full effect of the hydrogen bond could not be studied. However, when Cu(I) was not bound various different structures were observed for the overall protein, any of which could account for the change in transcription. Adding DNA to the calculations caused the protein to be less mobile and showed that the change in dihedral angle of the protein does directly affect the bend of the DNA. Many of the calculations also showed an increase in C112-S: S77-O distance

leading to a change in dihedral angle which would thus affect DNA transcription. Models of mutants were built that should affect the proposed hydrogen bonding chain and the results of the calculations were compared to biological data. The proline mutant that should have removed the hydrogen bonding chain caused a change in the structure of the protein, showing that it may not be the disruption of the hydrogen bonding chain that stops this mutant from working. The aspartate mutant that was predicted to recover the activity of the protein showed similar results to the wild type. Finally a resting state was predicted for CueR when Cu(I) is not present.

Uses of MerR and related proteins were discussed in Chapter 6, along with designing alternative binding sites. Since these proteins have such high specificity and selectivity they have already been employed as ultra high sensitivity metal detectors. However, it is possible that they could be used as metal ion sequesters to clean up contaminated land or water.

Since CueR and ZntR were shown to have such good binding sites in Chapters 3 and 4, calculations were performed to see whether binding sites could be designed for other metal ions. A variety of different metal ions and potential binding sites were evaluated, and several binding sites were designed. However, hydrogen bonding leading to polarisation of residues was not considered, and as such sites similar to CueR's (with a large delta negative charge on the thiol sulfur of cysteine) could not be designed. As such the binding affinity was not quite as good as was seen for CueR, showing how important hydrogen bonding may be.

Experiments to be considered in the future should include performing more MD simulations using modified force fields. As discussed in Chapter 5, modified force fields

could be designed with highly polarised cysteine residues to mimic the effect of the hydrogen bonding chain, thus holding the Cu(I) in place even when one of the cysteine residues is a thiol. An alternative modified force field would include the Cu(I) bound directly to the thiol sulfur of cys112.

To bring together both the QM and MM parts of these experiments, QM/MM calculations could be used, modelling the binding site and metal ion as a QM component and the rest of the protein as MM. Using QM/MM calculations would avoid problems with the MM force field not modelling the interaction between Cu(I) and the cysteine thiol correctly.

Further calculations should be performed on the structures proposed in Chapter 6 to confirm that the predicted structures would bind the metal ions well. Finally these sites should be overlaid on known MerR family structures, such as CueR and ZntR. If it is possible to mutate residues in the binding site to those suggested in Chapter 6, mutants should be grown to see whether such proteins could be used.

REFERENCES

(1) Fosmire, G. J. *Am. J. Clin. Nutr.* **1990**, *51*, 225-7.

(2) Franzblau, A.; Lilis, R. *Arch. Env. Health* **1989**, *44*, 385-390.

(3) Sloane, J. In *Mercury: Element of the Ancients*; Dartmouth Toxic Metals Research Program: Dartmouth, U.S.A., 2001.

(4) *Mercury In Your School and the Community: A National Issue*; Brachman, S.; Jones, B.; Skavroneck, S.; Stenstrup, A.; Thiry, M., Eds.; University of Wisconsin: Wisconsin, 2002.

(5) Morel, F. M. M.; Kraepiel, A. M. L.; Amyot, M. *Annu. Rev. Ecol. Syst.* **1998**, *29*, 543-566.

(6) Fleishman, B. In *Copper: An Ancient Metal*; Dartmouth Toxic Metals Research Program: Dartmouth, U.S.A., 2001.

(7) Solioz, M.; Stoyanov, J. V. *FEMS Microbiol. Rev.* **2003**, *27*, 183-195.

(8) Berg, J. M. *Annu. Rev. Biophys. Biophys. Chem.* **1990**, *19*, 405-421.

(9) Berg, J. M.; Tymoczko, J. L.; Stryer, L. *Biochemistry, fifth edition*; W, H. Freeman and Company: New York, 2002.

(10) Laity, J. H.; Lee, B. M.; Wright, P. E. *Curr. Opin. Struct. Biol.* **2001**, *11*, 39-46.

(11) Hambridge, K. M.; Casey, C. E.; Krebs, N. F. *Trace elements in human and animal nutrition*; 5th ed.; Academic Press: Orlando, 1986; Vol. 2.

(12) Prasad, A. S.; Brewer, G. J.; Schoomaker, E. B.; Rabbani, P. *J. Am. Med. Assoc.* **1978**, *240*, 2166-2168.

(13) Porter, K. G.; McMaster, D.; Elmes, M. E.; Love, A. H. *Lancet* **1977**, *2*, 774.

(14) Sohn, E. In *The facts on lead*; Dartmouth Toxic Metals Research Program: Dartmouth, U.S.A., 2001.

(15) Byers, R. K.; Lord, E. E. *Am. J. Dis. Child.* **1943**, *66*, 471-494.

(16) Dameron, C. T.; Harrison, M. D. *Am. J. Clin. Nutr.* **1998**, *67*, 1091s-1097s.

(17) Klausner, R. D.; Dancis, A. *FEBS Lett.* **1994**, *355*, 109-113.

(18) Germann, U. A.; Lerch, K. *Biochem. J.* **1987**, *245*, 479-484.

(19) Dameron, C. T.; Dance, I. G. *Biomimetic Materials Chemistry*; VCH Publishers, Inc.: New York, 1996.

(20) Schiering, N.; Kabash, W.; Moore, M. J.; Distefano, M. D.; Walsh, C. T.; Pai, E. F. *Nature* **1991**, *352*, 168-172.

(21) Kille, P.; Hemmings, A.; Lunney, E. A. *Biochim. Biophys. Acta.* **1994**, *1205*, 151-161.

(22) Hamer, D. H. *Annu. Rev. Biochem.* **1986**, *55*, 913-951.

(23) Dameron, C. T.; Winge, D. R.; George, G. N.; Sansone, M.; Hu, S.; Hamer, D. *Proc. Natl. Acad. Sci. U. S. A.* **1991**, *88*, 6127-6131.

(24) Brown, N. L.; Stoyanov, J. V.; Kidd, S. P.; Hobman, J. L. *FEMS Microbiol. Rev.* **2003**, *27*, 145-163.

(25) Christie, G. E.; White, T. J.; Goodwin, T. S. *Gene* **1994**, *146*, 131-132.

(26) Lund, P. A.; Ford, S. J.; Brown, N. L. *J. Gen. Microbiol.* **1986**, *132*, 465-480.

(27) Lund, P. A.; Brown, N. L. *J. Mol. Biol.* **1989**, *205*, 343-353.

(28) Brown, N. L.; Pridmore, R. D.; Fritzinger, D. C. *Biochem. Soc. Trans.* **1984**, *12*.

(29) Brown, N. L.; Misra, T. K.; Winnie, J. N.; Schmidt, A.; Seiff, M.; Silver, S. *Mol. Gen. Genet.* **1986**, *202*, 143-151.

(30) Barrineau, P.; Gilbert, P.; Jackson, W. J.; Jones, C. S.; Summers, A. O.; Wisdom, S. *Basic Life Sci.* **1985**, *30*, 707-718.

(31) O'Halloran, T. V.; Walsh, C. T. *J. Cell. Biochem.* **1986**, 104-104.

(32) Heltzel, A.; Gambill, D.; Jackson, W. J.; Totis, P. A.; Summers, A. O. *J. Bacteriol.* **1987**, *169*, 3379-3384.

(33) O'Halloran, T. V.; Frantz, B.; Shin, M. K.; Ralston, D. M.; Wright, J. G. *Cell* **1989**, *56*, 119-129.

(34) Lund, P.; Brown, N. L. *Nucleic Acids Res.* **1989**, *17*, 5517-5527.

(35) Parkhill, J.; Brown, N. L. *Nucleic Acids Res.* **1990**, *18*, 5157-5162.

(36) Hobman, J. L.; Wilkie, J.; Brown, N. L. *BioMetals* **2005**, *18*, 429-436.

(37) Newberry, K. J.; Brennan, R. G. *J. Biol. Chem.* **2004**, *279*, 20356-20362.

(38) Ralston, D. M.; O'Halloran, T. V. *Proc. Natl. Acad. Sci. U. S. A.* **1990**, *87*, 3846-3850.

(39) Condee, C. W.; Summers, A. O. *J. Bacteriol.* **1992**, *174*, 8094-8101.

(40) Adaikkalam, V.; Swarup, S. *Microbiology* **2002**, *48*, 2857-2867.

(41) Outten, F. W.; Outten, C. E.; Hale, J.; O'Halloran, T. V. *J. Biol. Chem.* **2000**, *275*, 31024-31029.

(42) Stoyanov, J. V.; Hobman, J. L.; Brown, N. L. *Mol. Microbiol.* **2001**, *39*, 502-511.

(43) Stoyanov, J. V.; Brown, N. L. *J. Biol. Chem.* **2003**, *278*, 1407-1410.

(44) Petersen, C.; Moller, L. B. *Gene* **2000**, *261*, 289-298.

(45) Brocklehurst, K. R.; Hobman, J. L.; Lawley, B.; Blank, L.; Marshall, S. J.; Brown, N. L.; Morby, A. P. *Mol. Microbiol.* **1999**, *31*, 893-902.

(46) Outten, C. E.; Outten, F. W.; O'Halloran, T. V. *J. Biol. Chem.* **1999**, *274*, 37517-37524.

(47) Khan, S.; Brocklehurst, K. R.; Jones, G. W.; Morby, A. P. *Biochem. Biophys. Res. Commun.* **2002**, *299*, 438-455.

(48) Borremans, B.; Hobman, J. L.; Provoost, A.; Brown, N. L.; Van der Lelie, D. *J. Bacteriol.* **2001**, *183*, 5651-5658.

(49) Ahmed, M.; Lyass, L.; Markham, P. N.; Taylor, S. S.; Vazquezlaslop, N.; Neyfakh, A. A. *J. Bacteriol.* **1995**, *177*, 3904-3910.

(50) Ahmed, M.; Borsch, C. M.; Taylor, S. S.; Vazquezlaslop, N.; Neyfakh, A. A. *J. Biol. Chem.* **1994**, *269*, 28506-28513.

(51) Heldwein, E. E. Z.; Brennan, R. G. *Nature* **2001**, *409*, 378-382.

(52) Champier, L.; Duarte, V.; Michaud-Soret, I.; Coves, J. *Mol. Microbiol.* **2004**, *52*, 1475-1485.

(53) Grass, G.; Rensing, C. *Biochem. Biophys. Res. Commun.* **2001**, *286*, 902-908.

(54) Changela, A.; Chen, K.; Xue, Y.; Holschen, J.; Outten, C. E.; O'Halloran, T. V.; Mondragon, A. *Science* **2003**, *301*, 1383-1387.

(55) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235-242.

(56) Gaballa, A.; Cao, M.; Helmann, J. D. *Microbiology-Sgm* **2003**, *149*, 3413-3421.

(57) Schewchuk, L. M.; Verdine, G. L.; Nash, H.; Walsh, C. T. *Biochemistry* **1989**, *28*, 6140-6145.

(58) Schewchuk, L. M.; Verdine, G. L.; Walsh, C. T. *Biochemistry* **1989**, *28*, 2331-2339.

(59) Helmann, J. D.; Ballard, B. T.; Walsh, C. T. *Science* **1990**, *247*, 946-948.

(60) Chen, K.; Yuldasheva, S.; Penner-Hahn, J. E.; O'Halloran, T. V. *J. Am. Chem. Soc.* **2003**, *125*, 12088-12089.

(61) Hitomi, Y.; Outten, C. E.; O'Halloran, T. V. *J. Am. Chem. Soc.* **2001**, *123*, 8614-8615.

(62) Schrödinger, E. *Phys. Rev.* **1926**, *28*, 1049-1070.

(63) Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory, Revised Edition*; Dover Publications Inc.: Mineola, N.Y., 1989.

(64) Born, M.; Oppenheimer, R. *Ann. Phys.* **1927**, *84*, 457-484.

(65) Eckart, C. *Phys. Rev.* **1935**, *47*, 552-558.

(66) Sutcliffe, B. T. *Fundamentals of Computational Quantum Chemistry*; D. Reidel Publishing Co.: Dordrecht-Holland, 1975.

(67) Roos, B. O.; Taylor, P. R.; Siegbahm, E. M. *Chem. Phys.* **1980**, *48*, 157-173.

(68) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618-622.

(69) Head-Gordon, M.; Pople, J. A.; Frisch, M. J. *Chem. Phys. Lett.* **1988**, *153*, 503.

(70) Frisch, M. J.; Head-Gordon, M.; Pople, J. A. *Chem. Phys. Lett.* **1990**, *166*, 275.

(71) Frisch, M. J.; Head-Gordon, M.; Pople, J. A. *Chem. Phys. Lett.* **1990**, *166*, 281.

(72) Head-Gordon, M.; Head-Gordon, T. *Chem. Phys. Lett.* **1994**, *220*, 122.

(73) Saebo, S.; Almlof, J. *Chem. Phys. Lett.* **1989**, *154*.

(74) Hohenburg, P.; Kohn, W. *Phys. Rev. B* **1964**, *136*, 864-871.

(75) Leach, A. R. *Molecular Modelling: Principles and Applications, Second Edition*; Pearson Education Limited: Harlow, Essex, 2001.

(76) Kohn, W.; Sham, L. J. *Phys. Rev. A* **1965**, *140*, 1133-1138.

(77) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648-5652.

(78) M. J. Frisch; G. W. Trucks; H. B. Schlegel; G. E. Scuseria; M. A. Robb; J. R. Cheeseman; V. G. Zakrzewski; J. A. Montgomery, J.; R. E. Stratmann; J. C. Burant; S. Dapprich; J. M. Millam; A. D. Daniels; K. N. Kudin; M. C. Strain; O. Farkas; J. Tomasi; V. Barone; M. Cossi; R. Cammi;

B. Mennucci; C. Pomelli; C. Adamo; S. Clifford; J. Ochterski; G. A. Petersson; P. Y. Ayala; Q. Cui; K. Morokuma; D. K. Malick; A. D. Rabuck; K. Raghavachari; J. B. Foresman; J. Cioslowski; J. V. Ortiz; A. G. Baboul; B. B. Stefanov; G. Liu; A. Liashenko; P. Piskorz; I. Komaromi; R. Gomperts; R. L. Martin; D. J. Fox; T. Keith; M. A. Al-Laham; C. Y. Peng; A. Nanayakkara; C. Gonzalez; M. Challacombe; P. M. W. Gill; B. G. Johnson; W. Chen; M. W. Wong; J. L. Andres; M. Head-Gordon; Replogle, E. S.; J. A. Pople; Gaussian, Inc.: Pittsburgh PA, 1998.

(79) Ditchfield, R.; Hehre, W. J.; Pople, J. A. *J. Chem. Phys.* **1971**, *54*, 724.

(80) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257.

(81) Hariharan, P. C.; Pople, J. A. *Mol. Phys.* **1974**, *27*, 209.

(82) Gordon, M. S. *Chem. Phys. Lett.* **1980**, *76*, 163.

(83) Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213.

(84) Blaudeau, J.-P.; McGrath, M. P.; Curtiss, L. A.; Radom, L. *J. Chem. Phys.* **1997**, *107*, 5016.

(85) Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; DeFrees, D. J.; Pople, J. A.; Gordon, M. S. *J. Chem. Phys.* **1982**, *77*, 3654.

(86) Binning, R. C. J.; Curtiss, L. A. *J. Comput. Chem.* **1990**, *11*, 1206.

(87) Rassolov, V. A.; Pople, J. A.; Ratner, M. A.; Windus, T. L. *J. Chem. Phys.* **1998**, *109*, 1223.

(88) Dunning, T. H. J.; Hay, P. J. *Modern Theoretical Chemistry*; Plenum: New York, 1976; Vol. 3.

(89) Hay, P. J.; Wadt, W. R. *J. Chem. Phys.* **1985**, *82*, 270.

(90) Wadt, W. R.; Hay, P. J. *J. Chem. Phys.* **1985**, *82*, 284.

(91) Hay, P. J.; Wadt, W. R. *J. Chem. Phys.* **1985**, *82*, 299.

(92) Hehre, W. J.; Stewart, R. F.; Pople, J. A. *J. Chem. Phys.* **1969**, *51*, 2657.

(93) Huzinga, S. *J. Chem. Phys.* **1965**, *42*, 1293-1302.

(94) Dunning, T. H. *J. Chem. Phys.* **1970**, *53*, 2823-2883.

(95) Fuentealba, P.; Preuss, H.; Stoll, H.; Szentpaly, L. v. *Chem. Phys. Lett.* **1989**, *89*, 418.

(96) Szentpaly, L. v.; Fuentealba, P.; Preuss, H.; Stoll, H. *Chem. Phys. Lett.* **1982**, *93*, 555.

(97) Fuentealba, P.; Stoll, H.; Szentpaly, L. v.; Schwerdtfeger, P.; Preuss, H. *J. Phys. B.* **1983**, *16*, 1323.

(98) Stoll, H.; Fuentealba, P.; Schwerdtfeger, P.; Flad, J.; Szentpaly, L. v.; Preuss, H. *J. Chem. Phys.* **1984**, *81*, 2732.

(99) Fuentealba, P.; Szentpaly, L. v.; Preuss, H.; Stoll, H. *J. Phys. B.* **1985**, *18*, 1287.

(100) *Quantum Chemistry: The Challenge of Transition Metals and Coordination Chemistry*; Wedig, U.; Dolg, M.; Stoll, H.; Preuss, H., Eds.; Reidel, and Dordrecht, 1986.

(101) Dolg, M.; Wedig, U.; Stoll, H.; Preuss, H. *J. Chem. Phys.* **1987**, *86*, 866.

(102) Igel-Mann, G.; Stoll, H.; Preuss, H. *Mol. Phys.* **1988**, *65*, 1321.

(103) Dolg, M.; Stoll, H.; Preuss, H. *J. Chem. Phys.* **1989**, *90*, 1930.

(104) Schwerdtfeger, P.; Dolg, M.; Schwarz, W. H. E.; Bowmaker, G. A.; Boyd, P. D. W. *J. Chem. Phys.* **1989**, *91*, 1762.

(105) Dolg, M.; Stoll, H.; Savin, A.; Preuss, H. *Theo. Chim. Acta* **1989**, *75*, 173.

(106) Andrae, D.; Haeussermann, U.; Dolg, M.; Stoll, H.; Preuss, H. *Theo. Chim. Acta* **1990**, *77*, 123.

(107) Kaupp, M.; Schleyer, P. v. R.; Stoll, H.; Preuss, H. *J. Chem. Phys.* **1991**, *94*, 1360.

(108) Kuechle, W.; Dolg, M.; Stoll, H.; Preuss, H. *Mol. Phys.* **1991**, *74*, 1245.

(109) Dolg, M.; Fulde, P.; Kuechle, W.; Neumann, C.-S.; Stoll, H. *J. Chem. Phys.* **1991**, *94*, 3011.

(110) Dolg, M.; Stoll, H.; Flad, H.-J.; Preuss, H. *J. Chem. Phys.* **1992**, *97*, 1162.

(111) Bergner, A.; Dolg, M.; Kuechle, W.; Stoll, H.; Preuss, H. *Mol. Phys.* **1993**, *80*, 1431.

(112) Dolg, M.; Stoll, H.; Preuss, H. *Theo. Chim. Acta* **1993**, *85*, 441.

(113) Dolg, M.; Stoll, H.; Preuss, H.; Pitzer, R. M. *J. Phys. Chem.* **1993**, *97*, 5852.

(114) Haeussermann, U.; Dolg, M.; Stoll, H.; Preuss, H. *Mol. Phys.* **1993**, *78*, 1211.

(115) Kuechle, W.; Dolg, M.; Stoll, H.; Preuss, H. *J. Chem. Phys.* **1994**, *100*, 7535.

(116) Nicklass, A.; Dolg, M.; Stoll, H.; Preuss, H. *J. Chem. Phys.* **1995**, *102*, 8942.

(117) Leininger, T.; Nicklass, A.; Stoll, H.; Dolg, M.; Schwerdtfeger, P. *J. Chem. Phys.* **1996**, *105*, 1052.

(118) Collins, J. B.; Schleyer, P. v. R.; Binkley, J. S.; Pople, J. A. *J. Chem. Phys.* **1976**, *64*, 5142.

(119) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785-789.

(120) Pople, J. A.; Head-Gordon, M.; Fox, D. J.; Raghavachari, K.; Curtiss, L. A. *J. Chem. Phys.* **1989**, *90*, 5622.

(121) Curtiss, L. A.; Jones, C.; Trucks, G. W.; Raghavachari, K.; Pople, J. A. *J. Chem. Phys.* **1990**, *93*, 2537.

(122) Miehlich, B.; Savin, A.; Stoll, H.; Preuss, H. *Chem. Phys. Lett.* **1989**, *157*, 200.

(123) Adamo, C.; Barone, V. *J. Chem. Phys.* **1998**, *108*, 664.

(124) Adamo, C.; Barone, V. *Chem. Phys. Lett.* **1997**, *274*, 242.

(125) *Electronic Density Functional Theory: Recent Progress and New Directions*; Burke, K.; Perdew, J. P.; Wang, Y., Eds.; Plenum, 1998.

(126) Perdew, J. P.; Wang, Y. *Phys. Rev. B* **1992**, *45*, 13244.

(127) Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiolhais, C. *Phys. Rev. B* **1992**, *46*, 6671 - 6687.

(128) Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiolhais, C. *Phys. Rev. B* **1993**, *48*, 4978.

(129) Perdew, J. P.; Burke, K.; Wang, Y. *Phys. Rev. B* **1996**, *54*, 16533.

(130) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. *J. Comput. Chem.* **2005**, *26*.

(131) Ponder, J. W.; Case, D. A. *Adv. Protein Chem.* **2003**, *66*, 27-85.

(132) Cheatham, T. E.; Young, M. A. *Biopolymers* **2001**, *56*, 232-256.

(133) Lennard-Jones, J. E. *Proc. Phys. Soc.* **1931**, *43*, 461-482.

(134) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; R., H. J. *J. Chem. Phys.* **1984**, *81*, 3684-3690.

(135) Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Wang, B.; Pearlman, D. A.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J. W.; Ross, W. S.; Kollman, P. A.; University of California: San Francisco, 2004.

(136) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327-341.

(137) Miyamoto, S.; Kollman, P. A. *J. Comp. Chem* **1992**, *13*, 952-962.

(138) Wulfsberg, G. *Principles of Descriptive Chemistry*; Brooks/Cole Publishing: Monterey, CA, 1987.

(139) Kraut, J. *Annu. Rev. Biochem.* **1977**, *46*, 331-358.

(140) Dodsona, G.; Wlodawer, A. *Trends Biochem. Sci.* **1998**, *23*, 347-352.

(141) Goettig, P.; Groll, M.; Kim, J.-S.; Huber, R.; Brandstetter, H. *EMBO J.* **2002**, *21*, 5343-5352.

(142) Zeng, Q.; Stalhanske, C.; Anderson, M. C.; Scott, R. A.; Summer, A. O. *Biochemistry* **1998**, *37*, 15885-15895.

(143) Kamerlin, S. C. L., The University of Birmingham, 2005.

(144) Shriver, D. F.; Atkins, P. W. *Inorganic Chemistry*; Oxford University Press: Oxford, 1999.

(145) In *InsightII*; Molecular Simulations, Inc.: San Diego, CA, 1997.

(146) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T. *J. Comput. Chem.* **2003**, *24*, 1999-2012.

(147) Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049-1074.

(148) Cieplak, P.; Caldwell, J. W.; Kollman, P. A. *J. Comp. Chem.* **2001**, *22*, 1048-1057.

(149) Dixon, R. W.; Kollman, P. A. *J. Comp. Chem.* **1997**, *18*, 1632-1646.

(150) Meng, E.; Cieplak, P.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1994**, *116*, 12061-12062.

(151) Shannon, R. D. *Acta Crystallogr., Sect. A* **1976**, *32*, 751-767.

(152) Hobman, J. L.; The University of Birmingham: Birmingham, 2005, p 1.

(153) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33-38.

(154) Corbisier, P.; van der Lelie, D.; Borremans, B.; Provoost, A.; de Lorenzo, V.; Brown, N. L.; Lloyd, J. R.; Hobman, J. L.; Csoregi, E.; Johansson, G.; Mattiasson, B. *Anal. Chim. Acta* **1999**, *387*, 235-244.

(155) Binet, M. R. B.; Poole, R. K. *FEBS Lett.* **2000**, *473*, 67-70.

(156) Lee, S. W.; Glickmann, E.; Cooksey, D. A. *Appl. Environ. Microbiol.* **2001**, *67*, 1437-1444.

(157) Bontidean, L.; Mortari, A.; Leth, S.; Brown, N. L.; Karlson, U.; Larsen, M. M.; Vangronsveld, J.; Corbisier, P.; Csöregi, E. *Environ. Poll.* **2004**, *131*, 255-262.

(158) Chen, P.; He, C. *J. Am. Chem. Soc.* **2004**, *126*, 728-729.

(159) Prendergast, F.; Mann, K. *Biochemistry* **1978**, *17*, 3448-53.

(160) Watras, C. J.; Back, R. C.; Halvorsen, S.; Hudson, R. J. M.; Morrison, K. A.; Wente, S. P. *Sci. Total Environ.* **1998**, *219*, 183-208.

(161) Boudou, A.; Delarche, A.; Ribeyre, F.; Marty, R. *Bull. Environ. Contam. Toxicol.* **1979**, *22*, 813-8.

(162) Black, H. *Environ. Health Perspect.* **1995**, *103*, 1106-1108.

(163) Zenk, M. H. *Gene* **1996**, *179*, 21-30.

(164) Salt, D. E.; Blaylock, M.; Kumar, N. P. B. A.; Dushenkov, V.; Ensley, B. D.; Chet, I.; Raskin, I. *Bio/Technology* **1995**, *13*, 468-474.

(165) McArthur, J. M.; Ravenscroft, P.; Safiullah, S.; Thirlwall, M. F. *Water Resour. Res.* **2001**, *15*, 109-117.

(166) Nickson, R. T.; McArthur, J. M.; Burgess, W. G.; Ahmed, K. M.; Ravenscroft, P.; Rahman, M. *Nature* **1998**, *395*, 338.

(167) Nickson, R. T.; McArthur, J. M.; Ravenscroft, P.; Burgess, W. G.; Ahmed, K. M. *Appl. Geochem.* **2000**, *15*, 403-413.

(168) Harvey, C. F.; Swartz, C. H.; Badruzzaman, A. B. M.; Keon-Blute, N.; Yu, W.; Ali, M. A.; Jay, J.; Beckie, R.; Niedan, V.; Brabander, D.; Oates, P. M.; Ashfaque, K. N.; Islam, S.; Hemond, H. F.; Ahmed, M. F. *Science* **2002**, *298*, 1602-1606.

(169) Blow, D. M. *Acc. Chem. Res.* **1976**, *9*, 145.

(170) Cox, H.; Akibo-Betts, G.; Wright, R. R.; Walker, N. R.; Curtis, S.; Duncombe, B.; Stace, A. J. *J. Am. Chem. Soc.* **2003**, *125*, 233-242.

(171) Pukar, L.; Cox, H.; Goren, A.; Aitken, G. D. C.; Stace, A. J. *Faraday Discuss.* **2003**, *124*, 259-273.

(172) Walker, N. R.; Wright, R. R.; Barran, P. E.; Cox, H.; Stace, A. J. *J. Chem. Phys.* **2001**, *114*, 5562.

(173) Hsin, K.; Sheng, Y.; Harding, M. M.; Taylor, P.; Walkinshaw, M. D. *J. Appl. Crystallogr.* **2008**, *41*, 963-968.

(174) Banci, L.; Bertini, I.; Ciofi-Baffoni, S.; Leontari, I.; Martinelli, M.; Palumaa, P.; Sillard, R.; Wang, S. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 15-20.

(175) Arnesano, F.; Banci, L.; Bertini, I.; Huffman, D. L.; O'Halloran, T. V. *Biochemistry* **2001**, *40*, 1528-1539.

(176) Botuyan, M. V.; Toy-Palmer, A.; Chung, J.; Blake 2nd., R. C.; Beroza, P.; Case, D. A.; Dyson, H. J. *J. Mol. Biol.* **1996**, *263*, 752-767.

(177) McGrath, M. E.; Haymore, B. L.; Summers, N. L.; Craik, C. S.; Fletterick, R. J. *Biochemistry* **1993**, *32*, 1914-1919.

(178) Krishna, T. S.; Kong, X. P.; Gary, S.; Burgers, P. M.; Kuriyan, J. *Cell* **1994**, *79*, 1233-1243.

(179) Dealwis, C. G.; Brennan, C.; Christianson, K.; Mandecki, W.; Abad-Zapatero, C. *Biochemistry* **1995**, *34*, 13967-13973.

(180) Murphy, J. E., Tibbitts, T.T., Kantrowitz, E.R. *J. Mol. Biol.* **1995**, *253*, 604-617.

(181) Wang, E., Koutsioulis, D., Leiros, H.K.S., Andersen, O.A., Bouriotis, V., Hough, E., Heikinheimo, P. *J. Mol. Biol.* **2007**, *366*, 1318-1331.

(182) Peterson, C. W.; Narula, S. S.; Armitage, I. M. *FEBS Lett.* **1996**, *379*, 85-93.

(183) Butts, C. A.; Swift, J.; Kang, S. G.; Di Costanzo, L.; Christianson, D. W.; Saven, J. G.; Dmochowski, I. J. 2008.

(184) Braun, W.; Vasak, M.; Robbins, A. H.; Stout, C. D.; Wagner, G.; Kagi, J. H.; Wuthrich, K. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 10124-10128.

(185) Kazmirski, S. L.; Isaacson, R. L.; An, C.; Buckle, A.; Johnson, C. M.; Daggett, V.; Fersht, A. R. *Nat. Struct. Biol.* **2002**, *9*, 112-116.

(186) Kaus-Drobek, M.; Czapinska, H.; Sokolowska, M.; Tamulaitis, G.; Szczepanowski, R. H.; Urbanke, C.; Siksnys, V.; Bochtler, M. *Nucleic Acids Res.* **2007**, *35*, 2035-2046.

(187) Kursula, P.; Majava, V. *Acta Crystallogr., Sect. A:* **2007**, *63*, 653-656.

(188) Touw, D. S.; Nordman, C. E.; Stuckey, J. A.; Pecoraro, V. L. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 11969-11974.

(189) Kishida, H.; Wada, T.; Unzai, S.; Kuzuyama, T.; Takagi, M.; Terada, T.; Shirouzu, M.; Yokoyama, S.; Tame, J. R.; Park, S. Y. *Acta Crystallogr., Sect. D* **2003**, *59*, 23-31.

(190) Teplyakov, A.; Sebastiao, P.; Obmolova, G.; Perrakis, A.; Brush, G. S.; Bessman, M. J.; Wilson, K. S. *EMBO J.* **1996**, *15*, 3487-3497.

(191) Ren, H.; Wang, L.; Bennett, M.; Liang, Y.; Zheng, X.; Lu, F.; Li, L.; Nan, J.; Luo, M.; Eriksson, S.; Zhang, C.; Su, X. D. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 303-308.

(192) Hohenester, E.; Keller, J. W.; Jansonius, J. N. *Biochemistry* **1994**, *33*, 13561-13570.

(193) Zobel, K.; Wang, L.; Varfolomeev, E.; Franklin, M. C.; Elliott, L. O.; Wallweber, H. J.; Okawa, D. C.; Flygare, J. A.; Vucic, D.; Fairbrother, W. J.; Deshayes, K. *Acs Chem.Biol.* **2006**, *1*, 525-533.

(194) Vucic, D.; Franklin, M. C.; Wallweber, H. J. A.; Das, K.; Eckelman, B. P.; Shin, H.; Elliott, L. O.; Kadkhodayan, S.; Deshayes, K.; Salvesen, G. S.; Fairbrother, W. J. *Biochem. J.* **2005**, *385*, 11-20.

(195) Enroth, C.; Eger, B. T.; Okamoto, K.; Nishino, T.; Nishino, T.; Pai, E. F. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 10723-10728.

(196) Fukunari, A.; Okamoto, K.; Nishino, T.; Eger, B. T.; Pai, E. F.; Kamezawa, M.; Yamada, I.; Kato, N. *J. Pharmacol. Exp. Ther.* **2004**, *311*, 519-528.

APPENDIX I – ADDITIONAL NUMERICAL DATA

(CHAPTERS 3 AND 6)

I a) Chapter 3 raw data

The following data is the numerical data, plotted in Chapter 3.5 to produce figures 3.15, 3.17, 3.20 and 3.22.

S-H dist(Å)	Energy (Hartree)
1.3	-2765.395555
1.4	-2765.397304
1.5	-2765.393438
1.6	-2765.389021
1.7	-2765.386430
1.8	-2765.384874
1.9	-2765.383484
2.0	-2765.381902
2.1	-2765.380029
2.2	-2765.378034
2.3	-2765.375929
2.4	-2765.380740
2.5	-2765.379148

Table I.1 - S-H plot 1 results (from section 3.5.2)

S-H dist(Å)	Energy (Hartree)
1.25	-687.0575534
1.3	-687.0619778
1.35	-687.0633914
1.4	-687.0625104
1.45	-687.0599074
1.5	-687.0560663
1.6	-687.0462433
1.7	-687.0363803
1.8	-687.0294382
1.9	-687.024414
2.0	-687.0202833
2.1	-687.0163732
2.2	-687.0126081
2.3	-687.0089836
2.4	-687.0055517
2.5	-687.0024001

Table I.2 - S-H plot 2 results (from section 3.5.3)

S-H dist. (Å)	Energy (Hartree)
1.2	-2992.634576
1.3	-2992.651695
1.5	-2992.660421
1.6	-2992.670892
1.65	-2992.673355
1.7	-2992.67551
1.75	-2992.677335
1.8	-2992.678813
1.9	-2992.68088
2.07	-2992.682025
2.3	-2992.680698
2.4	-2992.679582
2.5	-2992.678304

Table I.3 - S-H plot 3 results (from section 3.5.4)

S-H dist. (Å)	Energy (Hartree)
1.2	-914.2999437
1.25	-914.308348
1.3	-914.312904
1.35	-914.3146264
1.4	-914.3145237
1.45	-914.3130406
1.5	-914.3108093
1.55	-914.308171
1.6	-914.3064704
1.65	-914.3050759
1.7	-914.3041557
1.75	-914.3035263
1.8	-914.3086856
1.9	-914.3087001
2	-914.3080486
2.1	-914.3068872
2.3	-914.3042763
2.4	-914.302376
2.5	-914.299874

Table I.4 - S-H plot 4 results (from section 3.5.5)

I b) Chapter 6

The following data had been analysed in Chapter 6 by a variety of methods to produce most of the figures in section 6.4.

#	aspartate	cysteine	histidine δ	histidine ε	serine	tyrosine
1	-0.3708823	-0.382808	-0.3454727	-0.3451984	-0.342404	-0.3518057
2	-0.3329595	-0.343513	-0.3176648	-0.3166629	-0.2652041	-0.3221618
3	-0.1085673	-0.1039766		-0.0876951	0.0635194	-0.1013606
4		0.2606232				

Table I.5 - Total binding enthalpies for negatively charged ligands for Cu(I)

#	cysteine	histidine δ	histidine ε	lysine	serine	tyrosine
1	-0.3029133	-0.3531135	-0.3540072	-0.3329948	-0.3099013	-0.3046107
2	-0.3692199	-0.4641153	-0.4664058	-0.4318255	-0.3917854	-0.3779057
3	-0.3996612	-0.4962349		-0.4602226	-0.4231047	-0.4077566
4	-0.4231826	-0.5137068		-0.4811304	-0.4458834	
5	-0.4382497				-0.4889373	

Table I.6 - Total binding enthalpies for charge neutral ligands for Cu(I)

#	aspartate	cysteine	histidine δ	histidine ε	serine	tyrosine
1						
2		-142.7593025	-572.3221177	-572.323327	-272.834002	-656.1936601
3	-728.0162539	-192.8697574	-837.2567978		-387.9674202	-963.0627269
4					-502.9466284	

Table I.7 - Enthalpies of formation for negatively charged ligands for Hg(II)

#	cysteine	histidine δ	histidine ε	lysine	serine	tyrosine
1	-92.5341637			-137.7744165	-157.58452	
2	-143.252412	-573.0370597	-573.0394237	-233.7467596	-273.3935852	
3	-193.9126927	-838.5464939		-329.6423589	-389.150114	
4	-244.559158			-425.5210643	-504.8925863	
5	-295.1872631				-620.6182677	
6	-345.8119658				-736.3362352	

Table I.8 - Enthalpies of formation for charge neutral ligands for Hg(II)

#	aspartate	cysteine	histidine δ	histidine ε	serine	tyrosine
1	-293.5897145	-115.1970995			-180.2432782	
2	-522.3508853	-165.5544006	-595.1495907	-595.1510324	-295.6720879	-679.0324817
3	-750.8708746	-215.6868983		-860.122163	-410.8380736	-985.934206
4	-979.2470273	-265.6585821			-525.8189187	
5		-315.6718311			-640.8468557	

Table I.9 - Enthalpies of formation of negatively charged ligands for Zn(II)

#	cysteine	histidine δ	histidine ε	lysine	serine	tyrosine
1	-115.3096553			-160.5676276	-180.3904971	
2	-166.0348687	-595.8601554	-595.8626138	-256.5570101	-296.2237739	
3	-216.7117741	-861.4007052		-352.4812295	-412.0079049	
4	-267.369924			-448.3765913	-527.7661062	
5	-317.9980639				-643.4921186	
6	-368.6254945			-640.0673582	-759.2117516	

Table I.10 - Enthalpies of formation of charge neutral ligands for Zn(II)

#	aspartate	cysteine	histidine δ	histidine ε	serine	tyrosine
1	-374.1924601	-195.8185531	-410.5980372	-410.5983225	-260.8445505	-452.5307769
2		-245.9517853	-675.5595352	-675.5594867	-376.0073508	-759.4154121
3	-831.0502561	-295.8969987	-940.3377188	-940.3370581	-490.9476776	-1066.133775
4		-345.7293054	-1205.028877	-1205.022788		-1372.750986

Table I.11 - Enthalpies of formation of negatively charged ligands for Ag(I)

#	cysteine	histidine δ	histidine ε	lysine	serine	tyrosine
1	-196.1612072	-411.0260322	-411.0263908	-241.4018142	-261.2393837	-452.9091739
2	-246.8134104	-676.5393286	-676.5401065	-337.2972114	-376.9734561	-760.311333
3	-297.4369444	-941.9880713		-433.1376899	-492.6773538	-1067.687413
4	-348.0556742			-528.9728618	-608.3746921	-1375.050716
5					-724.1007187	
6	-449.2742358				-839.7794663	

Table I.12 - Enthalpies of formation of charge neutral ligands for Ag(I)

#	aspartate	cysteine	histidine δ	histidine ε	serine	tyrosine
1		-185.4986126	-400.2769218	-400.2771952		-442.1976983
2	-592.3907011	-235.6494066		-665.2662706	-365.7077302	-749.1034813
3		-285.5727942			-480.6093464	

Table I.13 - Enthalpies of formation of negatively charged ligands for Au(I)

#	cysteine	histidine δ	histidine ε	lysine	serine	tyrosine
1	-185.8121331	-400.6863529	-400.6862023	-231.0577877	-250.8727285	
2	-236.4910518	-666.2369104	-666.2373975	-326.9904661	-366.6358151	-749.9709247
3	-287.105183			-422.8378677	-482.3544992	
4	-337.7167992			-518.6638856	-598.0188737	
5	-388.3309495				-713.75584	
6	-438.9422294				-829.4406126	

Table I.14 - Enthalpies of formation of charge neutral ligands for Au(I)

#	aspartate	cysteine	histidine δ	histidine ε	serine	tyrosine
1	-276.1158679	-97.7335326			-162.7698312	
2	-504.8605082	-148.0758722	-577.6514715	-577.6529092	-278.1703658	-661.5327584
3	-733.3789032	-198.2015517	-842.606481	-842.6101973	-393.3245525	-968.4226222
4	-961.7433094	-248.176698	-1107.438533		-508.310482	
5					-623.1442516	

Table I.15 - Enthalpies of formation of negatively charged ligands for Cd(II)

#	cysteine	histidine δ	histidine ε	lysine	serine	tyrosine
1	-97.8619395	-312.768953	-312.7684121	-143.1095927	-162.9351684	
2	-148.5743367	-578.3714875	-578.374013	-239.0769604	-278.7446391	
3	-199.2432657			-334.985754	-394.5132588	-969.5439522
4	-249.8962376	-1109.392675		-430.8726684	-510.2628875	
5	-300.5268114				-625.9908627	
6	-351.1550668			-622.5741401	-741.7119132	

Table I.16 - Enthalpies of formation of charge neutral ligands for Cd(II)

#	aspartate	cysteine	histidine δ	histidine ε	serine	tyrosine
1	-231.6496586	-53.2512777	-268.0335847	-268.036704		
2	-460.355564	-103.5679769		-533.1531518	-233.6941699	
3		-153.6839342			-348.8421862	-923.8910979
4		-203.6389495			-463.7958546	

Table I.17 - Enthalpies of formation of negatively charged ligands for Pb(II)

#	cysteine	histidine δ	histidine ε	lysine	serine	tyrosine
1	-53.3775829	-268.2969735	-268.2979861	-98.6320968	-118.4728466	-310.1830035
2	-104.06661		-533.8657926	-194.5680937	-234.2565935	-617.6275583
3	-154.7346066	-799.3738642		-290.4758769	-350.0177986	-925.0452099
4	-205.3730111			-386.3230955	-465.7459575	
5					-581.4788971	
6	-306.6300575			-578.0324942	-697.1963064	

Table I.18 - Enthalpies of formation of charge neutral ligands for Pb(II)

#	aspartate	cysteine	histidine δ	histidine ε	serine	tyrosine
1	-233.6006753		-270.1724722	-270.1735816	-120.3142224	-312.0676697
2	-462.6219976	-105.8797569	-535.4604791	-535.4733553	-236.0203334	-619.3716396
3	-691.3335274				-351.4047503	-926.4162058
4	-919.8234196	-206.2819449			-466.5222646	
5		-256.2365305			-581.4696183	
6		-306.0818135			-696.5785093	

Table I.19 - Enthalpies of formation of negatively charged ligands for As(III)

#	cysteine	histidine δ	histidine ε	lysine	serine	tyrosine
1		-270.0927785	-270.0988953	-100.2893849		
2		-535.8312189	-535.8446892	-196.4010421		
3		-801.4700856	-801.4828817	-292.4394589	-351.9131263	
4				-388.357805		
5	-258.01189				-583.4495736	
6	-308.6621522			-580.1054021		

Table I.20 - Enthalpies of formation of charge neutral ligands for As(III)

#	aspartate	cysteine	histidine δ	histidine ε	serine	tyrosine
1	-231.7990984		-268.6042539	-268.6279359		
2				-534.4029607		-618.3618367
3	-690.4920993	-155.4192888			-350.5436958	-925.721658
4	-919.3841537	-205.9297438			-466.1827548	-1232.86135
5		-256.1759066			-581.5402455	
6		-306.2802168			-696.6954445	

Table I.21 - Enthalpies of formation of negatively charged ligands for As(V)

#	aspartate	cysteine	histidine δ	histidine ε	serine	tyrosine
1	-229.0067253	-50.5810872	-265.367941	-265.3687121	-115.658273	-307.3379388
2	-457.7707706	-100.9308938	-530.5347346	-530.5363949	-231.091301	-614.4546407
3	-686.2880068	-151.071858	-795.5156827	-795.5205484	-346.2638325	-921.3636526
4	-914.7010957	-201.0551581	-1060.35715		-461.2497545	
5		-250.8888689			-576.063707	

Table I.22 - Enthalpies of formation of negatively charged ligands for Mg(II)

#	cysteine	histidine δ	histidine ε	lysine	serine	tyrosine
1	-50.7282957	-265.6524215	-265.6515272	-95.9900771	-115.833627	-307.5425996
2	-101.4339732	-531.2561779	-531.2587242	-191.955066	-231.650637	-615.0268317
3	-152.1092607	-796.8029666	-796.8115093	-287.8809716	-347.4379665	-922.473491
4	-202.7694226		-1062.328423	-383.7808088	-463.2010068	
5	-253.4049897				-578.932148	
6	-304.0385242				-694.6565107	

Table I.23 - Enthalpies of formation of charge neutral ligands for Mg(II)

#	aspartate	cysteine	histidine δ	histidine ε	serine	tyrosine
1	-235.9770875	-57.5546963	-272.3579123	-272.35886	-122.6298443	-314.3186271
2	-464.4898688	-107.6788365	-537.3043679	-537.3054032	-237.7847905	-621.2028552
3	-692.8641542	-157.6387802		-802.1063879	-352.7308126	-927.930005
4				-1066.792825	-467.5422071	

Table I.23 - Enthalpies of formation of negatively charged ligands for Li(I)

#	cysteine	histidine δ	histidine ε	lysine	serine	tyrosine
1	-57.9204173	-272.7953272	-272.7965637	-103.1692464	-123.0259428	-314.6969742
2	-108.5583019	-538.2939597	-538.2961305	-199.0493499	-238.7643791	-622.1034228
3		-803.7641198	-803.767905	-294.9077862	-354.4849277	
4	-209.8097542			-390.7509175	-470.1890699	
5	-260.420443					
6	-311.0304016			-582.4181244	-701.5973828	

Table I.24 - Enthalpies of formation of charge neutral ligands for Li(I)

#	aspartate	cysteine	histidine δ	histidine ε	serine	tyrosine
1	-264.7894217	-86.3518071	-301.1506969	-301.1500812	-151.435042	-343.1150571
2	-493.5122645	-136.6496621	-566.2611111	-566.2622484		-650.1799164
3	-722.0476068	-186.7879134	-831.2288133	-831.2321079	-381.9766579	-957.0868103
4	-950.4076914	-236.793027	-1096.077062			
5		-286.6660582			-611.8225746	

Table I.25 - Enthalpies of formation of negatively charged ligands for Ca(II)

#	cysteine	histidine δ	histidine ε	lysine	serine	tyrosine
1	-86.5680052	-301.4691628	-301.4696438	-131.8193085	-151.6738636	-343.3510679
2	-137.2344757	-567.0191771	-567.0218192	-227.7334105	-267.4446509	-650.8116982
3	-187.8901083	-832.5413025		-323.6316069	-383.2033057	-958.2380702
4			-1098.052258	-419.5164331	-498.949858	
5					-614.6807094	
6	-339.8100173			-611.2383127	-730.4056214	

Table I.26 - Enthalpies of formation of charge neutral ligands for Ca(II)

APPENDIX II –ADDITIONAL GRAPHS (CHAPTER 5)

II a) Dihedral angle graphs

The following graphs show how dihedral angle varied *vs.* time for each of the experiments described in sections 5.4 and 5.6. Graphs are shown below, one per section.

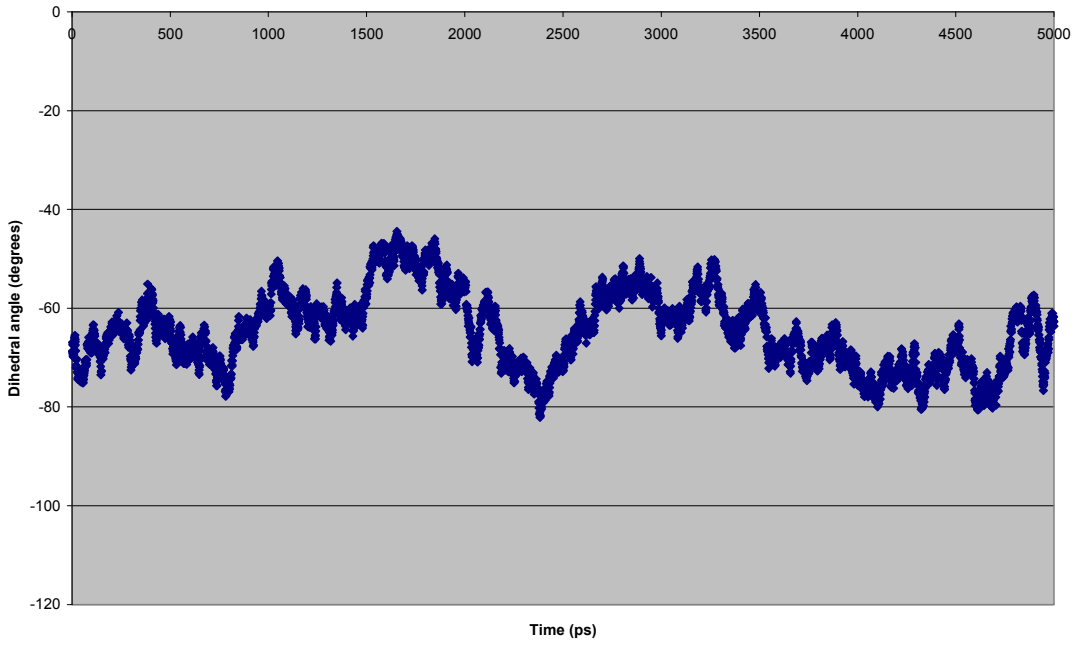


Figure II.1 – CueRH dihedral angle *vs.* time (5 ns)

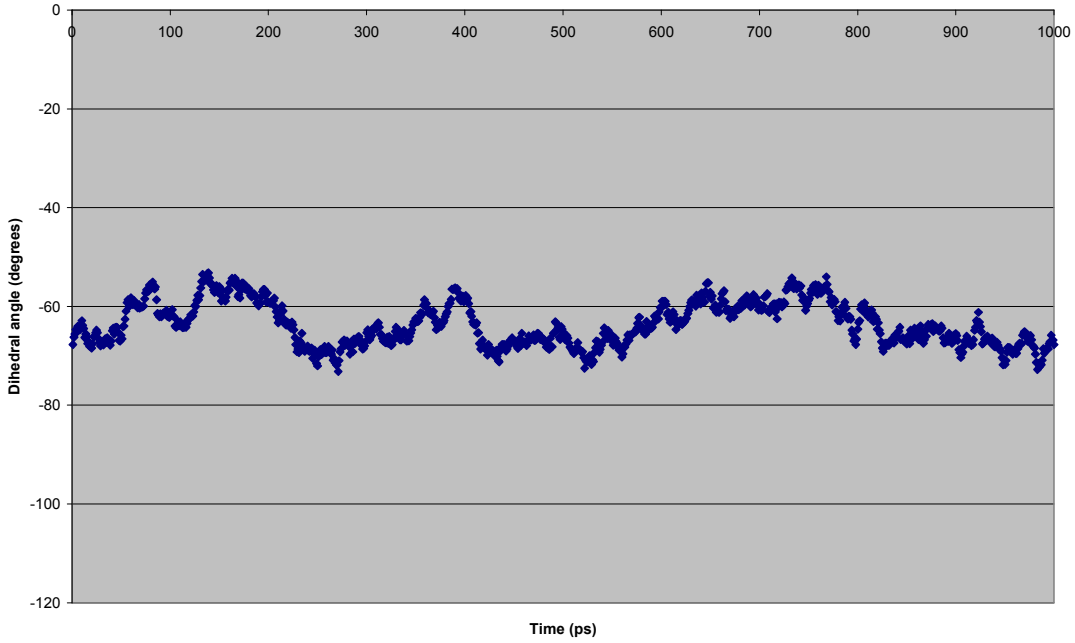


Figure II.2 – CueRHep dihedral angle *vs.* time

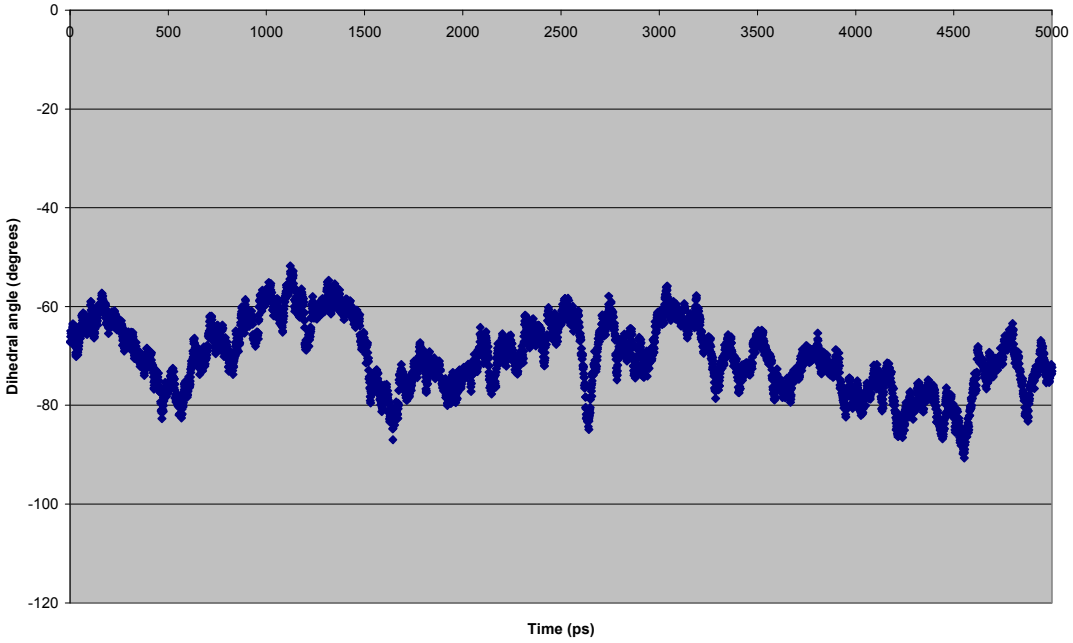


Figure II.3 – CueR- dihedral angle *vs.* time (5 ns)

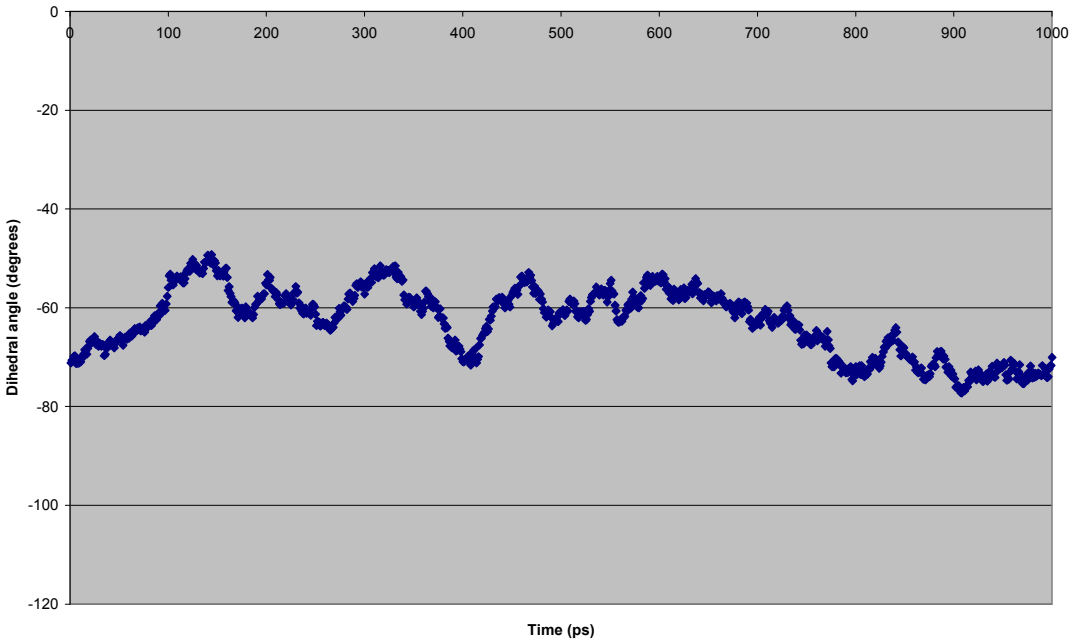


Figure II.4 – CueR-ep dihedral angle *vs.* time

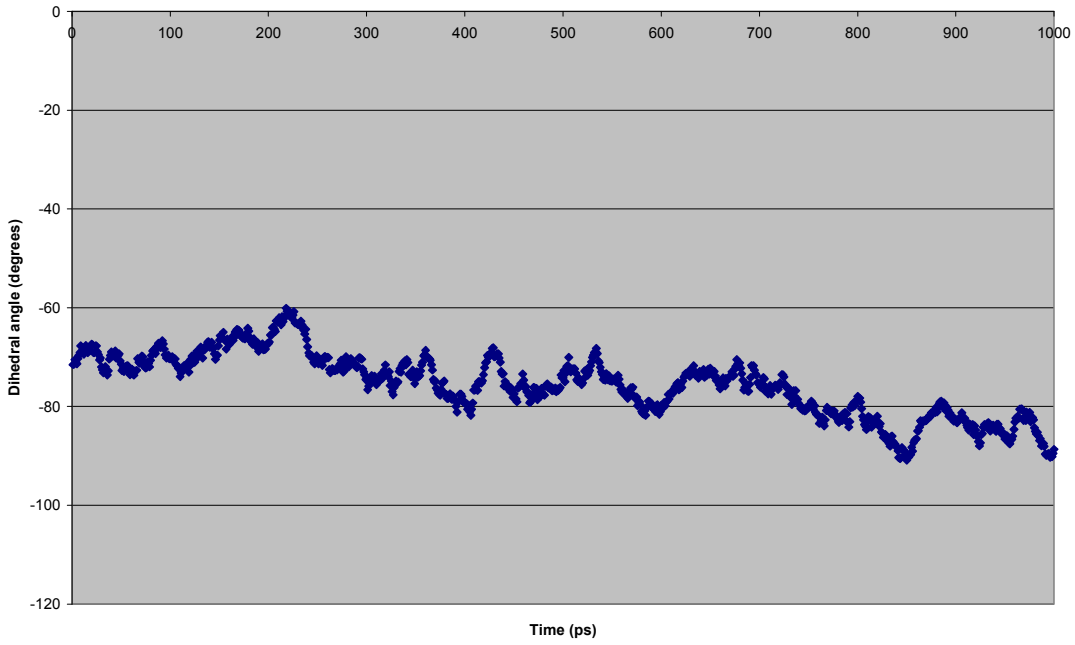


Figure II.5 – CueRnc dihedral angle *vs*. time

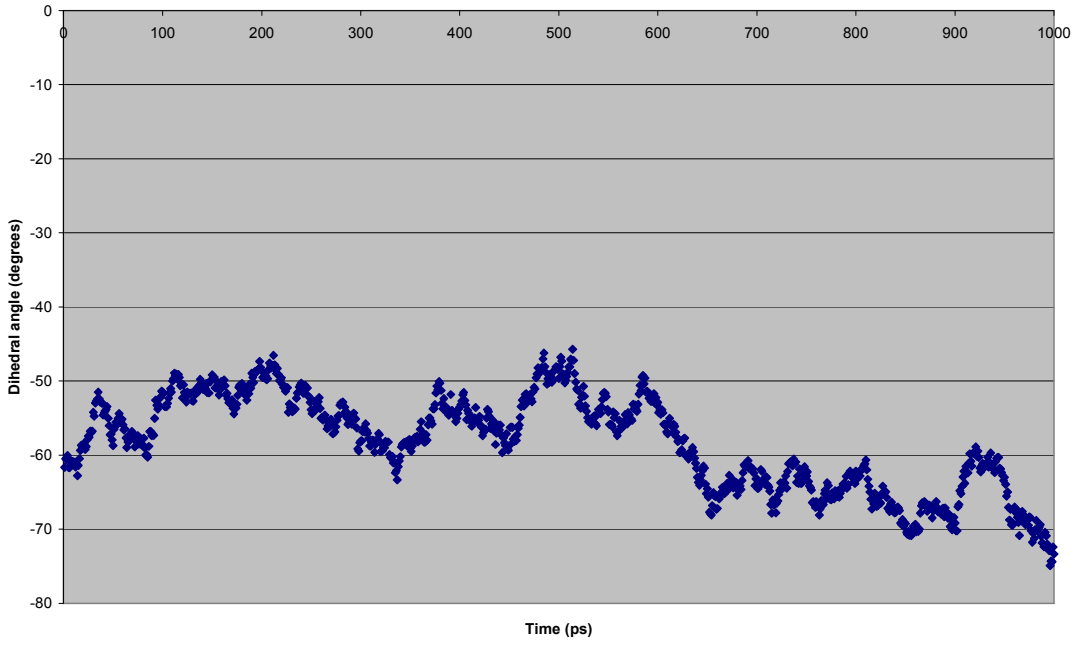


Figure II.6 – A78P dihedral angle *vs*. time

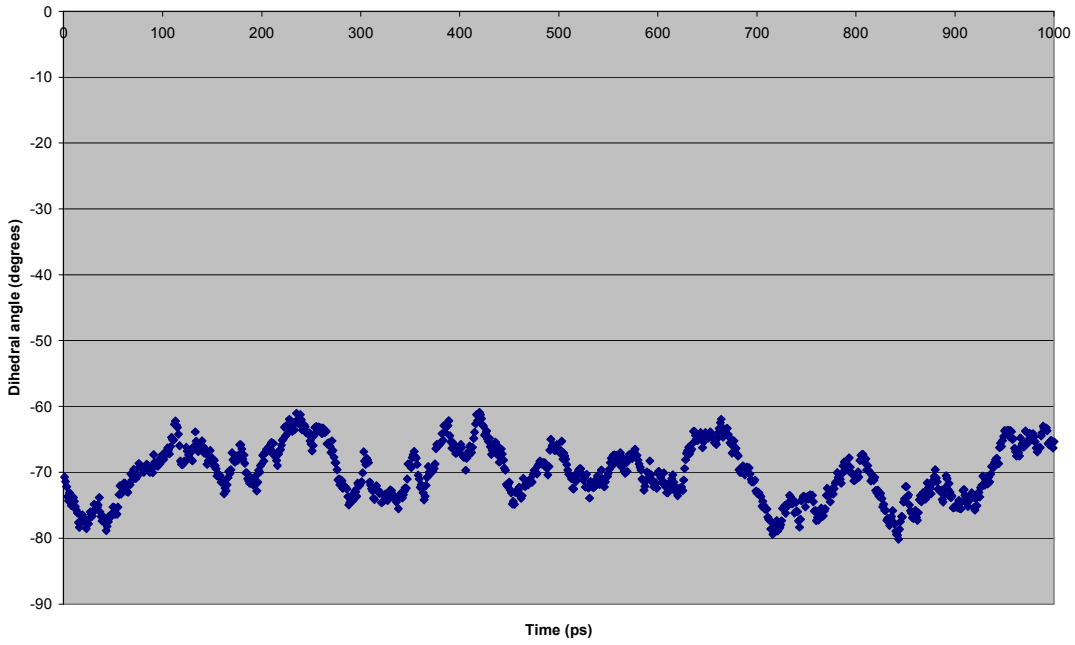


Figure II.7 – A78D dihedral angle *vs.* time

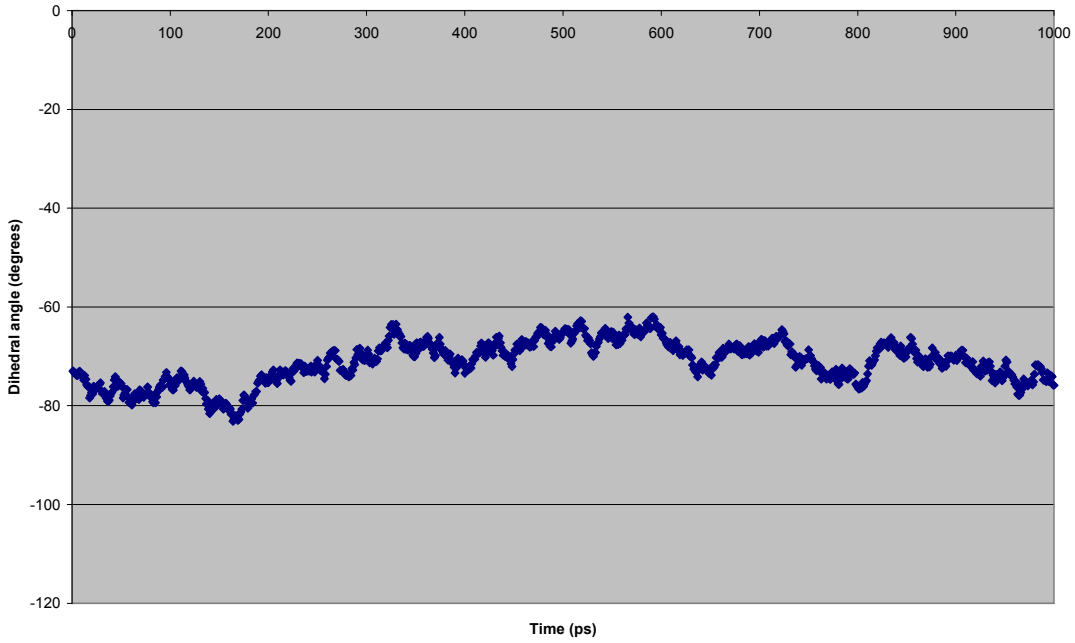


Figure II.8 – CueR.DNA dihedral angle *vs.* time

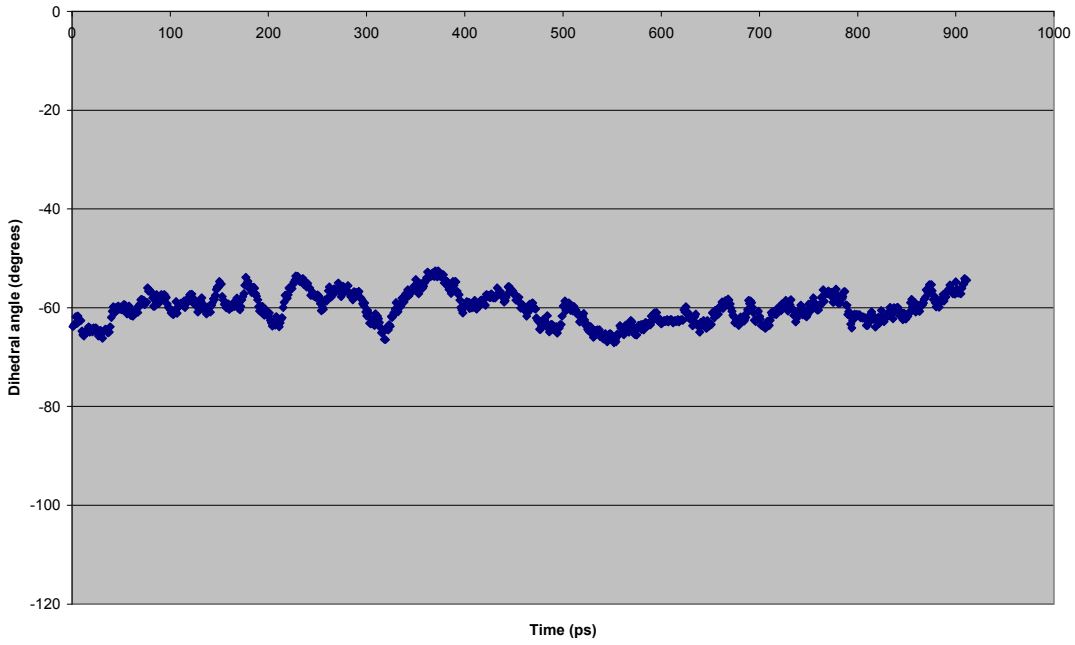


Figure II.9 – CueRncDNA dihedral angle *vs*. time

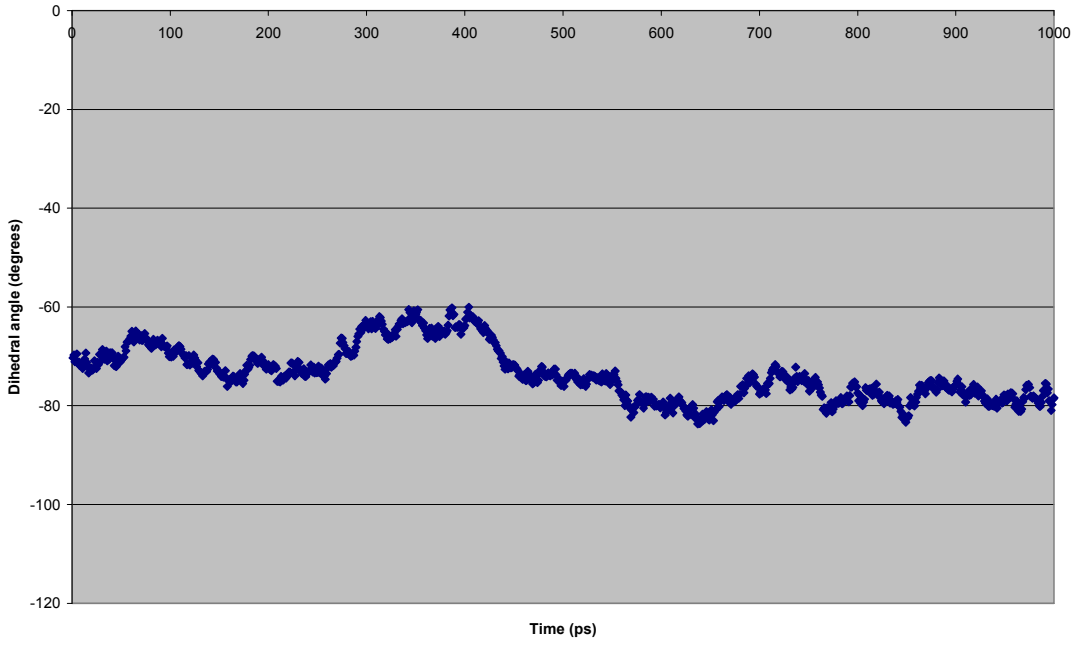


Figure II.10 – A78P.DNA dihedral angle *vs*. time

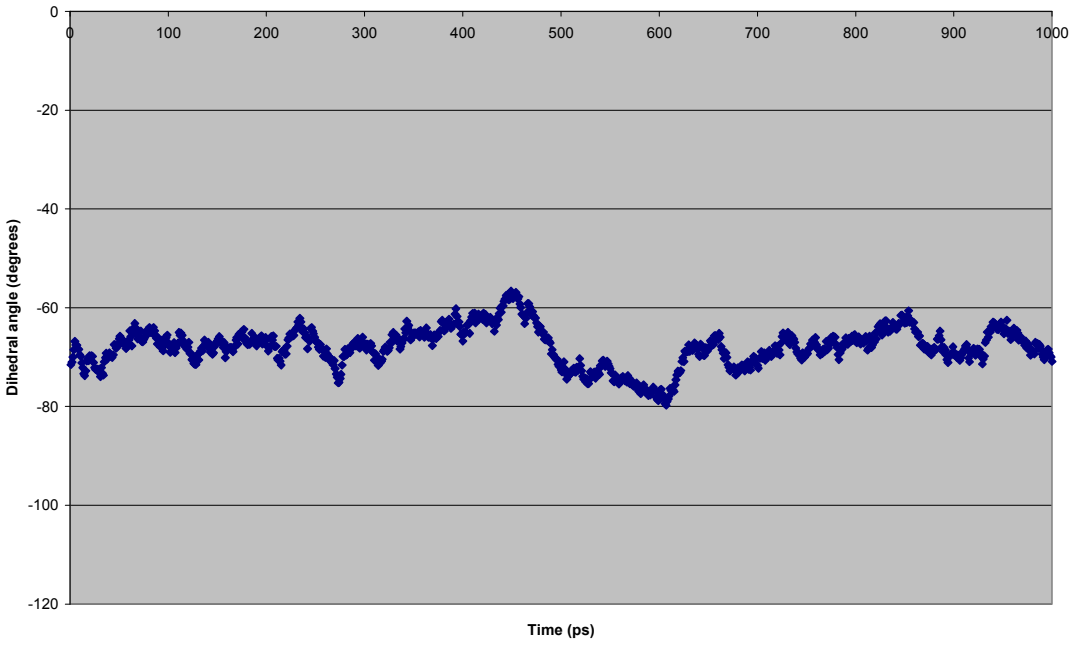


Figure II.11 – A78D.DNA dihedral angle *vs.* time

II b) C112-S: S77-O distance graphs

The following graphs show how the C112-S: S77-O distance varied *vs.* time for each of the experiments described in sections 5.4 and 5.7. Graphs are shown below, one per section.

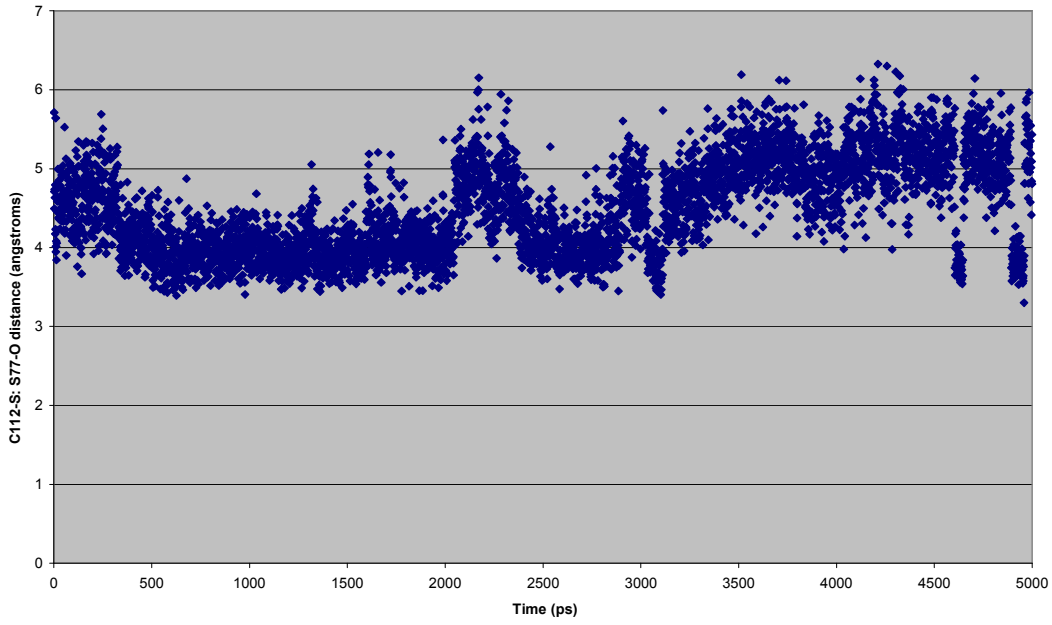


Figure II.12 – CueRH C112-S: S77-O distance *vs*. time 1 (5 ns)

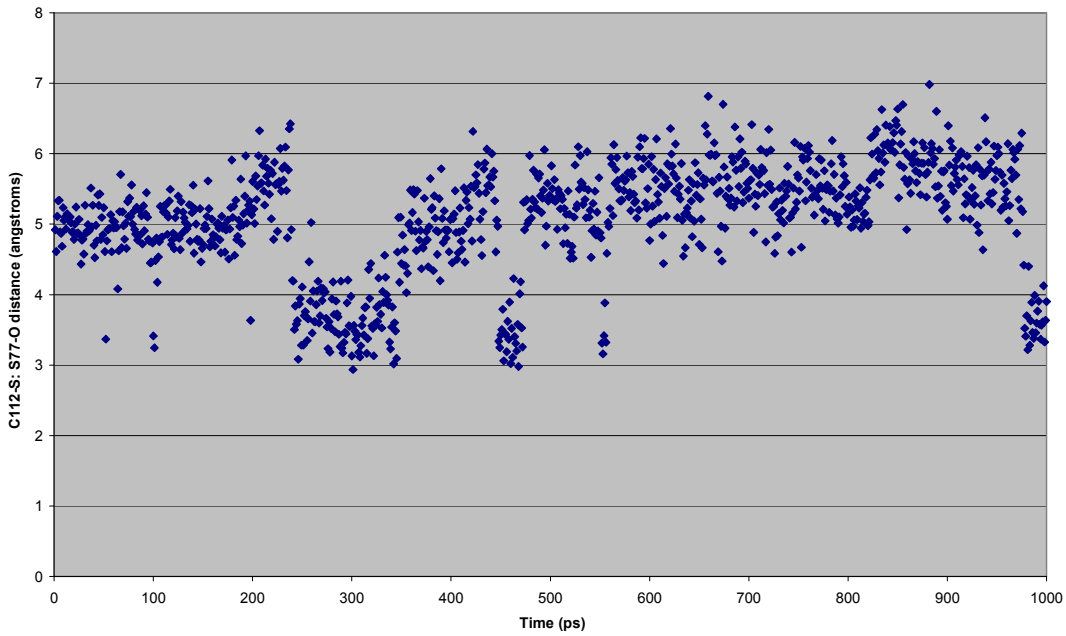


Figure II.13 – CueRHep C112-S: S77-O distance *vs*. time

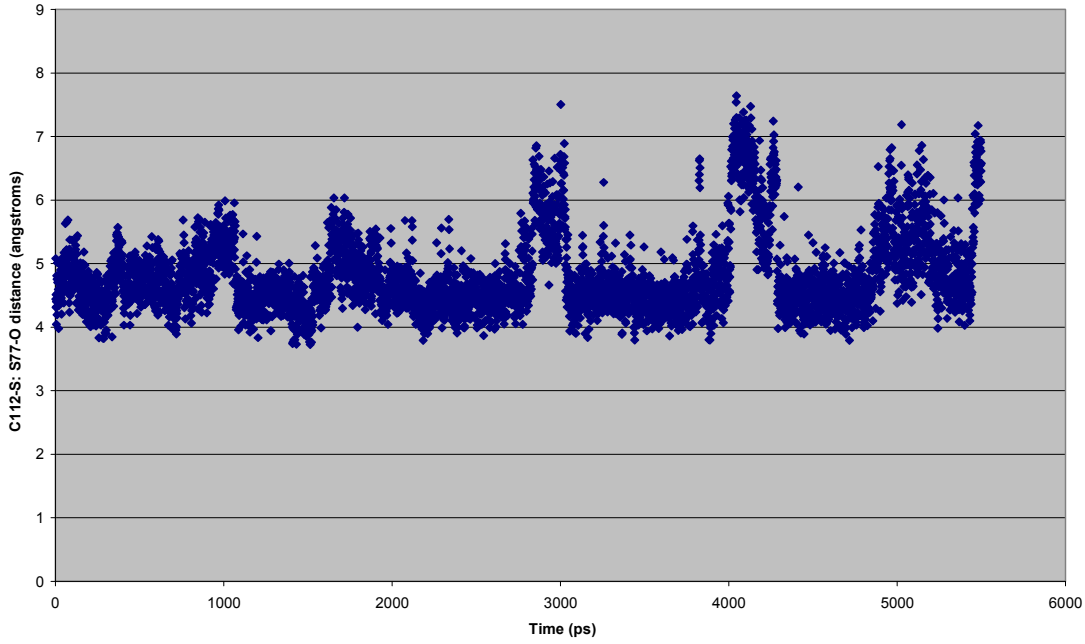


Figure II.14 – CueR- C112-S: S77-O distance *vs*. time 1 (5.5 ns)

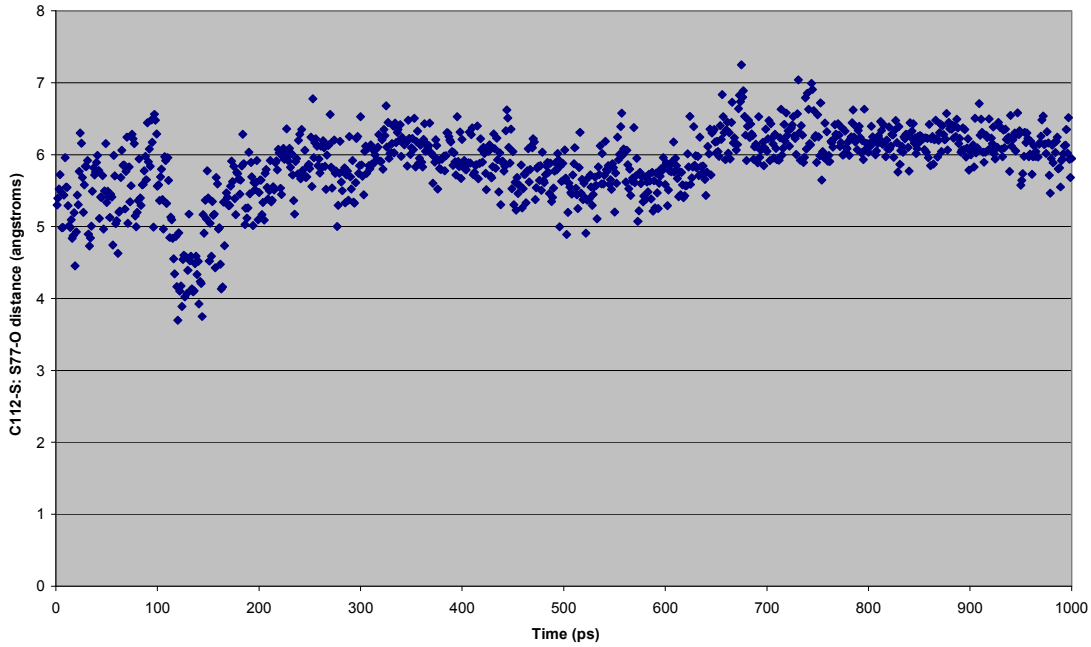


Figure II.15 – CueR-ep C112-S: S77-O distance *vs*. time

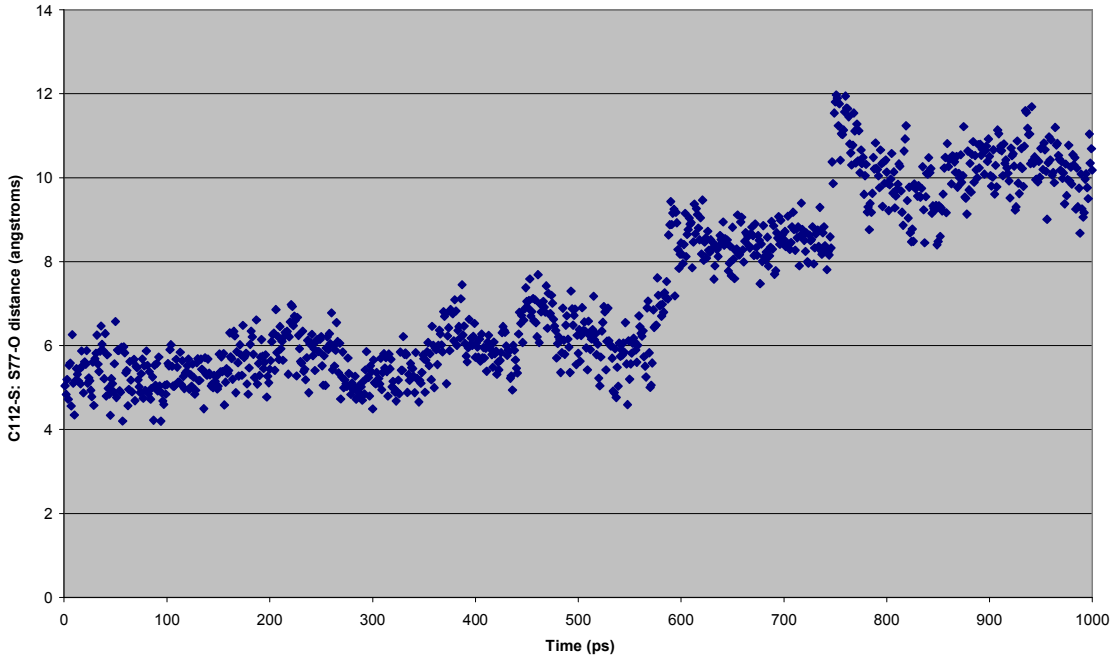


Figure II.16 – CueRnc C112-S: S77-O distance *vs*. time

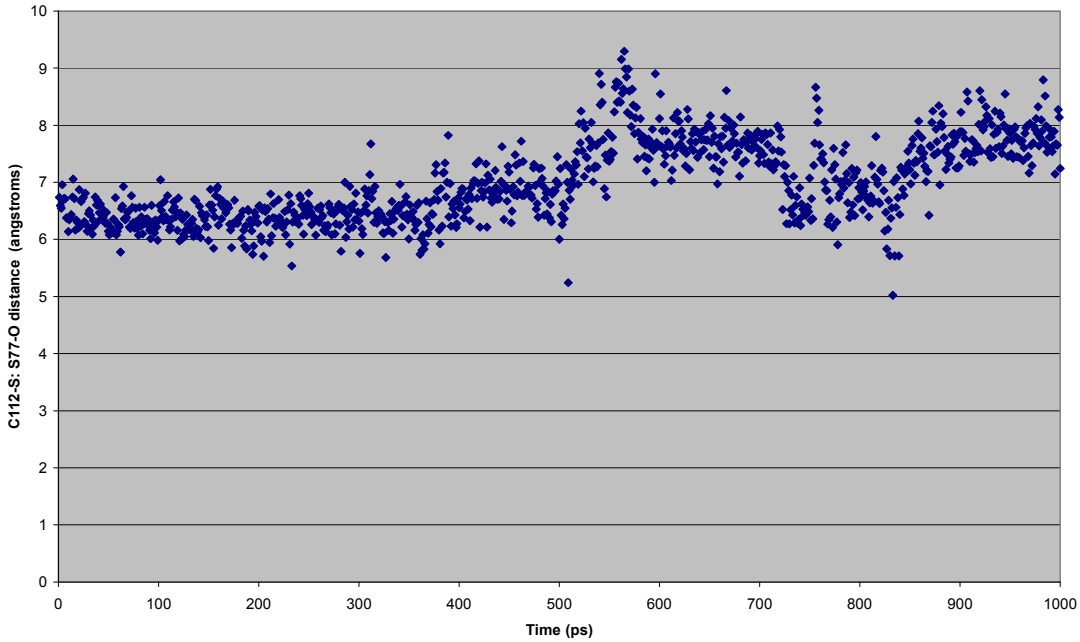


Figure II.17– A78P C112-S: S77-O distance *vs*. time

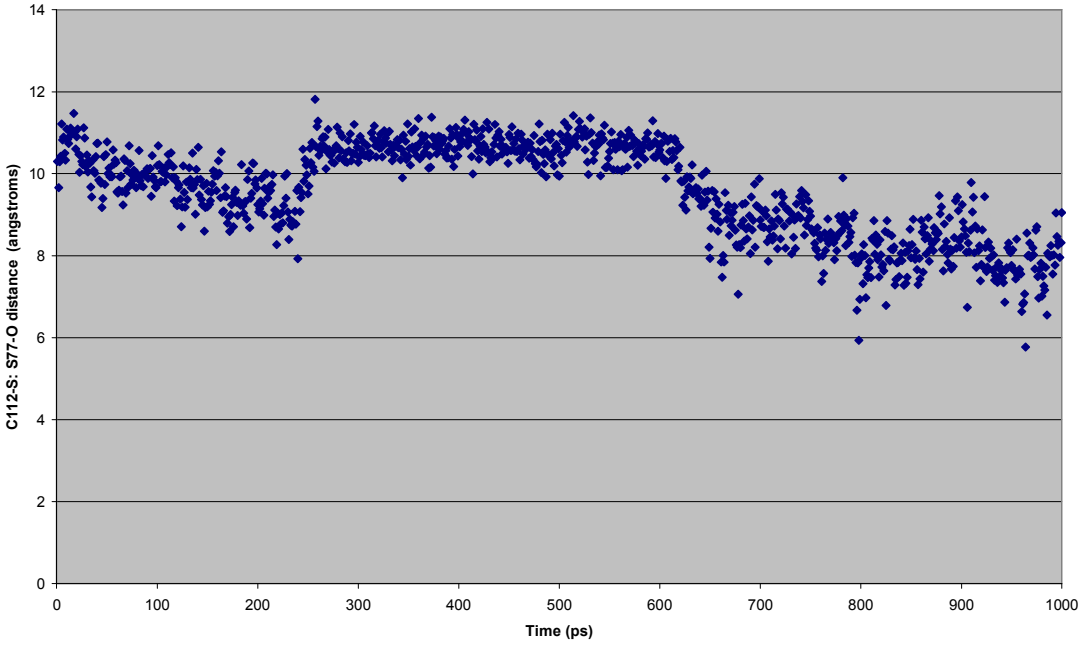


Figure II.18 – A78D C112-S: S77-O distance *vs*. time

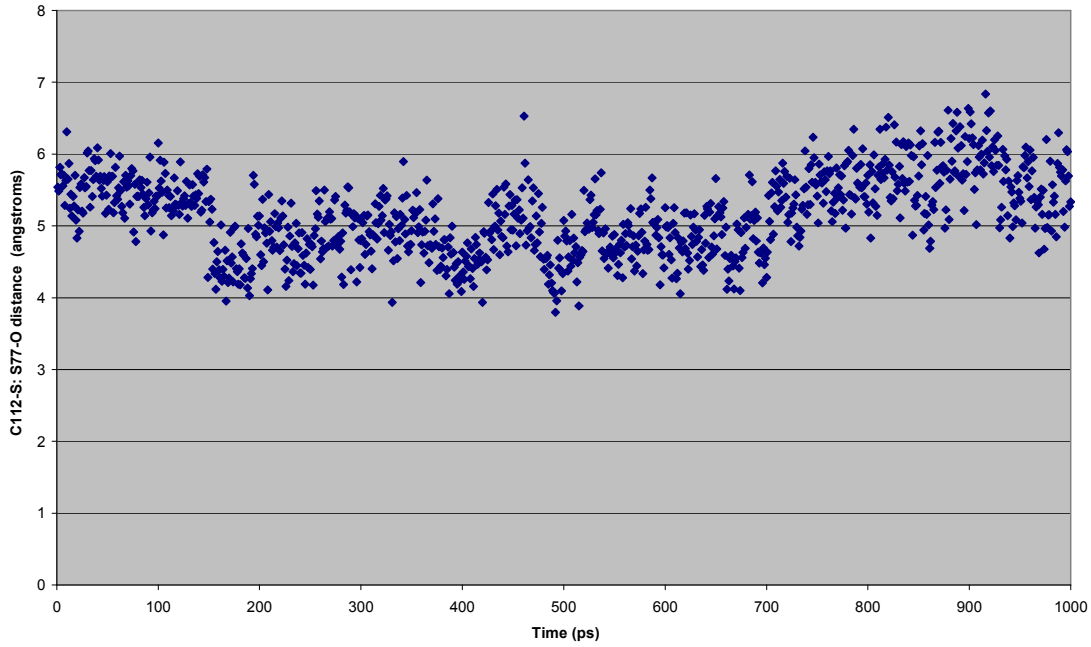


Figure II.19 – CueR.DNA C112-S: S77-O distance *vs*. time

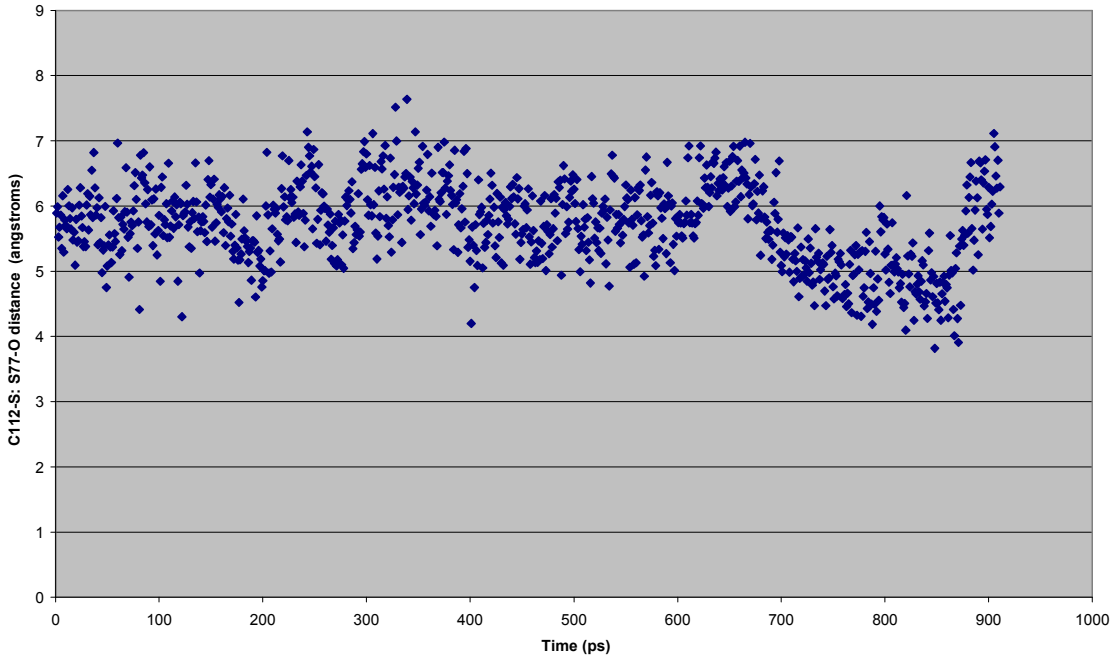


Figure II.20 – CueRncDNA C112-S: S77-O distance *vs.* time

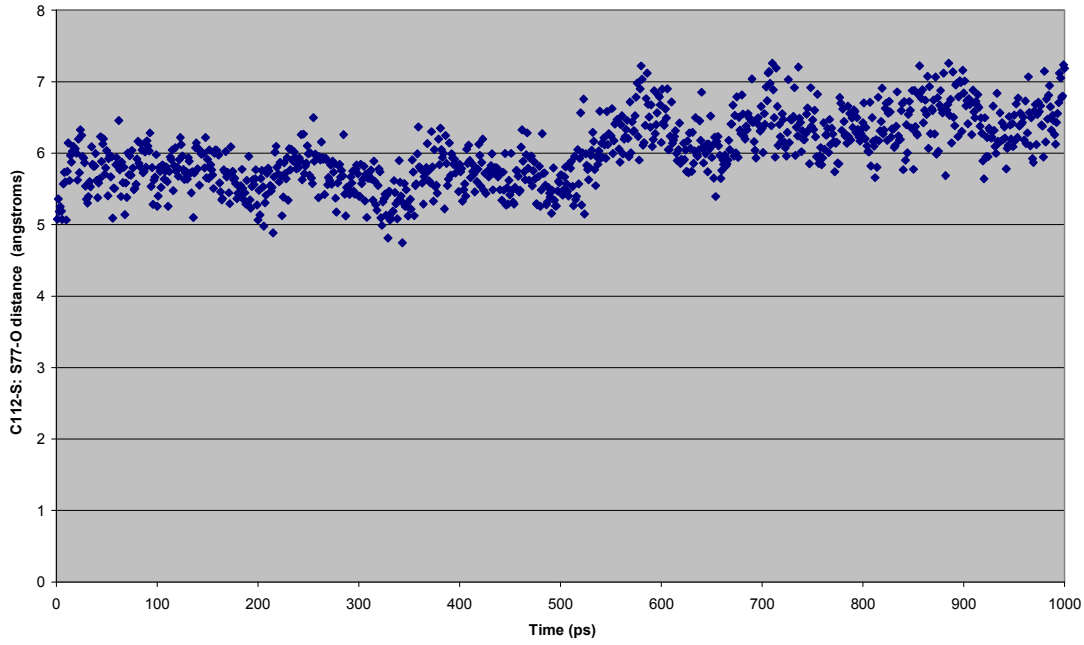


Figure II.21 – A78P.DNA C112-S: S77-O distance *vs*. time

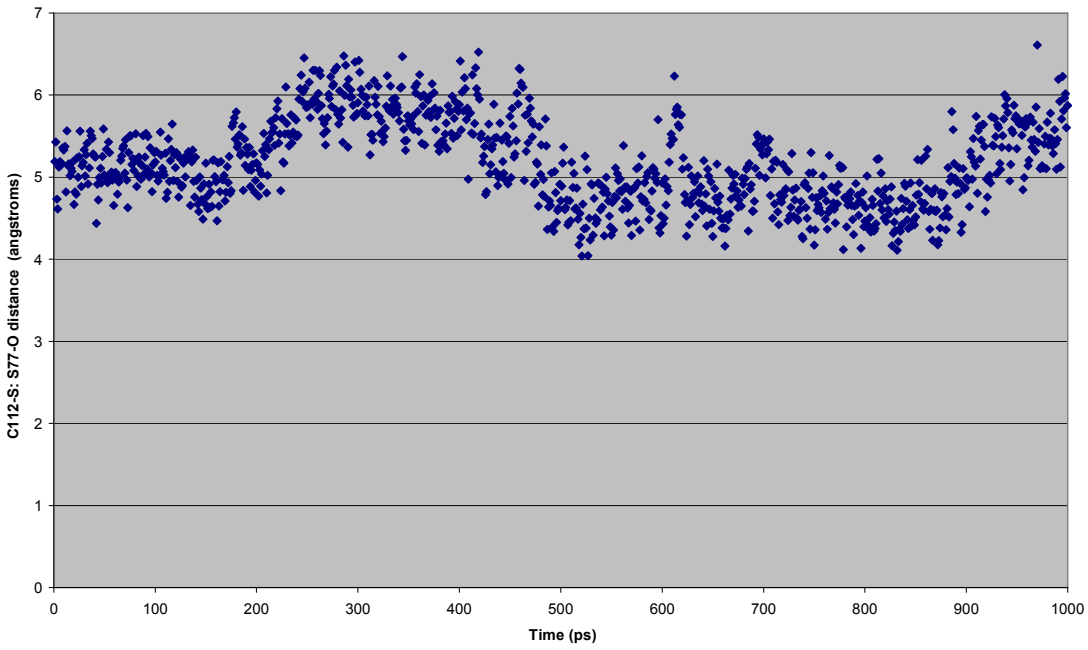


Figure II.22– A78D.DNA C112-S: S77-O distance *vs*. time

II c) RMSd graphs for all atoms

The following graphs show how the RMSd for all atoms varied *vs.* time for each of the experiments described in sections 5.4 and 5.8. Graphs are shown below, one per section.

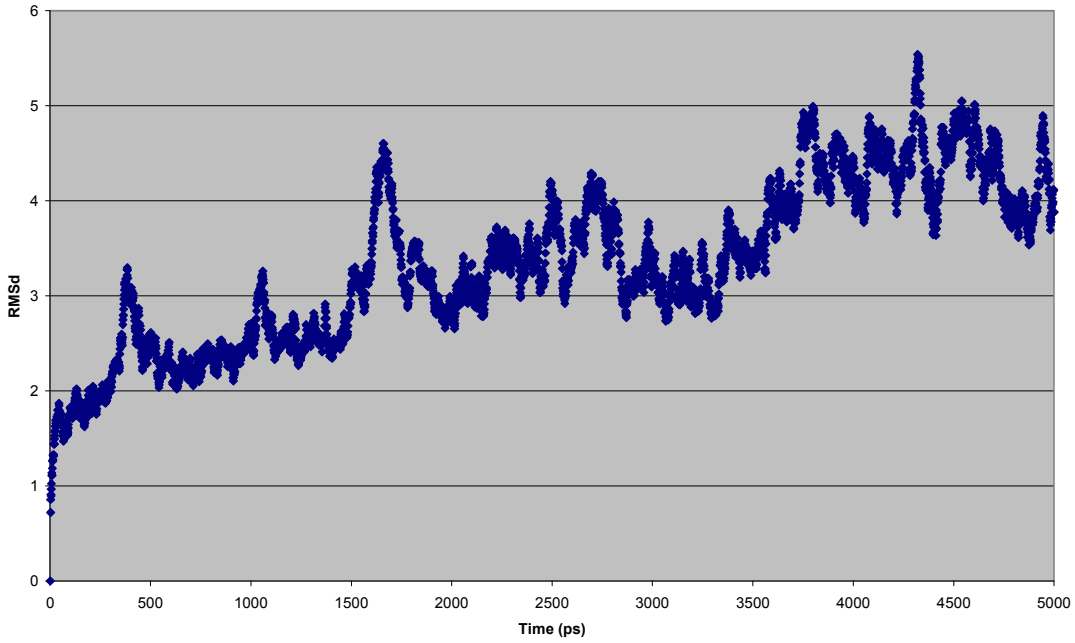


Figure II.23 – CueRH RMSd all atoms *vs*. time (5 ns)

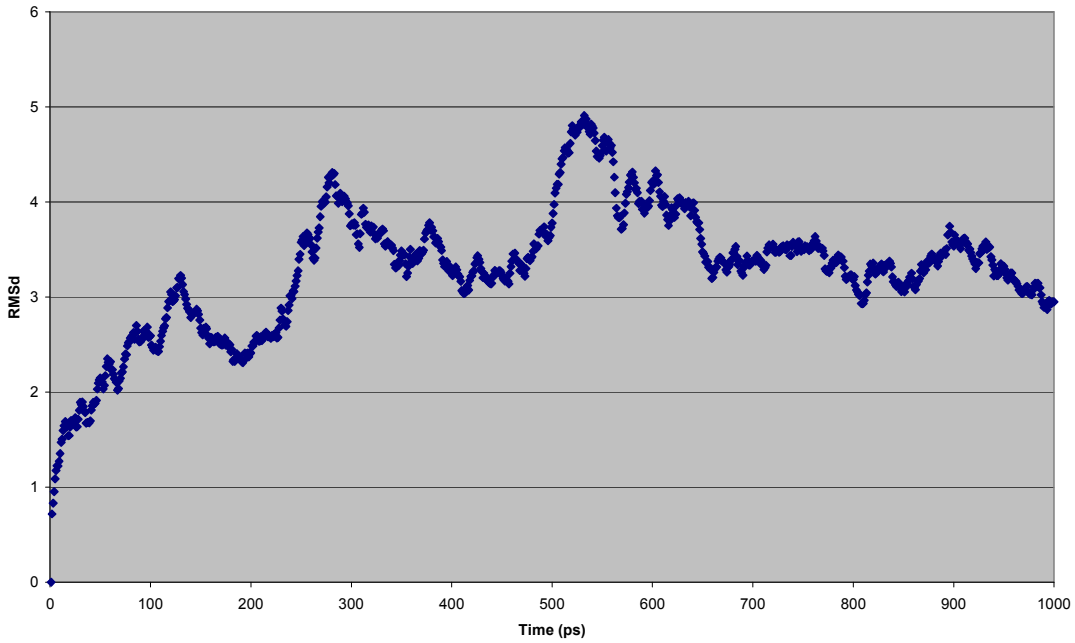


Figure II.24 – CueRHep RMSd all atoms *vs*. time

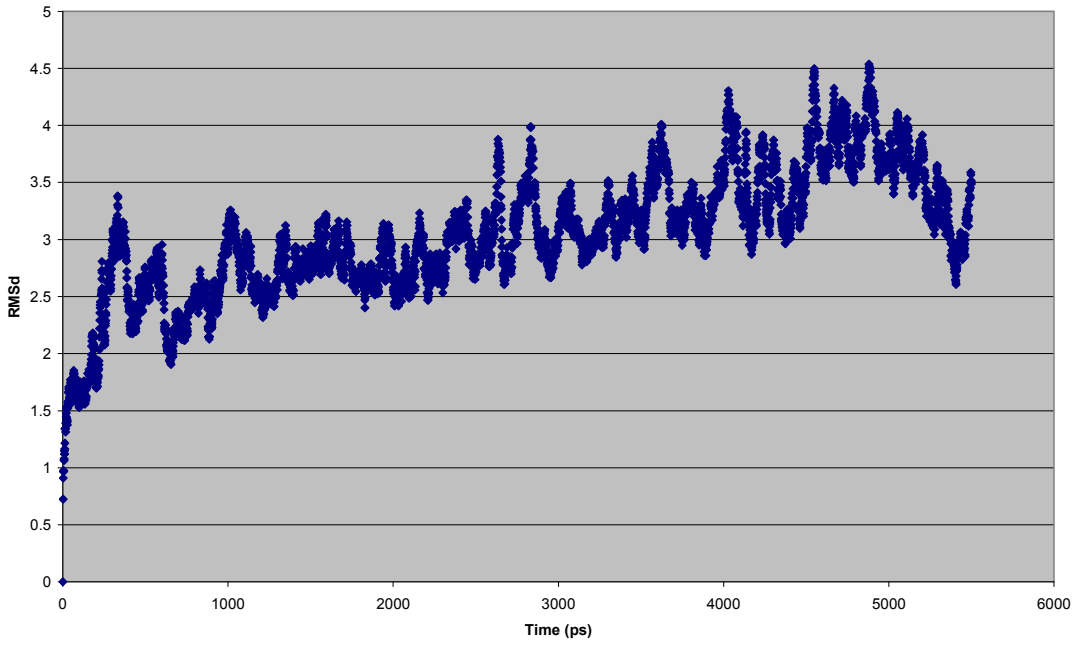


Figure II.25 – CueR- RMSd all atoms *vs*. time 1 (5.5 ns)

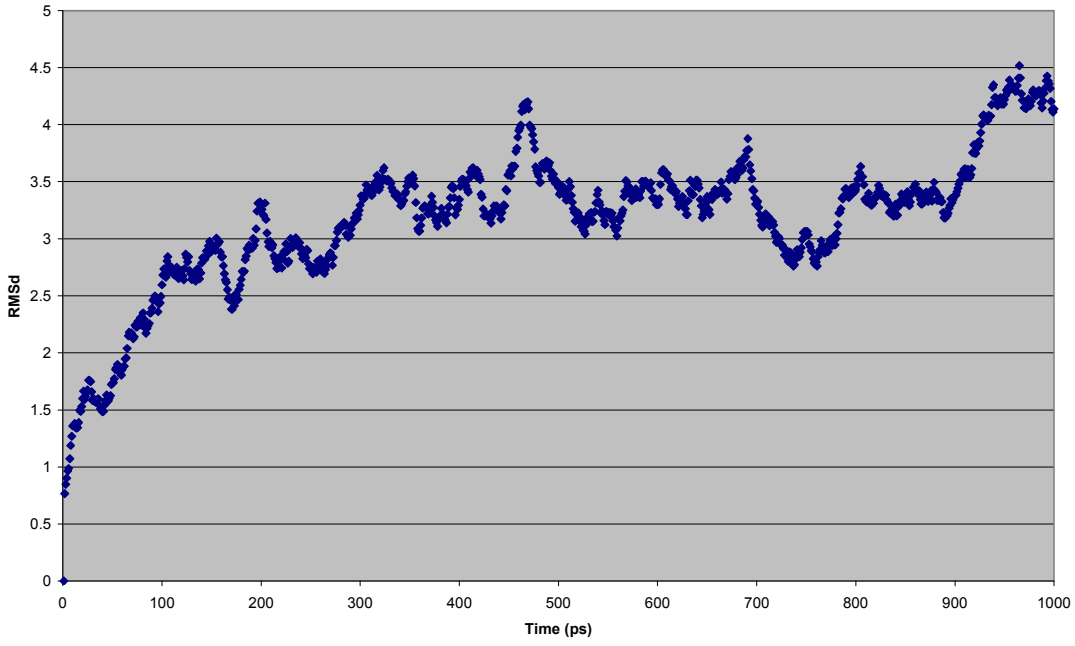


Figure II.26 – CueR-ep RMSd all atoms *vs*. time

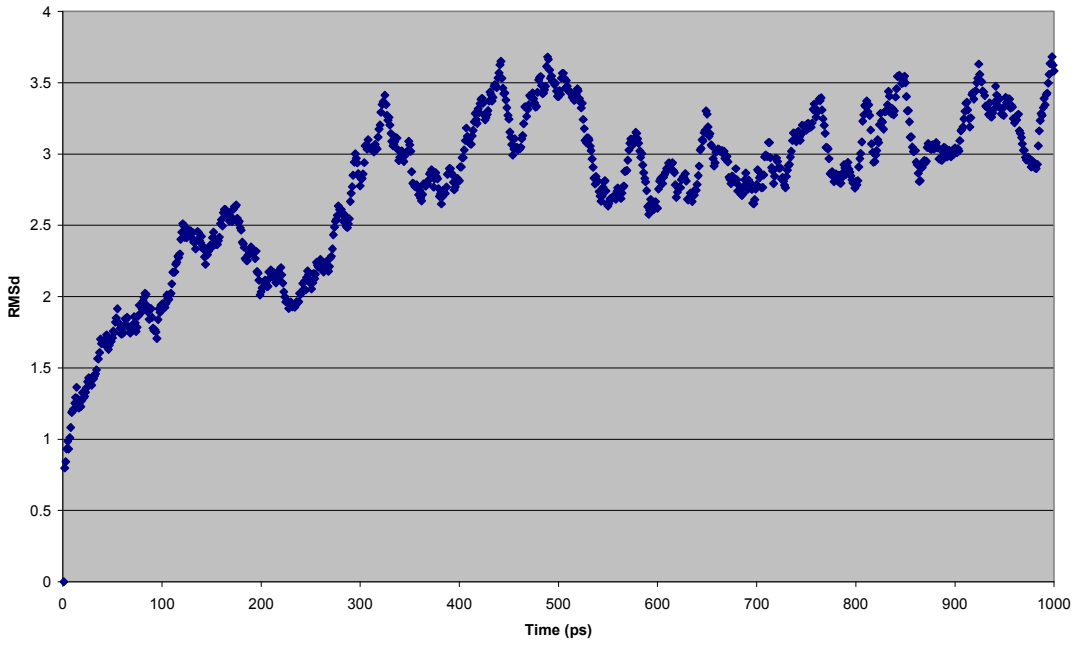


Figure II.27 – CueRnc RMSd all atoms *vs*. time

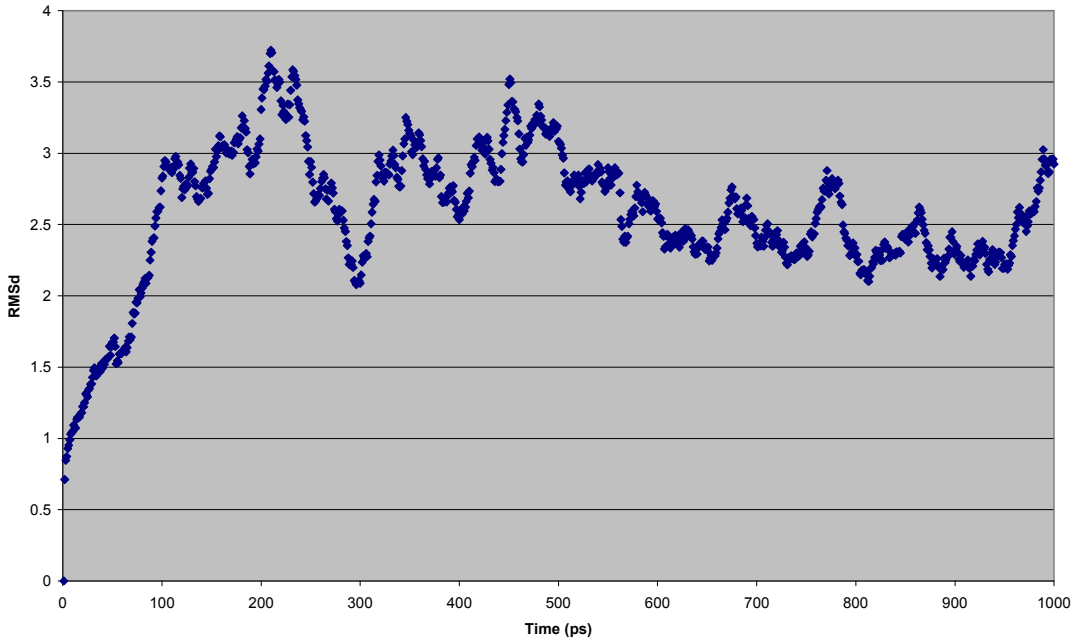


Figure II.28 – A78P RMSd all atoms *vs*. time

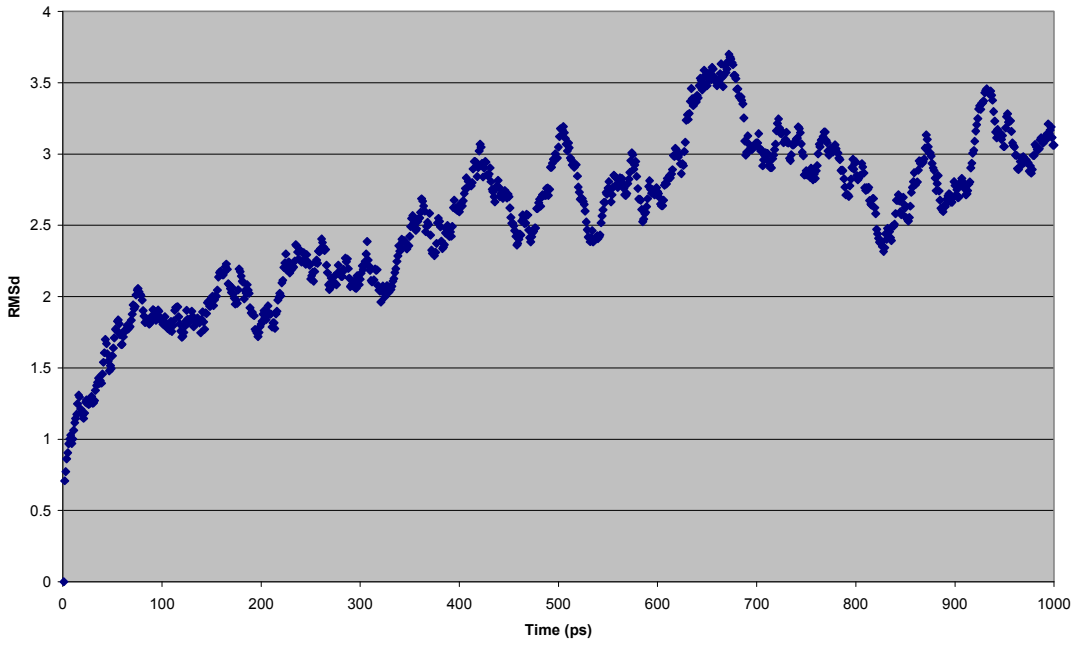


Figure II.29 – A78D RMSd all atoms *vs*. time

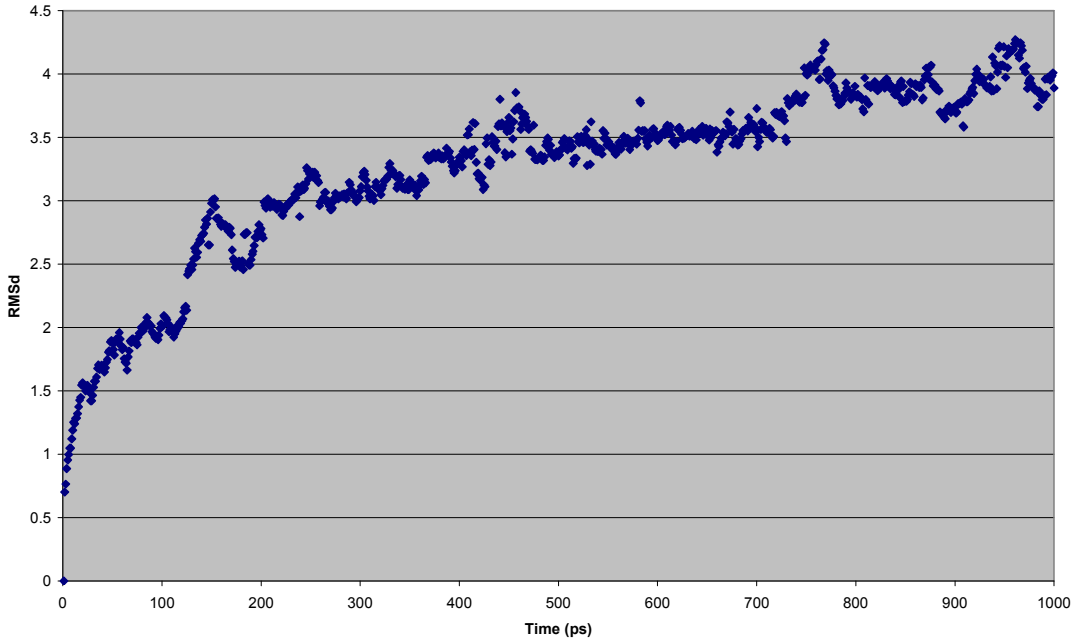


Figure II.30 – CueR.DNA RMSd all atoms *vs*. time

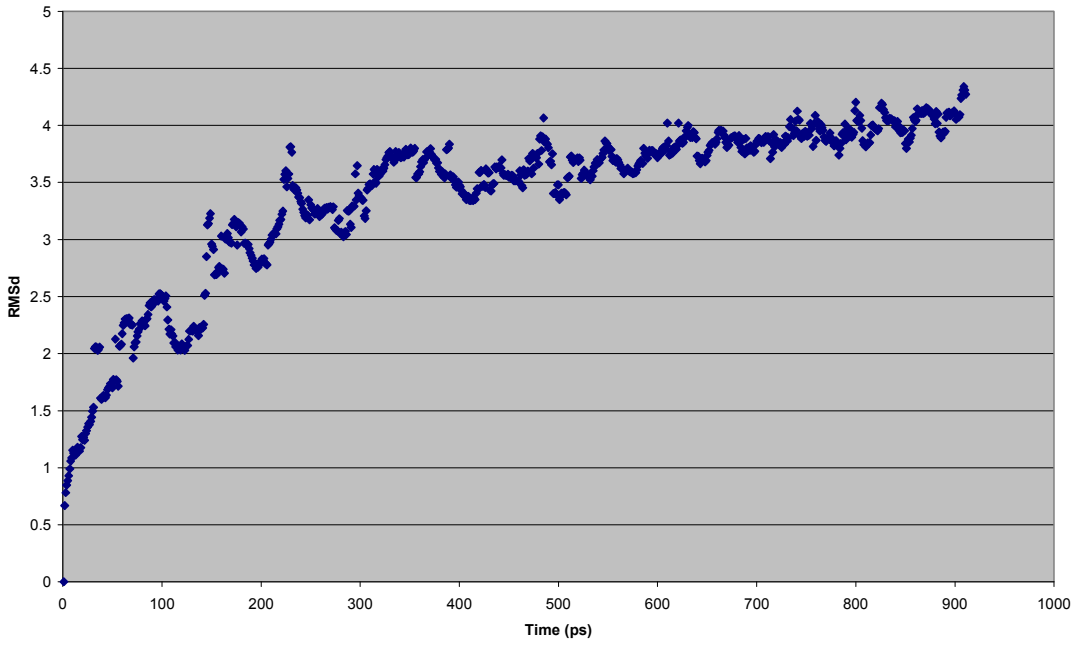


Figure II.31 – CueRncDNA RMSd all atoms *vs*. time

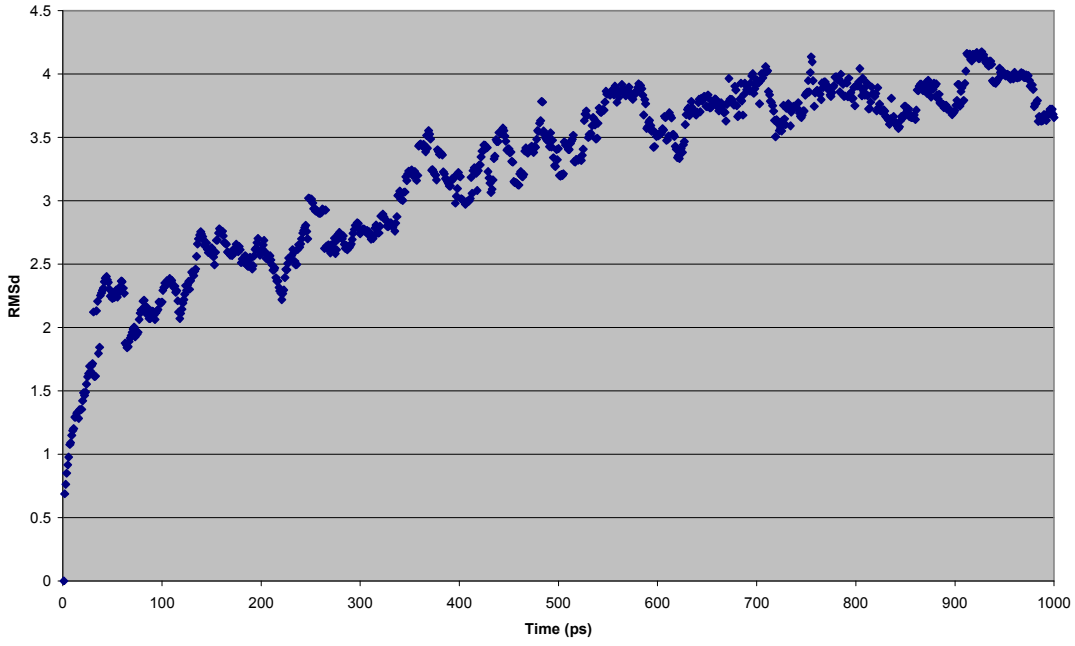


Figure II.32 – A78P.DNA RMSd all atoms *vs*. time

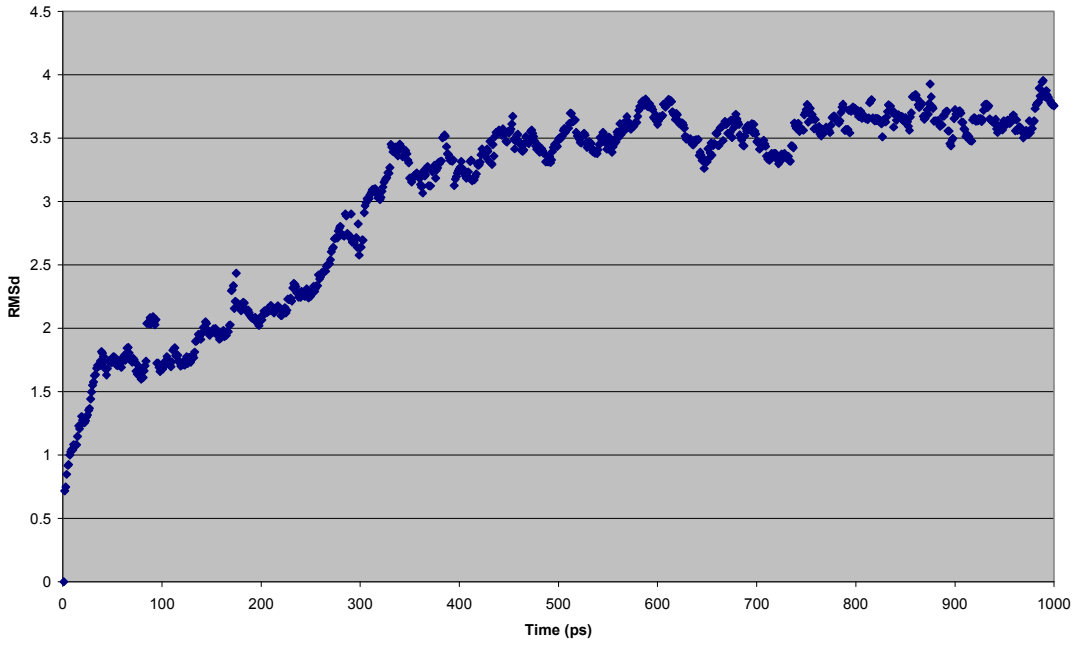


Figure II.33 – A78D.DNA RMSd all atoms *vs*. time

II d) RMSd graphs for alpha carbons

The following graphs show how the RMSd for the alpha carbons varied *vs.* time for each of the experiments described in sections 5.4 and 5.8. Graphs are shown below, one per section.

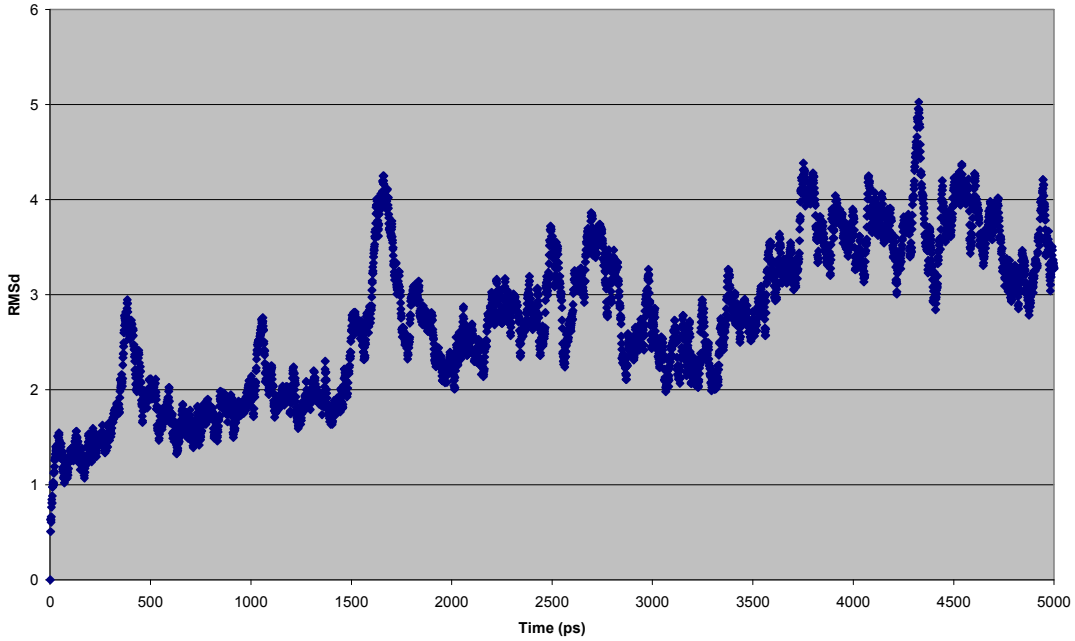


Figure II.34 – CueRH RMSd alpha carbons *vs*. time (5 ns)

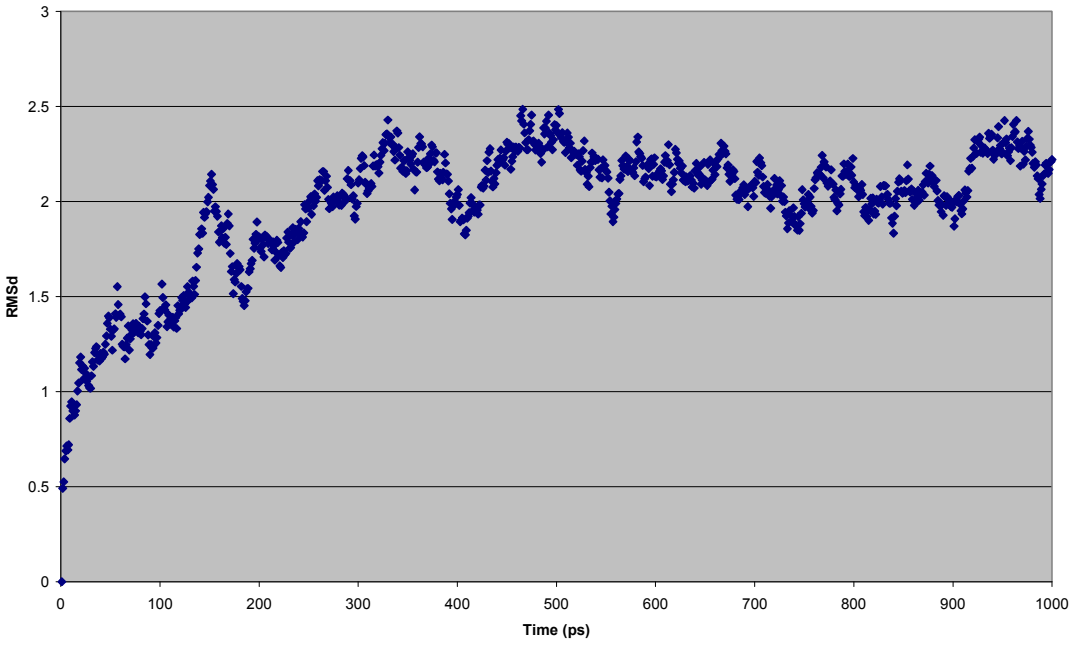


Figure II.35 – CueR.DNA RMSd alpha carbons *vs*. time

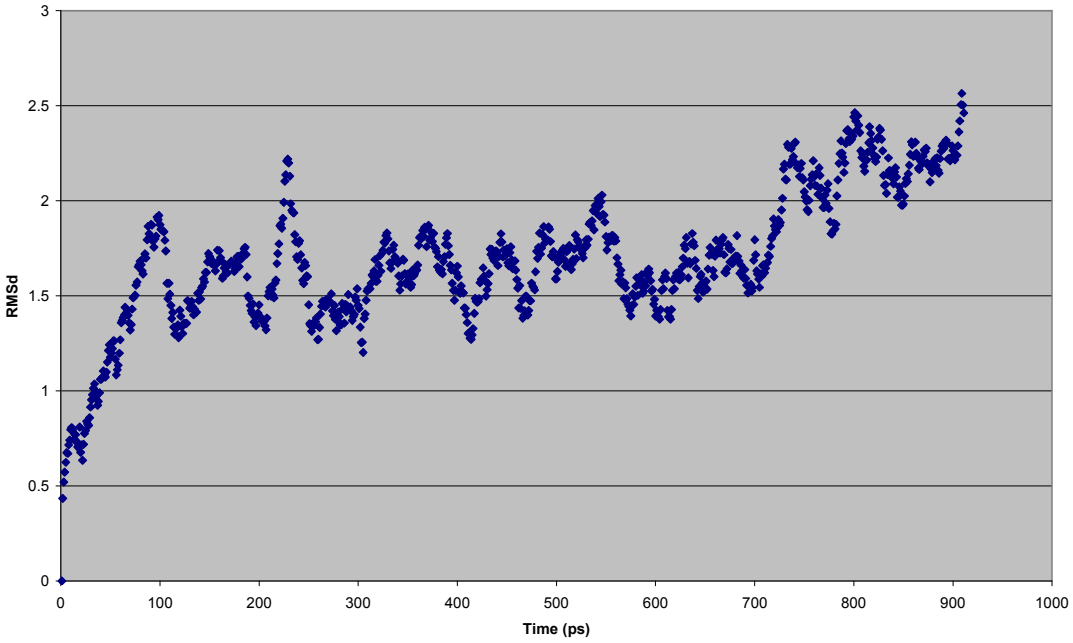


Figure II.36 – CueRncDNA RMSd alpha carbons *vs*. time

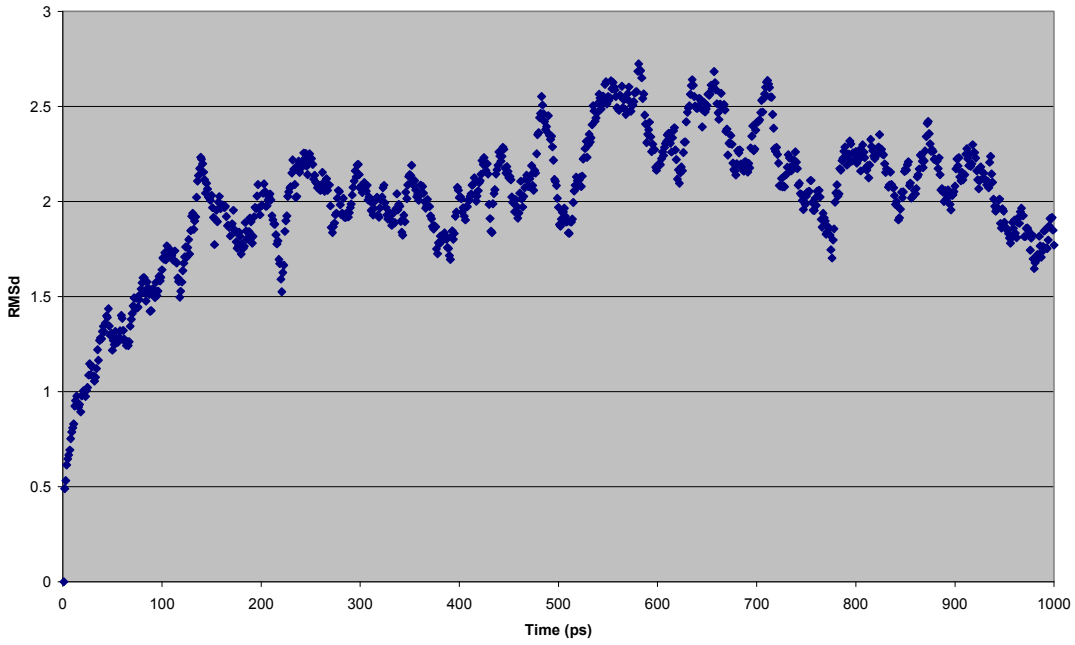


Figure II.37 – A78P.DNA RMSd alpha carbons *vs*. time

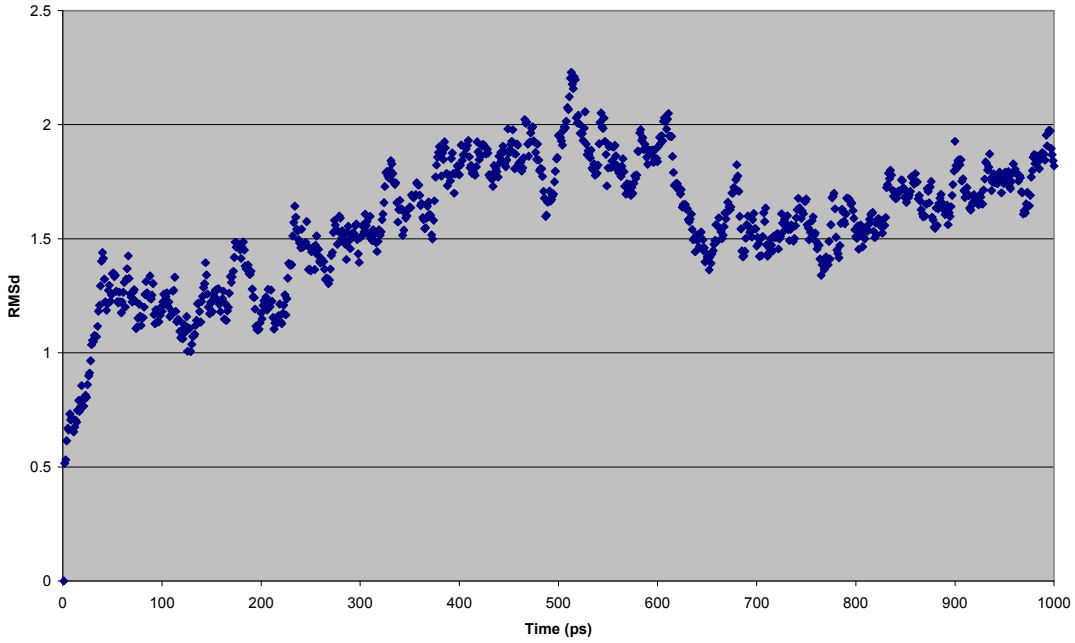


Figure II.38 – A78D.DNA RMSd alpha carbons *vs*. time

II e) DNA angle graphs

The following graphs show how the DNA angle varied *vs.* time for each of the experiments described in sections 5.4 and 5.9. Graphs are shown below, one per section.

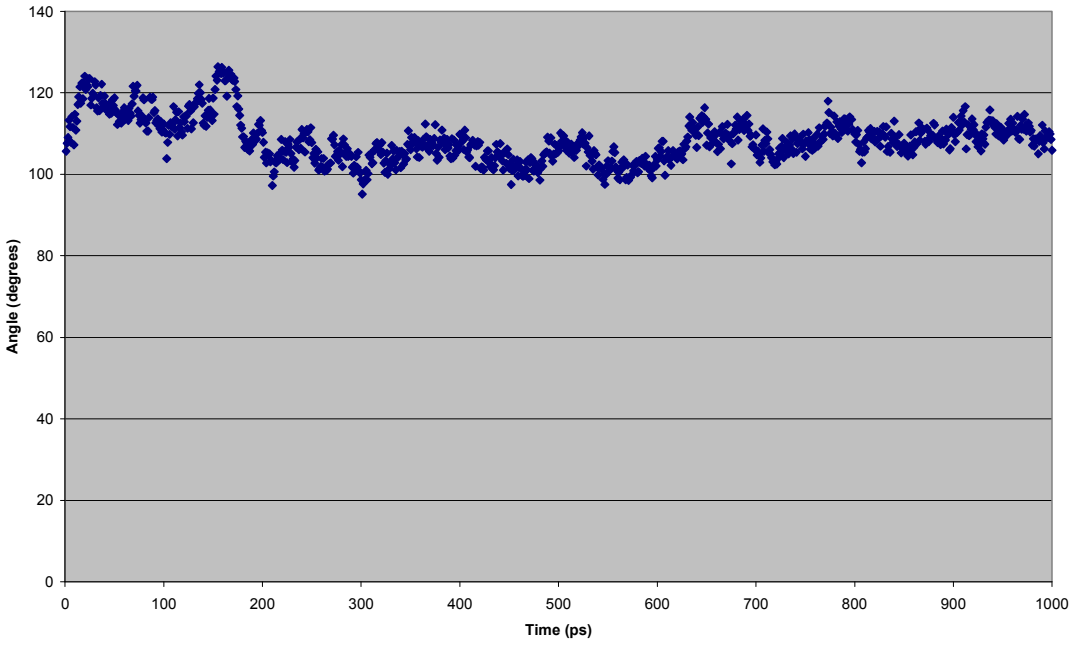


Figure II.39 – CueR.DNA DNA angle *vs*. time

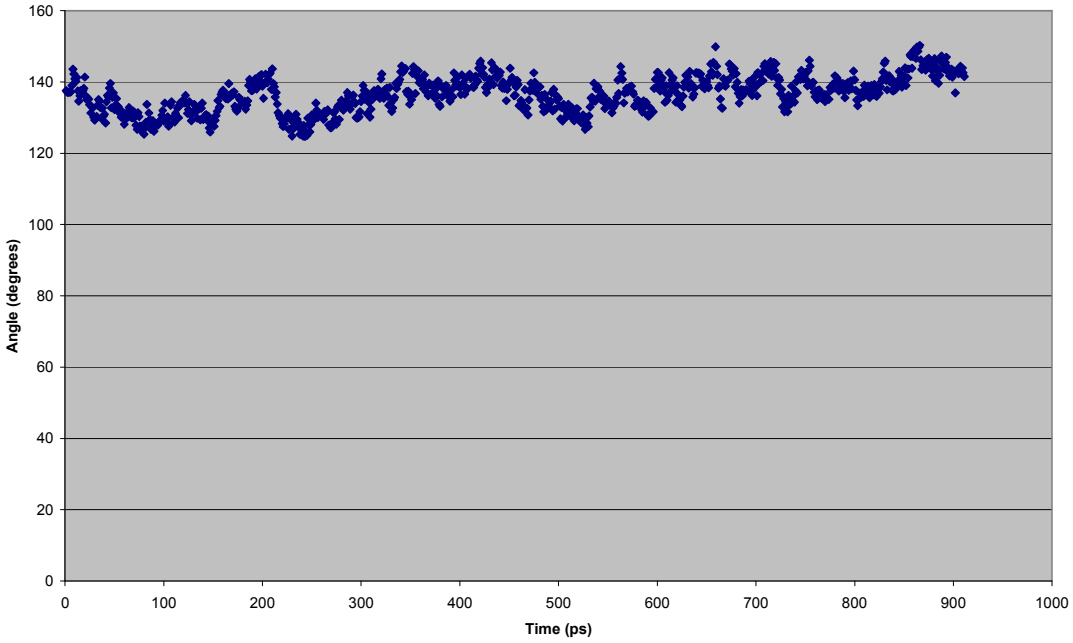


Figure II.40 – CueRncDNA DNA angle *vs*. time

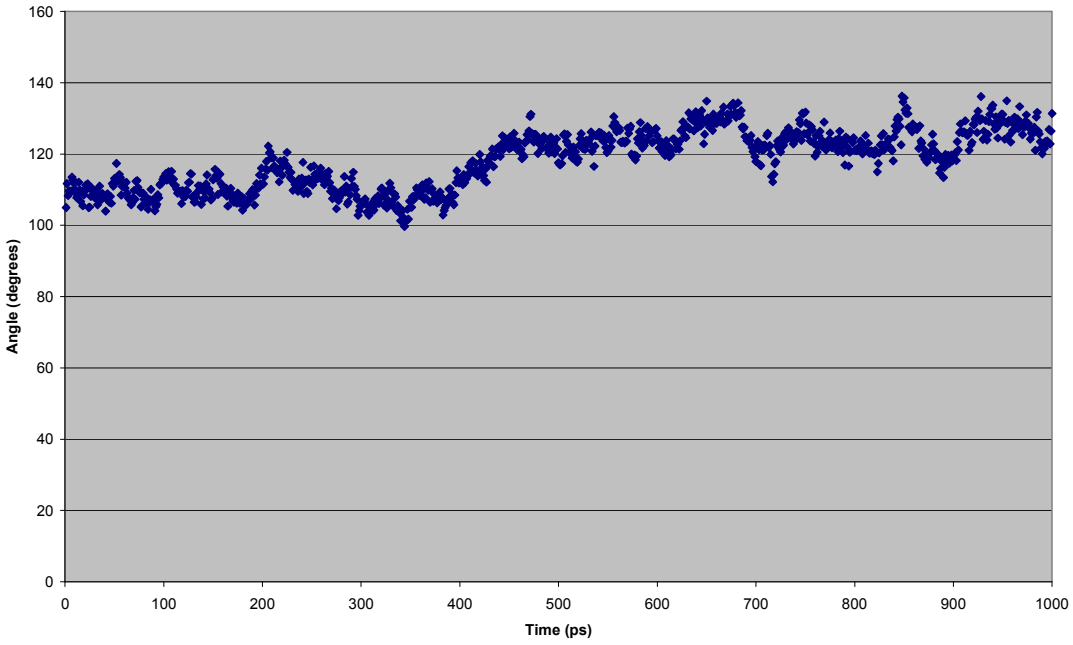


Figure II.41 – A78P.DNA DNA angle *vs*. time

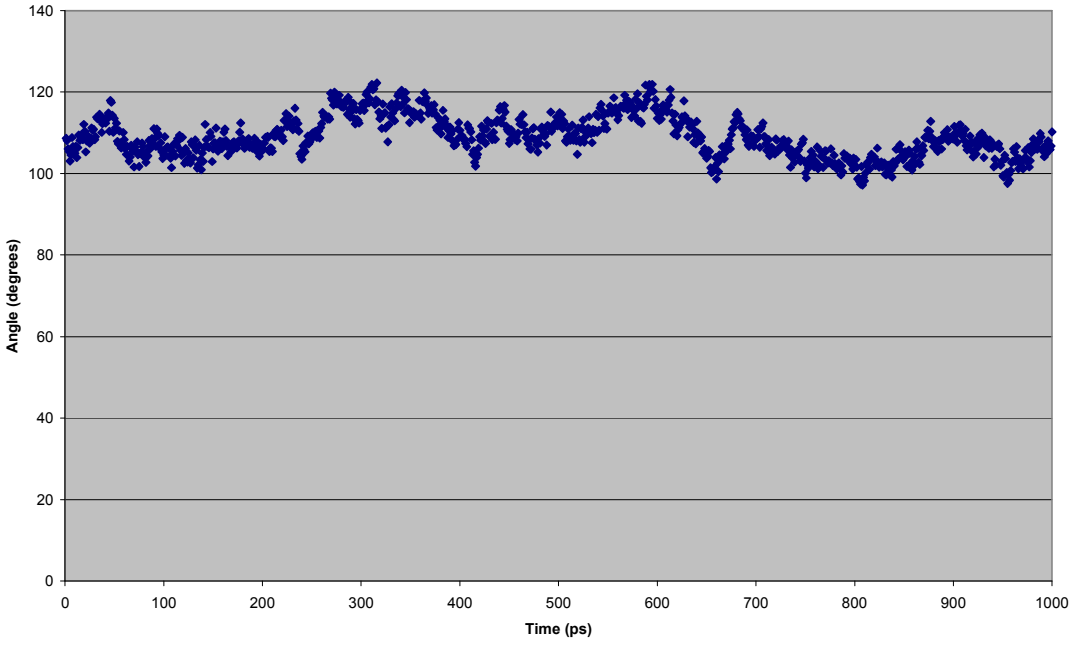


Figure II.42 – A78D.DNA DNA angle *vs*. time