Copper as a Biological Yin-Yang Element: 
Structural Dynamics, Protein-Protein Interactions and Transfer 
Mechanisms of Copper Transport Proteins

by

Agustina Rodriguez-Granillo

A THESIS SUBMITTED IN PARTIAL FULFILLMENT 
OF THE REQUIREMENTS FOR THE DEGREE 
Doctor of Philosophy

APPROVED, THESIS COMMITTEE:

Pernilla Wittung-Stafshede, Adjunct Professor 
Biochemistry and Cell Biology

John S. Olson, Ralph and Dorothy Looney Professor 
Biochemistry and Cell Biology

Kathleen M. Beckingham, Professor 
Biochemistry and Cell Biology

Yousif Shamoo, Associate Professor 
Biochemistry and Cell Biology

Luay K. Nakhleh, Assistant Professor 
Computer Science

HOUSTON, TEXAS
NOVEMBER 2009
ABSTRACT

Copper as a Biological Yin-Yang Element:
Structural Dynamics, Protein-Protein Interactions and Transfer Mechanisms of
Copper Transport Proteins

by

Agustina Rodriguez-Granillo

Copper (Cu) can be considered as a biological yin-yang element, because it is essential but toxic at the same time. To manage this paradox, cells have evolved complex molecular Cu transport pathways, in which Cu chaperones bind and shuttle Cu to specific intracellular targets. In humans, Atox1 delivers Cu(I) to the metal-binding domains (MBDs) of Cu-ATPases, the Menkes and Wilson disease proteins, for subsequent incorporation into cuproenzymes. Both metallochaperones and target MBDs adopt the same ferredoxin-like fold and bind Cu(I) via two Cys residues in a conserved motif. In this thesis, we have employed a wide selection of state-of-the-art computational schemes, including quantum mechanics and molecular mechanics methodologies in combination with molecular dynamics simulations, to broaden our understanding on the structure-function relationships of Cu chaperones, MBDs, and their interactions. This work reports on a thorough study of the structural dynamics, Cu-binding properties, protein-protein interactions and Cu-transfer mechanisms of key Cu transport proteins, and how conserved residues modulate these properties.
We found that residues framing the Cu loop have evolved differently in prokaryotic and eukaryotic Cu chaperones to tune the flexibility and provide an optimal stabilization of the Cu loop. Some of these residues are also key for metallochaperone-MBD interactions and subsequent Cu(I) transfer. We further found that the MBDs are not equivalent at the molecular level, and propose that backbone flexibility together with electrostatic complementarity are important factors to guide Atox1 interactions. We propose that Atox1 interacts with its partner MBDs via a "weak" interface that can be disrupted by at least one substitution in Atox1. Finally, we have elucidated for the first time the Cu(I) transfer mechanism from holo-Atox1 to an apo-MBD, and propose that the reaction proceeds with the existence of two trigonal intermediates. Our results suggest that the reaction is kinetically feasible but not energetically favorable, pointing to the apparent absence of a thermodynamic gradient for Cu(I) transfer. The structural, dynamic, thermodynamic and mechanistic details obtained here with atomic resolution are difficult to obtain by *in vitro* experiments, and can be used both as a complement to experiments and as predictive tools for functional insights.
ACKNOWLEDGEMENTS

I would like to acknowledge many people without whom their advice and support this thesis would have not been possible. First, I would like to thank my advisor, Prof. Pernilla Wittung-Stafshede, for allowing me to pursue non-typical “experiments” and for her constant support and encouragement.

I am indebted to Dr. Alejandro Crespo, who introduced me to the Linux or “ls” world and taught me almost everything I know about computational chemistry and modeling. Thank you for your endless patience and invaluable input, for answering my million questions 24-7, and for the exciting scientific discussions. Thank you for reading and editing all my scientific documents, including manuscripts and this thesis. You own the title of my unofficial co-advisor.

I would also like to acknowledge the Chair of my committee and my local mentor, Prof. John Olson, for always being there, for his insightful comments and words of encouragement. I am also grateful to the other members of my committee, Profs. Kathleen Beckingham and Yousif Shamoo for their constructive input.

I would like to thank former members of the Pernilla’s lab: Drs. Faiza Hussain, Mike Perham, Erik Sedlak and Loren Stagg as well as David Culbertson, for making every work day a more pleasant one. In particular, I am thankful to Dr. Faiza Hussain for sharing her experimental data with me, and for reading and editing some sections of this thesis. I am also thankful to all the people in Keck 3rd floor, especially to Prof. Kevin MacKenzie, the Olson’s and Shamoo’s lab. Specifically, I would like to thank Kasia Walkiewicz for the teatime talks.
I would like to express my appreciation to my collaborator and former professor, Prof. Dario Estrin from University of Buenos Aires, for his scientific and personal valuable input. Also, I am deeply grateful to all my professors at University of Buenos Aires and Rice University, who educated me in all the past years.

Finally, I would like to thank those who, in person, by chat or by phone, helped me survive graduate school in a foreign country. First of all, I would like to thank my husband, Ale. Gracias por acompañarme y estar a mi lado siempre, nada de esto sería posible sin vos. Gracias por soportar mis horarios, los fines de semana perdidos y por darme tu apoyo incondicional en los tiempos más difíciles. Gracias por insistirme en terminar, mas allá de la distancia (que ahora por fin se termina!!). I would also like to thank the Walkiewicz sisters, who opened their houses for us, and in home-sick times, like Christmas, made us feel like being at home. I would like to thank all my family and friends from Argentina. A mamá, por estar siempre presente (en espíritu o en persona!), por tus palabras de aliento y de esperanza de que todo iba a salir bien.... A papá, por creer en mí, y haber venido a visitarnos varias veces (y a mi tesis!!!) y haber hecho nuestra estadía en Houston mas amena, y por haberlos traído a Tomi y Cler!! A Matt, Gato, Tomi, Cler, Chachi, Feli, Santi, Iné, Lucas/Jaime por ser una familia increíble y estar siempre presente, mas allá de la distancia. A mi familia política, Elena, Jorge y Pablo, por escucharnos y apoyarnos siempre. Y por último, a las chicas, por divertirme con sus mails todos los días. En especial, a Luli por haber venido a visitarnos, y por guardar cada uno de mis papers!! Gracias a todos!!
LIST OF PUBLICATIONS

Publications or manuscripts currently in review discussed in this thesis:


Other publications not discussed in this thesis:

TABLE OF CONTENTS

CHAPTER 1:

THE REQUIREMENT OF AN INTRICATE COPPER BALANCE 1

1.1 PROTEINS AS THE CENTRAL CELLULAR MACHINERIES 1
1.2 COPPER AS A COMMON TRANSITION METAL IN BIOLOGY 2
1.3 COPPER AS A BIOLOGICAL YIN-YANG ELEMENT 4
1.4 EUKARYOTIC COPPER TRANSPORT PATHWAYS 6
1.5 COPPER ATPASES 8
1.6 COPPER TRANSPORT IN HUMANS AND ASSOCIATED PATHOLOGIES 10
1.7 COPPER CHAPERONES AND METAL-BINDING DOMAINS SHARE THE SAME FOLD 12

CHAPTER 2:

COMPUTATIONAL METHODS 15

2.1 INTRODUCTION TO COMPUTER SIMULATIONS 15
2.2 THEORETICAL FOUNDATIONS 16
2.2.1 QUANTUM MECHANICS 16
2.2.1.1 THE SCHRODINGER EQUATION 16
2.2.1.2 THE BORN-OPPENHEIMER APPROXIMATION 19
2.2.1.3 THE VARIATIONAL PRINCIPLE 19
2.2.1.4 THE HARTREE-FOCK APPROXIMATION 22
2.2.1.5 CHOOSING THE BEST BASIS SET 26
2.2.1.6 DENSITY FUNCTIONAL THEORY 27
2.2.2 MOLECULAR MECHANICS 33
2.2.2.1 CLASSICAL FORCE FIELDS 34
2.2.2.1.1 PARTIAL ATOMIC CHARGES 39
2.2.2.1.2 WATER MODELS 40
2.2.3 HYBRID QUANTUM MECHANICS-MOLECULAR MECHANICS METHODOLOGIES 40
2.2.3.1 FRONTIER BETWEEN SUBSYSTEMS 42
2.2.4 EXPLORING THE POTENTIAL ENERGY SURFACE 43
2.2.4.1 ENERGY MINIMIZATION 44
2.2.4.2 REACTION PATHWAY SEARCH 45
2.2.4.3 MOLECULAR DYNAMICS SIMULATION 46
2.2.4.3.1 PERIODIC BOUNDARY CONDITIONS 51
2.2.4.4 SIMULATED ANNEALING 52
2.2.5 PROTEIN STRUCTURE PREDICTION 52
2.3 METHODOLOGY 54
2.3.1 QM-MM SIMULATION DETAILS 54
2.3.2 PARAMETERIZATION OF COPPER CENTER 56
2.3.3 MM-MD SIMULATIONS DETAILS 57
2.3.4 POST-SIMULATION ANALYSIS 58
2.3.4.1 ROOT-MEAN SQUARE DEVIATIONS 59
2.3.4.2 ROOT-MEAN SQUARE FLUCTUATIONS 59
2.3.4.3 RADIUS OF GYRATION 60
2.3.4.4 RADIAL DISTRIBUTION FUNCTION 60
2.3.4.5 CROSS-CORRELATION ANALYSIS 62
2.3.4.6 ELECTROSTATIC POTENTIAL 63
2.3.4.7 FREE ENERGY CALCULATIONS 64
2.3.4.7.1 POTENTIAL OF MEAN FORCE 65
2.3.4.7.2 FREE ENERGY OF PROTEIN-PROTEIN INTERACTION 66

CHAPTER 3:

ROLE OF CONSERVED RESIDUES IN STRUCTURAL DYNAMICS OF COPPER CHAPERONES 69

3.1 ATOX1 AND COPZ: SIMILAR FOLD YET DISTINCT SEQUENCE 69
3.2 COMPUTATIONAL METHODS 73
3.2.1 QM-MM CALCULATIONS 73
3.2.2 MM-MD SIMULATIONS 74
3.3 STRUCTURE AND DYNAMICS OF COPPER BINDING TO ATOX1 AND COPZ 77
3.3.1 Cu(I) COORDINATION AND NEARBY ENVIRONMENT: QM-MM OPTIMIZATIONS 77
3.3.2 MM-MD SIMULATIONS OF APO- AND HOLO-METALLOCHAPERONES 81
3.3.2.1 BACKBONE FLUCTUATIONS 81
3.3.2.2 CROSS-CORRELATION ANALYSIS 85
3.3.2.3 CYS-CYS DISTANCE DISTRIBUTION IN THE APO-FORMS 87
3.3.2.4 INTERACTIONS AS A FUNCTION OF COPPER COORDINATION 89
3.3.2.4.1 PROTEIN-SOLVENT INTERACTIONS 89
3.3.2.4.2 SPECIFIC ANALYSIS OF CONSERVED RESIDUES 91
3.3.3 BIOLOGICAL RELEVANCE 94
3.4 SEQUENCE DIFFERENCES BETWEEN EUKARYOTIC AND PROKARYOTIC COPPER CHAPERONES 96
3.4.1 DIFFERENTIAL ROLES OF MET10, THR11, AND LYS60 IN ATOX1 STRUCTURAL DYNAMICS 97
3.4.1.1 MET10ALA ATOX1 VARIANT 97
3.4.1.2 THR11ala AND THR11SER ATOX1 VARIANTS 104
3.4.1.3 LYS60ALA AND LYS60TYR ATOX1 VARIANTS 108
3.4.1.4 CROSS-CORRELATION ANALYSIS OF ATOX1 VARIANTS 111
3.4.1.5 BIOLOGICAL RELEVANCE 112
3.4.2 TUNING OF COPPER LOOP FLEXIBILITY IN COPZ: ROLE OF MET11, SER12 AND TYR65 118
3.4.2.1 MET11ALA COPZ VARIANT 118
3.4.2.2 SER12ALA AND SER12THR COPZ VARIANTS 125
3.4.2.3 TYR65ALA AND TYR65LYS COPZ VARIANTS 130
3.4.2.4 CROSS-CORRELATION ANALYSIS OF COPZ VARIANTS 134
3.4.2.5 COMPARISON TO HUMAN COPPER CHAPERONE HOMOLOG 135
3.5 CONCLUSIONS 140

CHAPTER 4:

CONFORMATIONAL DYNAMICS OF ATP7B METAL-BINDING DOMAINS 142

4.1 ARE THE METAL-BINDING DOMAINS EQUIVALENT? 142
4.2 ARE THE METAL-BINDING DOMAINS INDEPENDENT? 145
4.3 COMPUTATIONAL METHODS 146
4.3.1 INITIAL STRUCTURES 146
4.3.2 Cu(I) CENTER PARAMETERIZATION 147
4.3.3 MM-MD SIMULATIONS 148
4.4 DIFFERENCES AMONG ATP7B METAL-BINDING DOMAINS AS POTENTIAL SELECTIVITY FILTERS FOR ATOXI INTERACTION 149
4.4.1 APO-FORMS OF WD1-6 149
4.3.1.1 WD1, WD2 AND WD6: STABLE DOMAINS WITH WELL-DEFINED Cu LOOPS 151
4.4.1.2 WD3: MOST DIVERGENT DOMAIN OF THE SIX WDs 156
4.4.1.3 WD4 AND WD5: FLEXIBLE DOMAINS WITH LARGE CHANGES IN HELIX α1 157
4.4.2 HOLO-FORMS OF WD2, WD4 AND WD6 159
4.4.3 SURFACE ELECTROSTATIC POTENTIAL ANALYSIS OF WDs 165
4.4.4 BIOLOGICAL RELEVANCE 167
4.5 INTER-DOMAIN INTERACTIONS MODULATE COLLECTIVE DYNAMICS OF ATP7B METAL-BINDING DOMAINS 171
4.5.1 PROTEIN DYNAMICS 173
4.5.1.1 WD12 173
4.5.1.2 WD34 180
4.5.1.3 WD56 184
4.5.2 CROSS-CORRELATED MOTIONS 188
4.5.2.1 INTRA-DOMAIN CORRELATIONS 191
4.5.2.2 INTER-DOMAIN CORRELATIONS 192
4.5.2.2.1 WD12 193
4.5.2.2.2 WD34 194
4.5.2.2.3 WD56 196
4.5.3 BIOLOGICAL RELEVANCE 197
4.6 CONCLUSIONS 201

CHAPTER 5:

INTER-PROTEIN INTERACTIONS AND COPPER TRANSFER MECHANISM BETWEEN ATOXI AND ATP7B METAL-BINDING DOMAINS 203

5.1 REQUIREMENT OF PROTEIN-PROTEIN SPECIFICITY 203
5.2 PROTEIN-PROTEIN INTERACTIONS 204
5.3 COPPER TRANSFER MECHANISM 206
5.4 COMPUTATIONAL METHODS 207
5.4.1 INITIAL STRUCTURES 207
5.4.2 MM-MD SIMULATIONS 208
5.4.3 INITIAL STRUCTURE FOR QM-MM CALCULATIONS 209
5.4.4 QM-MM CALCULATIONS 210
5.5 INVESTIGATING THE STABILITY OF ATOXI-WD AND WD-WD HETERO-COMPLEXES 210
5.5.1 CHOOSING THE BEST REACTANT: ATOXI-WD HETERO-COMPLEXES 210
5.5.2 PROBING INTER-DOMAIN INTERACTIONS: WD4-WD5/6 HETERO-COMPLEXES 217
5.5.3 BIOLOGICAL RELEVANCE 218
5.6 ROLE OF CONSERVED RESIDUES IN ATOXI-WD4 INTERACTIONS 222
5.6.2 BIOLOGICAL RELEVANCE 227
5.7 COPPER TRANSFER REACTION PATHWAY SEARCH: FROM HOLO-ATOXI TO WD4 230
5.7.1 FINDING THE FIRST INTERMEDIATE 232
5.7.2 FINDING THE SECOND INTERMEDIATE 239
CHAPTER 6:

OVERVIEW, CONCLUSIONS AND PERSPECTIVES

6.1 Present understanding of copper transport pathways 254
6.2 Residues in and near the copper loop are conserved differently in prokaryotic and eukaryotic copper chaperones to fine tune their flexibility 255
6.3 ATP7B metal-binding domains are unique, interact differently with ATOX1, and “communicate” with each other 258
6.4 The interaction interface between ATOX1 and a target domain is “weak” 262
6.5 Copper transfer from ATOX1 to a target domain appears to be energetically unfavorable but kinetically feasible 263
6.6 Future perspectives 264

REFERENCES 267
LIST OF FIGURES

FIGURE 1.1 7
FIGURE 1.2 9
FIGURE 1.3 13
FIGURE 2.1 35
FIGURE 2.2 61
FIGURE 2.3 66
FIGURE 3.1 70
FIGURE 3.2 79
FIGURE 3.3 81
FIGURE 3.4 82
FIGURE 3.5 83
FIGURE 3.6 86
FIGURE 3.7 88
FIGURE 3.8 89
FIGURE 3.9 90
FIGURE 3.10 93
FIGURE 3.11 99
FIGURE 3.12 101
FIGURE 3.13 103
FIGURE 3.14 104
FIGURE 3.15 106
FIGURE 3.16 109
FIGURE 3.17 118
FIGURE 3.18 120
FIGURE 3.19 121
FIGURE 3.20 123
FIGURE 3.21 124
FIGURE 3.22 127
FIGURE 3.23 131
FIGURE 4.1 143
FIGURE 4.2 150
FIGURE 4.3 153
FIGURE 4.4 154
FIGURE 4.5 155
FIGURE 4.6 160
FIGURE 4.7 162
FIGURE 4.8 163
FIGURE 4.9 164
FIGURE 4.10 166
FIGURE 4.11 173
FIGURE 4.12 174
FIGURE 4.13 176
FIGURE 4.14 177
FIGURE 4.15 179
FIGURE 4.16 182
FIGURE 4.17 186
FIGURE 4.18 189
FIGURE 4.19 190
FIGURE 5.1 212
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 2.1</td>
<td>56</td>
</tr>
<tr>
<td>Table 3.1</td>
<td>76</td>
</tr>
<tr>
<td>Table 3.2</td>
<td>82</td>
</tr>
<tr>
<td>Table 3.3</td>
<td>98</td>
</tr>
<tr>
<td>Table 3.4</td>
<td>113</td>
</tr>
<tr>
<td>Table 3.5</td>
<td>118</td>
</tr>
<tr>
<td>Table 3.6</td>
<td>136</td>
</tr>
<tr>
<td>Table 4.1</td>
<td>149</td>
</tr>
<tr>
<td>Table 4.2</td>
<td>152</td>
</tr>
<tr>
<td>Table 4.3</td>
<td>161</td>
</tr>
<tr>
<td>Table 4.4</td>
<td>172</td>
</tr>
<tr>
<td>Table 4.5</td>
<td>178</td>
</tr>
<tr>
<td>Table 5.1</td>
<td>214</td>
</tr>
<tr>
<td>Table 5.2</td>
<td>226</td>
</tr>
<tr>
<td>Table 5.3</td>
<td>231</td>
</tr>
<tr>
<td>Table 5.4</td>
<td>232</td>
</tr>
<tr>
<td>Table 5.5</td>
<td>234</td>
</tr>
<tr>
<td>Table 5.6</td>
<td>245</td>
</tr>
<tr>
<td>Table 5.7</td>
<td>251</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>AO</td>
<td>atomic orbital</td>
</tr>
<tr>
<td>ATPBD</td>
<td>ATP binding domain</td>
</tr>
<tr>
<td>BCA</td>
<td>bicinchoninic acid</td>
</tr>
<tr>
<td>CD</td>
<td>circular dichroism</td>
</tr>
<tr>
<td>COX</td>
<td>cytochrome c oxidase</td>
</tr>
<tr>
<td>CP</td>
<td>ceruloplasmin</td>
</tr>
<tr>
<td>Cu</td>
<td>copper</td>
</tr>
<tr>
<td>DFT</td>
<td>density functional theory</td>
</tr>
<tr>
<td>DTT</td>
<td>dithiothreitol</td>
</tr>
<tr>
<td>GGA</td>
<td>generalized gradient approximation</td>
</tr>
<tr>
<td>g(r)</td>
<td>radial distribution function</td>
</tr>
<tr>
<td>GSH</td>
<td>glutathione</td>
</tr>
<tr>
<td>HB</td>
<td>hydrogen bond</td>
</tr>
<tr>
<td>HF</td>
<td>Hartree-Fock</td>
</tr>
<tr>
<td>MBD</td>
<td>metal-binding domain from Cu-ATPase</td>
</tr>
<tr>
<td>MK</td>
<td>metal-binding domain from ATP7A</td>
</tr>
<tr>
<td>MM</td>
<td>molecular mechanics</td>
</tr>
<tr>
<td>MO</td>
<td>molecular orbital</td>
</tr>
<tr>
<td>MT</td>
<td>metallothionein</td>
</tr>
<tr>
<td>PB</td>
<td>Poisson-Boltzmann</td>
</tr>
<tr>
<td>PBE</td>
<td>Perdew, Burke and Ernzerhof</td>
</tr>
<tr>
<td>PDB</td>
<td>protein data bank</td>
</tr>
<tr>
<td>PES</td>
<td>potential energy surface</td>
</tr>
<tr>
<td>PMF</td>
<td>potential of mean force</td>
</tr>
<tr>
<td>QM</td>
<td>quantum mechanics</td>
</tr>
<tr>
<td>QM-MM</td>
<td>quantum mechanics-molecular mechanics</td>
</tr>
<tr>
<td>RESP</td>
<td>restrained electrostatic potential</td>
</tr>
<tr>
<td>Rg</td>
<td>radius of gyration</td>
</tr>
<tr>
<td>rmsd</td>
<td>root-mean square deviation</td>
</tr>
<tr>
<td>rmsf</td>
<td>root-mean square fluctuation</td>
</tr>
<tr>
<td>SA</td>
<td>simulated annealing</td>
</tr>
<tr>
<td>SASA</td>
<td>solvent accessible surface area</td>
</tr>
<tr>
<td>SOD1</td>
<td>Cu/Zn superoxide dismutase</td>
</tr>
<tr>
<td>SPLAM</td>
<td>scaled position link atom method</td>
</tr>
<tr>
<td>TGN</td>
<td>trans-Golgi network</td>
</tr>
<tr>
<td>TMD</td>
<td>transmembrane domain</td>
</tr>
<tr>
<td>vdW</td>
<td>van der Waals</td>
</tr>
<tr>
<td>WD</td>
<td>metal-binding domain from ATP7B</td>
</tr>
<tr>
<td>WT</td>
<td>wild type</td>
</tr>
<tr>
<td>XAS</td>
<td>X-ray absorption spectroscopy</td>
</tr>
</tbody>
</table>
"He who knows nothing, loves nothing. He who can do nothing understands nothing. He who understands nothing is worthless. But he who understands also loves, notices, sees... The more knowledge is inherent in a thing, the greater the love... Anyone who imagines that all fruits ripen at the same time as the strawberries knows nothing about grapes."

Paracelsus

“A person who never made a mistake never tried anything new.”

Albert Einstein
Chapter 1:

The requirement of an intricate copper balance

1.1 Proteins as the central cellular machineries

Proteins are essential macromolecular components of the cell, with great structural and functional diversity and complexity. This allows them to participate in most of the vital cellular functions, including catalysis, metabolism, transport, mechanical and structural support. Proteins are linear sequences of amino acids, which fold into a unique three-dimensional structure called the “native state”. We may often think about this “native state” as a single rigid structure (as seen for example in a X-ray crystallographic structure). However, proteins are highly flexible, which significantly increases the available conformational space that they are able to sample. Therefore, the native state is more correctly represented as a collection of different conformations with similar energies called the “conformational ensemble”. In fact, if proteins were rigid molecules, they would not be able to perform most of the aforementioned biological tasks.

The native state of a given protein, including both its unique overall structure and intricate dynamics, dictates its physical-chemical properties and therefore its function. The understanding of the structure-function relationship of proteins, and how their physical-chemical properties are modulated by the surrounding media and/or cellular partners, is one of the major efforts of the biological sciences, in which physics, chemistry and biology merge. In this thesis, we will focus on the physical-chemical properties of a particular group of proteins that bind and transport copper.
1.2 Copper as a common transition metal in biology

If proteins were constrained to employ their natural amino acids alone, they would not be able to catalyze all the reactions required to sustain life (Bertini et al., 2007). Therefore, Nature has exploited the great chemical diversity of inorganic elements, such as Na, Mg, K, Ca, Cl, Mn, Fe, Co, Ni, Cu and Zn, to allow proteins to perform a wider number of biochemical functions. In general, it is commonly believe that 30% of proteins bind metals as cofactors (Waldron et al., 2009), which define them as “metalloproteins”. However, a recent report showed that almost half of the enzymes in the Protein Data Bank (PDB) bind a metal cofactor (Andreini et al., 2008b). Among these, copper (Cu) was found in 93% of oxidoreductases (Andreini et al., 2008b), which are enzymes that catalyze oxidation-reduction reactions (Voet et al., 2002). In a related study, Cu was found to bind to ~1% of the proteome of bacteria, archaea and eukaryotes (Andreini et al., 2008a). Cu is the third most abundant trace element in humans (after Fe and Zn) (Elam et al., 2002). It binds to proteins with a great variety of functions, which include Cu homeostasis (Cu transport and storage; between 38% and 45% of the Cu proteome), catalysis of redox reactions or electron transfer (between 37% and 49% of the Cu proteome), and transport of oxygen (1% of the Cu proteome) (Andreini et al., 2008a; Malkin and Malmstrom, 1970).

To understand why Cu is able to perform such functions, it is useful to recall some fundamental concepts about the chemistry of this element. Cu is one of the most prevalent biological transition metals (Malmstrom and Leckner, 1998), characterized by having partially filled d-orbitals in some of its forms and more than one stable valence state (Malkin and Malmstrom, 1970). Cu has an atomic number of 29 and therefore an
electronic configuration of \( [\text{Ar}] \ 3d^{10} \ 4s^1 \) (one unpaired electron, doublet\(^1\)). In biological systems, Cu (without taking into account polymetallic clusters) can exist in two valence states: Cu(I) (reduced, cuprous) and Cu(II) (oxidized, cupric) (Malkin and Malmstrom, 1970). Cu(II) has two electrons less than Cu, so it has one unpaired electron in its fundamental state (\( [\text{Ar}] \ 3d^9 \), doublet). On the other hand, Cu(I) has one electron less than Cu, and thus can potentially exist in two spin configurations: high spin (triplet, two unpaired electrons, \( [\text{Ar}] \ 3d^9 \ 4s^1 \)) or low spin (singlet, no unpaired electron, \( [\text{Ar}] \ 3d^{10} \)). However, because the singlet constitutes an electronic state in which the \( 3d \) level is filled, it is more stable than the triplet, as we will corroborate in Chapter 2.

Cu(I) and Cu(II) have different ligand preferences. Whereas Cu(I) prefers “soft”\(^2\) sulfur ligation (such as cysteine thiolate and methionine thioeter) with lower coordination number (between two and four), Cu(II) prefers to bind “stronger” ligands, such as nitrogen (histidine imidazole) and oxygen (carboxylate of aspartate and glutamate) with higher coordination number (up to six) (Davis and O’Halloran, 2008; Holm et al., 1996).

In aqueous solution, the predominant copper species is Cu(II), because of the poor solubility of Cu(I) (Malkin and Malmstrom, 1970). However, because of the highly reductive intracellular environment (Schafer and Buettner, 2001), Cu is believed to exist in its reduced form inside cells (Davis and O’Halloran, 2008). Nevertheless, there is

---

\(^1\) Elementary particles have a quantum number of spin \( (s) \) associated with them; for electrons \( s=1/2 \). The total spin \( S \) of atoms and molecules is the sum of the spins of the unpaired electrons; therefore atoms with no, one and two unpaired electrons have \( S=0 \), \( S=1/2 \) and \( S=1 \), respectively. Spin multiplicity \( (M) \) is the number of possible quantum states of a system with a given \( S \), and is equal to \( 2S + 1 \). So, an atom with no unpaired electrons has \( M=1 \) and is a singlet, one unpaired electron has \( M=2 \) and is a doublet, and two unpaired electrons has \( M=3 \) and is a triplet (Levine, 2000).

\(^2\) The hard-soft character of a chemical species depends on the size, oxidation state and polarizability. Hard species are small, with high oxidation state and low polarizability, and the opposite is true for soft species (Andreini et al., 2008b). Therefore, Cu(II) is harder than Cu(I) (Bertini et al., 2007). Whereas harder metals prefer binding harder ligands, softer metals prefer to bind softer ligands.
virtually no "free" (i.e., aqua ions) Cu in the cytosol of living cells because most of it is bound to ligands.

1.3 Copper as a biological yin-yang element

The indispensability of Cu for all living organisms arises from its role in key metabolic pathways as an electron transfer intermediate, and also as a cofactor in a number of different proteins essential for cell survival (Lalioti et al., 2009). Because of its redox cycling ability, Cu is a versatile cofactor for many enzymes involved in a myriad of essential pathways, including cellular respiration (cytochrome c oxidase or COX), free radical defense (Cu/Zn superoxide dismutase or SOD1), cellular iron metabolism (ceruloplasmin or CP), synthesis of connective tissue (lysyl oxidase), pigmentation (tyrosinase), blood clotting (clotting factors V and VIII), angiogenesis (angiogenin), peptide hormone production (peptidylglycine monooxygenase), and synthesis of neurotransmitters (dopamine β-hydroxylase) (Lalioti et al., 2009; Pena et al., 1999).

Although Cu is essential, the free Cu concentration inside cells is virtually non-existent: \(~10^{-21}\) M in *Escherichia coli* (Changela et al., 2003), \(~10^{-18}\) M in yeast (Rae et al., 1999), and \(~10^{-13}\) M in blood plasma (Tapiero et al., 2003). The overcapacity of Cu chelation in the cytosol arises principally from the presence of abundant quantities of Cu scavengers, such as the low molecular weight thiol, glutathione (GSH) (Freedman et al., 1989), and the small Cys-rich protein metallothionein (MT), which bind Cu with high affinity and stoichiometry (MT) (Kagi and Kojima, 1987; Steinebach and Wolterbeek, 1994).
Two main characteristics of Cu may help explain why evolution has limited Cu levels in the cytosol of living cells. First, according to the Irving-Williams series for divalent metals, Cu(II) complexes are the most stable species, whereas Mg(II) and Ca(II) are the least (Irving and Williams, 1948). Thus, Cu(II) has high affinity for protein ligands in general (although it certainly has ligand preferences, as stated above), and therefore, this divalent ion is highly competitive (Waldron and Robinson, 2009). At the same time, Cu(I) is also the most effective monovalent ion for binding to organic ligands (Frausto da Silva and Williams, 2001). If Cu concentrations in the cell were high, most metalloproteins would bind Cu. Therefore, in order for lower affinity metals to bind to their cognate protein sites, free Cu must be limited (Waldron and Robinson, 2009).

Second, in aqueous solution, the Cu(II)/Cu(I) couple has a redox potential of 159 mV (Bard et al., 1985), which is in between the negative redox potential of the cell (Schafer and Buettner, 2001) and the highly positive one of hydrogen peroxide (H₂O₂) (Naskalaski and Bartosz, 2001). Thus, Cu is an ideal candidate for catalysis of reactive oxygen species via Fenton-like reactions (Halliwell and Gutteridge, 1984):

\[
\text{Cu}(I) + \text{H}_2\text{O}_2 \rightarrow \text{Cu}(II) + \text{HO}^+ + \text{HO}^-
\]

The generated hydroxyl radical (HO⁺) is highly reactive and therefore destructive and toxic, causing severe damage to macromolecules, including protein oxidation, lipid peroxidation and DNA cleavage (Halliwell and Gutteridge, 1990; Stohs and Bagchi, 1995). Because intracellular H₂O₂ is always available (Halliwell and Gutteridge, 1984), if intracellular free Cu concentrations were high, Cu(II) ions would constantly be reduced to Cu(I) because of the reducing environment, thus producing a positive feedback of the above reaction. The same redox potential that allows Cu to serve as an inducer of free
radical damage, allows Cu to catalyze free radical detoxification in SOD1 (Uriu-Adams and Keen, 2005). Given the paradox, Cu is both essential and detrimental for cell function, and so we could say that Cu is a biological yin-yang element, and hence the title of this thesis.

Because of this functional duality of Cu, a family of proteins called Cu transport proteins tightly regulates Cu homeostasis and bioavailability. These pathways prevent the existence of free Cu ions and at the same time, guarantee that the metal is delivered to the proper cellular targets.

1.4 Eukaryotic copper transport pathways

In eukaryotes, Cu homeostasis is maintained by a complex cellular machinery, which includes Cu uptake, scavenging, storage, delivery to functional targets, and excretion (Figure 1.1). Cu enters the cells via the highly specific Cu transporter Ctrl, which is conserved from yeast to humans (Dancis et al., 1994a; Zhou and Gitschier, 1997), post reduction to Cu(I) by membrane reductases (Knight et al., 1996). Once inside the cells, Cu is sequestered by GSH, MT and by a family of soluble proteins called Cu chaperones that bind, guide and shuttle Cu to different intracellular destinations (Culotta et al., 1999; Dancis et al., 1994b; Pfahlg et al., 1997). The Cu chaperone for cytochrome c oxidase (COX17) delivers Cu to the mitochondria, where it is incorporated into COX (Horng et al., 2004), whereas the Cu chaperone for superoxide dismutase (CCS) transfers Cu to SOD1 (Rae et al., 1999). The Cu chaperone or Cu transporter that delivers Cu to the nucleus has not yet been identified (Lutsenko et al., 2007a).
Figure 1.1. Intracellular Cu pathways in humans. Cu enters the cell via CTR1, and is then incorporated into different proteins including MT and the Cu chaperones COX17, CCS and ATOX1, for incorporation into COX, SOD1 and ATP7A/B, respectively. The proteins involved in the secretory path highlighted in red are conserved from prokaryotes to eukaryotes, and are the focus of this thesis. Different enzymes acquire Cu in the trans-Golgi network (TGN), such as CP, for subsequent secretion out from the cell.

The third most important destination of Cu is the secretory pathway (Figure 1.1), in which specialized Cu chaperones deliver Cu to soluble metal-binding domains of membrane-bound Cu-ATPases (Hamza et al., 1999; Huffman and O'Halloran, 2000), which couple ATP hydrolysis to Cu transport into the lumen of the trans-Golgi network (TGN) for incorporation into cuproenzymes. In yeast, the antioxidant protein Atx1 (Dancis et al., 1994b) shuttles Cu to Ccc2, which pumps Cu into the lumen of the TGN for incorporation into the ferroxidase Fet3, which is then transported to the plasma membrane (Yuan et al., 1995; Yuan et al., 1997). In the homologous human system, Atox1 (sometimes referred as HAH1) (Klomp et al., 1997) delivers Cu to the Menkes (or ATP7A) and Wilson (or ATP7B) disease proteins, for subsequent Cu incorporation into
hephaestin (Vulpe et al., 1999) or CP (Hellman and Gitlin, 2002) (both Fet3 homologs), which are then transported to the plasma membrane or secreted into the blood, respectively. Because these Cu chaperones and Cu-ATPases are highly conserved from prokaryotes to humans (Arnesano et al., 2002), we focus on the study of this pathway. For example, in the Gram-positive bacterium Bacillus subtilis, the Cu chaperone CopZ delivers Cu(I) to the metal-binding domains of CopA, the corresponding Cu-ATPase in this organism, which is located at the plasma membrane for Cu excretion (Banci et al., 2003a; Radford et al., 2003).

1.5 Copper ATPases

P-type ATPases are a family of more than 200 multi-domain membrane-bound proteins that actively pump cations across membranes in all organisms, such as the well known Na\(^{+}/K\(^{+}\), H\(^{+}\)- or Ca\(^{2+}\)-ATPases (Futai et al., 2004; Lutsenko and Kaplan, 1995). A hallmark of P-type ATPases is the formation of an acyl-phosphate intermediate during the catalytic cycle, by transfer of the γ-phosphate of ATP to an Asp residue in a conserved DKTG motif located in the phosphorylation or P-domain (Lutsenko and Kaplan, 1995; Solioz and Vulpe, 1996) (Figure 1.2). Other soluble domains include the actuator or A-domain and the nucleotide binding or N-domain (Figure 1.2). The P- and N-domain are sometimes collectively known as the ATP binding domain (ATPBD). Metal translocation is driven by ATP hydrolysis and profound conformational changes, with the concerted action of all domains (Kuhlbrandt, 2004).
Figure 1.2. Schematic representation of human Cu-ATPases located in the membrane of the Golgi. ATP7A and ATP7B have eight transmembrane domains (with the CPC sequence as the putative Cu channel) and different cytosolic domains, which include six N-terminal metal-binding domains (MBDs), an actuator or A-domain, a phosphorylation or P-domain and a nucleotide binding or N-domain. The P- and N-domain are sometimes collectively known as the ATP binding domain (ATPBD).

Figure 1.2. Schematic representation of human Cu-ATPases located in the membrane of the Golgi. ATP7A and ATP7B have eight transmembrane domains (with the CPC sequence as the putative Cu channel) and different cytosolic domains, which include six N-terminal metal-binding domains (MBDs), an actuator or A-domain, a phosphorylation or P-domain and a nucleotide binding or N-domain. The P- and N-domain are sometimes collectively known as the ATP binding domain (ATPBD).

Among this family, the P1B- or CPX-type subfamily of ATPases specifically transport heavy metals (Cu, Ag, Zn, Cd, Co, or Pb), and are characterized by the presence of eight transmembrane domains (TMDs), various numbers of soluble metal-binding domains and characteristic sequence motifs (Arguello et al., 2007; Futai et al., 2004; Kuhlbrandt, 2004; Solioz and Vulpe, 1996) (Figure 1.2). A CPX motif is always located in TMD6, in which position X is a Cys residue most of the time, and is proposed to participate in metal coordination during transport, acting as a metal channel (Arguello et al., 2007; Solioz and Vulpe, 1996). In the case of Cu-ATPases, the metal-binding
domains (MBDs) (hereafter, MBDs refers to MBDs from Cu-ATPases only) are located at the N-terminus, and the number varies from one or two in archaea, prokaryotes and yeast, three or four in insects and plants, to six in some mammals and humans (Arnesano et al., 2002; Futai et al., 2004). In humans, the MBDs are termed MKi in ATP7A and WDi in ATP7B, respectively, where \( i \) is the number of the domain 1 to 6, with 1 being the most N-terminal domain.

Although structurally conserved, the function of Cu-ATPases can vary, depending on the hierarchy of the organism and its intrinsic needs. In bacteria, the primary role of Cu-ATPases is Cu detoxification by excreting excess Cu from the cell; in yeast, the primary role of Ccc2 is to translocate Cu into the TGN (Futai et al., 2004). In humans, Cu-ATPases have a dual function depending on the intracellular Cu levels: to deliver Cu to the secretory pathway for mettallation of cuproenzymes at basal conditions (biosynthetic) and to export excess Cu from the cell (homeostatic) at high Cu concentrations (Futai et al., 2004; Lutsenko et al., 2007b).

### 1.6 Copper transport in humans and associated pathologies

In humans, the two Cu-ATPases, ATP7A (~163 kDa) and ATP7B (~157 kDa) (Lutsenko et al., 2008), are structurally and functionally related, with an overall sequence identity of 57% (Harris, 2003). However, ATP7A is expressed in most of the tissues, including the intestine and blood-brain barrier, except for the liver (Vulpe et al., 1993), whereas ATP7B is mainly expressed in the liver, but also in the kidney and placenta, and at lower levels in brain, heart and lungs (Bull et al., 1993; Tanzi et al., 1993).
Most dietary Cu is absorbed in the small intestine or duodenum, via Ctrl located at the apical membrane of enterocytes (Kuo et al., 2006). Cu is then picked up by Atoxl and delivered to ATP7A, which has the dual role of supplying Cu to cuproenzymes, including hephaestin (Vulpe et al., 1999), and exporting Cu to the blood. The latter process requires trafficking of the ATPase to the basolateral membrane (Monty et al., 2005). Cu is then transported in blood by albumin, transcuprein and histidine to the liver, which is the central organ that maintains Cu homeostasis (Lalioti et al., 2009; Pena et al., 1999). In hepatocytes, Cu also enters the cell via Ctrl and is picked up by Atoxl, but now is delivered to ATP7B, which at normal Cu levels delivers Cu to cuproenzymes, in particular to CP, before it is secreted into the blood. Holo-CP accounts for ~90% of the Cu content in the serum (Hellman and Gitlin, 2002). Under high intracellular Cu concentrations, ATP7B exports excess Cu out of the hepatocyte into the bile, a process that requires the trafficking of the ATPase into vesicles in proximity to the canalicular membrane (Schaefer et al., 1999).

In humans, the importance of maintaining Cu within physiological levels, enough to sustain life but not to induce toxicity, is underscored by the many diseases associated with the malfunctioning of Cu homeostasis pathways (Harris, 2003; Tapiero et al., 2003). In particular, the absence or misfunction of ATP7A and ATP7B lead to the severe genetic disorders Menkes and Wilson diseases, respectively, whose clinical manifestations are mainly associated with the different expression patterns of both ATPases. Menkes disease is a rare yet fatal X-linked disorder, with an incidence of 1 in 300,000 live-born babies in European countries (Tonnesen et al., 1991). It is characterized by a generalized Cu deficiency, especially in the brain, due to decreased export of Cu from the enterocyte and
decreased delivery of Cu to metalloenzymes (Kaler, 1998). On the other hand, Wilson disease is a more common autosomal recessive disorder that affects about 1 in 30,000 humans worldwide (Gitlin, 2003). It is characterized by accumulation of Cu primarily in the liver and brain, leading to a severe hepatoneurological disorder and cirrhosis (Gitlin, 2003; Harris, 2003). Over 200 mutations result in Menkes disease (Kodama and Fujisawa, 2009; Lalioti et al., 2009), whereas more than 300 disease causing mutations have been reported for Wilson disease (Kenney and Cox, 2007), and most of them are located in the respective ATPBD (Hsi and Cox, 2004). In Menkes disease, only 17% of the disease-causing variants are missense mutations (Hsi and Cox, 2004). In Wilson disease most of the variants are missense mutations (~60 %) (Hsi and Cox, 2004), and these are located throughout the protein, including WDs 1, 2, 5 and 6, all eight TMDs, the A-, P- and N-domains. Although there are no missense mutations reported for WD3 and WD4, they present several other types of mutations, which are also found in other domains, including nonsense substitutions, deletions, insertions and splice-site variants. Wilson disease mutations are constantly being deposited in public databases.

1.7 Copper chaperones and metal-binding domains share the same fold

The most striking feature shared by Atxl-like Cu chaperones and MBDs from Cu-ATPases is the ferredoxin-like fold with a compact \( \beta1\alpha1\beta2\beta3\alpha2\beta4 \) structure and a conserved metal-binding motif MXCXXC located in the solvent-exposed \( \beta1-\alpha1 \) loop (or Cu loop) (Arnesano et al., 2002) (Figure 1.3). Cu chaperones and MBDs bind one Cu(I) via the two Cys residues of the conserved motif, although other residues located close to

\[ \text{http://www.wilsondisease.med.ualberta.ca/database.asp} \]
\[ \text{http://www.uniprot.org/uniprot/P35670} \]
the Cu loop are believed to provide a unique environment (Anastassopoulou et al., 2004; DiDonato et al., 1997; DiDonato et al., 2000; Lutsenko et al., 1997; Rosenzweig et al., 1999). The surface accessible metal-binding motif appears optimized for facile Cu(I) transfer between the Cu chaperone and the MBDs.

**Figure 1.3.** Representative structures of Cu chaperones and MBDs of Cu-ATPases. Solution structures of holo-Atox1 (left, 1TL4.pdb) and MK1 (right, 1KVJ.pdb) revealing the conserved ferredoxin-like fold with the different structural elements (β1α1β2β3α2β4) labeled, and the Met and two Cys residues from the conserved MXCXXC motif in sticks. The Cu atom, coordinated by the two Cys only, is also shown.

Key questions arise from the conserved topology of these proteins and domains, which we will address in this thesis. Although structurally conserved, Cu chaperones are poorly conserved at the sequence level between prokaryotes and eukaryotes (Arnesano et al., 2002). However, there is significantly greater sequence conservation within each kingdom of life. It is not clear why distinct residues surround the metal-site in prokaryotic and eukaryotic Cu chaperones, which will be one of the subjects of this thesis (Chapter 3). The biological reason for the presence of six structurally similar MBDs in
ATP7A/B and some mammals is still not clear either. MBDs also share little sequence identity, although a higher homology is found between corresponding domains (for example, between the same MBD in ATP7A and ATP7B) (Arnesano et al., 2002). This points to functional diversification of MBDs based on intrinsic molecular differences of each MBD, which will be another subject of this thesis (Chapter 4). Moreover, given the number of MBDs in humans separated by different linkers (Figure 1.2), it is not clear if they behave as independent units, or if specific interactions exist amongst them which may result in cooperative Cu(I) binding and/or regulation of the ATPase. This issue will also be addressed (Chapter 4). Last but not least is the final piece of this puzzle that joins the first two questions: how is Cu(I) actually transferred from the metallochaperone to the MBDs? In Chapter 5, we will address this question by analyzing metallochaperone-MBDs interactions and the Cu-transfer mechanisms from the metallochaperone to a MBD. To address all these questions, we employed a wide range of computational methodologies, described in detail in Chapter 2, which allowed us to obtain more detailed molecular information than that afforded by experimental techniques. Our main system of study will be Atox1 and CopZ Cu chaperones and the six human WDs. We chose to study ATP7B over ATP7A because there is a greater body of in vivo and in vitro data available for the ATP7B protein to compare with our computations. However, the WDs have been more poorly described than the MKs from a structural point of view, which increases the originality of our computer simulations. Also, as opposed to ATP7A, there are several missense mutations in the WDs, which by observing the behavior of the wild type (WT) variant we may potentially be able to predict their function and misfunction in disease-causing variants.
Chapter 2:

Computational Methods

2.1 Introduction to computer simulations

As a complement to experimental physical-chemical techniques, computer simulation methodologies are employed to explore the microscopic behavior of biomolecules. Computational schemes have been recently adopted by the biological community and have been successfully implemented in a wide variety of examples, allowing us to obtain structural, thermodynamic, spectroscopic and dynamic information that in some cases would not be accessible by experimental techniques. At the same time, computer simulations may also serve as predictable tools in cases were experiments are too costly, both timely and economically. This impressive advance in computer simulation techniques applied to the study of complex biochemical systems is due to the increasing advance of efficient and cheap computers and the development of new methodologies and algorithms. This has been evidenced by the 1998 Chemistry Nobel prize awarded to Professors John Pople and Walter Kohn for their fundamental contributions in this field.

Computational chemistry is then a fundamental tool for elucidating protein structure and function and providing microscopic information hardly accessible by other techniques. Needless to say, computer simulations are best interpreted in conjunction with experimental observations, since these are what we ultimately want to explain. Conversely, the advance of many experimental techniques was possible with the simultaneous advance of computational algorithms. A simple example is the determination of a three-dimensional protein structure by X-ray crystallography or NMR
spectroscopy methods, in which the final structure is usually refined with computational methods. Therefore, in the present work, whenever possible, we will connect our computations to available experimental data, or use our computations as prediction tools for future experimental data collection. In the present work, we used computational chemistry and modeling techniques to study a group of essential copper proteins.

In the following sections we will introduce the computational methods used in this thesis. We will first introduce the theoretical foundations, followed by the methodology and the details specifically used in this thesis.

2.2 Theoretical foundations

2.2.1 Quantum mechanics

We will briefly review the principles of quantum mechanics and computational chemistry as described in a standard quantum chemistry and computational chemistry book. In this context, different quantum-mechanical techniques to solve electronic structure problems, such as Hartree-Fock and density functional theory, are explained. Here, we followed four different books: two quantum mechanics books, (Levine, 2000) and (Szabo and Ostlund, 1989), and two computational chemistry/modeling books, (Leach, 2001) and (Cramer, 2004).

2.2.1.1 The Schrödinger equation

To study the microscopic behavior of atomic or molecular systems from first principles we must frame our study within the theory of quantum mechanics (QM), whose fundamental equation is the Schrödinger equation (Levine, 2000). If the potential
is independent of time $V(r,t) = V(r)$ (as will be the case for all problems in this thesis),
we can use the time-independent Schrödinger equation (Levine, 2000):

$$-\frac{\hbar^2}{2m} \nabla^2 \Psi(r) + V(r) \Psi(r) = E \Psi(r) \quad (2.1)$$

where $\hbar$ is the reduced Planck constant and $\nabla^2$ is the Laplacian operator
($\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$). This is a differential equation whose resolution, which only
depends on the spatial coordinates $r$ of the particles, gives us the wave function $\Psi(r)$
and the energy $E$ of the system with mass $m$ (Levine, 2000). The wave function $\Psi(r)$ is
the mathematical description of a quantum particle system, subject to the external
potential $V(r)$, and contains all the information to obtain any system property (Levine,
2000). If we define the Hamiltonian operator $\hat{H}$ as (Levine, 2000):

$$\hat{H} = -\frac{\hbar^2}{2m} \nabla^2 + V(r) \quad (2.2)$$

the Schrödinger equation reduces to:

$$\hat{H} \Psi(r) = E \Psi(r) \quad (2.3)$$

The Schrödinger equation is then a partial differential eigenvalue equation in
which an operator $\hat{H}$ acts on the eigenfunction $\Psi(r)$ and returns the same function
multiplied by a scalar, the eigenvalue $E$ (Levine, 2000). In vectorial (bracket or Dirac)
otation, operators (such as $\hat{H}$) are represented as matrices and eigenfunctions $\Psi(r)$
become eigenvectors $|\Psi\rangle$ (Szabo and Ostlund, 1989). For the case of a system containing
$N$ electrons and $M$ nucleus, the $\hat{H}$ in atomic units\(^4\) is (Szabo and Ostlund, 1989):
\[
\hat{H} = -\sum_{i=1}^{N} \frac{1}{2} \nabla_i^2 - \sum_{k=1}^{M} \frac{1}{2m_k} \nabla_k^2 - \sum_{i=1}^{N} \sum_{k=1}^{M} \frac{Z_k}{r_{ik}} + \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{1}{r_{ij}} + \sum_{k=1}^{M} \sum_{l=k+1}^{M} \frac{Z_kZ_l}{r_{kl}}
\]
(2.4)

where $i$ and $j$ are the electron indexes and $k$ and $l$ are the nuclear ones, $m_k$ is the ratio of the mass of nucleus $k$ to the mass of an electron, $Z$ is the atomic number (number of protons) and $r_{AB}$ is the distance between particles $A$ and $B$. The first two terms are the electronic and nuclear kinetic energy, respectively, the third term is the nuclear-electron attraction, and the fourth and fifth terms are the electron-electron and nuclear-nuclear repulsion, respectively.

By multiplying both sides of the Schrödinger equation by the complex conjugate of the wave function $\Psi^*$ (since it may be a complex number), and integrating both sides over all space, we can obtain the energy of the system:
\[
E = \frac{\int \Psi^*(\mathbf{r}) \hat{H} \Psi(\mathbf{r}) d\mathbf{r}}{\int \Psi^*(\mathbf{r}) \Psi(\mathbf{r}) d\mathbf{r}}
\]
(2.5)

or in Dirac notation:
\[
E = \frac{\langle \Psi | \hat{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle}
\]
(2.6)

The main goal of quantum chemistry is to solve the Schrödinger equation for a system with nuclei and electrons, e.g. a molecule. Unfortunately, there is not an exact solution for these kinds of systems, so we need to find an approximate one.

\(^4\) To simplify the notation, we will use atomic units in the QM section of this thesis. In this convention the charge unit is the electron charge ($e$), the mass unit is the electron mass ($m_e$) and the angular momentum is defined in $\hbar$ multiples. Then, the energy and distance units are, respectively (Szabo and Ostlund, 1989):
\[
1 \text{hartree} = e^2 / 4\pi\varepsilon_0 a_0, \quad 1 \text{bohr} = \hbar^2 / m_e e^2
\]
2.2.1.2 The Born-Oppenheimer approximation

As a first simplification to the problem, we will use the Born-Oppenheimer approximation, which consists in uncoupling the electron and nuclear movements (Szabo and Ostlund, 1989). This approximation makes sense; since the nuclei are much heavier than the electrons, they move more slowly (Szabo and Ostlund, 1989). This means that the electrons are subject to a static potential of punctual nuclear charges (Szabo and Ostlund, 1989) or that they instantaneously respond to any nuclear motion. In this way, we can divide the problem into two: the nuclear problem and the electronic problem. The first problem is trivial because it is constant for a set of fixed nuclear positions, and depends on the kinetic energy of the nuclei and the nuclear-nuclear repulsion (Szabo and Ostlund, 1989). We can now solve the electronic problem for a set of fixed nuclear positions, and so the electronic Hamiltonian reduces to (Szabo and Ostlund, 1989):

\[ \hat{H}_{el} = -\sum_{i=1}^{N} \frac{1}{2} \nabla_i^2 - \sum_{i=1}^{N} \sum_{k=1}^{M} \frac{z_k}{r_{ik}} + \sum_{i=1}^{N} \sum_{j>i}^{N} \frac{1}{r_{ij}} \]  

(2.7)

which has an associated electronic wave function that is a function of the electrons coordinates only and depends parametrically (meaning that they are parameters, constants, not variables) on the nuclear coordinates (Szabo and Ostlund, 1989).

2.2.1.3 The variational principle

The variational principle or theorem is key to finding approximate solutions to an eigenvalue problem, such as the Schrödinger equation (Szabo and Ostlund, 1989). Given a Hamiltonian operator, there is an infinite set of solutions to the Schrödinger equation (Szabo and Ostlund, 1989):
\[ \hat{H} |\Psi_i\rangle = \varepsilon_i |\Psi_i\rangle \quad i = 0,1,2,... \quad \varepsilon_0 \leq \varepsilon_1 \leq \varepsilon_2 \leq ... \]  \hfill (2.8)

where \( \varepsilon_i \) are the true eigenvalues (\( \varepsilon_0 \) is the energy of the ground state, \( \varepsilon_1 \) is the energy of the first excited state, etc.) and \( |\Psi_i\rangle \) are the orthonormal (orthogonal and normal) eigenvectors, meaning that:

\[
\langle \Psi_i | \Psi_j \rangle = \delta_{ij} \quad \delta_{ij} = 0 \quad i \neq j \\
\delta_{ij} = 1 \quad i = j
\]  \hfill (2.9)

Because the set of eigenvectors \( \{ |\Psi_i\rangle \} \) is a complete set of functions, any other function \( |\Phi\rangle \) can be written as a linear combination of \( |\Psi_i\rangle \):

\[
|\Phi\rangle = \sum_i c_i |\Psi_i\rangle \quad \text{and} \quad c_i = \langle \Psi_i | \Phi \rangle
\]  \hfill (2.10)

The variational principle states that when we evaluate the energy with the trial function \( |\Phi\rangle \) (that is different to the true solution to the Schrödinger equation, which is \( |\Psi\rangle \)), we will always obtain a ground state energy that is greater or equal to the true ground state energy \( \varepsilon_0 \) (Szabo and Ostlund, 1989). So, if \( |\Phi\rangle \) is normalized \( \langle \Phi | \Phi \rangle = 1 \), then:

\[
\langle \Phi | \hat{H} | \Phi \rangle \geq \varepsilon_0
\]  \hfill (2.11)

What this means is that we want to find a function \( |\Phi\rangle \) that minimizes the energy of the system \( \langle \Phi | \hat{H} | \Phi \rangle \) (Szabo and Ostlund, 1989). In practice, we construct the trial wave function \( |\Phi\rangle \) as a linear combination of known real functions (Szabo and Ostlund, 1989):

\[
|\Phi\rangle = \sum_i c_i |\phi_i\rangle
\]  \hfill (2.12)
These $N$ known functions $\{\phi_i\}$ are called the basis set. The idea now is to find the set of coefficients $c_i$ that minimize the expected value of the energy, i.e. we set the derivative of the energy with respect to the coefficients equal to zero (Szabo and Ostlund, 1989). Mathematically, the expected value of the energy using the trial function as a linear combination of the basis set is (Szabo and Ostlund, 1989):

$$
\langle \Phi | \hat{H} | \Phi \rangle = \sum_i c_i \langle \phi_i | \hat{H} | \phi_j \rangle c_j = \sum_i c_i c_j H_{ij} 
$$

(2.13)

where $H_{ij} = \langle \phi_i | \hat{H} | \phi_j \rangle$ are the Hamiltonian matrix elements in the $\{|\phi_i\rangle\}$ base. Because $|\Phi\rangle$ is normalized, there is a restriction in the value of the coefficients:

$$
\langle \Phi | \Phi \rangle = \sum_i c_i c_j \langle \phi_i | \phi_j \rangle = \sum_i c_i^2 = 1
$$

(2.14)

In order to minimize the energy in equation (2.13) subject to the constraint imposed in equation (2.14), the Lagrange’s method of undetermined multipliers is used, which gives us a set of $N$ equations with $N$ unknowns (the coefficients) (Szabo and Ostlund, 1989):

$$
\sum_j c_j H_{ij} - Ec_i = 0 \quad i = 1, 2, \ldots, N
$$

(2.15)

This is a standard eigenvalue problem, which is solved by diagonalizing the matrix with elements $H_{ij} - E$. This is achieved by setting the determinant of this matrix equal to zero. The result will give us $N$ eigenvalues $E$, and each value of $E$ will give rise to a different set of coefficients, by solving the set of linear equations (2.15). The lowest eigenvalue is the best possible approximation to the ground state energy $\varepsilon_0$ (Szabo and Ostlund, 1989). With $E_0$ we can now obtain the coefficients, to build the wave function of the ground electronic state $|\Phi_0\rangle$, which is the corresponding eigenvector.
2.2.1.4 The Hartree-Fock approximation

The first effective method for finding an approximate solution to the *Schrödinger* equation for a many-electron problem is the Hartree-Fock (HF) approximation.

We should first build trial wave functions that satisfy certain requisites. If we take into account the spin of the electrons (\(s=1/2\)), we must satisfy the antisymmetric quantum mechanical principle for fermions, which demands that any well behaved wave function must be antisymmetric when interchanging the coordinates of any two electrons (Szabo and Ostlund, 1989):

\[
\Psi(x_1,\ldots,x_j,\ldots,x_N) = -\Psi(x_1,\ldots,x_j,\ldots,x_N)
\]

where \(x_i\) denotes the position and spin of electron \(i\). If we define the spin-orbital \(\chi_i\) as a one-electron wave function, the antisymmetric requirement for the all electron wave function \(\Psi\) is accomplished by using a Slater determinant (Szabo and Ostlund, 1989):

\[
\Psi(x_1,x_2,\ldots,x_N) = \frac{1}{\sqrt{N!}} \left| \begin{array}{ccc} \chi_1(x_1) & \chi_2(x_1) & \cdots & \chi_N(x_1) \\ \chi_1(x_2) & \chi_2(x_2) & \cdots & \chi_N(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(x_N) & \chi_2(x_N) & \cdots & \chi_N(x_N) \end{array} \right|
\]

This determinant involves \(N\) indistinguishable electrons distributed over \(N\) spin orbitals \((\chi_1,\chi_2,\ldots,\chi_N)\), without specifying which electron is in which orbital (Szabo and Ostlund, 1989). Each spin-orbital \(\chi_i(x_i)\) is the product of the spatial function that depends on the coordinates of electron \(i\) \((x_i, y_i, z_i)\) and the spin function that depends on the spin of electron \(i\) (Szabo and Ostlund, 1989). Each spatial orbital can accommodate two electrons with opposite spins (up and down) (Szabo and Ostlund, 1989).

In the HF approximation, using the electronic Hamiltonian as defined in equation (2.7), and choosing a Slater determinant as the electronic wave function in equation
(2.17), the expression of the electronic energy for \( N \) electrons is (Szabo and Ostlund, 1989):

\[
E = \langle \Psi | \hat{H}_{el} | \Psi \rangle = \sum_{i=1}^{N} H_{i}^{\text{core}} + \frac{1}{2} \sum_{i=1}^{N} \sum_{j>i}^{N} (J_{ij} - K_{ij})
\]  

(2.18)

where \( H_{i}^{\text{core}} \) is an integral that depends on the coordinates of only one electron, and \( J_{ij} \) and \( K_{ij} \) are the Coulomb and Exchange integrals, respectively, which depend on the coordinates of two electrons. These are written as a function of the spin-orbitals \( \chi_i \) (Szabo and Ostlund, 1989):

\[
H_{i}^{\text{core}} = -\frac{1}{2} \left[ \langle \chi_i(1) | \nabla_i^2 | \chi_i(1) \rangle - \langle \chi_i(1) | \sum_{\alpha} Z_{\alpha} | \chi_i(1) \rangle \right]
\]

(2.19)

\[
J_{ij} = \langle \chi_i(1) \chi_j(2) | r_{i\alpha}^{-1} | \chi_i(1) \chi_j(2) \rangle
\]

(2.20)

\[
K_{ij} = \langle \chi_i(1) \chi_j(2) | r_{i\alpha}^{-1} | \chi_i(1) \chi_j(2) \rangle
\]

(2.21)

\( H_{i}^{\text{core}} \) is a reduced Hamiltonian of an electron in the orbital \( i \) associated to the electronic kinetic energy and its Coulombic interaction with the nucleus (note that \( H_{i}^{\text{core}} \) makes a favorable, negative, contribution to the energy) (Szabo and Ostlund, 1989). In the absence of electron-electron interactions (i.e., single electron particle), \( H_{i}^{\text{core}} \) would be the only contribution to the energy, corresponding to the motion of a single electron moving in the field of the nuclei. The Coulomb integral \( J_{ij} \) represents the electrostatic repulsion (unfavorable, positive, contribution to the energy) between the electrons in the orbitals \( i \) and \( j \) (Szabo and Ostlund, 1989). The exchange integral \( K_{ij} \) arises from the antisymmetric requirement of the wave function and has no relationship with any quantity from classical mechanics (Szabo and Ostlund, 1989). In order to calculate the
total electronic energy, we sum these integrals for all $N$ electrons, as in equation (2.18). This means that the electronic repulsion is calculated as the sum of pairwise interactions, which can be thought of as a mean field approximation. The basic idea of HF approximation then, is to replace the complicated many-electron problem by a one-electron problem, in which electron-electron repulsion is averaged for all the electrons (Szabo and Ostlund, 1989).

According to the variational principle, we must find the spin-orbitals $\chi_i$ that minimize the energy of the system. If we minimize the expression in equation (2.18) subject to the restriction that the spin-orbitals must be orthonormal, and after performing some mathematical transformations, we obtain an eigenvalue system of equations known as the HF equations (Szabo and Ostlund, 1989):

$$\hat{F}(1)|\chi_i(1)> = \epsilon_i|\chi_i(1)> \quad i = 1,2,\ldots,N$$

(2.22)

where the eigenfunctions are the $N$ spin-orbitals $\chi_i$, also known as molecular or HF orbitals, whose energies are given by the eigenvalues $\epsilon_i$. $\hat{F}$ is the Fock operator, and has three contributions, corresponding to the three contributions to the energy described above in equation (2.18) (core, Coulomb and exchange) (Szabo and Ostlund, 1989). One important consequence of this is that the Fock operator depends explicitly on its eigenfunctions $\chi_i$, which are unknown (Szabo and Ostlund, 1989). This means that in order to solve the eigenvalue equation, we must first know its solution, that is, the resolution must be done in an iterative way (Szabo and Ostlund, 1989). We start by expanding the molecular orbitals (MOs) into a finite basis set of known functions $|\phi_\mu>$:
\[ |\chi_i\rangle = \sum_{\mu=1}^{K} C_{\mu} |\phi_{\mu}\rangle \]  

(2.23)

So now the problem is reduced to finding the values of the coefficients \( C_{\mu} \) (as explained in the previous section). In practice, we guess a starting tentative set of coefficients to build our MOs and thus build the Fock matrix (Szabo and Ostlund, 1989). We then solve equation (2.22) using matrix algebra, similarly as described in section 2.2.1.3, to find a new set of coefficients that are now used to build a new Fock matrix. This procedure is repeated until the coefficients do not differ significantly from the ones found in the step before, that is, when self-consistency has been reached (Szabo and Ostlund, 1989).

The energy and the wave function quality obtained by the HF method are directly related with the size and quality of the basis set functions (Szabo and Ostlund, 1989). A more complete set of basis functions increases the flexibility to expand the MOs, thus obtaining a lower and better value for the energy (Szabo and Ostlund, 1989). The weak point of the HF approximation is in the description of the electronic repulsion, which is done by means of a mean field approximation (i.e., between the one-electron orbitals), thus lacking the electron correlation arising from their instantaneous positions (Levine, 2000; Szabo and Ostlund, 1989). It is reasonable to assume that there will be an exclusion region surrounding each electron, which will lower the probability of encountering other electrons in that same region (Levine, 2000). This is neglected in the mean field approximation, leading to an overestimation of the total energy of the system (Levine, 2000). Therefore, even if we use a complete (i.e., infinite) basis set, the energy obtained by the HF method will always be greater than the exact energy of the system; the difference between them is the correlation energy (Szabo and Ostlund, 1989).
2.2.1.5 Choosing the best basis set

The question is how to choose the functions $|\phi_\mu\rangle$ in equation (2.23), i.e., the known basis set, in order to have a good representation of the MOs of our molecular system. Since we know the exact solution of the Schrödinger equation for the hydrogen atom, we could write our MOs as a linear combination of the hydrogenic atomic orbitals (1s, 2s, 2p, etc), each atomic orbital (AO) centered at the nucleus of each atom (Cramer, 2004). However, it is computationally very expensive to evaluate the integrals involving these types of functions (which are proportional to $e^{-ar}$) (Cramer, 2004). Therefore, we need to find a compromise between the accuracy of the basis set and the computational cost. Generally, the functions that are used for ab initio calculations are Gaussians (which have the form $e^{-ar^2}$) because their analytical integration is well known (Leach, 2001). In practice, more than one Gaussian function is required to correctly represent the hydrogenic AOs, therefore, the latter are written as a linear combination of Gaussian functions (Leach, 2001). The linear combination of Gaussian functions is called a contracted basis function, and each individual Gaussian is called a primitive Gaussian (Cramer, 2004).

The smallest number of basis functions (minimal basis set, also known as single-$\zeta$) for a given molecule is the one that can accommodate all the electrons (filled orbitals) in the molecule (Leach, 2001) (only one basis function per occupied orbital). However, as the number of electrons of an atom increases, a minimal basis set can lead to inaccuracies (Leach, 2001). If we use a large number of basis set functions, this will lead us to a more correct representation of the MOs. One common approach is to double the number of functions in the minimal basis set (double-$\zeta$) to describe the valence electrons (to
increase the flexibility, since these are the electrons most affected by bonding), but to keep a single function (single-\(\zeta\)) for the inner shells (core orbitals) (Leach, 2001). These types of basis sets are called split valence double-\(\zeta\) basis sets (Leach, 2001). For example, in the commonly used 6-31G basis set, one contracted basis function (composed of 6 primitive Gaussians) is used for the core orbitals, and two different basis functions (one contracted function composed of three primitives, and one primitive function) are used for the valence orbitals (Cramer, 2004). Also, another common addition is the incorporation of diffuse functions to the basis set, in order to include polarization, to better describe the electron density away from the nucleus (Leach, 2001). This is usually indicated by an * or a (d). In this thesis, all HF calculations were performed using the 6-31G(d) basis set, using the Gaussian program (Frisch et al., 2004).

### 2.2.1.6 Density functional theory

An alternative formalism to HF was developed by Hohenberg and Kohn in the sixties, by means of two key theorems (Levine, 2000). In this formalism, called Density Functional Theory (DFT), the electronic problem is described by means of the electronic charge density \(\rho\), which is only function of the spatial coordinates \(\rho(x,y,z) = \rho(r)\), and is related to a real physical observable (Levine, 2000). In this way, we avoid dealing with the wave function, which is a function of the coordinates of \(N\) electrons. In the context of DFT, electrons interact between each other and with an external potential \(V(r_i)\), which is the energy of the interaction between the electron \(i\) and the nucleus, in atomic units (Levine, 2000):
The first Hohenberg and Kohn theorem establishes that the wave function, and therefore the energy and any electronic property of the system, are uniquely determined by the electronic density (Levine, 2000). They also demonstrated that the electronic density of the ground state \( \rho_0(r) \) determines the external potential \( V(r) \) and the number of electrons, and thus the Hamiltonian (Levine, 2000). Therefore, the energy can be thought as a functional\(^5\) of the density (Levine, 2000):

\[
E_0 = E_V[\rho_0(r)]
\]  

(2.25)

where the \( V \) subscript indicates that \( E_0 \) depends on \( V(r) \). The different contributions to the total energy can be written as (Levine, 2000):

\[
E[\rho] = T[\rho] + V_{ee}[\rho] + V_{ne}[\rho] 
\]  

(2.26)

where the functionals correspond to the kinetic energy \( T \), and the potential energy arising from electron-electron interaction \( V_{ee} \) and nucleus-electron interaction \( V_{ne} \).

The second Hohenberg and Kohn theorem is called variational theorem (Levine, 2000), and is analogous to the variational principle of MO theory explained in section 2.2.1.3. The theorem establishes that for any particular well-behaved candidate density \( \rho' \), \( i.e., \) that satisfies the requisites \( \int \rho'(r) dr = N \) (the density integrated over all space should give the total number of electrons \( N \)) and \( \rho'(r) \geq 0 \) (the density should be positive in all space), then (Levine, 2000):

\[
E[\rho'] \geq E_0
\]  

(2.27)

---

\(^5\) A functional is a function that operates on another function to give a scalar.
where \( E_0 \) is the true ground state energy. The equality holds when \( \rho' \) is the ground state density \( \rho_0 \) (Levine, 2000). As before, in which the aim was to find the trial function that minimizes the energy, we now search for the \( \rho' \) that minimizes the energy functional \( E[\rho'] \) (Levine, 2000). The problem is then to find adequate expressions for the functionals \( T, V_{se} \) and \( V_{ne} \).

Kohn and Sham made a great contribution to DFT by considering a fictitious reference system (s sub index) composed of non-interacting electrons, whose density \( \rho_s \) is equal to the true one \( \rho_0 \) (Levine, 2000). By restricting \( \rho_s \) to be equal to \( \rho_0 \), we ensure that the external potential \( V_s(r_l) \), in which the reference system is subjected, defines the true density of the system (Levine, 2000). For this reference system, in which there is no electron-electron interaction, the Hamiltonian \( \hat{H}_s \) can be written as the sum of one-electron Hamiltonians called the Kohn-Sham operators \( h_i^{KS} \) (Levine, 2000):

\[
\hat{H}_s = \sum_{i=1}^{N} \left( -\frac{1}{2} V^2_i + V_s(r_l) \right) = \sum_{i=1}^{N} h_i^{KS}
\]  

(2.28)

We can now express the wave function of the reference system \( \Psi_s \) as a Slater determinant as defined in equation (2.17) of the spin-orbitals of the reference system \( \chi_i^{KS} \), which are called the Kohn-Sham orbitals (Levine, 2000). Note that as opposed to the HF approximation, it is strictly correct to express the non-interacting electron wave function as a Slater determinant of one-electron spin-orbitals (Levine, 2000). Then, because the probability of finding an electron is equal to the module of the square of the wave function, the electronic density is given by (Levine, 2000):
\[ \rho_s = \sum_{i=1}^{N} \langle \chi_i^{KS} | \chi_i^{KS} \rangle = \sum_{i=1}^{N} |\chi_i^{KS}|^2 \]  

(2.29)

The advantage of using a non-interacting electrons system is that the kinetic energy functional is then calculated by adding the individual kinetic energy of all the electrons (Levine, 2000):

\[ T_s = -\frac{1}{2} \langle \Psi_s | \sum_i \nabla_i^2 | \Psi_i \rangle = -\frac{1}{2} \sum_i \langle \chi_i^{KS}(1) | \nabla_i^2 | \chi_i^{KS}(1) \rangle \]  

(2.30)

Using this kinetic energy functional, we can define the term \( \Delta T \) as the functional that allows calculating the difference in the kinetic energy between the true system and the reference system (Levine, 2000):

\[ \Delta T[\rho] = T[\rho] - T_s[\rho] \]  

(2.31)

On the other hand, we can make a classical approximation of the electron-electron interaction, and define \( \Delta V_{ee} \) as the functional that allows calculating the difference in the electron-electron interaction energy between the true system \( V_{ee} \) and the system with the classical electron-electron interaction energy (Levine, 2000):

\[ \Delta V_{ee}[\rho] = V_{ee} - \frac{1}{2} \int \int \frac{\rho(r_1)\rho(r_2)}{r_{12}} dr_1 dr_2 \]  

(2.32)

\( \Delta V_{ee} \) can then be thought of as the non-classical contribution to the electron-electron interaction. If we now group both terms together, we can define a new functional \( E_{xc}[\rho] \) called the exchange-correlation functional (Levine, 2000):

\[ E_{xc}[\rho] = \Delta T[\rho] + \Delta V_{ee}[\rho] \]  

(2.33)

The last term of the functional that we need to define is the functional \( V_{ne} \), the electron-nuclei attraction, which given the density and the potential \( V(r) \) defined in equation (2.24), is defined as (Levine, 2000):
Using all the expressions for the functionals, the final formula for the energy functional is (Levine, 2000):

\[
E[\rho] = T_s[\rho] + V_{\text{ne}}[\rho] + \frac{1}{2} \int \frac{\rho(r_1)\rho(r_2)}{r_{12}} \, dr_1 dr_2 + E_{\text{xc}}[\rho]
\]  

(2.35)

With this expression, we can calculate the energy for all the terms, except for \( E_{\text{xc}}[\rho] \), for which we do not know the functional form (Levine, 2000).

As opposed to HF, which is an approximate theory, DFT involves a functional that is exact by definition, in which the exchange and correlation effects due to the quantum nature of the electrons are described by \( E_{\text{xc}}[\rho] \) (Cramer, 2004). However, the explicit dependence of \( E_{\text{xc}} \) with the density is unknown and then, the accuracy of DFT calculations will be related to the approximation used for this functional (Levine, 2000).

Because the density can be calculated from the Kohn-Sham orbitals from equation (2.29), if we know the orbitals and the exchange-correlation functional, we can calculate the energy (Levine, 2000). The problem then is how to find these orbitals and how to estimate an accurate functional. Using the variational theorem of Hohenberg and Kohn (i.e. we set the derivative of the energy with respect to the density equal to zero), with the restriction of orthonormality for the orbitals, we can find the equations that must fulfill the orbitals that minimize the energy of the system (Levine, 2000). The result, analogous to the HF equations, is a set of eigenvalue equations known as the Kohn-Sham equations (Levine, 2000). Similarly to HF, we expand the Kohn-Sham orbitals as a linear combination of a known basis set functions, and using matrix algebra we can obtain the coefficients (Levine, 2000). As in HF, the procedure must be done in an iterative way,
until self-consistency is reached, because the Kohn-Sham matrix \( h^{KS} \) depends on the density, and the density on the coefficients (Levine, 2000).

The accuracy of DFT calculations is limited by the quality of the exchange-correlation functional (Levine, 2000). The first approximation for the exchange-correlation functional is called Local Density Approximation or LDA (Levine, 2000). These functionals are useful when the density changes slowly with position, because it is based on a homogeneous and uniform electron gas model (Levine, 2000). However, this model is not accurate for describing large changes in the density and therefore, these functionals have the tendency to overestimate bond energies (Levine, 2000).

To improve the quality of the functional, new functional forms that incorporate an explicit dependency with the density gradient have been proposed, called Generalized Gradient Approximation (GGA) functionals (Levine, 2000). Among the most popular functionals we can mention the Perdew, Burke and Ernzerhof (PBE) exchange-correlation functional (Perdew et al., 1996), which is the one employed in this thesis.

2.2.1.6.1 The SIESTA method

Most of the electronic structure calculations performed in this thesis were done with a very efficient implementation of DFT using the program SIESTA (Spanish Initiative for the Electronic Structure of Thousands of Atoms) (Sánchez-Portal et al., 1997; Soler et al., 2002). One advantage of this code is that the nuclei and inner electrons are represented by atomic pseudopotentials (Sánchez-Portal et al., 1997). This strategy avoids the computation of core states, so only the interactions between the valence electrons are explicitly calculated, speeding up computations (Sánchez-Portal et al.,
Another benefit of SIESTA is the use of numerical basis sets defined in a three-dimensional and finite grid, which allows the calculations to be solved numerically (Artacho et al., 1999). An advantage of using numerical basis set functions is that the basis set is not constrained to any analytical functional form and can therefore adopt any form, increasing flexibility (Sánchez-Portal et al., 1997). The quality of the numerical basis set functions is controlled by setting the AO energy shift parameter (Soler et al., 2002). This parameter is kept fixed for all AOs, leading to a well-balanced basis (Soler et al., 2002). For typical medium size molecules, values close to 100 meV give a similar accuracy as the one obtained with standard Gaussian basis set calculations (Artacho et al., 1999). For transition metals complexes containing, for example, Cu, it is recommended to set the energy shift to lower values such as 20-30 meV (Crespo et al., 2003). The use of a finite grid also speeds up the calculations, since the AOs basis set (and therefore all the electronic properties, including MOs and \(p\)) are exactly zero outside the grid. The grid precision is defined by choosing the energy of the higher frequency plane-wave that can fit the grid (Soler et al., 2002). This energy is called grid cut-off energy and values between 150 and 200 Rydbergs ensure a precise grid (Crespo et al., 2003).

2.2.2 Molecular mechanics

Now, we will briefly review the principles of classical mechanics and computational modeling. In this case, we followed two books: (Leach, 2001) and (Cramer, 2004).

Because a medium-size protein has generally thousands of atoms, the treatment of these systems at the QM level is essentially impossible from a computational point of
view. Therefore, the physical-chemical study of proteins by computational techniques is generally done by classical mechanical methods. These classical models allow the efficient description of systems of thousands of atoms, such as a protein in water. The main limitation is that these models cannot provide a description of properties that depend upon the electronic distribution in a molecule, since they lack the ability to describe reactive processes that involve bond breaking and forming, or treat different electronic and spin states (Leach, 2001).

2.2.2.1 Classical force fields

Classical methods (also known as molecular mechanics or MM) ignore the quantum nature of electrons, and electrons themselves, and calculate the energy of a system as a function of the nuclear positions only (Leach, 2001). The potential energy function, which depends on the nuclear positions only, is called the force field, because the energy gradient allows the calculation of the forces that act over each of the atoms. Similar to ab initio methods, the Born-Oppenheimer approximation is implicitly part of classical models. This approximation allows us to define the potential energy surface (PES), which is the potential energy as a function of the nuclear coordinates (Leach, 2001).

In classical force fields, the interactions between the nuclei are described by simple equations of relatively easy evaluation, and are divided into two terms: bonded and non-bonded (Leach, 2001). The bonded terms describe the interaction between atoms that are directly bonded or separated by a distance of up to two atoms (Leach, 2001). In
this way, the bonded energy is decomposed into contributions from bond stretching, angle bending and bond rotation or torsion (Figure 2.1).

![Diagram](image)

**Figure 2.1.** Scheme of the bonded interactions in a classical force field.

The non-bonded terms, on the other hand, have an electrostatic contribution and an additional contribution arising from van der Waals (vdW) interactions (Leach, 2001). Presently, there are several available classical force fields such as CHARM, GROMOS and AMBER, which have a similar energy functional form, but differ in the parameterization (Leach, 2001). In this thesis, we have used the AMBER force field parameterization.

In the AMBER force field, the potential energy, which is a function of the positions of the $N$ atoms, is written as (Cornell et al., 1995):

$$E(r^N) = \sum_{\text{bonds}} K_b (b - b_0)^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_0)^2$$

$$+ \sum_{\text{torsions}} \frac{V_a}{\gamma} \left[1 + \cos(n\phi - \gamma)\right] + \sum_{i,j=1}^{N} \sum_{j=1}^{N} \left[ \frac{A_{ij}}{R_{ij}^6} - \frac{B_{ij}}{R_{ij}^{12}} + \frac{q_i q_j}{\varepsilon R_{ij}} \right]$$

(2.36)

It is easy to recognize the aforementioned interactions in this energy function:

1) **Bond stretching** (first term in equation 2.36): The potential energy curve for a typical bond is modeled as a harmonic oscillator using the Hooke's law (Leach, 2001). The parameter $b_0$ is the equilibrium bond distance of the particular bond considered, and is the value of the bond when all other terms in the force field are zero (Leach, 2001). The
parameter $K_b$ represents the force constant of the harmonic oscillator, related to the energy required to cause a bond to deviate from its equilibrium value (Leach, 2001). Typical $K_b$ values vary from 100 to 500 kcal/(mol Å²) (Cornell et al., 1995). Even though the potential energy curve for true bond stretching is not harmonic, this approximation is reasonably good, taking into account that in a MM calculation the bonds do not deviate significantly from their equilibrium values (Leach, 2001). It is clear that this type of potential does not allow bond breaking or forming.

2) **Angle bending** (second term in equation 2.36): The deviation of an angle between two bonds from its equilibrium value is also described with a harmonic potential, in which $\theta_0$ is the equilibrium angle and $K_\theta$ is the force constant (Leach, 2001). Typically, angles are more susceptible to distortion than bonds, so the force constants are smaller: 30-80 kcal/(mol rad²) (Cornell et al., 1995).

3) **Torsions** (third term in equation 2.36): The torsion energy of dihedral and improper angles describes the potential energy along the rotational axe defined by a bond (Leach, 2001). For example, **Figure 2.1** shows the dihedral angle defined for the atoms 1 and 4, along the axe of the bond between atoms 2 and 3. In the potential energy form, $\phi$ is the value of the angle between the planes that form atoms 123 and 234 in **Figure 2.1**. The parameter $V_n$ is related to the barrier high between two adjacent minima; $n$ is the multiplicity, which gives the number of minima as the angle is rotated from 0 to 360 degrees; and $\gamma$ is the phase, which determines where the angle passes through its first minimum (Leach, 2001).

4) **Non-bonded interactions** (last term in equation 2.36): The non-bonded interactions are only calculated between atoms in different molecules or atoms within the same
molecule but separated by at least three bonds (Leach, 2001). Those non-bonded interactions separated by exactly three bonds (1,4 interactions) are reduced by the application of a scale factor (Leach, 2001). The non-bonded interactions are through-space interactions, and therefore they have a functional dependence that is inversely proportional to the distance that separates the atoms involved (Leach, 2001). These terms are usually divided into two contributions: electrostatic and vdw interactions.

The electrostatic interaction arises from the asymmetric distribution of charge density in a molecule. In the AMBER force field, the electrostatic potential of each molecule is represented by an arrangement of fractional point charges $q_i$ centered in each atom of the molecule (Leach, 2001). These fractional charges are selected to reproduce the electrostatic potential generated by the molecule over its environment. After the point charges are assigned to all the atoms, the pairwise interactions are calculated using Coulomb's law (Leach, 2001).

If we consider the interaction between two atoms of opposite charge, and we only take into account the electrostatic contribution, these atoms would collapse to the same point in space leading to an infinite value of the energy. Moreover, how do we explain the interactions between the atoms in a rare gas, in which all the multipole moments are zero (Leach, 2001)? When the atoms are at an infinite distance, the potential energy is zero (Leach, 2001). As the atoms come closer in space, the energy decreases (attractive or dispersive forces) until it reaches a minimum point (equilibrium distance) (Leach, 2001). If the atoms come closer than the equilibrium distance, the energy rapidly increases (repulsive forces) (Leach, 2001). These types of interactions, with long-range attractive forces and short-range repulsive forces, are known as vdw interactions, and are
typically described by the 12-6 Lennard-Jones potential (Leach, 2001). The 12-6 Lennard-Jones potential contains two adjustable parameters for each pair of interacting atoms: the separation at which the energy passes through a minimum \( R_{ij}^* \) and the value of the energy at that point \( \varepsilon_{ij} \) (vdW well depth) (Leach, 2001). These parameters are tabulated for each atom type, so, for the interaction between atoms \( i \) and \( j \), they are calculated according to the following rules: \( R_{ij}^* = (R_i^* + R_j^*) \) and \( \varepsilon_{ij} = \sqrt{\varepsilon_i \varepsilon_j} \) (Leach, 2001). Then, the values for \( A_{ij} \) and \( B_{ij} \) in equation (2.36) are calculated according to \( A_{ij} = \varepsilon_{ij}(R_{ij}^*)^{12} \) and \( B_{ij} = 2\varepsilon_{ij}(R_{ij}^*)^6 \).

The non-bonded term is the most time-consuming part of a MM simulation, because in the ideal case it involves computing the interactions of \( N(N-1)/2 \) pair of atoms, whereas the bonded terms are proportional to the number of atoms (\( N \)) (Leach, 2001). To reduce the computational cost, the non-bonded interactions are only computed between pair of atoms that are within a certain distance (chosen by the user), which we call the non-bonded cutoff (Leach, 2001). Usually, the non-bonded cutoff is chosen to be close to 10 Å (Leach, 2001). The use of a cutoff, further from which all pairwise non-bonded interactions are set to zero, introduces a discontinuity in both the potential energy and the force near the cutoff value (Leach, 2001). To overcome this discontinuity, it is common to use a switching function, which assures that the non-bonded potential smoothly reaches zero at the cutoff (Leach, 2001). In this way, two instead of one cutoff values are used: the lower and upper cutoff distances. Non-bonded interactions are computed as usual until the lower cutoff is reached (Leach, 2001). Between the lower and upper cutoff, the potential is scaled by the switching function, which gradually takes the
potential to zero at the upper limit (Leach, 2001). This type of functions ensures that the first derivative of the energy is zero at the upper cutoff, so the forces approach zero smoothly (Leach, 2001). In this thesis, all calculations were performed using a switching function where the lower and upper cutoff distances were 8 and 10 Å, respectively.

Although some force fields have an explicit term that accounts for hydrogen bond (HB) interactions, many force fields, including AMBER, rely upon the electrostatic and vDW interactions to describe hydrogen bonding (Leach, 2001).

2.2.2.1 Partial atomic charges

We have seen in DFT that the charge distribution (i.e., the density) determines the wave function of a system, and therefore, all molecular properties. Often, however, we would like to take a more classical approach and ignore the wave character of electrons to calculate what we call partial atomic charges. This is particularly useful for force field methods, in which the electrostatic interaction is calculated between atom-centered charge-charge interactions (Cramer, 2004). There are many methods that allow calculation of partial atomic charges (Cramer, 2004). Restrained electrostatic potential (RESP) (Bayly et al., 1993) is the most popular method employed to calculate atomic charges for a force field parameterization (Cramer, 2004), as done for example in AMBER (Cornell et al., 1995). Therefore, we used this method to calculate partial charges to parameterize the Cu site.
2.2.2.1.2 Water models

Since macromolecules are not isolated entities and are typically immersed in a liquid environment, often we want to study a protein in solvent, with explicit water molecules. Despite the small size and relatively simple chemistry of an isolated water molecule, much effort has been placed in developing accurate explicit water models that reproduce the properties of bulk water (Leach, 2001). Because the number of water molecules required to properly solvate a protein of medium size is generally large, the number of solvent atoms is much larger than the number of protein atoms. For example, for a protein of 70 residues and ~1,000 atoms, ~3,700 water molecules (11,100 atoms) are required to solvate the protein with a typical cell extending 10 Å from the surface of the protein. Therefore, most of the computing time in a MM calculation is devoted to modeling the solvent. This means that we need to find a compromise between the accuracy of the model and the computational cost of its evaluation. In this thesis, we have used the commonly adopted TIP3P water model (Jorgensen et al., 1983), in which the geometry of the water molecule is fixed and the interaction between water molecules is described using Coulombic (three partial charges centered on each atom) and Lennard-Jones (only one interaction between pair of molecules, centered in the oxygen atom) expressions.

2.2.3 Hybrid quantum mechanics-molecular mechanics methodologies

So far, we have briefly explained the theoretical foundations of QM and MM methods. The complexity and precision of QM methods allow the study of reactive processes, by describing in detail the electronic structure of the system. However, due to
the high computational cost, it is virtually impossible to treat systems of more than 100 atoms at the QM level. On the other hand, MM methods provide a simple description of systems with thousands of atoms, such as proteins. However, they are unable to describe reactive processes in which bond breaking and forming occur. What should we use if we want to study a reactive process within a system of thousands of atoms, such as a chemical reaction in the active site of a protein? The answer is to use a method that combines the benefits from both theories, what is called hybrid or QM-MM methods (Cramer, 2004; Warshel and Levitt, 1976). Hybrid QM-MM schemes are adequate for the investigation of reactions taking place in complex environments, such as proteins or condensed phases (Cramer, 2004; Warshel and Levitt, 1976). The method combines an electronic structure description of the solute or protein active site (QM subsystem) with a less expensive molecular mechanical treatment of the environment, such as the protein and/or solvent (MM subsystem).

The total Hamiltonian, which operates over the whole system wave function, and the total energy of the system have three contributions (Leach, 2001):

$$E_{TOTAL} = E_{QM} + E_{MM} + E_{QM-MM}$$  \hspace{1cm} (2.37)

In this thesis, the quantum contribution is calculated at the DFT level, so that $E_{QM}$ is given by equation (2.35). Only the quantum atoms (i.e., QM subsystem) are treated at this level of theory, although still considering the electrons immersed in an external potential generated by both the quantum nuclei and also the partial charges of the MM atoms. The classical contribution ($E_{MM}$) is calculated with an appropriate force field, such as the one presented in equation (2.36), without any modifications. The critical term then is the
quantum-classical coupling $E_{QM-MM}$, which is composed of three contributions (Crespo et al., 2003):

$$E_{QM-MM} = \sum_{i=1}^{C} q_i \int \rho(\mathbf{r}) \frac{d\mathbf{r}}{|\mathbf{r} - \mathbf{\tau}_i|} dr + \sum_{i=1}^{C} \sum_{\alpha=1}^{Q} \frac{q_i Z_\alpha}{|\mathbf{R}_\alpha - \mathbf{\tau}_i|} + E_{LM}^{LJ}$$ (2.38)

where the $i$ and $\alpha$ indexes run for the classical and quantum atoms, respectively; $\mathbf{\tau}$, $\mathbf{R}$ and $\mathbf{r}$ represent the classical atoms coordinates, quantum nuclei coordinates and electron coordinates, respectively; $C$ is the number of classical atoms with partial charges $q_i$; $Q$ is the number of nuclei inside the quantum subsystem with core charges $Z_\alpha$ (Crespo et al., 2003). The first term represents the electrostatic interaction between the quantum electronic density and the partial charges of the classical atoms (Crespo et al., 2003). The second term represents the electrostatic interaction between the nuclei in the quantum subsystem and the classical partial charges (Crespo et al., 2003). The last term, $E_{LM}^{LJ}$, represents the vdw interaction between the atoms in the quantum and classical regions, and is modeled through a 6-12 Lennard-Jones potential, as in classical force fields (Crespo et al., 2003).

2.2.3.1 Frontier between subsystems

In the QM-MM study of chemical reactions in solution, there is typically not a problem with the partition of the system, because in general all the solute atoms are included in the QM subsystem and the solvent is included in the MM subsystem. In proteins, however, often certain amino acids in the active site participate actively in the chemical reaction, so they must be included in the QM subsystem. However, if we do not want to include the entire protein at the QM level, two atoms covalently bonded should
be considered in different subsystems (Cramer, 2004). This means that the frontier
between both QM and MM subsystems involves the breaking of a covalent bond.
Different methods exist to treat this QM-MM frontier (Cramer, 2004).

In this thesis, we have employed the scaled position link atom method (SPLAM)
(Eichinger et al., 1999) adapted to the program SIESTA. In this method, the bonds
between a carbon atom of the QM subsystem ($C_{QM}$) and a carbon of the MM subsystem
($C_{MM}$) are replaced with a carbon-hydrogen bond (Crespo et al., 2003). This hydrogen
link atom ($H_{link}$) fills the valence of the QM subsystem, and its position is superimposed
on the $C_{QM}$-$C_{MM}$ bond (Crespo et al., 2003). Forces exerted on the $H_{link}$ are added to the
$C_{MM}$ and $C_{QM}$ (Crespo et al., 2003). All classical bonded terms (bonds, angles, and
dihedrals) involving the $C_{QM}$-$C_{MM}$ bond are considered with the classical force field
(Crespo et al., 2003). Lennard-Jones interactions between the $C_{QM}$ and $H_{link}$ atoms and
classical atoms separated by less than three bonds from the $C_{QM}$-$C_{MM}$ bond are omitted
(Crespo et al., 2003).

2.2.4 Exploring the potential energy surface

In this section we will present four methods that, using different approximations,
allow obtaining properties of interest for a particular molecular system. The way in which
the energy varies with the coordinates is usually referred to as the PES (Leach, 2001). If
we are able to know the complete PES of a system, we will be able to obtain all
thermodynamic and kinetic properties. However, for a system with $N$ atoms the energy is
a function of $3N-6$ internal coordinates, so it is virtually impossible to visualize the entire
PES from a computational point of view (Leach, 2001). Therefore, alternative methods
that explore the PES have been adopted, which allow obtaining thermodynamic and kinetic parameters in a computationally more efficient way.

2.2.4.1 Energy minimization

Typically, the most relevant spatial conformations or points on the PES are those that have the least energy. These minima points correspond to stable states or conformations of the system, those in which it is more probable to find the system and therefore, that are more representative of it. These points are stationary points on the PES, meaning that the first derivative is zero with respect to all coordinates (Leach, 2001). There may be a very large number of local minima on the PES; the one with lowest energy is known as the global energy minimum (Leach, 2001).

To identify the conformations at these minima on the PES, we use a minimization algorithm (Leach, 2001). These algorithms allow the exploration of the PES until they find a point (i.e., a minimum), in which any movement away from that point in any direction increases the energy of the system. The basic idea is that, given a function (e.g., the energy function) that depends on a number of independent variables (e.g., the coordinates of the \(N\) atoms), the function will reach a minimum point when its first and second derivatives with respect to each of the variables are zero and positive, respectively (Leach, 2001). As we may predict, finding the minimum of a multidimensional function, such as the PES, is a complex mathematical problem and different algorithms exist to tackle this problem. Among the most commonly used minimizations algorithms employed by computational chemistry programs are the steepest descend and conjugate gradient methods (Leach, 2001), the latter being the one implemented within the SIESTA
and AMBER codes. In general, the minimization methods are good for finding minima that are close (in coordinate space) to the starting point conformation (Leach, 2001).

2.2.4.2 Reaction pathway search

In chemical and biochemical problems, we are not only interested in the thermodynamic properties of a system, but also in the kinetic properties associated with reactive processes or conformational exploration. Whereas thermodynamic properties are related to energy minima and the relative energy between these minima, the kinetics of a process is determined by the barrier height (activation energy) that separates the relevant minima. For example, the minima could be the reactants and products of a given chemical reaction, or simply two alternative conformations of a protein. So, we may be interested in how the system changes from one minimum to the other. The lowest-energy path that separates these two minima is known as reaction pathway (Leach, 2001). As the system moves along the reaction pathway, the energy increases until it reaches a maximum. This point, which corresponds to an energy maximum of that particular reaction pathway but it is a minimum in all other perpendicular directions, is known as the saddle point and corresponds to the transition state structure or conformation (Leach, 2001).

Frequently, the reaction pathway can be described as a simple coordinate, for example the distance between two atoms or the angle formed by three atoms. The combination of atomic coordinates that allow the description of a given process is called the reaction coordinate (Leach, 2001). In such cases we can obtain an approximation to the reaction pathway called restrained minimization (Crespo et al., 2003). In this
approximation, the system is pushed "up-hill" in the PES along the given reaction coordinate and at the same time we perform a minimization of the rest of the conformational space. In other words, we gradually change the reaction coordinate allowing the system to relax at each step while keeping the chosen reaction coordinate fixed. To perform this type of calculation, an additional term is added to the potential energy (Crespo et al., 2003):

$$V_K = k(\xi - \xi_0)^2$$  \hspace{1cm} (2.39)

where $k$ is an adjustable force constant to force the system "up-hill" in the PES, $\xi$ is the value of the reaction coordinate of the system, and $\xi_0$ is the reference value of the reaction coordinate. The force constants employed in this thesis for this type of calculations were the corresponding force constants employed in the classical force field parameterization. By changing $\xi_0$ along the desirable reaction coordinate from reactants to products and performing energy minimizations at each step, the system is forced to follow the minimum reaction path along the given coordinate, and in this way, the reaction path is mapped out by adding the $V_R$ to the potential energy. The point of highest energy on the path is an approximation of the transition state structure (Leach, 2001).

### 2.2.4.3 Molecular dynamics simulation

In complex systems such as solvated proteins, because of the large number of atoms and degrees of freedom involved, the PES is very complex with many local minima. Computer simulation methods enable the study of such systems by generating representative configurations of the system in a way that accurate thermodynamics and
structural values can be obtained (Leach, 2001). One of the most widely used simulation
technique is the Molecular Dynamics (MD) method.

MD calculates the “real” temporal progression (dynamics) of the system (Leach,
2001), providing a detailed picture of the way in which it changes from one conformation
to another. In this way, MD simulations can provide detailed information on the
fluctuations and conformational ensemble of proteins, from which thermodynamic
properties can be derived. The atomic positions of the atoms in the system are propagated
in time by integration of Newton’s equations of motion (Leach, 2001). The initial atomic
positions can be obtained from an experimentally determined structure (NMR or X-ray)
or from a structural model. At each step of MD, the energy and the forces acting over
each atom are computed, and together with the current positions and velocities, the new
positions a short time ahead can be calculated (Leach, 2001). The forces are assumed to
be constant during this time interval (Leach, 2001). The atoms are then moved to these
new positions, the energy and forces are computed again, and the sequence continue
(Leach, 2001). In this way, a MD simulation generates a ‘trajectory’ (set of snapshots)
that describes the time-dependent behavior of the system, i.e., how the dynamic variables
change with time (Leach, 2001).

In a MD simulation, the trajectory is obtained by solving Newton’s second law:

\[- \frac{dE_{x_i}}{dx_i} = F_{x_i} = m_i a_i = m_i \frac{d^2 x_i}{dt^2}\]  \hspace{1cm} (2.40)

where \(F_{x_i}\) is the force on the particle \(i\) with mass \(m_i\) that is moving along the coordinate
\(x_i\), and \(a_i\) is the acceleration of the particle \(i\). In the QM or MM interaction potentials,
the force on each atom changes when the atoms itself changes position or any other
interacting atom changes position (Leach, 2001). Under the influence of such a continuous potential the motions of all the atoms are coupled together, giving rise to a many-body problem that cannot be solved analytically (Leach, 2001). Therefore, the equations of motion are integrated using a finite difference method (i.e., they are solved numerically), in which the integration is broken down into many small stages, each separated by a fixed time or time step $\delta t$ (Leach, 2001). The idea is the following: the total force on each atom is calculated as the negative gradient of the potential energy function (derivative of the energy with respect to the atomic coordinates). Using Newton’s second law, from the force we can determine the accelerations, which are combined with the positions and velocities to calculate them at a time ahead $t + \delta t$ (Leach, 2001). The forces in their new positions are then determined again, leading to new positions and velocities at time $t + 2\delta t$, and so on (Leach, 2001). There are many algorithms for integrating numerically Newton’s equations, but they all assume that the positions, velocities and accelerations can be approximated as Taylor series expansions (Leach, 2001). The Verlet algorithm is one of the most widely used method for integrating the equations of motion in a MD simulation (Leach, 2001), and is the method used for all MD simulations in this thesis, as it is included in the AMBER package. It uses the positions and accelerations at time $t$, and the positions from the previous step, $r(t - \delta t)$, to calculate the new positions at $t + \delta t$, $r(t + \delta t)$ (Leach, 2001):

$$r(t + \delta t) = r(t) + \delta t v(t) + \frac{1}{2} \delta t^2 a(t) + ...$$ (2.41)

$$r(t - \delta t) = r(t) - \delta t v(t) + \frac{1}{2} \delta t^2 a(t) - ...$$ (2.42)
where $v$ is the velocity (the first derivative of the positions $r$ with respect to time) and $a$ is the acceleration (the second derivative of the positions $r$ with respect to time). Adding these two equations gives:

$$r(t + \delta t) = 2r(t) - r(t - \delta t) + \delta t^2 a(t)$$

(2.43)

The velocities do not explicitly appear in the Verlet integration algorithm, however they can be calculated by dividing the difference in positions at times $t - \delta t$ and $t + \delta t$ by $2\delta t$ (Leach, 2001):

$$v(t) = \frac{r(t + \delta t) - r(t - \delta t)}{2\delta t}$$

(2.44)

How do we pick an appropriate time step? This value should be sufficiently small to be able to describe in an accurate way the natural motion of the system, as if integration of Newton’s equations was exact. However, as we decrease the time step, more calculations (and more computer time) would have to be performed to simulate a given total time. In general, a good compromise is to use a time step that is one-tenth the characteristic time of the shortest period (highest frequency) of motion of the system (Leach, 2001). In molecules, the highest-frequency vibrations correspond to bond stretches, especially those including hydrogen atoms, which vibrate with a period of approximately 10 fs (Leach, 2001). The requirement that the time step is approximately one order of magnitude smaller than the shortest motion is clearly a severe restriction (Leach, 2001). One solution to this problem is to “freeze” these high-frequency vibrations by constraining all bonds involving a hydrogen atom to their equilibrium values (Leach, 2001), by applying for example, as in this thesis, the SHAKE algorithm (Ryckaert et al., 1977). This enables a longer time step to be used, which in this thesis was of 2 fs in all cases.
In any MD simulation, in which the system is allowed to explore its phase space (multidimensional space), although each individual snapshot has different microscopic properties (i.e., the conformation or flexibility of the entire protein or even of some particular sidechains may vary), the collection of these different microscopic states altogether form a particular macroscopic state, which we call thermodynamic ensemble. These connections between microscopic and macroscopic states (framed within statistical mechanics) allow the calculation of time-average of properties that we can relate to the ones obtained experimentally (ensemble averages). There are several different types of thermodynamic ensembles. In biology, most of the processes occur at a constant pressure and temperature. Therefore, in this thesis, all MD simulations were performed in the isobaric-isothermal ensemble \(NPT\), in which the number of atoms \((N)\), pressure \((P)\) and temperature \((T)\) are maintained fixed, being the volume \((V)\) allowed to vary.

The number of atoms is always fixed in a MD simulation, but how can we fix the temperature and the pressure? The temperature of a system is the macroscopic manifestation of a microscopic property, the average kinetic energy \(K\) of the \(N\) particles (Leach, 2001):

\[
K = \frac{1}{2} \sum_{i=1}^{N} \frac{p_i^2}{2m_i} = \frac{k_B T}{2} (3N - N_c)
\]

(2.45)

where \(p_i = m_i v_i\) is the momentum of atom \(i\), \(k_B\) is the Boltzmann constant, \(N_c\) is the number of constraints and \(3N - N_c\) is the total number of degrees of freedom of the system. One of the most widely used methods to maintain the temperature fixed is the Berendsen thermostat algorithm (Berendsen et al., 1984), which is also the one used in this thesis. This algorithm couples the system to an external bath that is fixed at the
desired temperature and acts as a source of thermal energy, supplying or removing heat from the system as appropriate (Leach, 2001). The velocities (and thus the kinetic energy) are scaled at each MD step, such that the rate of change of temperature is proportional to the difference in the temperature between the bath and the system (Leach, 2001).

Whereas a macroscopic system maintains a constant temperature by changing the kinetic energy of the particles, the pressure is maintained by changing the volume. Therefore, in this thesis we used a modified version of the Berendsen algorithm, which couples the system to a pressure bath and scales the volume of the simulation cell appropriately (Leach, 2001).

2.2.4.3.1 Periodic boundary conditions

As stated before, when we simulate a protein in explicit solvent, much of the computing time is devoted to modeling the typically thousands of solvent atoms. We have already seen that one way of decreasing the computational cost is by using simpler explicit water models. We could potentially decrease the cost of modeling the solvent by decreasing the number of water molecules used to solvate our protein. However, this would introduce a discontinuity in the system, since there would be a distance, further from which the protein would be in vacuum. Moreover, experiments are usually performed in bulk water. Therefore, one way of getting around this problem is to simulate our system using periodic boundary conditions (Leach, 2001). In this way, the atoms experience forces as if they were in bulk solvent (Leach, 2001). The basic idea is that the simulating cell is replicated in all directions to give a periodic array (Leach, 2001). If a
given particle leaves the central cell, it is replaced by an image particle in the opposite side of the cell, maintaining a constant number of atoms (Leach, 2001). One of the most widely used cells in MD simulations is the truncated octahedron box (Leach, 2001), which is also the one employed for all simulations in this thesis.

2.2.4.4 Simulated annealing

Simulated annealing (SA) is another way of exploring a complex PES with many different minima, and it is usually employed in conformational analysis to find the global minimum in the PES (Leach, 2001). In a MD simulation at constant temperature, the temperature serves as an entropic resource for the system to explore multiple minima points in the PES and thus multiple conformations. However, we may sometimes be interested in the global minimum on the PES. We have stated before that minimization algorithms are only good at finding minima close to the starting point, introducing a bias. In a SA calculation, the system is first allowed to reach thermal equilibrium using a MD simulation, after which the temperature of the system is slowly decreased to zero (Leach, 2001). Ideally, if we decrease the temperature in an infinite number of temperature steps, the system will be able to reach thermal equilibrium in each step and therefore, it will reach the global minimum at 0 K (Leach, 2001). This is an impossible task, so SA cannot guarantee to find the global minimum (Leach, 2001). However, if we use a relatively large number of steps, we increase the probability of finding it.

2.2.5 Protein structure prediction

In order to simulate a given protein, the computer modeler needs a starting point
structure, *i.e.*, a three dimensional structure with the coordinates of the $N$ atoms. In the ideal case, we would like to use structures experimentally determined by NMR spectroscopy or X-ray crystallography. Despite the enormous number of protein structures currently and continuously being deposited in the PDB, many times there is no available experimentally determined structure for a protein of interest. Predicting the three dimensional structure of a protein from the linear sequence of amino acids is a very difficult and still an open problem in biology. However, an alternative and very useful approach is to estimate approximately the three-dimensional structure of a protein, if we know that this particular protein shares the same fold as another protein whose structure was experimentally determined. Comparative modeling is a method in which a single or multiple sequence alignment is performed between structurally related proteins to create a three-dimensional model of our protein of interest (Leach, 2001). The protein of interest with unknown structure is the target and the other proteins with known structures are the templates.

As we can imagine, these type of algorithms based solely on the sequence alignment are only accurate when the target and the template share significant sequence identity, typically 70 % (Leach, 2001). However, many times, two proteins with very little sequence identity can fold into the same structure. In such cases in which sequence identity is not high enough (for example 30 %), we may still be able to predict a model of interest by homology threading (Leach, 2001). In this method, the model is built by aligning or "threading" (hence the name) each amino acid in the target protein to a position in a given template structure (Leach, 2001). The program creates multiple models of the target and, using a scoring function, evaluates how well the target fits the
template (Leach, 2001). The user then picks the model or models with lowest energy. In this thesis, we have used the program Modeller (Eswar et al., 2007; Fiser et al., 2000; Marti-Renom et al., 2000; Sali and Blundell, 1993) to perform such calculations.

2.3 Methodology

2.3.1 QM-MM simulation details

Throughout this thesis, QM-MM calculations were performed in different contexts: i) to explore the effect of the environment on Cu(I) binding to Atox1 and CopZ (Chapter 3); ii) to parameterize the Cu(I) center in all holo-protein forms (Atox1, CopZ, WD2, WD4 and WD6) (Chapters 3-5); and iii) to investigate the reaction mechanism of Cu(I) transfer from Atox1 to WD4 (Chapter 5). All QM-MM calculations were performed with a QM-MM implementation (Crespo et al., 2003; Crespo et al., 2005; Crespo et al., 2006) of the SIESTA code (Soler et al., 2002) in which the QM subsystem is treated at the DFT level and the MM subsystem is treated using the AMBER force field. The QM-MM SIESTA implementation has shown excellent performance for medium and large systems and has also proved to be appropriate for studying chemical reactivity in biomolecules, in particular for Cu coordinated systems (Crespo et al., 2006).

Unless indicated otherwise in each Chapter, in all cases the Cu(I) atom plus the coordinated side chains (methylthiolate group: CH$_3$S$^-$) of the two functional Cys residues were selected as the QM subsystem, which comprises 11 atoms. Only the Cys atoms were included in the QM subsystem since it has been reported that these are the intrinsic Cu(I)-coordinating residues (Anastassopoulou et al., 2004; Banci et al., 2001; Dalosto, 2007). The rest of the protein, counter ions and water molecules (MM subsystem) were treated
classically using the Amber99 force field parameterization (Wang et al., 2000) for protein and counter ions atoms and TIP3P (Jorgensen et al., 1983) for water molecules. We allowed free motion for all QM atoms and for some MM atoms, as follows. We computed the center of mass of the QM subsystem and its distance to the furthest QM atom \(d_{\text{furth}}\). Only MM atoms inside the sphere of radius \(d_{\text{furth}} + 12\ \text{Å}\) from the QM subsystem center of mass were allowed to move. The frontier between the QM and MM portions of the system (Cys C\(_\alpha\)-C\(_\beta\) bond) was treated by the SPLAM method (Eichinger et al., 1999).

For all QM atoms, basis sets of double-\(\zeta\) plus polarization quality were employed, with an AO energy shift of 25 meV and a grid cutoff of 150 Ry (Soler et al., 2002). All calculations were performed with the spin-unrestricted approximation (open-shell scheme), in which the energy of two electrons in the same MO but with different spin is allowed to be different (in a close-shell scheme, the spin-orbitals of two electrons with different spin are degenerate) (Szabo and Ostlund, 1989). All calculations were performed using the GGA-PBE functional (Perdew et al., 1996). This combination of functional, basis sets, and grid parameters were validated for the isolated (in vacuum) model systems \([\text{Cu(I)}(\text{CH}_3\text{S}^\delta)_2]^{-1}\), \([\text{Cu(I)}(\text{CH}_3\text{S}^\delta)_3]^{2-}\) and \([\text{Cu(I)}(\text{CH}_3\text{S}^\delta)_4]^{3-}\). Relevant geometrical and energetic parameters are reported in Table 2.1. To perform any QM calculation, one “only” needs to set the overall charge and the total spin of the molecule. Because the three model systems favor the low-spin configuration (see \(\Delta E_{HS:LS}\) values in Table 2.1), and the Cu(I) electronic configuration is \([\text{Ar}]\ 3d^{10}\) with \(\Delta E_{HS:LS}\) of 49.21 kcal/mol, all QM-MM calculations were performed in the low-spin (singlet) state.
Table 2.1. Structural parameters (distances in Å and angles in degrees) of \([\text{Cu}(I)(\text{CH}_3\text{S}^-)_2]^+\), \([\text{Cu}(I)(\text{CH}_3\text{S})_3]^2^-\) and \([\text{Cu}(I)(\text{CH}_3\text{S})_4]^3^-\) isolated model systems in their low-spin configuration. The energy difference (\(\Delta E_{\text{HS-LS}}, \text{in kcal/mol}\)) between high-spin (triplet) and low-spin (singlet) configurations for each model system is also shown. Ideally a perfect linear, trigonal planar and tetrahedral geometry would have S-Cu-S angles of 180, 120 and 109 degrees, respectively. When more than one value is available for the model, the average and standard deviation are reported.

<table>
<thead>
<tr>
<th></th>
<th>([\text{Cu}(I)(\text{CH}_3\text{S}^-)_2]^+)</th>
<th>([\text{Cu}(I)(\text{CH}_3\text{S})_3]^2^-)</th>
<th>([\text{Cu}(I)(\text{CH}_3\text{S})_4]^3^-)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cu-S</td>
<td>2.162 ± 0.001</td>
<td>2.29 ± 0.02</td>
<td>2.49 ± 0.03</td>
</tr>
<tr>
<td>S-Cu-S</td>
<td>178.029</td>
<td>120 ± 6</td>
<td>109 ± 2</td>
</tr>
<tr>
<td>(\Delta E_{\text{HS-LS}})</td>
<td>60.99</td>
<td>41.64</td>
<td>51.87</td>
</tr>
</tbody>
</table>

For all QM-MM calculations, we performed energy minimizations (optimizations) using a conjugate gradient algorithm, and with this we obtained the optimized protein structure with its particular Cu(I) center geometry and the corresponding energy.

### 2.3.2 Parameterization of copper center

As opposed to standard residues, there are not parameters available for a Cu(I) atom and coordinated Cys residues in the AMBER force field. Therefore, these atoms should be parameterized by the user, using the same algorithms with which the AMBER force field was parameterized, for consistency. To perform MD simulations of all holo-proteins, the Cu(I) centers were parameterized using the QM-MM optimized geometries. Although it is common, and simpler, to parameterize using the isolated model systems, we had to use another approach. Because the S-Cu-S geometry of some of the holo-proteins deviated significantly from that of the isolated model system, a parameterization based on this model system geometry would yield unrealistic results. Therefore, to include the effect of the protein environment, the geometrical parameters were obtained.
using the QM-MM optimized geometries. In all cases, the atomic charges of Cu(I) coordinated to two \( \text{CH}_3\text{S}^- \) groups (using the geometry from the QM-MM optimized structures) were determined using RESP (Bayly et al., 1993) and HF/6-31G(d) single-point wave functions with Gaussian 03, revision D.01 (Frisch et al., 2004), following the protocol recommended in the Amber web page\(^6\). These RESP charges calculated over the QM atoms were combined with the charges of an Amber classical cysteinate residue (CYM) modified to maintain an overall integer charge of -1.

### 2.3.3 MM-MD simulations details

Generally, we can divide a MD simulation into four stages: 1) generation of the initial structure, 2) equilibration phase, 3) production phase and 4) post-simulation analysis (Leach, 2001). Because in a MM-MD simulation atoms are described with models, an equilibration phase is first performed to allow the system to reach equilibrium from the initial configuration. Several parameters can be monitored during this phase to assess whether the system reached equilibrium or not (Leach, 2001). In this thesis, we monitored the root-mean-square deviation with respect to the initial structure as a function of time. After the equilibration phase, a production phase is run, and is this final phase the one used for calculating the system properties (post-simulation analysis).

All MM-MD simulations were performed using Amber9 (Case et al., 2005; Case et al., 2006; Cornell et al., 1995; Pearlman et al., 1995), starting from the corresponding initial structures indicated in each Chapter. The initial structures were immersed in a pre-equilibrated truncated octahedral box of TIP3P explicit water molecules (Jorgensen et al.,

---

\(^6\) http://ambermd.org
1983) and counter ions (Na\(^+\) or Cl\(^-\) as needed) were added to neutralize the systems (Case et al., 2006). The size of the box varied for different proteins, and is indicated in each Chapter. Parameters of the coordinating Cys and Cu(I) in the holo-forms are indicated in each Chapter. The rest of the protein atoms were described with the parm99SB force field parameterization (Hornak et al., 2006). The protonation state of the titratable residues corresponds to the stable form at pH 7. In all apo-systems, the relevant Cys residues were simulated in their protonated physiological form.

The initial solvated structures were optimized with conjugate gradient and equilibrated for 200 ps at 300 K in the NVT ensemble. The resulting structures were the starting points of the MD simulations in the NPT ensemble, employing periodic boundary conditions. Constant pressure of 1 atm and temperature of 300 K were maintained using the Berendsen coupling scheme (Berendsen et al., 1984). A SHAKE algorithm was employed to keep bonds involving hydrogen atoms at their equilibrium length (Ryckaert et al., 1977), which allowed us to employ a 2 fs time step for the integration of Newton's equations. The particle-mesh Ewald method was used for evaluating long-range electrostatic interactions (Darden et al., 1993).

### 2.3.4 Post-simulation analysis

Before computing any property, the trajectory is processed using the ptraj module of Amber9: first, a centering of the center of mass of the protein to the box origin is performed to discard any translational motions, followed by imaging of the solvent due to periodic boundary conditions (which brings solvent molecules outside the primary unit cell back into it), and finally a root-mean square fitting with respect to the initial frame is
performed to discard any rotational motions.

2.3.4.1 Root-mean square deviations

The root-mean square deviation (rmsd) with respect to a reference structure as a function of time is:

\[
\text{rmsd}(t) = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left[ x_i(t) - x_i(\text{ref}) \right]^2 + \left[ y_i(t) - y_i(\text{ref}) \right]^2 + \left[ z_i(t) - z_i(\text{ref}) \right]^2}
\]

(2.46)

where the \( i \) index runs for the number of atoms \( N \); and \( x_i, y_i, \) and \( z_i \) are the Cartesian coordinates of atom \( i \). The rmsd (in Å) defined as above is a measure of how much the coordinates of a corresponding set of atoms in a given snapshot deviate from a reference snapshot. In all cases, the rmsd as a function of time was calculated for the backbone heavy atoms (N, Ca, C, O) using the ptraj module of Amber9.

2.3.4.2 Root-mean square fluctuations

The root-mean square fluctuation (rmsf) per residue with respect to the average structure, for an MD simulation of \( M \) frames is:

\[
\text{rmsf} = \frac{1}{M} \sum_{k=1}^{M} \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left( x_{ki} - \langle x_i \rangle \right)^2 + \left( y_{ki} - \langle y_i \rangle \right)^2 + \left( z_{ki} - \langle z_i \rangle \right)^2}
\]

(2.47)

where the \( i \) and \( k \) indexes run for the number of atoms \( N \) in the residue and the number of frames \( M \), respectively; \( x_{ki}, y_{ki}, \) and \( z_{ki} \) are the Cartesian coordinates of atom \( i \) in the frame \( k \); and \( \langle x_i \rangle, \langle y_i \rangle \) and \( \langle z_i \rangle \) are the Cartesian coordinates of atom \( i \) in the average structure. The rmsf (in Å) defined as above is a measure of how much the coordinates of a residue deviate from the average structure, or in another words, fluctuate in the MD
simulation. In all cases, the rmsf as a function of residue was calculated for the backbone heavy atoms (N, Cα, C, O) using the ptraj module of Amber9.

2.3.4.3 Radius of gyration

The radius of gyration (Rg) as a function of time is:

\[
R_g(t) = \frac{1}{N} \sum_{i=1}^{N} [x_i(t) - g_{ci}(t)]^2 + [y_i(t) - g_{ci}(t)]^2 + [z_i(t) - g_{ci}(t)]^2
\]  

(2.48)

where the \( i \) index runs for the number of atoms \( N \); \( x_i, y_i, \) and \( z_i \) are the Cartesian coordinates of atom \( i \); and the geometric center coordinates are defined as:

\[
g_{ci}(t) = \frac{1}{N} \sum_{i=1}^{N} \xi_i(t) \quad \xi = x, y, z
\]  

(2.49)

The Rg (in Å) defined as above is a measure of how distant the coordinates of a set of atoms are from the geometric center of the protein in a given snapshot, or in another words, the size of the protein. In all cases, the Rg as a function of time was calculated for the backbone heavy atoms (N, Cα, C, O) using the ptraj module of Amber9.

2.3.4.4 Radial distribution function

In this thesis we will use the radial distribution function as a useful way to visualize the solvation of a given atom, in other words, to estimate the average number of water molecules that surround the atom within a certain distance. If we consider a spherical shell of thickness \( dr \) at a distance \( r \) from a chosen atom (Figure 2.2), the volume \( V \) of the shell is (Leach, 2001):

\[
V = \frac{4}{3} \pi (r + dr)^3 - \frac{4}{3} \pi r^3 = 4\pi r^2 dr + 4\pi r dr^2 + \frac{4}{3} \pi dr^3 \approx 4\pi r^2 dr
\]  

(2.50)
and the last equality holds if $dr$ is very small. If the number of water molecules per unit volume (density) is $\rho$, then the total number of water molecules in the shell is $4\pi \rho r^2 dr$.

Figure 2.2. Scheme of a spherical shell of thickness $dr$ at a distance $r$ from a given atom.

The radial distribution function $g(r)$ between a given atom and the solvent is then calculated as:

$$
g(r) = \frac{\sum_{k=1}^{M} \#W_k[r,r+dr]}{M \pi \rho r^2 dr}$$

(2.51)

where the $k$ index runs for the number of frames $M$, and $\#W_k[r,r+dr]$ is the number of water molecules in the spherical shell between $r$ and $r+dr$ for frame $k$. In this way, $g(r)$ gives the probability of finding a water molecule at a distance $r$ from a given atom. At very short distances (less than the sum of the vdw radii of the two atoms), the probability is zero, and at long distances, the probability is independent of $r$ and equal to 1 (Cramer, 2004). If $g(r) > 1$ and $g(r) < 1$ it is more and less probable, respectively, to find a water molecule than in bulk water (Cramer, 2004). In all cases, the $g(r)$ was calculated between an atom of a given residue and the oxygen atom of water, with a maximum $r$ equal to 15 Å and $dr$ equal to 0.2 Å, using the ptraj module of Amber9.
2.3.4.5 Cross-correlation analysis

Knowing how the atomic motion of proteins are cross-correlated between each other is useful for understanding long-range communication and large domain movements relevant to protein function (Lin et al., 2008). The covariance $c_{ij}$ between the atoms $i$ and $j$ is defined as the product of the displacement vectors $\Delta r_i$ and $\Delta r_j$ (Ichiye and Karplus, 1991):

$$c_{ij} = \frac{1}{M} \sum_{k=1}^{M} [r_{ik} - \langle r_i \rangle] [r_{jk} - \langle r_j \rangle] = \frac{1}{M} \sum_{k=1}^{M} (\Delta r_{ik} \cdot \Delta r_{jk}) = \langle \Delta r_i \cdot \Delta r_j \rangle$$  \hspace{1cm} (2.52)

where the $k$ index runs for the number of frames $M$; $r_{ik} - \langle r_i \rangle$ indicates how the position (or coordinates $x$, $y$, $z$) of atom $i$ in frame $k$ deviates from its average position; and the brackets in $\langle \Delta r_i \cdot \Delta r_j \rangle$ indicate an ensemble average computed for the $M$ frames of the MD simulation. The cross-correlation or normalized covariance is calculated by dividing the covariance by the mean-square atomic fluctuations of each atom (Ichiye and Karplus, 1991):

$$C_{ij} = \frac{c_{ij}}{\sqrt{c_{ii}c_{jj}}} = \frac{\langle \Delta r_i \cdot \Delta r_j \rangle}{\sqrt{\langle \Delta r_i^2 \rangle \langle \Delta r_j^2 \rangle}} \quad -1.0 \leq C_{ij} \leq 1.0$$ \hspace{1cm} (2.53)

For completely correlated and anticorrelated motions, $C_{ij} = 1$ and $C_{ij} = -1$, respectively.

The cross-correlated matrix is then built by calculating the cross-correlation coefficients $C_{ij}$ between all pair of $\alpha$ atoms of the $R$ residues:

$$\begin{pmatrix}
C_{11} & C_{12} & \cdots & C_{1R} \\
C_{21} & C_{22} & \cdots & C_{2R} \\
\vdots & \vdots & \ddots & \vdots \\
C_{R1} & C_{R2} & \cdots & C_{RR}
\end{pmatrix}$$ \hspace{1cm} (2.54)
Note that this matrix is symmetrical, because $C_{ij} = C_{ji}$, and also all the elements of the diagonal are equal to 1, because they represent the correlation of the atom with itself $C_{ii}$.

In this thesis, the cross-correlations matrix for the Cα atoms was calculated using the ptraj module of Amber9, and all matrix plots were generated with Matlab version 7.2.

2.3.4.6 Electrostatic potential

The electrostatic potential at a point is defined as the work done to bring unit positive charge from infinity to the point (Leach, 2001). One of the most popular methods to estimate the electrostatic potential of proteins in a continuum implicit solvent is based upon the Poisson-Boltzmann (PB) equation (Leach, 2001). In this method, instead of representing the charge distribution of the solvent explicitly, we represent it by a continuous electric field (Cramer, 2004). In the regions of space occupied by the protein (the solute), this continuous electric field is usually called the reaction field, since it describes the reaction of the solvent to the presence of the solute (Cramer, 2004). The PB equation describes the variation in the electrostatic potential as a function of the charge density of the solute, and the dielectric constant and ionic strength of the solvent (Cramer, 2004; Leach, 2001). The charge distribution of the solute is thought of as being a cavity that displaces a homogenous dielectric medium, the solvent (Cramer, 2004). For certain ideal cavities, the PB equation can be solved analytically (Cramer, 2004). However, this is not true in most of the cases, including proteins, in which the PB equation is thus solved numerically (Cramer, 2004). In these cases, the space (containing the protein) is divided by a three-dimensional grid, in which the charge density, dielectric constant and ionic strength are defined at each point (Cramer, 2004; Leach, 2001). These values are
then used to calculate the electrostatic potential at each point of the grid; and because the potential at each point is influenced by the potential at the next point, the calculation is done iteratively (Cramer, 2004; Leach, 2001). Inside the cavity, the dielectric is small (typically between 1 and 4), and outside the cavity the dielectric is high (typically 80 for water) (Cramer, 2004; Leach, 2001). In this thesis, we used the Adaptive PB Solver (APBS) program (Baker et al., 2001), as implemented in the Visual Molecular Dynamics (VMD) program (Humphrey et al., 1996), to calculate the electrostatic potential of Atox1 and WD1-6, in Chapter 4. In all calculations, we used the default parameters, in which the ionic strength is 150 mM and the dielectric constant inside and outside the cavity is 1 and 78.54, respectively.

2.3.4.7 Free energy calculations

The free energy is usually expressed as the Helmholtz function $A$, or the Gibbs function $G$ (Leach, 2001). The Helmholtz free energy is defined within the canonical ensemble ($NVT$), whereas the Gibbs free energy is defined within the isobaric-isothermal ensemble ($NPT$) (Leach, 2001). However, as long as the number of atoms $N$ is constant, all simulations assume that $d(PV)$ is zero, in which case both free energies functions are identical$^7$ (Cramer, 2004).

The free energy is very difficult to estimate for systems such as flexible proteins, in which the PES is composed of many different minimum energy configurations separated by low energy barriers (Leach, 2001). The reason behind this is that in a standard MD simulation, many low probability regions (high energy conformations) are

\[ G = A + PV \rightarrow dG = dA + d(PV) \]
never sampled, and these correspond to regions of phase space that make important contributions to the free energy (Leach, 2001). Many methods have been developed to estimate the free energy of a given process from a MD simulation (Cramer, 2004; Leach, 2001). We will now explain the two methods for free energy estimation employed in this thesis.

2.3.4.7.1 Potential of mean force

We may be interested to know how the free energy changes as a function of some inter- or intra-molecular coordinate, such as the distance between two atoms within a molecule (Leach, 2001). The free energy surface along the chosen coordinate is known as a potential of mean force (PMF), and the point of higher energy on the free energy profile corresponds to the transition state for the process (Leach, 2001). We can calculate the PMF using the following expression for the Helmholtz free energy as a function of a given coordinate $r$ (Cramer, 2004; Leach, 2001):

$$A(r) = -k_B T \ln \zeta(r) + \text{constant} \quad (2.55)$$

where $\zeta(r)$ is the probability of the coordinate taking on a particular value $r$; and the constant is often chosen so that the most probable distribution corresponds to a free energy of zero (Leach, 2001). Although the PMF is accurate in the limit of infinite simulation time, in which the system is able to sample the entire phase space (Leach, 2001), if we simulate enough we can use PMF to compare the profiles for the same process in two different proteins, for example.
2.3.4.7.2 Free energy of protein-protein interaction

Free energy cycles are very useful to evaluate, for example, the energy of interaction between two proteins (Cramer, 2004). From a standard MD simulation of the complex (i.e., the two proteins interacting in a complex) with explicit solvent, and employing the thermodynamic cycle shown in Figure 2.3, we can estimate the binding or interaction free energy in solution $\Delta G_{\text{bind}}^{\text{solution}}$ between proteins A and B by (Leach, 2001):

$$\Delta G_{\text{bind}}^{\text{solution}} = \Delta G_{\text{bind}}^{\text{vacuo}} + \Delta G_{\text{AB}}^{\text{solvation}} - \Delta G_{A+B}^{\text{solution}}$$  \hspace{1cm} (2.56)

![Figure 2.3](image)

**Figure 2.3.** Scheme of the thermodynamic cycle employed to estimate the binding free energy in solution between protein A and B.

Although we could calculate the entropic cost of complex formation, using the usual statistical mechanics expression of entropy, in our approach we will not take into account entropic effects explicitly. The idea of this approximation is that we are not interested in the absolute change of free energy $\Delta G$, but rather we are interested in relative changes of free energy $\Delta(\Delta G)$ between two different binding processes, in which we assume that the entropic change is the same. For example, we would like to compare the free energy of interaction between protein A and two mutant versions of protein B, or between protein A and two different proteins that share many similar properties, like size and fold. So, if we neglect explicit entropic effects then:
Because in our MD simulations the pressure is constant, and because the change in volume is small for condensed phases, then 
$\Delta H_{\text{bind}}^{\text{vacuo}} = \Delta E_{\text{bind}}^{\text{vacuo}}$, the latter being the non-bonded interacting energy between molecules A and B as given by the force field AMBER (sum of the inter-molecular vdW and Coulomb interactions). Now, we need to estimate the free energy of solvation, which is the free energy to transfer a molecule from vacuum to solvent (Leach, 2001). Even though we perform the MD simulation in explicit solvent, we estimate solvation free energies with a continuous implicit model, to make the calculations faster. In particular, there are two main contributions to the solvation free energy: polar (or electrostatic) and non-polar contributions (Leach, 2001). The non-polar solvation free energy has in turn two contributions: one arising from the vdW interactions between the protein and the solvent, and the other one related to the free energy of creating a cavity in the solvent (entropy change due to reorganization of the solvent around the protein) (Leach, 2001). This combined non-polar term

$\Delta G_{\text{non-polar}} = G_{A}^{\text{non-polar}} - G_{A+B}^{\text{non-polar}}$ is proportional to the change in solvent accessible surface area (SASA) in the complex with respect to the separate proteins (Leach, 2001). On the other hand, the polar contribution, also sometimes referred to as the change in reaction field energy, is related to the free energy of removing all charges from the vacuum and adding them to a continuum solvent environment, and can be estimated by solving the PB equation (Leach, 2001).

In this thesis, the free energy of interaction between two proteins or domains was estimated by a thermodynamic cycle like the one above, with the program sietraj (Cui et al., 2008; Naim et al., 2007). The reported free energy values correspond to the average
calculation over a given number of frames from the MD simulation (indicated in each case). In the *sietraj* program, the free energy of interaction is calculated as above, and this value is then scaled by an empirically determined factor based on binding affinities in solution (Nairn et al., 2007). The scaling can be considered a crude treatment of entropy-enthalpy compensation, but still contains the caveats of implicit solvation (Chen et al., 2004; Nairn et al., 2007). The *sietraj* method has been shown to predict experimental absolute binding affinities with a mean absolute error of 1.29 kcal/mol (Nairn et al., 2007), which is comparable to current state-of-the-art binding affinity scoring functions (Wang et al., 2004).
Chapter 3:

Role of conserved residues in structural dynamics of copper chaperones

3.1 Atoxl and CopZ: similar fold yet distinct sequence

Despite having little sequence identity, Cu chaperones and target MBDs from various organisms adopt a similar ferredoxin-like fold (Arnesano et al., 2002). A hallmark of these proteins is a metal-binding motif MX_1C_1X_2X_3C_2 located in a surface exposed loop (Figure 3.1). Although all Cu chaperones and MBDs coordinate Cu(I) via the two Cys in the conserved motif, residues in the motif and near the Cu-binding loop vary distinctly among the different proteins and domains (Arnesano et al., 2002). These residues, although not Cu ligands, create a unique environment around the metal site, which may dictate their different Cu-binding properties and protein partners within each organism. Although in the past years our understanding of the function and structure of Cu chaperones and MBDs has advance impressively, the role of different residues in and surrounding the Cu loop is still not clear. In this Chapter, we have studied the role of conserved residues in two prototype metallochaperones from eukaryotes and bacteria.

Human Atoxl and B. subtilis CopZ metallochaperones are small (68 and 69 residues, respectively) homologous proteins with a similar topology (backbone rmsd of 2.2 Å), but they share only 22% sequence identity (Figure 3.1). Although Cu stabilizes the folded state in both cases, Atoxl is more stable than CopZ against thermal and chemical-induced denaturation (Hussain and Wittung-Stafshede, 2007). The molecular reason behind the difference in stability is likely linked to sequence differences, as they share the same fold. However, no residue-specific analysis has yet been performed to explain the in vitro stability differences.
Figure 3.1. A: Superimposition of NMR structures of apo-Atoxl (green, 1TL5.pdb) and apo-CopZ (orange, 1P8G.pdb). The Cys residues are shown in Licorice in Atoxl and in CPK in CopZ. B: Blow-up of the Cu-binding loop area of each holo-structure (up: Atoxl, 1TL4.pdb; bottom: CopZ, 1K0V.pdb): Met10/11, Thr11/Ser12, Cys12/13, Cys15/16, Lys60/Tyr65 and Cu(I) are labeled in Licorice for each structure. C: Sequence alignment of Atoxl and CopZ. Residues highlighted in blue are completely conserved, yellow have similar chemical properties, and pink indicate the position of residue 60/65 in Atoxl/CopZ. Residues M, X1, C1 and C2 of the metal binding motif MX1C1X2X3C2, together with position 60/65 are highlighted with a black box. The secondary structure elements ($\beta$1-$\beta$4, $\alpha$1-$\alpha$2 and loop regions L1-L5) and residue numbering are indicated based on apo-Atoxl structure.
Although both proteins bind a single Cu(I) atom with only the two Cys as the protein ligands (Anastassopoulou et al., 2004; Banci et al., 2001; Banci et al., 2003c; Ralle et al., 2003), the coordination of the metal seems to vary. Atoxl is proposed to have a S-Cu-S linear coordination geometry (Anastassopoulou et al., 2004; Ralle et al., 2003). In the case of CopZ, although NMR data showed that the S-Cu-S geometry significantly deviated from linear, a third Cu-ligand was not identified (Banci et al., 2001). X-ray absorption spectroscopy (XAS) experiments from the same lab suggest that this deviation might be due to an exogenous thiol present in the sample buffer, such as dithiothreitol (DTT) (Banci et al., 2003c). Although Met (Met10 and 11 in Atoxl and CopZ, respectively) is an appropriate ligand for Cu(I) coordination and is part of the conserved signature motif in all organisms, it appears not to be directly involved in metal ligation (Arnesano et al., 2002). Instead, it has been proposed to act as a tether that modulates the Cu-binding loop structure (Arnesano et al., 2002; Poger et al., 2005).

Important differences between the eukaryotic and bacterial Cu chaperones are residue X1 of the metal-binding motif and residue 60/65 in Atoxl/CopZ (Figure 3.1). Residue X1, located in the Cu-binding loop, is a Thr in eukaryotic Cu chaperones (including Atoxl, Thr11) and MBDs (except in MBD3 that is a His) and in prokaryotic MBDs (Arnesano et al., 2002). However, this position is filled with a Ser in prokaryotic Cu chaperones (including B. subtilis CopZ, Ser12) (Arnesano et al., 2002). In eukaryotic Cu chaperones, this residue is proposed to HB with the first Cu-coordinating Cys of the partner MBD during Cu(I) transfer (Wernimont et al., 2000). However, its role in prokaryotic Cu chaperones has not yet been explored.
Residue 60/65, located in the α2-β4 loop, is distant in sequence but structurally close to the metal-binding loop. This position is an invariant Lys in eukaryotic Cu chaperones (including Atoxl, Lys60) but an invariant Tyr in prokaryotic Cu chaperones (including CopZ, Tyr65) and MBDs (Arnesano et al., 2002). In the solution structure of holo-Atoxl, the distance between Lys60(NZ) and Cu is 4.7 ± 0.6 Å (Anastassopoulou et al., 2004), whereas in the solution structure of holo-CopZ (Banci et al., 2001) the distance between Tyr65(CE2) and Cu is 8.4 ± 1.5 Å (Figure 3.1B). All eukaryotic MBDs have a Phe at this position, except MBD3, which has a Pro (Arnesano et al., 2002). Because of its positive charge, Lys60 was proposed to neutralize the overall negative charge of the Cu(I) bis-thiolate center in the holo-metallochaperone (Arnesano et al., 2002; Wernimont et al., 2000). However, the role of a Tyr here (as in prokaryotes) is not clear, as this residue lacks the positive charge of Lys.

To learn more about the role of Cu(I) in metallochaperone structure and dynamics, we have performed a set of MM-MD simulations in combination with QM-MM calculations of Atoxl and CopZ with and without Cu(I). First, to assess the molecular origins of the different in vitro protein stabilities between Atoxl and CopZ (Hussain and Wittung-Stafshede, 2007) we have performed a set of MM-MD simulations on WT Atoxl and CopZ. For the MD simulations of the holo-forms, QM-MM calculations were first performed to explore the effect of the environment in Cu(I) coordination. Next, we extended our computational work to address the roles of three key residues (Met10/11, Thr11/Ser12 and Lys60/Tyr65) in Atoxl/CopZ structural dynamics, by performing a set of similar MM-MD simulations of different in silico generated point mutants. The aim of this second part was not only to determine the roles of these three
key residues in Cu chaperone structural dynamics but also to obtain insights into the reasons of why Met is completely conserved but Thr and Lys are only preserved in eukaryotes.

### 3.2 Computational Methods

#### 3.2.1 QM-MM calculations

The initial structures correspond to the NMR structures of both holo-proteins: Atox1, 1TL4.pdb (Anastassopoulou et al., 2004) and CopZ, 1K0V.pdb (Banci et al., 2001). These structures were solvated, and counter ions were added to neutralize the systems, as explained in Chapter 2. Water molecules extended at least 9 Å from the surface of the proteins. Atox1 and CopZ required 1 and 9 Na\(^+\) atoms to be neutral, respectively. To obtain correct QM-MM starting structures, the NMR-based solvated structures were optimized and equilibrated by performing 500 ps of MM-MD simulations at 300 K, prior to QM-MM geometry optimization. These short MD simulations were performed only to equilibrate the environment (i.e., protein and solvent) that surrounds the QM-subsystem active site. Because the active site has not yet been parameterized, the Cu(I) atom and coordinating Cys residues (that will become the QM-subsystem) were maintained fixed during these equilibrating short MD simulations.

QM-MM calculations were carried out to explore the effect of the environment on Cu(I) binding in both Atox1 and CopZ, using SIESTA with conventional parameters as described in Chapter 2. The Cu(I) atom plus the coordinated side chains (CH\(_3\)S\(^-\)) of residues Cys12 and Cys15 in Atox1, and Cys13 and Cys16 in CopZ were selected as the QM-subsystem, which comprises 11 atoms. The rest of the protein, counter ions and
water molecules were treated at the MM level. To investigate whether Met11 was able to coordinate Cu(I) in CopZ, similar QM-MM calculations were performed including residue 11 as part of the QM-subsystem. We also performed QM-MM optimizations on the \textit{in silico} generated Met11Ala mutant, treating residue 11 as part of the MM- or QM-subsystem.

To determine the optimal Cys(S)-Cu-Cys(S) angle configuration in both holo-proteins, restrained energy optimizations along this angle were computed, as described in \textbf{Chapter 2}. For this purpose, an additional term is added to the potential energy according to equation 2.39. In this case, the Cys(S)-Cu-Cys(S) angle was changed from 110 to 180 degrees, and the force constant was 40 kcal/(mol rad\textsuperscript{2}), as reported for a similar Cu chaperone in (Dalosto, 2007).

\textbf{3.2.2 MM-MD simulations}

The initial apo-structures were obtained from the PDB: Atox1, 1TL5.pdb (Anastassopoulou et al., 2004) and CopZ, 1P8G.pdb (Banci et al., 2003b). The initial holo-structures were the QM-MM optimized structures of holo-Atox1 and holo-CopZ. CopZ in both apo- and holo- forms contains four additional non-native residues at the C-terminus as a result of how the proteins were cleaved during purification (Banci et al., 2001; Banci et al., 2003b); however, this probably has little effect on the fold and biophysical parameters. Although there is experimental evidence suggesting that CopZ may bind Cu(I) in a three coordinated fashion, via an exogenous non-protein ligand present in the media (Banci et al., 2001) or by the result of protein dimerization (Kihlken et al., 2002; Zhou et al., 2008), the nature of this third ligand has not yet been identified.
Therefore, the effect of the presence of a third Cu(I) ligand originating from an exogenous donor or as a result of protein dimerization was not taken into account. Also, the lack of a CopZ high-resolution structure in which Cu(I) is coordinated via more than two ligands, makes it difficult to address these issues computationally. Therefore, we decided to analyze the two-coordinated Cu(I) center in both CopZ and Atoxl. Besides WT Atoxl and CopZ, we simulated five mutants of each protein: Met10Ala, Thr11Ala, Thr11Ser, Lys60Ala and Lys60Tyr for Atoxl and Met11Ala, Ser12Ala, Ser12Thr, Tyr65Ala and Tyr65Lys for CopZ. Because PDB structures of the mutants are not available, the mutations were generated in silico using the corresponding WT apo- and holo-structures of Atoxl and CopZ as templates.

To perform MM-MD simulations of the holo-proteins, the Cu(I) centers were parameterized using the QM-MM optimized geometries. For Atoxl, the QM-MM optimized Cu(I) geometry is similar to the one obtained from the isolated model system (compare Tables 2.1 and 3.1), yielding a similar set of parameters. However, because the QM-MM optimized Cu(I) geometry in CopZ differs from the isolated model system, a MM parameterization based on this model system geometry would yield unrealistic results. The atomic charges of Cu(I) coordinated to two CH₃S⁻ groups were determined with RESP and combined with the CYM charges modified to maintain an overall integer charge of -1 (as explained in Chapter 2): qCu(I) = 0.4124 e, qS = -0.7525 e for holo-Atoxl and qCu(I) = 0.3527 e, qS = -0.7316 e for holo-CopZ. The equilibrium parameters were taken from the QM-MM optimized structures: Cu-S = 2.15 Å, S-Cu-S = 178.6° for holo-Atoxl and Cu-S = 2.19 Å, S-Cu-S = 154.4° for holo-CopZ. The vdW parameters for Cu(I) were taken from (Fuchs et al., 2006). The bond and angle force constants involving
the Cu(I) atom were taken from (Dalosto, 2007). This combination of parameters are similar to those reported in (Holt and Merz, 2007).

**Table 3.1.** Relevant structural parameters (distances in Å and angles in degrees) of QM-MM optimized holo-forms of Atoxl and CopZ, and comparison with experiments (NMR and XAS). For CopZ, Met11 (in WT) or Ala11 (in the Met11Ala variant), was either treated classically (MM) or as part of the QM-subsystem (QM). SG1: SG atoms of Cys12 and 13 in Atoxl and CopZ, respectively; SG2: SG atoms of Cys15 and 16 in Atoxl and CopZ, respectively; SE: SE atoms of Met10 and Met11 in Atoxl and CopZ, respectively.

<table>
<thead>
<tr>
<th></th>
<th>holo-Atoxl</th>
<th>holo-CopZ</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>QM-MM</td>
<td>XAS/ NMR(^a)</td>
<td>QM-MM</td>
<td>XAS/ NMR(^b)</td>
</tr>
<tr>
<td></td>
<td>MM</td>
<td>QQ</td>
<td>MM</td>
<td>QQ</td>
</tr>
<tr>
<td>SG1-SG2</td>
<td>4.29</td>
<td>2.16</td>
<td>4.29</td>
<td>2.16</td>
</tr>
<tr>
<td>SG1-Cu</td>
<td>2.15</td>
<td>2.16</td>
<td>2.19</td>
<td>2.16</td>
</tr>
<tr>
<td>SG2-Cu</td>
<td>2.05</td>
<td>2.16</td>
<td>2.19</td>
<td>2.16</td>
</tr>
<tr>
<td>SG1-Cu-SG2</td>
<td>178.8</td>
<td>160 ± 25</td>
<td>154.4</td>
<td>152.3</td>
</tr>
<tr>
<td>SE-Cu</td>
<td>7.76</td>
<td>5.70</td>
<td>5.87</td>
<td>-</td>
</tr>
</tbody>
</table>

\(^a\) NMR data from (Anastassopoulou et al., 2004) and XAS data from (Ralle et al., 2003).
\(^b\) NMR data from (Banci et al., 2001) and XAS data from (Banci et al., 2003c).

MM-MD simulations were performed for the different proteins using AMBER, as described in **Chapter 2**. The structures were simulated (equilibration phase) until the backbone rmsd as a function of time was stable for at least 20 ns. In all cases these converged last 20 ns (production phase) were used for data analysis. To assure a proper equilibration of the 8-9 Na\(^+\) atoms required to neutralize CopZ WT, this protein was simulated for longer times than Atoxl. Previous simulations in the presence of a high number of explicit counterions were shown to be equilibrated after 50 ns (Fogolari et al., 2005; Mukhopadhyay et al., 2004; Pande and Nilsson, 2008; Partay et al., 2007; Walton and Vanvliet, 2006). In the case of the mutants, since the starting point structures were generated *in silico*, they were simulated longer than the corresponding WT proteins.
3.3 Structure and dynamics of copper binding to Atoxl and CopZ

3.3.1 Cu(I) coordination and nearby environment: QM-MM optimizations

The QM-MM optimized geometrical parameters of the Cu(I) center are reported in Table 3.1. In holo-Atoxl the Cys(S)-Cu distances are typical of digonal Cu(I) (Pickering et al., 1993; Ralle et al., 1998) and agree with that determined by XAS (Ralle et al., 2003) (Table 3.1) and the isolated linear model system (Table 2.1). In holo-CopZ the Cys(S)-Cu distance is shorter than that observed by XAS in the presence of DTT (Banci et al., 2003c), and is between the linear and trigonal model systems (Tables 2.1 and 3.1). The optimized S-Cu-S angles differ significantly: 178.6° and 154.4° for Atoxl and CopZ, respectively (Figure 3.2.A). This results in a linear Cu(I) coordination in Atoxl, in agreement with XAS (Ralle et al., 2003) and NMR (Anastassopoulou et al., 2004) studies (Table 3.1). In CopZ, our results indicate the presence of a distorted linear Cu(I) coordination, whereas XAS (Banci et al., 2003c) and NMR estimations (Banci et al., 2001) suggested a trigonal coordination (Table 3.1). Based on these earlier findings, it was proposed that Cu(I) in CopZ has a third non-protein ligand, possibly an exogenous thiol (such as DTT) present in the buffer (Banci et al., 2003c). In vacuum, we find that two CH₃S⁻ groups and a Cu(I) favor a linear coordination (∼178°, Table 2.1). However, in the protein, the geometry and particular environment can change this preference, and this seems to be the case in CopZ. The fact that CopZ optimized in a non-linear coordination in the absence of a third ligand, suggests that the protein structure imposes this Cu(I) coordination.

The S-Cu-S angle discrepancy between the two holo-forms was further corroborated by performing restrained energy optimization along the Cys(S)-Cu-Cys(S)
angle (Figure 3.2.B). With this approach, we found that the optimal angle configuration in CopZ (~150°) differs from that in Atoxl (~180°). This observation also suggests that the QM-MM optimized angles are independent from the initial values. This is important, as the QM-MM calculations do not sample all energy space, but only regions around local minima.

In the holo-Atoxl QM-MM optimized structure, the conserved Met10 is buried, far away from the active site (Table 3.1, Figure 3.2.A), and in close contact with Val40 located in β3 allowing for vdW interactions. Both Cys are solvent exposed, and Cys12 forms a HB with Thr11 (H-S = 3.07 Å and O-H-S = 173.4°) and one water molecule (H-S = 2.23 Å and O-H-S = 169.4°), whereas Cys15 interacts with Lys60 (H-S = 2.12 Å and N-H-S = 153.3°) (Figure 3.2A). In holo-CopZ, Met11 is solvent exposed, closer to the active site (Table 3.1, Figure 3.2.A), and interacts with Cys13. In holo-CopZ, only Cys13 HBs with two waters (H-S = 2.11 and 2.08 Å and O-H-S = 168.1° and 162.1°), and Tyr65 points to the solvent.

To further investigate the molecular basis of the particular CopZ Cu(I) coordination and to test whether Met11 is capable of binding Cu(I), we included the Met11 sidechain as part of the QM-subsystem, as opposed to before when we treated it at the MM level. In this new simulation, the electronic description of Met11 allows Cu(I) binding as a possibility. Furthermore, we mutated in silico this conserved residue to Ala and analyzed the variant (Met11Ala) classically (MM) and as part of the QM-subsystem. As shown in Table 3.1 and Figure 3.2C there are no major geometrical changes between the different models, suggesting that Met11 is not a Cu(I) ligand.
Figure 3.2. A, C: Cu center of the optimized QM-MM holo-structures for Atoxl (A, left), CopZ WT (A, right; C, left) and Met11Ala mutant (C, right). The QM-subsystem [A, Cu(I) and Cys; C, Cu(I), Cys and residue 11] is in Licorice and relevant residues in the MM layer (A, Met10/11, Thr11/Ser12, Lys60/Tyr65; C, Ser12 and Tyr65) are shown in CPK. B: Restrained energy optimization profile along the Cys(S)-Cu-Cys(S) angle for holo-Atoxl (green) and holo-CopZ (orange). The reaction coordinate was varied from 110° to 180°, but only the 130°-180° relevant range is shown.
When Met11 is included in the QM-subsystem, instead of moving closer to the Cu-center, as expected if this was a Cu(I) ligand, it moves further away from the active site (Table 3.1). The S-Cu-S angle does not change significantly between the WT and mutant proteins, suggesting that the particular orientation of Met11 in CopZ is not responsible for the deviation of linearity in the S-Cu-S angle.

To gain insight into differences in the active site (i.e., the Cu-binding loop) that may cause the difference in Cu(I) coordination between Atox1 and CopZ, we superimposed the QM-MM optimized structures and analyzed only the active sites (residues 9 to 18 in Atox1 and residues 10 to 19 in CopZ) (Figure 3.3). There is clearly a different structural arrangement of this region in the two proteins. This can quantitatively be represented in a Ramachandran plot of the residues in the conserved metal-binding motif (MX1C1X2X3C2). The plot reveals that residues M, X1, C1 and X3 have significantly different torsion angles in the two proteins, allowing for different conformations of this region. In Atox1, all residues lay within allowed energy regions; whereas in CopZ, Ser12 and Gln14 lay within disallowed energy regions. This may suggest that in CopZ, the residues in the conserved motif collectively induce steric hindrance that results in the distorted S-Cu-S angle. In particular, the two bulkier residues between the coordinating Cys in CopZ (Gln and His in positions X2 and X3, respectively) as opposed to the two Gly residues in Atox1, may predispose the distorted Cys arrangement. Moreover, the Ramachandran plots of the QM-MM optimized structures of CopZ WT and Met11Ala mutant overlap, suggesting that Met11 is not responsible for the particular loop conformation (data not shown).
Figure 3.3. **Left:** Superimposition of the Cu-active site (residues 9 to 18 in Atoxl, green; and residues 10 to 19 in CopZ, orange) of the QM-MM optimized holo-structures. The functional Cys residues and Cu(I) are shown in Licorice in Atoxl and in CPK in CopZ. **Right:** Ramachandran plot for the residues of the Cu-binding motif (MX\textsubscript{1}C\textsubscript{1}X\textsubscript{2}X\textsubscript{3}C\textsubscript{2}) of Atoxl (circles) and CopZ (squares). The corresponding residues in Atoxl and CopZ are depicted with the same color.

3.3.2 MM-MD simulations of apo- and holo-metallochaperones

3.3.2.1 Backbone fluctuations

Next, apo- and holo-forms were subjected to MM-MD simulations to assess structural dynamics. After ~22 ns of MD simulations, both apo- and holo-forms of Atoxl were found to be stable in this time period, as testified by the backbone rmsd time evolution (Figure 3.4A, Table 3.2), suggesting no major conformational changes. Because of the presence of 8-9 Na\textsuperscript{+} ions in CopZ, we simulated ~100 ns and analyzed the last 20 ns, to be consistent with the Atoxl results. Holo-CopZ is equilibrated in the first 20 ns whereas apo-CopZ is equilibrated after approximately 60 ns (Figure 3.4, inset). Although in the last 20 ns both apo- and holo-CopZ are stable, apo-CopZ exhibits greater flexibility as shown by the mean rmsd and standard deviation for the whole 100 ns run (Figure 3.4, Table 3.2). The greater flexibility of apo-CopZ is more likely due to an
intrinsic property of the protein than to the lack of equilibration of the Na$^+$ ions, because holo-CopZ is already equilibrated in the first 20 ns.

![Figure 3.4. Rmsd (in Å) of the backbone heavy atoms as a function of the 20 ns MD production time for Atoxl (left) and CopZ (right). Inset: Rmsd of the 100 ns MD for CopZ. Apo: black; holo: blue.](image)

**Figure 3.4.** Rmsd (in Å) of the backbone heavy atoms as a function of the 20 ns MD production time for Atoxl (left) and CopZ (right). Inset: Rmsd of the 100 ns MD for CopZ. Apo: black; holo: blue.

**Table 3.2.** Total simulation time (in ns) and backbone rmsd (in Å) for the total (rmsd$_{total}$, with respect to the first structure) and the last 20 ns of the simulation (rmsd$_{20}$, with respect to the average structure).

<table>
<thead>
<tr>
<th></th>
<th>time</th>
<th>rmsd$_{total}$</th>
<th>rmsd$_{20}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>apo</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atoxl-WT</td>
<td>23</td>
<td>1.4 ± 0.1</td>
<td>0.8 ± 0.2</td>
</tr>
<tr>
<td>CopZ-WT</td>
<td>100</td>
<td>1.9 ± 0.6</td>
<td>1.3 ± 0.2</td>
</tr>
<tr>
<td>holo</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atoxl-WT</td>
<td>22</td>
<td>1.2 ± 0.1</td>
<td>0.6 ± 0.1</td>
</tr>
<tr>
<td>CopZ-WT</td>
<td>99</td>
<td>2.2 ± 0.2</td>
<td>0.9 ± 0.2</td>
</tr>
</tbody>
</table>

The distribution of structures obtained from the last 20 ns MD simulations was compared to the initial structures by computing the rmsf per residue (**Figure 3.5A**). As expected, the $\alpha$-helices and $\beta$-sheets have more restricted backbone motion than less structured regions, such as loops. The Cu-binding loop (residues 10-14 in Atoxl and 9-15 in CopZ) is flexible and relatively unstructured in both chaperones, and the first Cys (Cys12 in Atoxl and Cys13 in CopZ) is more mobile than the second Cys (Cys15 in Atoxl and Cys16 in CopZ).
Figure 3.5. A: Average fluctuations (rmsf in Å) of backbone heavy atoms per residue for Atoxl (left) and CopZ (right) with (blue) and without (black) Cu(I). The secondary structure elements are indicated. B: The Δrmsf (holo minus apo) is mapped in the apo PDB structures (arbitrary scale), where red color shows increased disorder when Cu(I) binds, and the other colors show various levels of decreased disorder upon Cu(I) binding, with blue indicating the greatest changes. C: Histograms (arbitrary units) of the Rg (in Å) distribution of the backbone heavy atoms for Atoxl (left) and CopZ (right) with (blue) and without (black) Cu(I).
Upon Cu(I) binding, in both cases the loop becomes more rigid and both Cys show restricted mobility. There is an overall reduced mobility upon Cu(I) binding that extends throughout the proteins; this effect is most dramatic for CopZ (blue and green regions in Figure 3.5B). Because apo-CopZ exhibits higher flexibility than apo-Atox1, and both holo-proteins have similar flexibility, this results in a higher entropic cost of Cu(I) binding to CopZ than to Atox1 (data not shown). The Rg data is consistent with the trend of both proteins becoming more compact in the presence of Cu(I), being the change in CopZ significantly larger than in Atox1 (Figure 3.5C). Interestingly, the loop between β3 and α2 in Atox1 and the same loop plus the α2 helix itself in CopZ show opposite trends: they become more dynamic when Cu binds (red in Figure 3.5B). We propose that these regions may specify interfaces for protein-protein interactions involved in Cu transport processes in bacterial and human cells.

The simulation data are in agreement with the reported solution NMR structures for both proteins. Major changes in hydrophobic interactions between secondary structure elements as a function of Cu were demonstrated by NMR in both CopZ (Banci et al., 2003b) and the yeast Cu chaperone Atx1 (Arnesano et al., 2001). Moreover, NMR experiments have revealed that, in contrast to the other chaperones, there are only minor alterations in the Atox1 structure upon Cu binding (Anastassopoulou et al., 2004). Thus, it is reasonable that the effects on conformational dynamics due to Cu(I) binding are not as dramatic in Atox1 as in CopZ.
3.3.2.2 Cross-correlation analysis

To gain insight into the effect of Cu(I) on the cooperative dynamics of Atox1 and CopZ structure, the cross-correlation matrix for the Ca atoms was calculated (as explained in Chapter 2) for apo- and holo-forms (Figure 3.6). In this analysis, positive or in-phase correlations are indicated in red (strong) and yellow (moderate), and negative or out-of-phase correlations are indicated in black (strong) and cyan (moderate). We find no (Atox1) or few (CopZ) strong cross-correlated regions, except for the diagonal that corresponds to the correlation of a residue with itself. This is consistent with no major conformational changes occurring during the MD simulations.

In both apo-forms (Figure 3.6A), moderate and strong (in CopZ) positive correlation are observed between residues forming secondary structure elements, such as helices α1 and α2, and strands β1-β3, β2-β3 and β1-β4, which hold the fold together. Interestingly, there is one distinct area of negative cross-correlation (blue rectangle in Figure 3.6A), which represents out-of-phase movements of residues located in the Cu-binding loop and its vicinity, and with residues that are distant in both sequence and space (Figure 3.6A, inset). In apo-Atox1, Cys12, which is the first Cys believed to bind Cu (Banci et al., 2006b), has large (in magnitude) anti-correlation coefficients with most residues of the blue region: residues 8 to 10 (Cu loop), 22 (helix α1), 26 (α1-β2 loop), 33 to 38 (strand β2 and β2-β3 loop) and 60 to 62 (α2-β4 loop).

In apo-CopZ, the entire metal binding motif (residues MX\textsubscript{1}C\textsubscript{1}X\textsubscript{2}X\textsubscript{3}C\textsubscript{2}) has large (in magnitude) anti-correlation coefficients with most residues of the blue region: residues 6 to 8 (strand β1), 19 to 21 and 24 (helix α1), 34 to 36 (strand β2), 39 to 43 (β2-β3 loop and strand β3), 55 to 68 (helix α2, α2-β4 loop and strand β4).
Figure 3.6. Cross-correlation matrices of fluctuations of Cα atoms from their average values for apo- (A) and holo- (B) forms of WT Atox1 (left) and CopZ (right). Correlations are color-coded: red, strong positive (0.6 ≤ C_{ij} ≤ 1); yellow, moderate positive (0.4 ≤ C_{ij} < 0.6); black, strong negative (0.6 ≤ |C_{ij}| < 1); cyan, moderate negative (0.4 ≤ |C_{ij}| < 0.6); white, weak or no correlation (-0.4 < C_{ij} < 0.4). Relevant negative cross-correlations regions (blue square) are mapped in the corresponding structures (Insets).

The observed flexibility of the Cu loop in both apo-proteins (Figure 3.5A) may allow for these concerted out-of-phase movements and interactions between the loop and distant parts of the protein. Also, the larger anti-correlated motions found in CopZ are
consistent with its larger fluctuations, with respect to Atoxl. This distinct anti-correlation pattern of the Cu loop may be important for Cu(I) binding, possibly by positioning the Cys in an appropriate conformation for Cu(I) uptake. Thus, it seems that whereas positive correlations hold the fold together, anti-correlations are predominant in regions with functional relevance. This is in agreement with previous simulations, in which regions with negative cross-correlation corresponded to a protein’s active site (Arnold and Ornstein, 1997).

In the holo-proteins, the in-phase movements are similar to those found in the apo-forms (Figure 3.6B), as expected since the fold is not greatly altered by the presence of Cu(I) (Figure 3.4A). However, in sharp contrast to the apo-forms, the Cu loop loses its negative cross-correlated motions with the rest of the protein (the blue regions disappear). This is consistent with the reduction in backbone fluctuations for both holo-proteins (Figure 3.5A). Overall, this analysis supports the hypothesis that the structural and dynamic changes occurring upon Cu(I) binding affect mainly the Cu loop, and these changes are more significant in CopZ.

3.3.2.3 Cys-Cys distance distribution in the apo-forms

The distance distribution between the S atoms of Cys12 and Cys15 in apo-Atoxl and Cys13 and Cys16 in apo-CopZ was calculated for the last 20 ns of the simulations (Figure 3.7A). In both proteins there are mainly two populations of Cys(S)-Cys(S) distances (I and II). In Atoxl, the first population (I) is divided into two (in CopZ there is only one), although one of them (centered at 3.6 Å) is visited with small probability. In general, the S-S distance in CopZ is shorter. In both cases, the most probable S-S distance
population centers between 4.5-5.5 Å, whereas the longer S-S distance population is found at ~ 8 Å for apo-Atoxl but at ~ 7 Å for apo-CopZ.

**Figure 3.7.** A: Histograms (arbitrary units) of the S-S distance distribution (in Å) between the two Cys in the Cu-binding loop in the apo-forms of Atoxl (green) and CopZ (orange). Shorter (I) and longer (II) Cys-Cys populations are labeled. B: Free energy as a function of the Cys(S)-Cys(S) distance for apo-Atoxl (green) and apo-CopZ (orange). Inset: Cu-binding loop conformation of apo-Atoxl at the two minima (Cys shown in CPK and distances between atoms in Å).

If we assume that in the last 20 ns of simulation the inter-conversion between the conformations has been sampled enough, this distance distribution is the probability of being at a certain S-S distance, or the S-S $\zeta(r)$ and thus, the free energy can be readily estimated as a function of the S-S distance (see section 2.3.4.7.1 for details) (Figure 3.7B). In both cases, the energy profile is smooth and the activation energy between the minima is low (between 0.5 and 1 kcal/mol), allowing for rapid inter-conversion between the two conformations. The lowest minimum is that for the short S-S distance (i.e. 4.5-5.5 Å). Since this distance is close to the sum of the parameterized S-Cu distances in the holo-forms (4.30 Å and 4.38 Å in Atoxl and CopZ, respectively), it appears that in this apo-conformation, Cu(I) could become incorporated without much perturbation. The
energy increases with distances lower than 4 Å, which is desirable if the Cys are not supposed to form a disulfide bond (the equilibrium length of a disulfide bond is 2.038 Å according to Amber parameterization (Case et al., 2006)).

3.3.2.4 Interactions as a function of copper coordination

3.3.2.4.1 Protein-solvent interactions

To determine the average proximity to solvent molecules of the Cys and Met residues in the metal-binding loop, protein-solvent radial distribution functions $g(r)$ were calculated between the S atoms of Cys and Met, and the O atom of water (Figure 3.8).

![Figure 3.8](image-url)

**Figure 3.8.** Protein-solvent $g(r)$ of Atox1 (A) and CopZ (B) of Cys1(S), Cys2(S) and Met(S) of the metal-binding loop, with and without Cu(I). Black, apo-Cys1; orange, apo-Cys2; green, apo-Met; dark blue, holo-Cys1; red, holo-Cys2; cyan, holo-Met.

In addition, solvent HB networks for the same residues were studied from protein structures throughout the simulations. The HBs that were maintained through the production phase were analyzed at the end of the simulations. In general, Cys(S) HBs are weaker in the apo- than in the holo-proteins, because the S partial charge of Cys is less
negative in the apo-form than in the holo-form (parameterized qS = -0.31e for apo; and -0.75e and -0.73e for holo Atoxl and CopZ, respectively).

In Atoxl, Cys12 is solvent exposed regardless the presence of Cu (Figure 3.8A). However, it forms a weak HB with a water in the apo-form, whereas it forms a strong such bond in the holo form (H-S = 2.20 Å and O-H-S = 174.7°). Cys15 is more buried in the apo-form, although still capable of forming a weak HB with water. Upon Cu(I) binding, Cys15 becomes more solvent exposed and now forms a strong HB with the solvent (H-S = 1.92 Å and O-H-S = 162.1°). Met10 is completely buried regardless of the presence of the metal, and interacts extensively with the protein core (Figures 3.8A and 3.9A). The interacting residues include the sidechain of Ile33, Leu35 and Val40, and the backbone of Asp34, Lys38 and Lys39, all located in the β2-β3 sheet, as well as with Cys15 sidechain.

**Figure 3.9.** Blow-up of residues 9-11 and 33-41 in apo-Atoxl (A), which has a similar conformation in the holo-form, and residues 10-12 and 35-43 in apo- and holo-CopZ (B and C, respectively). Met10, Ile33, Leu35, Lys38 and Val40 in Atoxl, and Met11, Val35, Leu37, Gly40 and Val42 in CopZ are shown in Licorice.
The corresponding residues in CopZ show a somewhat different pattern. Notably, Met11 is entirely solvent exposed in the apo-form (Figures 3.8B and 3.9B), although it is not able to form HB because carbon-divalent S is a poor HB acceptor (Gregoret et al., 1991; Wierzejewska and Saldyka, 2004). Upon Cu(I) binding, Met11 becomes completely buried and interacts extensively with residues in strands β2 and β3 (and their connecting loop), including Val35, Asn36, Leu37, Gly40, Lys41 and Val42, but most strongly with Val35, Leu37 and Val42 (Figures 3.8B and 3.9C). Thus, in the absence of Cu, Met11 appears to move around freely towards the solvent, but in the presence of Cu it is forced to adopt a defined position into the hydrophobic core. This is in contrast to what is seen in the holo-NMR structure, where Met11 is solvent exposed and points toward the metal and interacts with Tyr65 (Banci et al., 2001). In apo-CopZ, Cys16 is more buried than Cys13 (Figure 3.8B), although it still is accessible to the solvent, and both form weak HBs with water. In the NMR structure of apo-CopZ, Cys16 is completely buried and interacts with Leu37 (Banci et al., 2001), but this interaction is not maintained in our MD simulation and both residues (Cys16 and Leu37) flipped towards the solvent during the MD run. Upon Cu(I) binding, both Cys become completely solvent exposed and form strong HBs with the solvent (H-S = 2.10 Å and 2.12 Å; O-H-S = 168.2° and 165.8°, for Cys13 and Cys16, respectively).

3.3.2.4.2 Specific analysis of conserved residues

In the absence of Cu(I), Thr11 and Ser12 in Atox1 and CopZ, respectively, are floppy and solvent exposed, and do not interact with Cys12/13 or Lys60/Tyr65 (Figure 3.10A, B). In CopZ, Ser12 points towards the β2-β3 loop, and interacts with Met11 and
Glu38. On the other hand, in the holo-forms, Thr11 and Ser12 in Atoxl and CopZ, respectively, are still exposed to solvent but their flexibility is significantly reduced. In holo-Atoxl, Thr11 forms a stable electrostatic interaction with Cys12 (Figure 3.10A), which is possible because of a strong and stable HB found between Thr11 and Lys60 (present in ~96% of the simulation time with a O-N distance less than 3 Å) (Figure 3.10B). In holo-CopZ, Ser12 no longer interacts with the β2-β3 loop or Met11, which is completely buried. In this case, Ser12 points towards the Cu site and adopts two conformations in which it interacts alternatively with Cys13 and Tyr65 (Figure 3.10A, B). Thus, it appears that Thr11 and Ser12 help to stabilize the Cu-bound state by facilitating a network of interactions near the Cu site.

Lys60 and Tyr65 in Atoxl and CopZ, respectively, are located in a surface exposed loop, and therefore, they are always solvent exposed throughout the simulation in both apo- and holo-forms. Although the Lys60 sidechain is close to the Cu-binding loop in apo-Atoxl, its flexibility impedes any stable interaction in the apo-form (Figure 3.10B, C). In apo-CopZ, Tyr65 is also flexible but very far away from the Cu loop (Figure 3.10B, C), and interacts with Glu9 and Gln63 located in the β1-α1 and α2-β4 loops, respectively. On the other hand, in the holo-forms, the flexibility of these residues is reduced, and now Lys60 and Tyr65 point to the Cu site. In holo-Atoxl, Lys60 interacts with the Cu-binding loop via strong HBs with Thr11 and Cys15 (Figure 3.10B, C). In holo-CopZ, Tyr65 also interacts extensively with Ser12 and Cys16 (Figure 3.10B, C). Therefore, it appears that Lys60 and Tyr65 undergo a conformational change upon Cu(I) binding, which greatly stabilizes the Cu-bound state in Atoxl and CopZ, respectively.
Figure 3.10. Histograms (arbitrary units) of distance distribution (in Å) between heavy atoms of Thr11/Ser12 (O) and Cys12/13 (S) (A), Thr11/Ser12 (O) and Lys60/Tyr65 (N/O) (B), and between Lys60/Tyr65 (N/O) and Cys15/16 (S) (C) of WT Atox1 (left) and CopZ (right) with (blue) and without Cu(I) (black).
3.3.3 Biological relevance

The computational data obtained here for Atoxl and CopZ correlates well with the recently reported in vitro stability data (Hussain and Wittung-Stafshede, 2007). The higher conformational dynamics and lack of core interactions due to exposure of Met11 found in apo-CopZ as compared to apo-Atoxl may explain why apo-CopZ is significantly less stable than Atoxl towards thermal and chemical perturbation in vitro (Hussain and Wittung-Stafshede, 2007). Moreover, CopZ is stabilized to a greater extent than Atoxl upon Cu(I) binding (Hussain and Wittung-Stafshede, 2007), in agreement with the in silico differences between apo- and holo-forms. The changes in rmsf (and entropic cost), Rg, cross-correlation data, as well as burial of Met11, demonstrate larger effects of Cu(I) binding for CopZ than for Atoxl.

Despite the divergences noted above, there are trends that are similar between Atoxl and CopZ. We propose that these similarities represent features that may be shared by Cu chaperones in all organisms. For both Atoxl and CopZ, although the Cu site is confined to a loop at the protein surface, Cu(I) binding rigidifies and shrinks the whole protein. Interestingly, the loop between $\beta$3 and $\alpha$2, which is located at the opposite side of the Cu-binding loop in the protein, becomes more floppy upon binding of Cu(I) in both proteins. It is tempting to speculate that this region defines the interface for protein-protein interactions in Cu-transfer pathways that may be shared between the bacterial and human cells. In Chapter 5 we will see that the $\beta$3-$\alpha$2 loop in Atoxl does not directly make contacts with residues in the WDs. Therefore, it is possible that this region is important for other partner recognition, for example, for metallochaperone Cu acquisition. Interestingly, we found that both apo-proteins exhibit bi-phasic distributions
of Cys(S)-Cys(S) distances. One may speculate that the ‘stretched’ structure (corresponding to the longer S-S distances) facilitates Cu(I) uptake via a ‘pull-in’ type mechanism (while returning to the shorter S-S distances, which have the lowest energy). The heterogeneity in terms of S-S distances may also be a key to Cu(I) transfer: to accommodate transient structures in the chaperone-target complex, the metallochaperone may need to extend one of its Cys(S) (thereby adopting the ‘stretched’ conformation) into the target’s Cu site.

We have found several differences between the two proteins that cannot easily be detected by experimental methods. The position and interactions of the conserved Met residue differ between the two proteins. Whereas this Met is solvent exposed in apo-CopZ, it is always buried in Atoxl. Also, the QM-MM optimized Cu(I) geometry significantly differs between CopZ and Atoxl. Whereas Cu(I) favors a linear coordination in Atoxl, it adopts a distorted linear coordination in CopZ. Notably, QM-MM optimization of the Cu(I) site in yeast Atxl resulted in a Cys(S)-Cu-Cys(S) angle intermediate between those found here for Atoxl and CopZ of 166° (Dalosto, 2007). As we will later see in Chapter 4, intermediate geometries are also adopted by the WDs. We propose that structural constraints within CopZ are responsible for this Cu(I) geometry, as the optimized structure for [Cu(I)(CH₃S)₂]⁺ is linear. This distortion is probably not due to the peculiar conformation of Met11 in CopZ, since mutation of this residue to an Ala followed by QM-MM optimization yielded similar results as for WT. The overall arrangement of the conserved metal-binding motif in CopZ is likely due to protein steric effects within this region that deviate the S-Cu-S angle from linearity, as opposed to Atoxl. This is probably due to sequence differences within the Cu loop, specifically, the
two residues between the coordinating Cys. We speculate that this distortion pre-disposes
the Cu(I) atom to readily pick up a third ligand; which may be a small thiol compound
from the solvent (like DTT in vitro or GSH in vivo) or a Cys(S) from a target MBD.
Moreover, this distorted coordination may favor CopZ dimerization in solution (Kihlken
et al., 2002) as opposed to Atoxl (Anastassopoulou et al., 2004). Dimerization of CopZ
may serve as an energetic compensation to overcome the high entropic cost of Cu(I)
binding. Alternatively, CopZ dimerization could be a way of protecting the Cu(I) atom,
as the Cu site is more exposed in holo-CopZ, compared to holo-Atoxl (Figure 3.2.A), as
previously suggested (Kihlken et al., 2002). The structural and dynamic differences
observed between Atoxl and CopZ may be related to unique aspects of the bacterial
versus the human cellular environments and their respective physiological partner
proteins.

3.4 Sequence differences between eukaryotic and prokaryotic copper chaperones

Although our simulations on WT Atoxl and CopZ have helped us reveal key
molecular differences between the eukaryotic and prokaryotic systems, and possibly
explain in vitro behavior between these two homologous proteins, it is still not clear why
Met10/11 is conserved but residues X1 and 60/65 are not. Although Met is conserved in
all Cu chaperones and MBDs, the behavior of this residue differs between Atoxl and
CopZ. By NMR, Met11 in CopZ was found to be solvent exposed (Banci et al., 2001;
Banci et al., 2003b), as opposed to Met10 in Atoxl, which is buried (Anastassopoulou et
al., 2004). However, we have identified a significant change of the dynamics and solvent
exposure of Met11 in CopZ depending on the presence/absence of Cu(I). In apo-CopZ,
Met11 is entirely solvent exposed and floppy, whereas it is buried in the protein core in the holo-form, similarly to what was found for Atoxl. It appears that Met11 serves as a Cu-dependent switch in CopZ, and its role as a hydrophobic anchor as appears to be the case in Atoxl is less obvious.

Residue X1 is not a Cu(I) ligand, and in eukaryotic Cu chaperones it was proposed to HB with the first Cys (Cys1) of the partner MBDs during Cu(I) transfer (Wernimont et al., 2000). However, its role in intrinsic Cu chaperone structure or dynamics is not clear. This position is always occupied by a Thr, Ser or His, suggesting that a residue capable of forming a HB is necessary (Arnesano et al., 2002). If both Thr and Ser can form HBs and only differ by a methyl group, why did eukaryotes preserve Thr but prokaryotes conserve a Ser in this position? We have seen that in both Atoxl and CopZ, residue X1 appears to be involved in a network of interactions that stabilize the Cu site in the holo-form. But its role in the apo-form is less clear. Similarly, Lys60 and Tyr65 stabilize the Cu-bound state by electrostatic interactions with the Cu loop. However, their role in the apo-proteins is still unclear. In the next section, we study the consequences of replacing Met10/Met11, Thr11/Ser12 and Lys60/Tyr65 with Ala or with the corresponding residue in the other organism, to shed light into the individual roles of such residues/positions in Atoxl/CopZ structural dynamics.

3.4.1 Differential roles of Met10, Thr11, and Lys60 in Atoxl structural dynamics

3.4.1.1 Met10Ala Atoxl variant

Met10Ala Atoxl mutant required longer times for appropriate equilibration, as it was found to be more flexible than WT (Table 3.3, Figure 3.11A, B).
Table 3.3. Total simulation time (ns) and backbone rmsd (Å) for the total (rmsd$_{total}$, with respect to the first structure) and the last 20 ns of the simulation (rmsd$_{20}$, with respect to the average structure).

<table>
<thead>
<tr>
<th></th>
<th>time</th>
<th>rmsd$_{total}$</th>
<th>rmsd$_{20}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>apo</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atox1-Met10Ala</td>
<td>82</td>
<td>1.8 ± 0.2</td>
<td>0.7 ± 0.1</td>
</tr>
<tr>
<td>Atox1-Thr11Ala</td>
<td>35</td>
<td>1.4 ± 0.1</td>
<td>0.5 ± 0.1</td>
</tr>
<tr>
<td>Atox1-Thr11Ser</td>
<td>84</td>
<td>1.6 ± 0.2</td>
<td>0.7 ± 0.1</td>
</tr>
<tr>
<td>Atox1-Lys60Ala</td>
<td>40</td>
<td>1.3 ± 0.1</td>
<td>0.5 ± 0.1</td>
</tr>
<tr>
<td>Atox1-Lys60Tyr</td>
<td>55</td>
<td>1.5 ± 0.2</td>
<td>0.7 ± 0.1</td>
</tr>
<tr>
<td>holo</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atox1-Met10Ala</td>
<td>115</td>
<td>1.6 ± 0.3</td>
<td>0.8 ± 0.2</td>
</tr>
<tr>
<td>Atox1-Thr11Ala</td>
<td>45</td>
<td>1.4 ± 0.1</td>
<td>0.7 ± 0.1</td>
</tr>
<tr>
<td>Atox1-Thr11Ser</td>
<td>60</td>
<td>1.2 ± 0.1</td>
<td>0.6 ± 0.1</td>
</tr>
<tr>
<td>Atox1-Lys60Ala</td>
<td>44</td>
<td>1.3 ± 0.1</td>
<td>0.6 ± 0.1</td>
</tr>
<tr>
<td>Atox1-Lys60Tyr</td>
<td>47</td>
<td>1.2 ± 0.1</td>
<td>0.5 ± 0.1</td>
</tr>
</tbody>
</table>

In general, the backbone is destabilized in the absence of Met10 in both apo- and holo-forms, especially in the Cu-binding loop and helix $\alpha$1 (Figure 3.11C, D). This result is in agreement with Met10 being completely buried in the protein core in apo-WT, as opposed to the mutant, in which Ala10 does not participate in any important interaction. In terms of structure and backbone dynamics, the largest effect of Met10 removal is found in the apo-form (Figure 3.11C). The dynamics of the Cys residues in the apo-mutant are largely different from in apo-WT: whereas Cys12 in the loop becomes less floppy, Cys15 in helix $\alpha$1 becomes more flexible in Met10Ala as compared to WT. Apart from the Cu-binding loop and helix $\alpha$1, whose fluctuations are larger than WT, the rest of the backbone dynamics of apo-Met10Ala is only somewhat increased as compared to apo-WT (Figure 3.11C). Interestingly, residues 40 to 57, which correspond to strand $\beta$3, the beginning of helix $\alpha$2 and their connecting loop, are significantly more rigid in the apo-form of Met10Ala as compared to apo-WT. This demonstrates that the effect of the mutation is not only local but extends through the protein.
Figure 3.11. A, B: Rmsd (in Å, with respect to the first structure) of the backbone heavy atoms as a function of simulation time for the different Atox1 variants in apo- (A) and holo- (B) forms. C, D: Average fluctuations (rmsf in Å) of backbone heavy atoms per residue (with respect to the average structure of the last 20 ns) for the different variants and WT in apo- (C) and holo- (D) forms. The secondary structure elements are indicated (α1, α2; β1-β4). Black, WT; blue, Met10Ala; green, Thr11Ala; orange, Thr11Ser; red, Lys60Ala; cyan, Lys60Tyr.

Although the differences are subtler in the presence of Cu(I), holo-Met10Ala also exhibits increased backbone dynamics as compared to holo-WT, and again the greatest change is found in the Cu loop and helix α1 (Figure 3.11D). The smaller defects observed in this holo-mutant as compared to the apo-mutant may be explained, in part, by Ala10 sidechain being buried in holo-Met10Ala. This enables some hydrophobic contacts between Ala10 and the protein core (as opposed to the apo-mutant), although not to the same extent as WT (see below). We note that this holo-mutant has significantly increased
flexibility at both ends of the protein, in particular at the C-terminus, which explains its rmsd time evolution (Figure 3.11B). Also similar to the apo-form, residues 43 to 48, which form the last part of strand β3 and the subsequent loop, have their backbone much more restricted in the holo-mutant as compared to holo-WT.

Although the ferredoxin-like fold is intact in apo-Met10Ala, the Cu loop is largely destabilized and changes conformation dramatically, and the beginning of helix α1 becomes distorted (Figure 3.12A). These changes involve residues 9 to 17 and appear to be the result of missing hydrophobic contacts between Met10 and residues in the protein core. In WT, Met10 sidechain is buried in the core and points towards strands β2 and β3, where it interacts extensively with residues 33-35 and 38-40 (Figure 3.9A). These interactions in WT provide proper packing of the core, which extends to facilitate proper Cu loop conformation, since Met10 also interacts with Cys15 in α1. In apo-Met10Ala, the Ala10 sidechain changes orientation, and instead of pointing towards strands β2-β3, it points towards the α2-β4 loop where it interacts with Thr58 and Lys60. These new interactions contribute to the bending of the Cu loop in this mutant (Figure 3.12A). Superimposition of the final equilibrated structures of apo-WT and apo-Met10Ala reveal a total backbone rmsd of 1.3 Å, which reduces to 0.6 Å if residues 9-17 are excluded from the calculation.

Surprisingly, the lack of key hydrophobic contacts due to Met10 replacement does not alter the conformation of the Cu loop in the holo-form significantly. Consequently, the entire structure of holo-Met10Ala is very similar to WT, with a total backbone rmsd of only 0.9 Å (Figure 3.12B). This is probably due to Cu-induced stabilization of the loop, which preserves a correct orientation of Ala10 towards strands β2-β3, as in WT.
Figure 3.12. Superimposition of the final equilibrated structures of Atox1 WT (black) and Met10Ala (blue) in the apo- (A) and holo- (B) forms. A blow-up of the Cu-binding loop is shown to the right; Met/Ala10, Thr11, Cys12, Cys15, Lys60 and Cu(I) for each variant are labeled.

However, the small Ala10 sidechain allows for fewer contacts than WT, extending to only Leu35 and Lys38, explaining the increased backbone dynamics in this region.
Greater structural changes in apo- versus holo-forms due to the absence of Met10, were earlier reported for an Atox1 Met10Ser mutant based on 5 ns MD simulations (Poger et al., 2005).

The dynamics and positions of the two Cu-binding Cys in the apo-form of Met10Ala differ from in apo-WT, as a result of the structural and dynamical alterations observed in this region. Importantly, the S(Cys)-S(Cys) distance distribution changes significantly: whereas WT has mainly two populations centered at ~5.5 Å and ~8 Å, Met10Ala has a more diffuse distribution that reaches values of up to ~12 Å (Figure 3.13A). This may affect the efficiency of Cu(I) uptake in the mutant, because the Cys might have to re-arrange significantly before coordinating Cu(I). Also, apo-Met10Ala gains a HB between Thr11 and Cys12, which is lacking in the apo-form of WT (Figure 3.13B). This interaction may somewhat counteract the overall destabilization of this region in the apo-mutant. On the other hand, the Thr11-Cys12 interaction in the holo-form, and the Lys60-Cys15 interaction in both apo- and holo-forms are not affected by the Met10Ala mutation (Figure 3.13C-E).

The average proximity to solvent of the Cu loop in Met10Ala Atox1 was analyzed by computing the $g(r)$ between S(Cys12) or S(Cys15) and water (O). Cys12 is solvent exposed in both apo- and holo-forms similarly to WT, although Cys12 in apo-Met10Ala is slightly more exposed (Figure 3.14A, B). The Met10Ala mutation does not affect solvent exposure of Cys15 in the holo-form, but it results in its greater exposure in the apo-form (Figure 3.14C, D). Therefore, in contrast to WT, Met10Ala buries Cys15 upon Cu(I) binding. Increased solvent exposure of Cys15 in apo-Met10Ala may explain the longer S(Cys)-S(Cys) distances observed for this mutant.
Figure 3.13. Histograms (arbitrary units) of distance distribution (in Å) between heavy atoms of Cys12 (S) and Cys15 (S) in the apo-form (A), Thr/Ser11 (O) and Cys12 (S) in apo- (B) and holo- (C) forms, and between Lys/Tyr60 (N/O) and Cys15 (S) in apo- (D) and holo- (E) forms, for the different variants and WT. Black, WT; blue, Met10Ala; green, Thr11Ala; orange, Thr11Ser; red, Lys60Ala; cyan, Lys60Tyr.
Figure 3.14. Protein-solvent \( g(r) \) of Cys12(S) in apo- (A) and holo- (B) forms and Cys15(S) in apo- (C) and holo- (D) forms, for the different Atoxl variants and WT. Black, WT; blue, Met10Ala; green, Thr11Ala; orange, Thr11Ser; red, Lys60Ala; cyan, Lys60Tyr.

3.4.1.2 Thr11Ala and Thr11Ser Atoxl variants

We found that introducing the Thr11Ser mutation has a greater effect than introducing Thr11Ala in the apo-forms, whereas in the holo-form, the opposite is true. As in Met10Ala, Thr11 mutations have greater effects on Atoxl backbone flexibility in the absence of Cu(I) (Figure 3.11C, D). The backbone motion of the entire protein, also in regions far away from the mutation, is greatly reduced by both substitutions in the apo-form as compared to apo-WT. The greatest changes are found in the Cu loop and in helix...
α1, which results in less mobility of the two Cu-coordinating Cys. In this region, the reduction in flexibility is largest for the Thr11Ala mutant; however in the rest of the protein both mutants have similar backbone dynamics. Similarly to what we observed for Met10Ala, residues 44 to 57 are significantly less flexible in the Thr11 mutants than in apo-WT. In WT apo-Atoxl, the methyl group (CG) of Thr11 is solvent exposed and close to the N of Lys38 in the β2-β3 loop. This proximity of a hydrophobic and a polar group may increase the entropy of the loop. In the Thr mutants, the loop becomes “frozen” due to the absence of this entropic effect. Despite the changes in backbone flexibility, the overall fold is essentially unaltered in the two apo-Thr mutants (Figure 3.15A), with a total backbone rmsd with respect to apo-WT of 0.8 Å and 1.2 Å, for Thr11Ala and Thr11Ser, respectively.

The effect of the mutations in terms of protein and loop dynamics is smaller in the holo-form (Figure 3.11D) because Cu(I) binding stabilizes the Cu loop in WT Atoxl. However, the backbones of both holo-mutants are more rigid than in holo-WT; the effect is largest for residues 40-55. Because of the dramatic reduction in flexibility of apo-Thr11Ala, this mutant undergoes only a slight stabilization of the Cu loop upon Cu(I) binding, whereas Thr11Ser is more markedly stabilized. The folds of the holo-variants are similar to holo-WT, with backbone rmsd with respect to WT of 0.9 Å and 0.8 Å for Thr11Ala and Thr11Ser, respectively (Figure 3.15B).

The S(Cys)-S(Cys) distance distribution is not greatly altered in the Thr11Ala apo-mutant: during most of the simulation time (~70 %) is ~5.5 Å, which is similar to the distance in WT (Figure 3.13A). On the contrary, the Thr11Ser mutant exhibits most of the time (~60 %) a shorter S-S distance of ~3.5 Å. We previously proposed that a
‘stretched’ Cys conformation, in which the S(Cys)-S(Cys) is ~8 Å, may be important for initial Cu(I) uptake. Shorter S(Cys)-S(Cys) distances, like the ones found here for Thr11Ser, may increase the probability of formation of a disulfide bond.

Figure 3.15. Superimposition of the final equilibrated structures of Atox1 WT (black), Thr11Ala (green) and Thr11Ser (yellow) in the apo- (A) and holo- (B) forms. A blow-up of the Cu-binding loop is shown to the right; Met10, Thr/Ala/Ser11, Cys12, Cys15, Lys60 and Cu(I) for each variant are labeled.
Thr11Ala lacks the ability of residue 11 to HB with Cys12 or Lys60, whereas Thr11Ser retains this ability. Similarly to Met10Ala but different from WT Atox1, in Thr11Ser, Ser11 forms a HB with Cys12 in the apo-form (Figure 3.13B). Formation of this HB in apo-Thr11Ser results in a different orientation of Cys12: this can explain the reduced S(Cys)-S(Cys) distance observed in this mutant. The Ser11-Cys12 interaction appears to be the principal cause of all alterations found for apo-Thr11Ser but not for apo-Thr11Ala (see also Cys exposure below). In holo-WT, Thr11 HBs to Lys60 ~96% of the simulation time; this facilitates an interaction between Thr11 and Cys12 (Figure 3.10A, B). Although Thr11Ser is still capable of forming HBs, in this case the HB between Ser11 and Lys60 is only present ~55% of the time. This in turn weakens the electrostatic interaction between Ser11 and Cys12 in holo-Thr11Ser, and they interact only ~50% versus ~90% of the time in WT (Figure 3.13C). The interaction between Lys60 and Cys15 found in holo-WT is not affected in Thr11Ser. However, in Thr11Ala, Lys60 sidechain fluctuates further away from the Cu site, which weakens the HB between this residue and Cys15 (Figure 3.13E). Thus, it appears that both Thr11 mutants result in weakening of the electrostatic network surrounding the Cu-center, which may destabilize the Cu-bound state.

In the apo-form, Cys solvent exposure changes only in Thr11Ser mutant: whereas Cys12 is more buried, Cys15 is significantly more exposed as compared to apo-WT (Figure 3.14A, C). These findings correlate well with the altered S(Cys)-S(Cys) distance distribution observed only for the Thr11Ser variant. In the holo-form, solvation of the Cu-coordinating Cys residues is not altered by mutation of Thr11 (Figure 3.14B, D).
Because of this, in Thr11Ser Cys15 becomes more buried, instead of more exposed upon Cu(I) binding.

### 3.4.1.3 Lys60Ala and Lys60Tyr Atoxl variants

The Lys60 mutants equilibrate quite quickly, except for apo-Lys60Tyr, which required longer time (Figure 3.11A, B and Table 3.3). Similar to the other Atoxl mutants, the greatest effect of the Lys substitutions is observed in the absence of Cu(I). For both apo-forms, the proteins become less flexible, especially in the Cu loop and helix α1 (Figure 3.11C). Lys60Ala flexibility is similar to that of the Thr11 mutants and is dramatically less than WT; the effect extending throughout the entire protein. On the other hand, Lys60Tyr flexibility is closer to the WT, except in the Cu loop, which is more rigid. Nonetheless, the folds of the apo-mutants are the same as WT, with backbone rmsd of only 0.9 Å and 1.3 Å for Lys60Ala and Lys60Tyr, respectively (Figure 3.16A).

In WT Atoxl, even though Lys60 does not HB to Cys15 and has increased conformational flexibility in the apo-form, its sidechain fluctuates near the Cu loop. Presence of this bulky, charged and flexible residue towards the Cu loop seems to increase the entropy of this loop in apo-WT. On the other hand, in apo-Lys60Ala, the sidechain of Ala60 points towards the protein core; this appears to rigidify the backbone. In apo-Lys60Tyr, Tyr60 is also close to the Cu loop, like Lys60 in WT, but it has reduced flexibility and interacts extensively with Ala18 and Cys15 in helix α1 (Figure 3.13D). This stable and strong HB between Tyr60 and Cys15 in apo-Lys60Tyr greatly stabilizes the Cu loop and the beginning of helix α1 but does not appear to affect much the core.
Thus, addition of Ala in position 60 affects the whole protein whereas addition of Tyr has mainly a local (but strong) effect in the loop.

Figure 3.16. Superimposition of the final equilibrated structures of Atox1 WT (black), Lys60Ala (red) and Lys60Tyr (cyan) in the apo- (A) and holo- (B) forms. A blow-up of the Cu-binding loop is shown to the right; Met10, Thr11, Cys12, Cys15, Lys/Ala/Tyr60 and Cu(I) for each variant are labeled.
In the presence of Cu(I), both Lys mutants appear rather similar to WT in terms of backbone fluctuations except for some regions (Figure 3.11D). Both holo-mutants have restricted motion of residues 40 to 50, which was also observed for the other Atox1 mutants. Also, the holo-form of Lys60Ala exhibits increased flexibility in loop α1-β2, resulting in a greater mobility of this region upon Cu(I) binding, as opposed to WT and Lys60Tyr. On the other hand, the Cu loop is stabilized in Lys60Ala upon Cu(I) binding but not as much as in Lys60Tyr, which is more WT-like. Thus, it appears that whereas an Ala in position 60 is not tolerated, the prokaryotic residue, Tyr, can maintain part of the WT behavior. Structurally, both holo-mutants are identical to holo-WT, with backbone rmsd values of 0.7 Å and 0.6 Å for Lys60Ala and Lys60Tyr, respectively (Figure 3.16B).

In apo-Lys60Ala the S(Cys)-S(Cys) distance is not much perturbed as compared to in WT (Figure 3.13A). On the other hand, in apo-Lys60Tyr, most of the simulation time (~85 %), the S-S distance is shorter, ~3.5 Å. In Lys60Tyr, but not in Lys60Ala, a HB is formed between Thr11 and Cys12 in the apo-form (Figure 3.13B). As in the Thr11Ser mutant, formation of this unique HB in apo-Lys60Tyr results in a different conformation of Cys12, which may explain the altered Cys-Cys distance distribution in this mutant (and not in Lys60Ala). On the other hand, there is a weakening of such electrostatic interaction around the Cu-center in holo-Lys60Tyr as compared to holo-WT (Figure 3.13C). Lys60Ala is unable to provide a HB to Cys15, and thus this holo-variant lacks an important electrostatic component that stabilizes the Cu-center. Whereas Tyr60 in holo-Lys60Tyr is capable of forming such bond, it does not occur as frequently as in holo-WT (Figure 3.13E). However, this interaction is present in apo-Lys60Tyr and not
in apo-WT (Figure 3.13D), which contributes to stabilize the Cu loop as mentioned above. Altogether, these findings point to a role of Lys60 in preserving an electrostatic network around the Cu site of holo-Atox1.

Solvation of the Cu-coordinating Cys residues changes in the two Tyr60 mutants. In the apo-forms, Cys12 is less solvent exposed in both mutants as compared to apo-WT (Figure 3.14A). Because Cys12 is proposed to be the first ligand to bind Cu(I) (Banci et al., 2006b), shielding of this residue may affect Cu-binding parameters. Whereas Cys15 exposure is not affected in apo-Lys60Tyr, it becomes more exposed in apo-Lys60Ala (Figure 3.14C). In the presence of Cu(I), Cys15 is much more exposed to solvent in both mutants (Figure 3.14D), accompanied by the gain of a HB with water. As a consequence, Cu(I) is less shielded in these mutants; this makes the Cu-Cys15 bond more susceptible for attack by water or another protein.

3.4.1.4 Cross-correlation analysis of Atox1 variants

The cross-correlation matrices of apo- and holo-forms of each Atox1 variant were also calculated (data not shown). The data indicate that all positive cross-correlations that maintain the ferredoxin-like fold in WT Atox1 are conserved in the variants, consistent with the small effects of all mutations seen in the rmsd time evolution data and fold. Interestingly, in all the apo-mutants, the out-of-phase coupling between structural elements within the blue region (involving residues in the Cu loop side of the protein) (Figure 3.6A) is completely (in Thr11Ala, Thr11Ser and Lys60Ala) or nearly (in Met10Ala and Lys60Tyr) lost. Also, none of the holo-mutants exhibit any significant negative cross-correlations. These results demonstrate that the specific pattern of anti-
correlation between residues in the Cu loop and nearby residues in WT apo-Atoxl is disrupted by the mutations. Because all mutations alter the Cu loop flexibility to some extent, we can conclude that the intrinsic flexibility of this loop in apo-Atoxl is necessary to induce specific concerted motions, which may be necessary for Cu(I) binding.

3.4.1.5 Biological relevance

We have assessed the consequences of removing, one by one, three residues in Atoxl. None of the residues are interacting with the Cu(I); nonetheless, they are all conserved. We have simulated Met10Ala, Thr11Ala, Thr11Ser, Lys60Ala and Lys60Tyr variants of Atoxl in apo- and holo-forms, resulting in a total simulation time of ~600 ns, to obtain a detailed description of the effects of Met10, Thr11 and Lys60 on Cu loop properties, overall protein flexibility, S(Cys)-S(Cys) distance, solvent exposure of the functional Cys and correlated protein fluctuations. The most important differences found in the variants with respect to WT Atoxl are summarized in Table 3.4.

Interestingly, all substitutions in Atoxl resulted in subtle effects, affecting mainly the Cu loop structure and dynamics but the overall fold is conserved. This points to the robustness of Atoxl ferredoxin-like fold, and is consistent with purified Met10Ala, Lys60Ala (Hussain et al., 2008), Lys60Tyr and Thr11Ala (Hussain et al., 2009) Atoxl variants having similar far-UV circular dichroism (CD) and fluorescence emission signals as WT Atoxl. We find here that, in terms of backbone flexibility, all substitutions have the largest effects in the apo-form. It appears thus that the binding of Cu(I) to the proteins overrules structural or dynamic “defects” in the apo-variants.
Table 3.4. Summary of most significant changes in the different mutants with respect to WT Atoxl. +, increase; -, decrease with respect to WT of backbone flexibility (rmsf), length of Cys-Cys distance and Cys solvent exposure. WT means wild type-like behavior of that property. In apo-WT, there is no HB between T11 and C12 and no HB between K60 and C15. In holo-WT, there is a HB between T11 and K60, electrostatic interaction between T11 and C12 and a HB between K60 and C15. In both WT forms, M10 interacts with the hydrophobic core. NA: not applicable.

<table>
<thead>
<tr>
<th></th>
<th>rmsf</th>
<th>Cys-Cys distance</th>
<th>interactions</th>
<th>Cys exposure</th>
</tr>
</thead>
<tbody>
<tr>
<td>apo</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Met10Ala</td>
<td>+</td>
<td>+</td>
<td>gain of HB T11-C12/loss of core interactions</td>
<td>C12 +</td>
</tr>
<tr>
<td>Thr11Ala</td>
<td>-</td>
<td>WT</td>
<td>WT</td>
<td>C15 +</td>
</tr>
<tr>
<td>Thr11Ser</td>
<td>-</td>
<td>-</td>
<td>gain of HB S11-C12</td>
<td></td>
</tr>
<tr>
<td>Lys60Ala</td>
<td>-</td>
<td>WT</td>
<td>WT</td>
<td></td>
</tr>
<tr>
<td>Lys60Tyr</td>
<td>-</td>
<td>-</td>
<td>gain of HB T11-C12 and Y60-C15</td>
<td>C12 -</td>
</tr>
<tr>
<td>holo</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Met10Ala</td>
<td>+</td>
<td>NA</td>
<td>loss of core interactions</td>
<td>WT</td>
</tr>
<tr>
<td>Thr11Ala</td>
<td>-</td>
<td>NA</td>
<td>loss of HB T11-K60 and T11-C12/weaker HB K60-C15</td>
<td>WT</td>
</tr>
<tr>
<td>Thr11Ser</td>
<td>-</td>
<td>NA</td>
<td>weaker HB S11-K60 and S11-C12</td>
<td>WT</td>
</tr>
<tr>
<td>Lys60Ala</td>
<td>-</td>
<td>NA</td>
<td>loss of HB K60-C15</td>
<td>C15 +</td>
</tr>
<tr>
<td>Lys60Tyr</td>
<td>-</td>
<td>NA</td>
<td>weaker HB T11-Y60, T11-C12 and Y60-C15</td>
<td>C15 +</td>
</tr>
</tbody>
</table>

Despite individual differences, all holo-forms are less floppy than the corresponding apo-forms, as also found for WT Atoxl. This reduced flexibility is accompanied by a gain of interactions between specific residues near the Cu site, stabilizing the Cu-bound state. This correlates with in vitro thermal and chemical unfolding experiments on purified proteins, which indicate that Met10Ala, Thr11Ala,
Lys60Ala and Lys60Tyr Atoxl variants are more stable in the presence of Cu(I) (Hussain et al., 2009), as was earlier found for WT Atoxl (Hussain and Wittung-Stafshede, 2007).

Each Atoxl variant studied here has a unique in silico behavior, which allows dissection of the role of each residue in Atoxl structure and dynamics. Nonetheless, general trends among the variants also emerge from our simulation data. Deletion of Met10 in Atoxl disrupts the core and the Met10Ala variant is more floppy than WT. We find that Met10 participates in several hydrophobic interactions in the core, which stabilizes the overall fold. Proper packing of the fold in turn contributes to an optimal Cu-binding loop conformation. In agreement with our MD data, in vitro studies on purified Met10Ala Atoxl in apo- and holo-forms demonstrate that both forms have lower thermal and chemical stability than WT Atoxl (Hussain et al., 2009). Moreover, the increase in backbone and Cu loop flexibility of Met10Ala may explain the reported increase in both rate and extent of Met10Ala Atoxl dissociation from an Atoxl-Cu-bicinchoninic acid (BCA) ternary complex (Hussain et al., 2008). It is tempting to hypothesize that the added flexibility and altered loop conformation in this mutant may hinder the docking of a bulky molecule, explaining the lower extent of formation of the Atoxl-Cu-BCA ternary complex in vitro in Met10Ala Atoxl (Hussain et al., 2008). Interestingly, Met10Ala Atoxl mutant has its Cys residues much further away from each other, which predicts that the Cys will have to re-arrange considerably to be able to coordinate Cu(I). This finding is consistent with the reported lower rate of binding of the Cu-BCA complex to apo-Met10Ala Atoxl (Hussain et al., 2008).

To our surprise, the Thr11 and Lys60 variants are more rigid than WT Atoxl. This finding suggests that presence of these two residues increase the entropy of WT
Atoxl; in particular, they significantly increase the backbone fluctuations of the Cu loop. Thr11 has its hydrophobic methyl group (CG) exposed to solvent and the Lys60 sidechain fluctuates close to the Cu loop; both these features seem to contribute to increasing Cu loop flexibility in apo-WT. Proper plasticity of the Cu loop and thereby an appropriate distance between and solvent exposure of the two Cys may be necessary for efficient Cu(I) uptake in vivo. In vitro, Lys60Ala (Hussain et al., 2008), Lys60Tyr and Thr11Ala (Hussain and Wittung-Stafshede, unpublished results) Atoxl variants exhibit a lower rate of binding of the Cu-BCA complex to the apo-protein, as also observed for Met10Ala (Hussain et al., 2008). Thus, a fine-tuning of the Cu loop flexibility and conformation in the apo-form of Atoxl, in which the Cys need to pre-organize to be able to receive the metal, appears to be important for Cu(I) uptake: a loop too flexible decreases the binding of an incoming Cu-BCA molecule, as in Met10Ala, but a loop too rigid also decreases the binding of an incoming Cu-BCA molecule, as in Thr11Ala and Lys60Ala/Tyr.

Though the effect is more pronounced in the apo-forms, in the presence of Cu(I), all mutants except Met10Ala are less flexible than WT. In the holo-form, the Cys are already bound to the Cu(I) atom, so there is no need to pre-organize the Cu site for the arrival of an incoming acceptor, such as BCA or a partner domain. In this case, whereas increase flexibility of the Cu loop may obstruct the docking of a bulky molecule to holo-Atoxl, reduced flexibility may result in the opposite effect, facilitating the docking. The higher rigidity of the Cu loop found here for the Thr11 and Lys60 variants may thus explain the faster rate and higher extent of a BCA molecule to bind to holo-Thr11Ala (Hussain and Wittung-Stafshede, unpublished results) and holo-Lys60Ala (Hussain et al.,
2008). However, similar experiments on Lys60Tyr mutant indicate that whereas the speed and extent of BCA binding to the holo-protein is reduced, the extent and rate of protein dissociation from the Atox1-Cu-BCA ternary complex is increased (Hussain and Wittung-Stafshede, unpublished results). Although we do observe a reduction in backbone fluctuations in this mutant, presence of the bulky Tyr sidechain near the Cu loop may impede the accommodation of a BCA molecule.

We have found that replacing of Thr11 and Lys60 for the corresponding prokaryotic residues results in gaining of interactions within the Cu loop in the apo-form, which are lacking in the WT protein and in the other mutants. In agreement, apo-Lys60Tyr was found to have higher thermal and chemical stability than apo-WT \textit{in vitro} (Hussain et al., 2009). Although not measured experimentally, since we have found that apo-Thr11Ser follows a similar trend in our simulations, we predict that this apo-protein will also be more stable than WT \textit{in vitro}.

Furthermore, from the simulation data on the holo-forms of the Thr and Lys mutants, it is clear that these residues are important in forming an electrostatic network around the Cu site. Both Thr11 and Lys60 stabilize the Cu-bound state by forming HBs with key residues in the metal-binding loop. Loss or weakening of these interactions in the holo-mutants agrees with the lower thermal and chemical stabilities (Hussain et al., 2009) and lower affinity for Cu(I) found for all holo-mutants \textit{in vitro} (Thr11Ser mutant was not tested experimentally) (Hussain and Wittung-Stafshede, unpublished results). Interestingly, Lys60 appears to protect Cys15 from solvent in the holo-form, shielding the Cu site. It appears that proper solvation of the Cys residues that coordinate Cu(I) is important to facilitate Cu(I) uptake and release. The Cys should be sufficiently solvent
exposed in the apo-form to assure timely uptake of Cu(I), but buried enough in the holo-
form to avoid premature loss of Cu(I).

Although the prokaryotic residues, Ser and Tyr, at first sight appear as good
replacements for the corresponding eukaryotic residues, Thr11Ser and Lys60Tyr variants
of Atox1 did not behave WT-like. In both cases, the prokaryotic residue is tolerated better
than an Ala substitution in the holo-form, whereas Ala is a better replacement in the apo-
form. In the apo-forms, Thr11Ser and Lys60Tyr are involved in HBs that are absent in
WT, Thr11Ala and Lys60Ala, which alter dynamics and positioning of the Cys. In the
holo-forms however, Ser11 and Tyr65, but not Ala11/60, can still provide a certain
degree of electrostatic stabilization of the Cu loop. This may explain presence of His in
position X1 in MBD3 of ATP7A and ATP7B (Arnesano et al., 2002). Importantly, in both
Thr11Ser and Lys60Tyr, but not in any other variant, the Cys-Cys distance shortened in
the apo-form. We previously speculated on a ‘stretched’ Cys conformation that would
allow for easy “pull-in” of Cu(I). It is possible that Thr11 and Lys60 play roles in
optimizing the Cys-Cys distance distribution in Atox1 by their interactions with the Cys
residues. Differences in Cu-transfer mechanisms, protein-partners and protein-metal
affinities, may explain the presence of Ser and Tyr in bacterial Cu chaperones.

Finally, we previously showed that in WT forms of Atox1 and CopZ the β3-α2
loop, which is located at the opposite side of the Cu loop, was the only region that
becomes more floppy upon Cu(I) binding and we proposed that conformational changes
in this loop may mediate protein-protein interactions. In all Atox1 mutants tested here, in
contrast to WT Atox1, this region either did not change at all, or became more structured,
upon Cu(I) binding. This may suggest that the studied residues, in addition to facilitating
Cu(I) uptake and release, are important in providing appropriate protein-protein interactions.

### 3.4.2 Tuning of copper loop flexibility in CopZ: Role of Met11, Ser12 and Tyr65

#### 3.4.2.1 Met11Ala CopZ variant

The holo-form of Met11Ala undergoes significant conformational changes and has increased flexibility as compared to apo-Met11Ala (Table 3.5 and Figure 3.17A, B).

**Table 3.5.** Total simulation time (ns) and backbone rmsd (Å) for the total (rmsd\text{total}, with respect to the first structure) and last 20 ns of the simulation (rmsd\text{20}, with respect to the average structure).

<table>
<thead>
<tr>
<th></th>
<th>time</th>
<th>rmsd\text{total}</th>
<th>rmsd\text{20}</th>
</tr>
</thead>
<tbody>
<tr>
<td>apo</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CopZ-Met11Ala</td>
<td>118</td>
<td>2.4 ± 0.2</td>
<td>0.7 ± 0.1</td>
</tr>
<tr>
<td>CopZ-Ser12Ala</td>
<td>113</td>
<td>3.1 ± 0.4</td>
<td>1.0 ± 0.2</td>
</tr>
<tr>
<td>CopZ-Ser12Thr</td>
<td>111</td>
<td>5 ± 1</td>
<td>1.8 ± 0.5</td>
</tr>
<tr>
<td>CopZ-Tyr65Ala</td>
<td>122</td>
<td>3 ± 1</td>
<td>1.0 ± 0.2</td>
</tr>
<tr>
<td>CopZ-Tyr65Lys</td>
<td>107</td>
<td>2.3 ± 0.3</td>
<td>0.7 ± 0.1</td>
</tr>
</tbody>
</table>

| holo      |      |                  |               |
| CopZ-Met11Ala | 120  | 3.1 ± 0.6        | 0.7 ± 0.2     |
| CopZ-Ser12Ala | 113  | 4.8 ± 0.4        | 1.2 ± 0.2     |
| CopZ-Ser12Thr | 120  | 3.5 ± 0.4        | 1.2 ± 0.2     |
| CopZ-Tyr65Ala | 131  | 5 ± 2            | 2.1 ± 0.6     |
| CopZ-Tyr65Lys | 120  | 2.6 ± 0.2        | 0.8 ± 0.1     |

**Figure 3.17.** Rmsd (in Å) of the backbone heavy atoms as a function of simulation time for the different CopZ variants in apo- (A) and holo- (B) forms. Blue, Met11Ala; green, Ser12Ala; orange, Ser12Thr; red, Tyr65Ala; cyan, Tyr65Lys.
In fact, after ~80 ns, holo-Metat11Ala undergoes a conformational change that is not present in the apo-form. This is evidenced by a "jump" in the rmsd time evolution, and is consistent with Met11 being exposed in the apo-form, not making any stabilizing interactions whereas in the holo-form it significantly stabilizes the core. Removal of this residue is thus likely more tolerated in the absence of Cu(I). During the last 20 ns of simulation, however, both forms are stable, suggesting no detectable conformational changes (Table 3.5). Whereas the apo-form of Met11Ala has significantly more restricted backbone than apo-WT, the mutant is more floppy than WT in the holo-form (Figure 3.18A, B). The net effect is that instead of becoming more structured, as in WT, Met11Ala gains flexibility upon Cu(I) binding. The most affected regions are residues 10 to 40, which include the Cu loop, helix a1, strand b2 and connecting loops.

Interestingly, even though Met11 is solvent exposed in apo-WT, it still has a significant effect on the backbone dynamics of the whole protein. An overlay of the equilibrated apo-structures of WT and Met11Ala revealed a backbone rmsd of 3.4 Å and significant structural differences (Figure 3.19A). It appears that solvent exposure of the bulky hydrophobic Met11 sidechain in apo-CopZ is important to assure flexibility of the Cu loop. Replacement of Met11 by Ala rigidifies the protein because the entropic effect is absent. In apo-WT, Met11 and Ser12, although highly mobile, interact with Ala39 and Glu38, located in the b2-b3 loop. This interaction probably assures that apo-CopZ is flexible enough to bind Cu(I) but not to destabilize the fold. On the other hand, in Met11Ala mutant, presence of the small Ala11 results in lack of an interaction with the b2-b3 loop, which weakens the interface between these loops. This loop-loop interaction helps maintain the ferredoxin-like fold in WT. In its absence, the Cu loop in Met11Ala
forms a turn that folds into the protein, markedly restricting the motion of the Cu(I)-coordinating Cys and completely burying Cys13 (Figure 3.19A).

Figure 3.18. Average fluctuations (rmsf in Å) of backbone heavy atoms per residue for the different CopZ variants and WT in apo- (A, C, E) and holo- (B, D, F) forms. The secondary structure elements are indicated. Black, WT; blue, Met11Ala; green, Ser12Ala; orange, Ser12Thr; red, Tyr65Ala; cyan, Tyr65Lys.
Figure 3.19. Superimposition of the final equilibrated structures of CopZ WT (black) and Met11Ala (blue) in the apo- (A) and holo- (B) forms. **Left:** overall structure. **Center:** blow-up of residues 7 to 15 and 35 to 43; Met/Ala11 and Ser12 (A) and Met/Ala11 (B) for each variant are labeled. **Right:** blow-up of residues 10 to 19 and 62 to 68 (A), and 8 to 17 and 62 to 68 (B); Met/Ala11, Ser12, Cys13, Cys16, Tyr65 and Cu(I) for each variant are labeled.
Holo-Met11Ala exhibits important structural changes comparing to holo-WT. In the mutant, strand β1 and helix α1 are shortened, so the Cu loop extends from residues 8 to 15 versus residues 10 to 13 in WT. This change increases the flexibility of the coordinating Cys (Figure 3.18B). The mutant is destabilized because important hydrophobic interactions in the protein core involving Met11 are absent. Because of the lack of these interactions in holo-Met11Ala, after ~80 ns of simulation the protein undergoes a large conformational change in which the interaction interface between strands β2 and β3 and helix α1 is significantly compromised, opening up the whole structure. Overlay of the final equilibrated structures shows these structural differences and reveals a backbone rmsd of 4.1 Å between WT and mutant structures (Figure 3.19B). Altogether, the destabilization observed in Met11Ala demonstrates the importance of Met11 in maintaining the integrity of the fold in holo-CopZ.

The structural and dynamical defects observed in apo-Met11Ala CopZ result in an alteration of the S(Cys)-S(Cys) distance distribution. Whereas WT exhibits two populations of distances centered at ~4.5 Å and ~7 Å, Met11Ala mutant has its Cys far from each other, most of the time at ~9 Å (Figure 3.20A). This alteration may affect the efficiency of Cu(I) uptake, as the Cys will have to re-arrange significantly to be able to coordinate Cu(I). As in WT, in Met11Ala Ser12 interacts with Cys13 only in the holo-form (Figure 3.20B, C). The interaction is more stable in the holo-mutant, which may counterbalance the increased destabilization of its Cu loop. On the other hand, the electrostatic stabilization of Cys16 by Tyr65 is lost in holo-Met11Ala (Figure 3.20E).

The gain of structure in the Cu loop in apo-Met11Ala is accompanied by complete burial of Cys13, whereas Cys16 becomes more exposed (Figure 3.21A, C).
Figure 3.20. Histograms (arbitrary units) of distance distribution (in Å) between heavy atoms of Cys13 (S) and Cys16 (S) in the apo-form (A), Ser/Thr12 (O) and Cys13 (S) in apo- (B) and holo- (C) forms, and between Tyr/Lys65 (O/N) and Cys16 (S) in apo- (D) and holo- (E) forms, for the different CopZ variants and WT. Black, WT; blue, Met11Ala; green, Ser12Ala; orange, Ser12Thr; red, Tyr65Ala; cyan, Tyr65Lys.
Figure 3.21. Protein-solvent $g(r)$ of Cys13(S) in apo- (A) and holo- (B) forms, Cys16(S) in apo- (C) and holo- (D) forms, and Met11(S)/Ala11(C) in apo- (E) and holo- (F) forms for the different CopZ variants and WT. Black, WT; blue, Met11Ala; green, Ser12Ala; orange, Ser12Thr; red, Tyr65Ala; cyan, Tyr65Lys.
If Cys13 is buried it may not be able to receive Cu(I) from the solvent or a donor. It is possible that in this case Cys16, which is more exposed, will be the first ligand to interact with Cu(I); Cys13 will then have to change conformation to accommodate Cu(I) coordination. In the presence of Cu(I), whereas Cys16 mobility and exposure is not affected by the mutation, Cys13 is more flexible and becomes significantly more exposed in Met11Ala (Figure 3.21B, D). Increased accessibility of Cys13 to the solvent in the holo-form may increase the probability of Cu(I) release to the solvent or Cu-chelators.

In conclusion, because of its different conformation in apo- and holo-CopZ, Met11 contributes to protein stability and dynamics in different ways in the presence and absence of Cu(I). In apo-CopZ, Met11 is important for assuring flexibility of the Cu loop and the coordinating Cys, which may be requirements for proper Cu(I) uptake. Met11 is essential for exposing Cys13 to solvent and maintaining an appropriate Cys-Cys distance that is probably optimal for Cu(I) binding. In holo-CopZ, a buried Met11 is necessary to make key hydrophobic contacts within the protein core. This in turn protects the Cu site by not exposing Cys13 to the solvent.

3.4.2.2 Ser12Ala and Ser12Thr CopZ variants

Ser12Thr has a greater negative effect than Ser12Ala in the apo-form, but a smaller one in the holo-form. In the case of Ser12Ala mutant, the holo-form exhibited greater conformational changes (Figure 3.17 and Table 3.5). On the other hand, apo-Ser12Thr has greater deviations and conformational dynamics than the holo-form (Figure 3.17 and Table 3.5). Thus, it appears that apo-Ser12Thr and holo-Ser12Ala undergo significant conformational changes throughout the simulations, pointing to the
importance of Ser12 in CopZ structure and dynamics. We notice that apo-Ser12Thr exhibits the greatest flexibility within all apo-mutants.

The two mutations have different effects on backbone dynamics in the two protein variants. In the absence of Cu(I), Ser12Ala CopZ exhibits restricted backbone motion in the Cu loop, helix α1 and β2-β3 loop, but the rest of the protein is destabilized as compared to WT (Figure 3.18C). The most significant structural changes that occur in this mutant are an extension of helix α1 (residues 17 to 26 versus 20 to 24 in Ser12Ala and WT, respectively), a shortening of the Cu loop (residues 9 to 16 versus 10 to 19 in Ser12Ala and WT, respectively) and a complete loss of strand β4, which in WT corresponds to residues 67 to 72. Overlay of the equilibrated apo-structures reveal a backbone rmsd of 4.0 Å between Ser12Ala and WT (Figure 3.22A).

On the other hand, greater changes are observed for apo-Ser12Thr; an overlay of the equilibrated structures results in a backbone rmsd of 5.7 Å with WT (Figure 3.22A). This mutant is extremely flexible; the complete structure of this variant is destabilized as compared to WT (Figure 3.18C), and its structure is opened up (Figure 3.22A). The increased flexibility found in apo-Ser12Thr may likely be due to an entropic effect due to incorporation of an exposed methyl group at position 12. Because CopZ WT is very floppy to begin with, apparently it cannot tolerate a residue in position 12 that increases entropy further. In apo-CopZ WT, Ser12 interacts with Met11 and Glu38 in the β2-β3 loop. These interactions assure a correct positioning of Met11 allowing it to interact with the β2-β3 loop, stabilizing the fold, as previously discussed. In Ser12Thr, because of its increased flexibility, Thr12 cannot interact with Met11 and in turn Met11 cannot interact with the β2-β3 loop. This results in destabilization of this region and weakening of the
loop-loop interface as also found in apo-Met11Ala. As in apo-Ser12Ala, in Ser12Thr the secondary structure of strand β4 is completely lost.

**Figure 3.22.** Superimposition of the final equilibrated structures of CopZ WT (black), Ser12Ala (green) and Ser12Thr (yellow) in the apo- (A) and holo- (B) forms. **Left:** overall structure. **Center:** blow-up of residues 7 to 14 and 35 to 43 (A), and 10 to 14 and 37 to 41 (B); Met11 and Ser/Ala/Thr12 (A) and Met11, Ser/Ala/Thr12 and Cys13 (B) for each variant are labeled. **Right:** blow-up of residues 9 to 19 and 64 to 66 (A), and 10 to 19 and 64 to 66 (B); Cys13, Cys16, Tyr65 and Cu(I) for each variant are labeled.
As opposed to the apo-forms, the Ser12Ala change in CopZ has a greater effect than the Ser12Thr change in the presence of Cu(I). Both mutants are more flexible than WT, in particular in the Cu loop and helix α1; this is most dramatic in the Ser12Ala CopZ mutant (Figure 3.18D). In holo-WT, the interactions between Ser12, Cys13 and Tyr65 allow for proper Cu loop conformation and burial of Met11. In Ser12Ala these interactions are abolished because of the inability of Ala to HB, Tyr65 moves away from the Cu loop (Figure 3.20E), and consequently the loop becomes distorted, flexible and extended (from residue 10 to 19); in addition, the entire fold is destabilized (Figure 3.22B). This conformational change prevents interactions between Met11 and the protein core (Figures 3.21F and 3.22B), which adds to the destabilization. Also, helix α2 and strand β4 move away from the central ferredoxin-like fold, extending the overall protein size. The overlay of the final structures in this case results in a backbone rmsd of 4.6 Å with respect to WT (Figure 3.22B). In holo-Ser12Thr, Thr12 is also unable to interact with Cys13 (Figure 3.20C) and again, Met11 cannot interact with the protein core (Figures 3.21F and 3.22B). However, in this case Met11 and Tyr65 interact with each other, which assures proximity between the α2-β4 loop and the Cu loop, conserving in part the fold (Figure 3.22B). Superimposition of the final structures of holo-forms of Ser12Thr and WT reveals a backbone rmsd of 3.6 Å with respect to WT.

Whereas in apo-Ser12Ala the S(Cys)-S(Cys) distance distribution is not greatly altered compared to WT (most of the time, ~5 Å), in apo-Ser12Thr this distance increases dramatically, reaching values up to 14 Å, with two dominant peaks centered at 10 and 12 Å (Figure 3.20A). This longer distance will probably affect the efficiency of Cu(I) uptake, as in the Met11Ala mutant. Formation of a HB between residue 12 and Cys13 in
apo-Ser12Thr, which is absent in apo-WT (Figure 3.20B), may in part explain the altered Cys-Cys distance distribution. In holo-WT, an interaction between Ser12 and Cys13 serves as an electrostatic stabilization of the Cu-center, but in holo-Ser12Thr this interaction is absent (Figure 3.20C). In this way, Ser12Thr mimics the behavior of Ser12Ala, which also is unable to form this interaction. Moreover, in both mutants Tyr65 moves away from the Cu loop in the holo-forms; this is more marked in Ser12Ala (Figure 3.20E and 3.22B). Therefore, it appears that both mutations result in weakening of electrostatic interactions of the Cu-bound state.

Solvation of the functional Cys is not greatly affected in apo-Ser12Ala (Figure 3.21A, C). However, in apo-Ser12Thr solvent accessibility of both Cys residues is greatly altered. Whereas Cys13 is more buried, Cys16 is much more solvent exposed, comparing to WT (Figure 3.21A, C). This difference further explains the longer S(Cys)-S(Cys) distances found in this mutant. Cys16 is more exposed in both holo-mutants, to the largest degree in Ser12Ala CopZ (Figure 3.21D). Consequently, Cu(I) is less protected in these mutants and therefore, may be more susceptible to dissociation. The structural changes that occur in the holo-mutants mentioned above result in a completely exposure of Met11, as opposed to WT (Figure 3.21F). Solvent exposure of Met11 in Ser12Ala mimics that of apo-WT, consistent with the capability of this residue to interact with the β2-β3 loop. In contrast, Ser12Thr has its Met11 even more exposed, without making any interaction with the rest of the protein. Based on our findings, it is clear that burial of Met11 in WT holo-CopZ is key to assure a proper loop and Cys conformation.

To conclude, Thr (the eukaryotic residue) cannot substitute Ser in position 12 in CopZ. Ser12 is important for assuring a proper Cu loop conformation and dynamics of
the coordinating Cys both in the apo- and holo-forms. Also, substitution of Ser12 by either an Ala or Thr results in a destabilization of the Cu-bound state due to loss of electrostatic stabilization of the Cu-center provided by residues 12 and 65. Importantly, in these holo-mutants Met11 is not buried, which is another source of destabilization.

3.4.2.3 Tyr65Ala and Tyr65Lys CopZ variants

In this case, the presence of the eukaryotic Lys residue in position 65 of CopZ is tolerated much better than an Ala substitution. Tyr65Ala and Tyr65Lys are the least and the most stable of all mutants, respectively (Figure 3.17 and Table 3.5). Tyr65Ala exhibited enhanced structural dynamics, and that is why it was simulated longer time. After ~60 ns, it undergoes a conformational change in both apo- and holo-forms, as testified by a “jump” in the rmsd time evolution (Figure 3.17), in which the protein loses significantly secondary and tertiary structure.

Both apo-mutants have restricted motion of the Cu loop and helix α1, and consequently of the coordinating Cys (Figure 3.18E). Apart from this region, apo-Tyr65Lys has similar backbone fluctuations than WT. On the contrary, except for the loop and helix α1, apo-Tyr65Ala is more floppy than WT. In the holo-forms, whereas the Cu loop in Tyr65Lys is slightly more flexible than WT, the entire backbone of Tyr65Ala is markedly destabilized (Figure 3.18F). Backbone fluctuations in holo-Tyr65Ala are on average three times greater than WT.

Except for the Cu loop and helix α1, which gains structure, the structure of apo-Tyr65Lys is similar to WT, with a backbone rmsd of 2.3 Å (Figure 3.23A). Presence of Lys in position 65 markedly stabilizes the Cu loop and helix α1, because this residue,
instead of pointing to the solvent as Tyr65 in apo-WT, points towards the loop and interacts extensively with Cys16 (Figures 3.20D and 3.23A).

**Figure 3.23.** Superimposition of the final equilibrated structures of WT (black), Tyr60Ala (red) and Tyr65Lys (cyan) in the apo- (A) and holo- (B) forms. **Left:** overall structure. **Center:** blow-up of residues 8 to 13 and 36 to 42 (A), and 10 to 14 and 37 to 41 (B); Met11 and Ser12 for each variant are labeled. **Right:** blow-up of residues 9 to 19 and 64 to 66 (A), and 10 to 19 and 64 to 66 (B); Cys13, Cys16, Tyr/Ala/Lys65 and Cu(I) for each variant are labeled.
On the contrary, apo-Tyr65Ala is largely destabilized and presents significantly less secondary structure than WT: helix α2 and strand β4 are unfolded (Figure 3.23A). Also, the Cu loop is shortened, going from residues 10 to 15, so that now Cys16 is part of helix α1, as opposed to WT in which it is part of the loop. Because of this, helix α1 adopts a different orientation and is shifted, extending from residues 16 to 19 and 20 to 23 in apo-Tyr65Ala and apo-WT, respectively. After ~60 ns this variant undergoes a large structural change (Figure 3.17A) in which it loses tertiary contacts: helix α2 and strand β4 no longer fold against strands β1-β3. In this way, the entire fold is destabilized. In apo-WT, Tyr65 forms a strong HB with Gln63, which positions the latter for a stable interaction with its consecutive residue Asp62. These residues are all located in the α2-β4 loop. The latter interaction allows for contacts between Asp62 and Lys18 located at the end of the Cu loop. This network of interactions present in WT apo-CopZ preserves tertiary contacts and conserves the fold. In apo-Tyr65Ala these contacts are not present and thus the stability and structure of the fold is compromised, with a final backbone rmsd of 4.8 Å compared to apo-WT (Figure 3.23A).

In the presence of Cu(I), the structural changes observed for the apo-forms are attenuated in Tyr65Lys but are markedly intensified in Tyr65Ala. The equilibrated WT and Tyr65Lys holo-structures are similar, with a backbone rmsd of 1.5 Å (Figure 3.23B). Differences include lengthening of the Cu loop by one residue (strand β1 is shorter) and lost of some secondary structure in strand β4 in the mutant. On the other hand, holo-Tyr65Ala is destabilized significantly, with a final backbone rmsd with WT of 5.8 Å (Figure 3.23B). As in apo-Tyr65Ala, lack of important secondary and tertiary contacts
because of lack of Tyr65 are the consequence of a similar conformational change that occurs after \( \sim 60 \) ns (Figure 3.17B). However, in this case the overall effect is greater, because Tyr65 (or Lys65 in case of Tyr65Lys) is an essential source of electrostatic stabilization of the Cu(I) bis-thiolate center. As a consequence, helix \( \alpha_2 \) and the \( \alpha_2-\beta_4 \) loop completely lose their interface with the Cu loop and helix \( \alpha_1 \) (Figure 3.23B). Opening of the fold exposes a great portion of the protein core to the solvent, which in turn destabilizes even more the backbone (Figure 3.18F). As a result, strand \( \beta_1 \) changes orientation and moves towards helix \( \alpha_1 \), protecting, in part, the hydrophobic core. Other important structural changes include the lengthening of the Cu loop (residues 9 to 19 in Tyr65Ala versus 10 to 13 in WT) and distortion of its conformation, and destabilization and altered orientation of helix \( \alpha_1 \). All the structural changes described explain the \(~\)three-fold increment in backbone fluctuation of this mutant compared to WT (Figure 3.18F), and probably suggest that this mutant will unfold and lose Cu(I) in solution.

The distance between the coordinating Cys is not greatly altered in apo-Tyr65Lys, but it is longer in Tyr65Ala (Figure 3.20A). In the latter, these values hardly reach 4 Å, as opposed to WT and Tyr65Lys. As mentioned above, Lys65 in apo-Tyr65Lys is close to Cys16 as opposed to WT (Figure 3.20D), and this appears to be key to prevent lost of tertiary contacts, as in Tyr65Ala. In the presence of Cu(I), as opposed to Tyr65Lys and WT, Ala65 in Tyr65Ala is not able to interact with Ser12, and in turn Ser12 does not interact with Cys13 (Figure 3.20C), which may contribute further to the destabilization of this holo-mutant. As mentioned above, the same is true for the electrostatic stabilization of the Cu-bound state provided by Tyr65, which is completely absent in Tyr65Ala. Lys65 in Tyr65Lys is still able to provide this interaction (Figure 3.20E), and
that is probably why this residue is tolerated in this position. Solvation of the Cys and Met residues is not greatly altered in the Tyr65 mutants (Figure 3.21).

In summary, in the absence of Cu(I), Tyr65 is essential for CopZ structure and dynamics, assuring appropriate Cu loop conformation and dynamics that may possibly aid in Cu(I) uptake. In this case, Lys may not be a good replacement for Tyr, as apo-Tyr65Lys exhibits a loop that is more structured and rigid. On the other hand, in the presence of Cu(I), Lys can substitute Tyr in this position. Presence of a residue capable of donating a HB to Cys16, as Tyr or Lys, in position 65 stabilizes the Cu-bound state and provides a platform for the interaction between helices α2 and α1.

**3.4.2.4 Cross-correlation analysis of CopZ variants**

The cross-correlation matrices of apo- and holo-forms of each CopZ variant were also calculated (data not shown). As opposed to what we observed for Atoxl, the collective motions are largely affected in CopZ mutants. Both negative and positive correlations are altered, consistent with the fold being affected in many of the CopZ mutants. In the apo-form, the out-of-phase coupling between structural elements within the blue region (involving residues in the Cu loop side of the protein) (Figure 3.6A) is reduced in all mutants except for Ser12Thr. This is consistent with all mutants, except Ser12Thr, having reduced backbone fluctuations of the Cu loop (Figure 3.18). In the holo-form, on the other hand, all mutants retain many of the out-of-phase motions of the apo-WT protein, which are absent in holo-WT. This is also consistent with all holo-mutants having a more flexible Cu loop than WT. The most floppy mutants, apo-Ser12Thr and holo-Tyr65Ala, are also the ones that present the strongest negative
correlations. Based on these results, we can conclude that the intrinsic flexibility of the Cu loop in apo-CopZ, tuned by the presence of conserved nearby residues, is necessary to induce specific concerted motions, which may be necessary for Cu(I) binding, as we concluded on Atox1.

3.4.2.5 Comparison to human copper chaperone homolog

We have simulated Met11Ala, Ser12Ala, Ser12Thr, Tyr65Ala and Tyr65Lys variants of CopZ in apo- and holo-forms until structural equilibrium was reached, resulting in an impressive total simulation time of ~1.2 μs. We have provided a detailed description of the effects of Cu loop structure and dynamics, overall protein flexibility, S(Cys)-S(Cys) distance, solvent exposure of key residues and correlated motions as a function of these mutations. The most important differences with respect to CopZ WT are summarized in Table 3.6.

CopZ mutants are much more unstable than the corresponding mutants in Atox1, and required longer simulations times to reach equilibrium. This is consistent with CopZ WT being a more floppy and unstable protein than Atox1 WT as evidenced by our previous MD simulations and in vitro protein stabilities (Hussain and Wittung-Stafshede, 2007). As opposed to in Atox1, where the changes are subtler and affect mainly the Cu loop structure and dynamics, mutations in CopZ have greater effects and alter significantly the protein structure regardless the presence of Cu(I). This suggests that the human protein is able to tolerate mutations much better than the bacterial one, at least at the structural level. As in Atox1, each CopZ variant studied here has a unique behavior, which allows dissection of the role of each residue in CopZ structure and dynamics.
Comparison with the corresponding mutations in Atoxl can further clarify the conservation of certain residues (like Met11) and the divergence of others (like Ser12 and Tyr65) in bacterial and human Cu chaperones.

Table 3.6. Summary of most significant changes in the different mutants with respect to WT CopZ. +, increase; -, decrease with respect to WT of backbone flexibility (rmsf), length of Cys-Cys distance and Cys solvent exposure. -/+ represents a mix of decreased flexibility in some regions and increased dynamics in other regions of the protein. WT means wild type-like behavior of that property. In apo-WT, there is no interaction between S12 and C13 and between Y65 and C16, and M11 is exposed. In holo-WT, there is an interaction between S12 and C13, S12 and Y65, and Y65 and C16; M11 is buried and interacts with the hydrophobic core. NA: not applicable.

<table>
<thead>
<tr>
<th></th>
<th>rmsf</th>
<th>Cys-Cys distance</th>
<th>interactions</th>
<th>Met/Cys exposure</th>
</tr>
</thead>
<tbody>
<tr>
<td>apo</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Met11Ala</td>
<td>-</td>
<td>+</td>
<td>weaker interaction between the Cu and β2-β3 loops</td>
<td>C13 -, C16 +</td>
</tr>
<tr>
<td>Ser12Ala</td>
<td>+/-</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
</tr>
<tr>
<td>Ser12Thr</td>
<td>+</td>
<td>+</td>
<td>weaker interaction between the Cu and β2-β3 loops, gain of S12-C13 interaction</td>
<td>C13 -, C16 +</td>
</tr>
<tr>
<td>Tyr65Ala</td>
<td>+/-</td>
<td>+</td>
<td>loss of core, and α2-β4 and β1-β3 interactions</td>
<td>WT</td>
</tr>
<tr>
<td>Tyr65Lys</td>
<td>-</td>
<td>WT</td>
<td>gain of K65-C16 interaction</td>
<td>WT</td>
</tr>
<tr>
<td>holo</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Met11Ala</td>
<td>+</td>
<td>NA</td>
<td>loss of core and Y65-C16 interactions</td>
<td>C13 +</td>
</tr>
<tr>
<td>Ser12Ala</td>
<td>+</td>
<td>NA</td>
<td>loss of core, S12-C13 and Y65-C16 interactions</td>
<td>M11 +, C16 +</td>
</tr>
<tr>
<td>Ser12Thr</td>
<td>+</td>
<td>NA</td>
<td>loss of core, S12-C13 and Y65-C16 interactions</td>
<td>M11 +, C13 +, C16 +</td>
</tr>
<tr>
<td>Tyr65Ala</td>
<td>+</td>
<td>NA</td>
<td>loss of core, α2-β4 and α1 and Cu loop, S12-C13, S12-Y65 and Y65-C16 interactions</td>
<td>WT</td>
</tr>
<tr>
<td>Tyr65Lys</td>
<td>+</td>
<td>NA</td>
<td>WT</td>
<td>WT</td>
</tr>
</tbody>
</table>
An important difference between CopZ and Atoxl is the behavior of Met11/10: always buried in Atoxl but only buried in holo-CopZ. Similar to the role of Thr11 in apo-Atoxl (section 3.4.1.5), solvent exposure of the hydrophobic sidechain of Met11 in apo-CopZ may serve as an entropic effect that increases the flexibility of this protein, thus explaining in part the difference in floppiness between Atoxl and CopZ. This is evidenced by the fact that whereas replacement of Met11 by Ala in apo-CopZ significantly reduces backbone motion, Met10Ala mutation in apo-Atoxl has the opposite effect. It thus appears that exposure of Met11 in apo-CopZ has evolved as a way of assuring protein and Cu loop flexibility possibly to facilitate Cu(I) uptake. On the other hand, some similarities are found between the role of Met11 in CopZ and Met10 in Atoxl. In both cases, Met is important for proper Cu loop dynamics and Cys exposure. In the apo-form, improper packing of the protein core as a consequence of Met removal alter the Cu loop dynamics significantly (although in opposite ways), and consequently affect the S(Cys)-S(Cys) distance distribution for optimal Cu(I) binding, as both Met mutants have their Cys much further away from each other. Because the Cys will have to significantly re-arrange to bind Cu(I) in these mutants, this is predicted to affect Cu(I) uptake. In the holo-form, the behavior of the conserved Met is similar in both CopZ and Atoxl. In both cases the Met assures stability of the fold by participating in several hydrophobic contacts with the protein core. In fact, several residues located in strands β2 and β3 that interact extensively with Met11, including Val35, Leu37 and Val42, are conserved between CopZ and Atoxl (Val35 in CopZ is replaced by another hydrophobic residue in Atoxl, Ile33, see Figure 3.9), pointing to the conservation of a network of interactions that contributes to the ferredoxin-like fold. This suggests that Met shares the
same function in both holo-proteins: to contribute to the stability of the fold. Thus, it appears that even if Met is conserved at the sequence level between prokaryotes and eukaryotes, its location and dynamics in the tertiary structure has adapted to the intrinsic “needs” of each system. Apo-CopZ may need to have greater conformational flexibility than Atox1 to interact with its partner proteins in vivo, and exposure of Met11 may have evolved to provide this entropic resource. On the other hand, Atox1 may be flexible enough for interaction with its partners, and exposure of Met is therefore detrimental for its stability.

As in Atox1, Ser12 and Tyr65 CopZ apo-variants (except for apo-Ser12Thr, see below) have also reduced flexibility of the Cu loop compared with apo-WT. The increased flexibility of the Cu loop due to the presence of Ser12 and Tyr65 in WT CopZ can be explained with similar arguments to those of Atox1. Ser12 and Tyr65 do not maintain any stable interaction in the apo-form, instead their sidechains are highly mobile that fluctuate close to the Cu loop. On the other hand, in agreement to what we observed between the apo-forms of WT and Thr11Ser Atox1, addition of a methyl group in the sidechain of residue 12 in CopZ (as in Ser12Thr variant) further increases the flexibility of the backbone, as the exposed hydrophobic group serves as an entropic booster.

Ser12 and Tyr65 are important stabilizing factors for holo-CopZ, as Thr11 and Lys60 in holo-Atox1, because both residues participate in an electrostatic network of interactions that ultimately stabilize the Cu-bound state. In CopZ, it appears that presence of a HB donor (Tyr, or Lys as tested here) in position 65 is essential for maintaining the integrity of the holo-protein. Holo-Tyr65Ala variant is extremely unstable in our MD simulations, and we predict that this protein will probably unfold and lose the Cu(I) in
vitro, if it is able to bind it in the first place. Surprisingly, Ser12 is essential to maintain Met11 buried in holo-CopZ, as mutation to either Ala or Thr completely exposes this residue to the solvent. Burial of Met11 in holo-CopZ is the most significant change that occurs upon Cu(I) binding and contributes to the stabilization of the Cu-bound state by providing core interactions.

As in Atoxl, Ser12Thr and Tyr65Lys variants of CopZ did not behave WT-like. The eukaryotic residue, Lys, is tolerated better than Ala in position 65 in both apo- and holo-CopZ. In the holo-form, Lys is able to keep the electrostatic network, but since this residue interacts with Cys16 also in the apo-Tyr65Lys variant, it stiffens the Cu loop in the apo-form. As noted above, the intrinsic flexibility of the Cu loop in apo-CopZ WT may be necessary for efficient Cu(I) uptake and/or interaction with its partner proteins in vivo. Therefore, the reduction of loop flexibility in Tyr65Lys may affect these abilities. On the other hand, whereas Thr12 was tolerated better than Ala in holo-CopZ, Ala was a better replacement in the apo-form, as observed for residue 11 in Atoxl. In the apo-form, Ser12Thr is the most flexible of all CopZ variants. These changes are possibly the result of an entropic effect due to the exposure of the methyl group of Thr12 in apo-Ser12Thr, just as appears to happen for apo-Atoxl WT. In Atoxl WT, in which Met10 is completely buried in the core, presence of Thr in position 11 assures proper Cu loop dynamics that would otherwise be too rigid for Cu(I) uptake. In apo-CopZ, however, exposure of Met11 significantly increases Cu loop fluctuations, so presence of another residue that increases backbone entropy further in position 12 is not tolerated and the fold is compromised. Furthermore, similar to what was seen with the Thr11Ser and Lys60Tyr mutants in apo-Atoxl, the Ser12Thr and Tyr65Lys mutants in apo-CopZ gain interactions
within the Cu loop. In the holo-form, whereas Tyr65Lys is able to retain key interactions within the Cu site that are found in the WT, involving Ser12, Cys13, Cys16 and residue 65, neither of the two Ser12 variants nor Tyr65Ala are able to provide such interactions. Thus, it appears that residues in position 12 and 65 in CopZ, and 11 and 60 in Atox1, have evolved differently to tune the flexibility of the Cu loop in the apo-form and provide electrostatic stabilization of the holo-form.

3.5 Conclusions

In this Chapter we have studied two homologous Cu chaperones that, although sharing the same fold and similar Cu-binding motifs, have little sequence identity and significantly different behavior in vitro. First, via QM-MM calculations, we found that despite sharing the same ligands and similar fold, the optimized Cu(I) geometries differ between the two proteins. Next, we have explored the conformational dynamics of Atox1 and CopZ WT and selected mutants, to understand conservation and diversity of key residues around the Cu loop. We conclude that residues in Cu chaperones have evolved to maintain a delicate balance of the Cu-binding loop flexibility and core rigidity as well as electrostatic stabilization. This fine tuning may likely be important for Cu(I) uptake and release, as corroborated experimentally for Atox1 (Hussain et al., 2008), and for interaction with partners, as we will see in Chapter 5. Our findings emphasize the importance of Cu-binding loop flexibility for Atox1 and CopZ biological activity. This property may be required in order to form Cu-bridged ternary complexes with target domains in vivo. NMR spin relaxation experiments on the same Atox1 and CopZ mutants could be performed in the future to experimentally validate our findings, as this type of
studies can report on protein dynamics in a per-residue basis within similar timescales as reported here (ps-ns).
Chapter 4:

Conformational dynamics of ATP7B metal-binding domains

4.1 Are the metal-binding domains equivalent?

The biological reason for the presence of six similar MBDs in humans, as opposed to one or two in bacteria and yeast (Arnesano et al., 2002), is still unclear. To date, solution structures of all apo- and holo-forms of MK1 to MK6 (Banci et al., 2004; Banci et al., 2005a; Banci et al., 2005b; Banci et al., 2006a; DeSilva et al., 2005; Gitschier et al., 1998) and apo-forms of the WD34 (Banci et al., 2008) and WD56 (Achila et al., 2006) two-domain constructs have been published. Although individual domains vary in sequence (between 32 and 50 % sequence identity), they all adopt a similar ferredoxin-like fold (Figure 4.1). All MBDs in ATP7A/B can bind Cu(I) in vitro (Cobine et al., 2000; DiDonato et al., 1997; Lutsenko et al., 1997) with similar affinities (Wernimont et al., 2004; Yatsunyk and Rosenzweig, 2007). However, genetic and biochemical data (Cater et al., 2004; Forbes et al., 1999; Iida et al., 1998) suggests that the six WDs are not functionally equivalent and possible roles include fine-tuning and regulation of the Cu-transfer activity. Nonetheless, it is still not clear whether each domain is unique because of intrinsic differences, or, despite the redundancy, its relative spatial location in the entire N-terminal domain of ATP7B dictates its function.
Figure 4.1. **Up:** Multiple sequence alignment of WD1-6. Residues highlighted in blue are completely conserved, orange are conserved in all but one domain, and yellow are conserved in some but are similar in others. Residues M, X₁, C₁ and C₂ of the metal binding motif MX₁C₁X₂X₃C₂, together with position 66 in WD1 are in bold and highlighted with a black box. The secondary structures elements (β₁-β₄, α₁-α₂ and loop regions L1-L5) and residue numbering are indicated based on WD1. **Bottom:** Scheme of the N-terminal domain of ATP7B with the six WDs. The length of each inter-domain linker (in residue number) is indicated. TMDs, transmembrane domains. **Right:** Structure of an individual MBD (holo-MK1, 1KVJ.pdb) revealing its ferredoxin-like fold. Residues M, X₁, C₁ and C₂, together with position 66 in WD1, and Cu(I) are shown in Licorice.
Furthermore, several studies have addressed the interaction between Atoxl and the WDs. Reported experimental data vary, likely due to the use of different constructs and experimental conditions. Using yeast two-hybrid assays, it has been reported that combinations of WD1-4 but not WD5 and WD6 can interact with Atoxl (Larin et al., 1999; van Dongen et al., 2004). In one such study, in addition to the dominant Atoxl interaction with WD4, Atoxl was found to interact also with WD1 and WD2 (van Dongen et al., 2004). Cys labeling studies have suggested that in the full N-terminal domain Cu-Atoxl preferentially delivers the metal to WD2 (Walker et al., 2004). In contrast, in Cu-chelator competition experiments using a six-domain WD construct, the affinity of Cu-Atoxl for individual Cu sites appeared similar and delivery could occur to all domains (Yatsunyk and Rosenzweig, 2007). On the other hand, NMR experiments showed that holo-Atoxl can form a Cu-dependent adduct with both WD2 and WD4 when presented as individual domains, although the interaction is stronger with WD4 (Achila et al., 2006). Similar NMR experiments revealed no Cu(I) transfer or adduct formation between holo-Atoxl and a WD56 two-domain construct (Achila et al., 2006). Instead, Cu-loaded WD4 preferentially delivered Cu(I) to WD6 in the WD56 construct (Achila et al., 2006). Subsequent NMR experiments between holo-Atoxl and the two-domain WD34 construct showed that, although Cu(I) was transferred to both domains, Cu-Atoxl formed a Cu-dependent ternary complex only with WD4 (Banci et al., 2008). Finally, NMR experiments on a six-domain WDs construct suggest that WD1, WD2 and WD4 are able to form adducts with Cu-Atoxl but the remaining domains, although metallated by Atoxl, cannot form detectable adducts (Banci et al., 2009b).
Taken together, it appears that all WDs have an ability to interact with Atox1, but that kinetic and thermodynamic properties (perhaps governed by intrinsic domain properties, inter-domain interactions, and the surrounding media) will act as selectivity filters at limiting conditions. In this Chapter, to gain insight into intrinsic molecular differences among individual WDs that may serve as selectivity filters for Atox1 recognition, we have investigated the conformational dynamics of the six individual apo-WDs (WD1-6), via MD simulations. To investigate the effect of Cu(I) coordination in individual domains, and shed light into possible inter-domain Cu-transfer selectivity, we simulated holo-WD2 (WD2c), holo-WD4 (WD4c) and holo-WD6 (WD6c). These three holo-domains were selected because they preferably interact with Atox1 (in the case of WD2 and WD4) (Achila et al., 2006; Banci et al., 2008; Banci et al., 2009b) and with WD4 (in the case of WD6) (Achila et al., 2006).

4.2 Are the metal-binding domains independent?

Because of the existence of multiple WDs, it has been suggested that inter-domain interactions regulate ATP7B function in vivo (Lutsenko et al., 2007b). Each WD comprises ~70 residues, whereas the complete N-terminus of ATP7B is about 620 residues long (Lutsenko et al., 2007b), which leaves about 200 residues in linker regions that connect the six domains (Figure 4.1). These inter-domain linkers vary in length: the longest connects WD4 with WD5, whereas the shortest linker is located between WD5 and WD6. This observation together with the presence of only one or two MBDs in bacterial Cu-ATPases, may lead us to believe that the most N-terminal domains (i.e., WD1-4) behave independently to the most C-terminal domains (i.e., WD56) (Lutsenko et
al., 2007b). However, Cu(I) translocation across the membrane by Cu-ATPases appears to be accompanied by profound conformational changes (Voskoboinik et al., 2001) and the entire N-terminal domain of ATP7B can interact specifically with the ATPBD (Figure 1.2) in a Cu dependent manner (Tsivkovskii et al., 2001). Moreover, secondary and tertiary structure changes take place in the entire N-terminal domain of ATP7B upon Cu(I) binding (DiDonato et al., 2000). This suggests that conformational changes occurring in one WD, as a consequence of Cu(I) binding for example, may be amplified through the linker regions to a consecutive or more distant WD, resulting in large-scale conformational changes of the entire N-terminal domain (Lutsenko et al., 2007b). In this Chapter, to explore the effect of domain-domain interactions and/or conformational changes amplified through the linkers among the WDs, we have further simulated the two-domain constructs WD12, WD34 and WD56 in the apo-form, and holo-forms of the same two-domain constructs, in which Cu(I) was incorporated into WD2 (WD12c), WD4 (WD34c) or WD6 (WD56c). These two-domain constructs were selected because there is PDB reported structure for WD34 (Banci et al., 2008) and WD56 (Achila et al., 2006) only, and because the linkers between WD2 and WD3 and between WD4 and WD5 are the longest (Figure 4.1).

4.3 Computational Methods

4.3.1 Initial structures

The initial structures for the apo-forms correspond to the NMR-reported PDB structures: WD34 (2ROP) (Banci et al., 2008) and WD56 (2EW9) (Achila et al., 2006). Apo-WD12 structure was generated by homology threading calculations (as explained in
Chapter 2) using the structure of the WD56 construct as template (2EW9.pdb, 36% sequence identity). The individual domain structures were generated from the corresponding two-domain structures. These structures were solvated as explained in Chapter 2; water molecules extended at least 10 Å (individual WDs) and 12 Å (WD12, WD34 and WD56) from the surface of the proteins. For the holo-forms WD2c, WD4c, WD6c, WD12c, WD34c and WD56c, Cu(I) was inserted between the Cys(S) atoms in structures obtained from the MD simulations of the corresponding apo-forms where the Cys S-S distance was lower than 5 Å. These holo-structures were re-solvated as explained in Chapter 2; water molecules extended at least 12 Å (WD12c and WD56c) and 20 Å (WD34c) from the surface of the proteins.

4.3.2 Cu(I) center parameterization

To perform the simulations of the holo-forms WD2c, WD4c, WD6c, WD12c, WD34c and WD56c, we employed the same parameterization for the Cu(I) center in all systems. As opposed to conventional QM vacuum parameterizations, we decided to perform QM-MM geometry optimizations of the three individual holo-WDs (WD2c, WD4c and WD6c) to include the effects of the protein and solvent environments in the calculation. We have already shown in Chapter 3 that in the Cu chaperone CopZ, the environment significantly affects the geometrical parameters of the Cu(I) center, in comparison with a sole QM optimization in vacuum of a Cu(I) atom plus two CH₃S⁻ groups.

Starting from the generated holo-forms WD2c, WD4c and WD6c (see before), we first performed short MM-MD simulations to relax and equilibrate these initial structures.
The solvated structures were optimized and then equilibrated by performing 700 ps of MM-MD simulations at 300 K, followed by a SA calculation (200 ps) in which the temperature was slowly decreased to 0 K. In these short simulations the parameters for the Cu(I) center were the same than those used in Chapter 3 for CopZ. The annealed structures were subsequently optimized at the QM-MM level using SIESTA with conventional parameters (Chapter 2), with the Cu(I) atom plus two CH₃S⁻ groups as the QM subsystem.

The equilibrium parameters for the Cu(I) center employed in the following MD simulations on the holo-forms were therefore taken as the average value of the QM-MM optimized Cu-forms of WD2c, WD4c and WD6c structures (Cu-S=2.154 Å, S-CuS=158.2° and C-S-Cu=101.1°). The atomic charges also correspond to the averaged RESP charges over the QM atoms combined with the CYM charges modified to maintain an overall integer charge of -1 (as explained in Chapter 2): \( q_{\text{Cu(I)}} = 0.388252 \) e and \( q_S = -0.730695 \) e. The vdW parameters for Cu(I) were taken from (Holt and Merz, 2007) and the bond and angle force constants involving the Cu(I) atom were taken from (Dalosto, 2007). This combination of parameters is similar to the ones employed in Chapter 3 for the metallochaperones and to the ones reported in (Holt and Merz, 2007).

### 4.3.3 MM-MD simulations

Simulations were performed for the different systems using AMBER, as described in Chapter 2. The coordinating Cys and Cu(I) in the holo forms were parameterized as reported above. The structures were simulated (total run) until the backbone rmsd as a function of time of each individual domain (WD1-6 and each domain
in the two-domain constructs) was stable for at least 100 ns (production run). Except indicated otherwise, the converged last 100 ns were used for data analysis.

4.4 Differences among ATP7B metal-binding domains as potential selectivity filters for Atox1 interaction

4.4.1 Apo-forms of WD1-6

Apo-forms of WD1, WD2, WD3, WD4, WD5 and WD6 were subjected to MD simulations for more than 100 ns each. The backbone rmsd data show that WD5 undergoes the largest conformational changes during the entire simulation (Table 4.1, Figure 4.2A). In the last 100 ns, which were used for data analysis, the rmsd suggests no significant conformational changes in any of the apo-domains (Table 4.1).

Table 4.1: Total simulation time (ns), backbone rmsd (Å) for the entire simulation (rmsd$_{\text{total}}$, with respect to the first structure) and for the last 100 ns of the simulation (rmsd$_{\text{100}}$, with respect to the average structure), and backbone rmsd (Å) between the initial and last structures from the entire simulations (Δrmsd).

<table>
<thead>
<tr>
<th></th>
<th>time</th>
<th>rmsd$_{\text{total}}$</th>
<th>rmsd$_{\text{100}}$</th>
<th>Δrmsd</th>
</tr>
</thead>
<tbody>
<tr>
<td>WD1</td>
<td>143</td>
<td>1.9 ± 0.2</td>
<td>1.1 ± 0.2</td>
<td>1.5</td>
</tr>
<tr>
<td>WD2</td>
<td>133</td>
<td>1.6 ± 0.3</td>
<td>1.2 ± 0.3</td>
<td>1.2</td>
</tr>
<tr>
<td>WD3</td>
<td>116</td>
<td>1.7 ± 0.3</td>
<td>1.1 ± 0.3</td>
<td>2.0</td>
</tr>
<tr>
<td>WD4</td>
<td>132</td>
<td>2.1 ± 0.3</td>
<td>1.4 ± 0.2</td>
<td>1.5</td>
</tr>
<tr>
<td>WD5</td>
<td>140</td>
<td>3.6 ± 0.2</td>
<td>1.3 ± 0.4</td>
<td>3.3</td>
</tr>
<tr>
<td>WD6</td>
<td>125</td>
<td>1.8 ± 0.2</td>
<td>0.9 ± 0.2</td>
<td>2.1</td>
</tr>
<tr>
<td>WD2c</td>
<td>126</td>
<td>1.4 ± 0.1</td>
<td>0.9 ± 0.2</td>
<td>1.2</td>
</tr>
<tr>
<td>WD4c</td>
<td>135</td>
<td>2.5 ± 0.2</td>
<td>1.3 ± 0.3</td>
<td>2.2</td>
</tr>
<tr>
<td>WD6c</td>
<td>127</td>
<td>1.0 ± 0.2</td>
<td>0.8 ± 0.2</td>
<td>0.7</td>
</tr>
</tbody>
</table>

Backbone fluctuations of WD4 and WD5 are similar, and the Cu loop and helix α1 are significantly more flexible than the other domains (Figure 4.2B). The trend in fluctuations in WD4 and WD5 is similar to that found for the bacterial Cu chaperone CopZ.
Figure 4.2. A: Rmsd (in Å, with respect to the first structure) of the backbone heavy atoms as a function of simulation time for the apo-domains. B: Average fluctuations (rmsf in Å) of backbone heavy atoms per residue (with respect to the average structure of the last 100 ns) for the apo-domains. The secondary structure elements are shown (reference is WD1 final structure). Black, WD1; red, WD2; green, WD3; blue, WD4; orange, WD5; cyan, WD6.

The fluctuations found in WD1, WD2 and WD6 are similar to each other and differ from those of WD4 and WD5 (Figure 4.2B); although the Cu loops are the most flexible regions (excluding C-terminal ends), they are much less floppy in these three
domains than the observed in the corresponding WD4 and WD5 Cu loops. WD6 is the least floppy of the six WDs; its fluctuation pattern is similar to that found in the human Cu chaperone Atox1. Backbone fluctuations in WD3 do not correspond to either of the aforementioned groups. This protein appears rigid in the Cu loop but more flexible in the β3-α2 and α2-β4 loops.

4.3.1.1 WD1, WD2 and WD6: stable domains with well-defined Cu loops

The final structures of WD1, WD2 and WD6 after the simulations are similar to each other, typical of a well-packed ferredoxin-like fold (Table 4.2, Figure 4.3), and do not differ significantly from the starting structures (Table 4.1). The secondary structure elements are well defined, helices α1 and α2 are parallel and properly packed against the β-sheet. Also, the conformation of key residues around the Cu loop is very similar, including Met, the Cu binding Cys, residues X1 (Thr in these cases) and X3 (Ser in all cases) of the MX1C1X2X3C2 motif, and the conserved Phe located in the α2-β4 loop (Figure 4.3). This Phe aligns with Lys60/Tyr65 in Atox1/CopZ that appear to have an important role in Cu chaperone structural dynamics (Chapter 3). The final structures of WD1, WD2 and WD6 are also very similar to that of Atox1. The conformation of helix α1 in WD1, WD2 and WD6 changes throughout the production MD simulation. Although it does not unwind completely, the beginning of this helix loosens (opens) up and returns to the original conformation several times (Figure 4.3). This movement is subtler in WD6, in which helix α1 appears more rigid.

Consistent with similar fluctuations and structures observed for WD1, WD2 and WD6, all of them exhibit similar interactions involving the Cu loop residues (Figure 4.4).
Both Cu-binding Cys sidechains are flexible and move around freely during the simulations. The first Cys (Cys1) adopts two main conformations during the simulation in which it interacts either with residue X1 (Thr) or with residue X3 (Ser) (Figure 4.4A, B). The second Cys (Cys2) also interacts with the Ser, although the interaction is weaker and less frequent (Figure 4.4C), and with the Phe (Figure 4.4D). In contrast, the conserved Met is buried in the core of all three domains (Figure 4.5A) and its sidechain makes several contacts with residues in the β2-β3 sheet and residues close to the Cu loop (Phe and Cys2), which appears to stabilize the protein core. The terminal methyl of the Met (CE) interacts most of the simulation time with residues in the β2-β3 sheet, including the sidechain of Val36 and Leu38 (numbering corresponds to WD1), which are fully conserved in all the MBDs (Figure 4.1), and the backbone of Gly41 and Ser42 in WD1 (Gln42 and Glu43 in WD2, and Ser39 and Lys40 in WD6, Figure 4.1), and with Val18 (fully conserved among the WDs, Figure 4.1) in helix α1. Moreover, the rest of Met sidechain (CB, CG and SD) interacts with the Phe ring and Cys2. Similar interactions were observed for the corresponding Met in Atox1 and holo-CopZ (Figure 3.9).

Table 4.2: Backbone rmsd (in Å) between the corresponding final structures.

<table>
<thead>
<tr>
<th></th>
<th>WD1</th>
<th>WD2</th>
<th>WD3</th>
<th>WD4</th>
<th>WD5</th>
<th>WD6</th>
<th>WD2c</th>
<th>WD4c</th>
<th>WD6c</th>
</tr>
</thead>
<tbody>
<tr>
<td>WD1</td>
<td>0</td>
<td>1.5</td>
<td>2.3</td>
<td>1.9</td>
<td>3.1</td>
<td>1.4</td>
<td>------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>WD2</td>
<td>----</td>
<td>0</td>
<td>2.4</td>
<td>1.5</td>
<td>2.6</td>
<td>1.4</td>
<td>1.1</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>WD3</td>
<td>----</td>
<td>----</td>
<td>0</td>
<td>2.6</td>
<td>2.9</td>
<td>2.3</td>
<td>------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>WD4</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>0</td>
<td>2.1</td>
<td>1.5</td>
<td>------</td>
<td>1.7</td>
<td>------</td>
</tr>
<tr>
<td>WD5</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>0</td>
<td>2.6</td>
<td>------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>WD6</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>0</td>
<td>------</td>
<td>------</td>
<td>1.4</td>
</tr>
<tr>
<td>WD2c</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>0</td>
<td>1.7</td>
<td>1.0</td>
</tr>
<tr>
<td>WD4c</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>0</td>
<td>1.7</td>
<td>------</td>
</tr>
<tr>
<td>WD6c</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>0</td>
<td>------</td>
<td>------</td>
</tr>
</tbody>
</table>
Figure 4.3. Top: Final structures of WD1 (black), WD2 (red), WD3 (green), WD4 (blue), WD5 (orange) and WD6 (cyan) in the apo-forms. For each domain, residues M, X₁, C₁, X₃ and C₂ from the conserved MX₁C₁X₂X₃C₂ motif and the conserved Phe are shown. Bottom: Superposition of 50 representative conformations of the Cu loop in the production run for the apo-domains.
Figure 4.4. Histograms (arbitrary units) of distance distribution (in Å) between heavy atoms of residues $X_1$ Thr/His (OG1/NE2) and Cys1 (SG) (A), Cys1 (SG) and residue $X_3$ Ser (OG) (B), residue $X_3$ Ser (OG) and Cys2 (SG) (C), Cys2 (SG) and Phe (CZ) (D), and between Cys1 (SG) and Cys2 (SG) (E) for the apo-domains. Black, WD1; red, WD2; green, WD3; blue, WD4; orange, WD5; cyan, WD6.
Figure 4.5. Protein-solvent g(r) of Met(SE) (A), Cys1(SG) (B) and Cys2(SG) (C) for the apo-domains. Black, WD1; red, WD2; green, WD3; blue, WD4; orange, WD5; cyan, WD6.

The Cys are equally exposed to solvent in WD1, WD2 and WD6, but Cys1 is always more exposed than Cys2 (Figure 4.5B, C). The S-S distance distribution between the Cu-binding Cys is also similar between these three domains, and consists of two peaks centered at ~3.5 and ~5 Å (Figure 4.4E), allowing for strong interactions between the residues. However, the weights of the S-S distance distribution peaks differ: the short distance is favored in WD6 but the longer distance is favored in WD2. Overall, in WD1, WD2 and WD6, the conserved residues Met, Thr, Ser, and Phe form a stable network of interactions with the Cu-binding Cys. This results in a well-defined Cu loop conformation, with the Cys exposed to solvent and in proximity to each other.
4.4.1.2 WD3: most divergent domain of the six WDs

Among the six WDs, WD3 has lowest sequence identity (average of 36.6 %, Figure 4.1), so it may not be surprising that this domain exhibits unique structural dynamics. Sequence differences around the Cu loop are residue $X_1$, which is a His in WD3 instead of a Thr, and position 66 in the multiple sequence alignment in Figure 4.1. Residue 66 is an invariant Phe in all MBDs except in WD3 and MK3, where it is a Pro. However, WD3 contains a Phe located three positions downstream, which is also close to the Cu loop in the folded structure (Figure 4.3). Therefore, this Phe in WD3 may have a similar role to the Phe in position 66 in the other domains.

WD3's final structure differs by more than 2 Å in backbone rmsd with all of the other structures; WD1 and WD6 are the closest matches and WD5 the most distant (Table 4.2). The rmsf as a function of residue number is unique in WD3: the fluctuations are more homogenous throughout the sequence and thus the Cu loop is not more floppy than other regions (Figure 4.2B). Particular flexibility (larger than in the Cu loop) is seen in helix $\alpha2$, which is shorter than in the rest of the WDs (because of the presence of two unique Pro at its C-terminus) and the $\beta3-\alpha2$ and $\alpha2-\beta4$ loops. Helix $\alpha1$ in WD3 is shorter at the C-terminus resulting in a longer $\alpha1-\beta2$ loop (Figure 4.3). Because of the unique position of the Phe that, although further away in sequence, still interacts with the Cu loop, the $\alpha2-\beta4$ loop is close to the N-terminus of helix $\alpha1$ (Figure 4.3).

The interactions around the Cu loop are altered in WD3 as compared to WD1, WD2 and WD6 (Figure 4.4). Instead of interacting alternately with residues $X_1$ (His) and $X_3$ (Ser), as in the aforementioned domains, in WD3, Cys1 interacts only with the Ser (Figure 4.4A, B), and both these residues are in close contact with the Phe ring (Figure
4.3). These interactions seem to stabilize the Cu loop. In WD1, WD2 and WD6 interactions of Cys2 with Ser and Phe maintain the Cys sidechain solvent exposed and close to the other Cys. In WD3, on the other hand, Cys2 is more buried (Figures 4.3 and 4.5C), and consequently does not interact with either Ser or Phe (Figure 4.4C, D). Instead, it points to the hydrophobic core (Figure 4.3) and interacts with Val18 in helix α1, and with Val36 and Leu38 in the β2 strand. Because of the interactions described above, the Cys-Cys distance distribution differs with respect to the other domains, with a peak centered at ~7 Å (Figure 4.4E). The conserved Met is completely solvent exposed in WD3 (Figures 4.3 and 4.5A). This does not correlate with increased backbone fluctuations in the Cu loop because: (i) the interface formed by Cys1, Ser and Phe stabilizes the Cu loop and restricts the motion of Met, (ii) Cys2 provides many core contacts that in WD1, WD2 and WD6 are provided by their buried Met, and (iii) the Met sidechain interacts with residue X1 (His) and Lys41 in the β2-β3 loop.

4.4.1.3 WD4 and WD5: flexible domains with large changes in helix α1

WD4 and WD5 exhibit similar fluctuations to each other during the simulation, with helix α1 observed as a particularly floppy region (Figure 4.2B). Comparison of the final structures reveals that WD4 is most similar to WD2 and WD6, and it differs the most with WD3 (Table 4.2). WD4’s final structure does not differ much from the initial one (Table 4.1) and the Cu loop appears well organized, as also found in WD1, WD2 and WD6 (Figure 4.3). However, helix α1 and the C-terminus of the Cu loop in WD4 present unique dynamics that is absent in the other WDs (Figure 4.3). Residues 19 (which corresponds to Cys2) to 24 at the N-terminus of WD4’s helix α1 completely unwind and
rewind during the simulation. This unique unwinding/rewinding process distinguishes WD4 from the other domains in the N-terminus of ATP7B.

Because of the dynamics of helix $\alpha_1$, the interactions around the Cu loop are weaker in WD4 (Figure 4.4). However, when this helix is folded, several interactions maintain a well-organized Cu loop. The Phe sidechain is buried and interacts with Met, Ser and Cys2, which stabilizes the Cu loop. The conserved Met is buried but not to the same extent as in WD1, WD2 and WD6 (Figure 4.5A): during the first ~70 ns of simulation it interacts with Leu40 and Gly43 in the $\beta_2$-$\beta_3$ loop, and in the last ~30 ns it moves deeper inside towards the $\beta_2$-$\beta_3$ sheet. Comparing all WDs, Cys1 is most exposed in WD4 (Figure 4.5B) and it does not interact with Thr or Ser (Figure 4.4A, B), instead it interacts with Leu40 in the $\beta_2$-$\beta_3$ loop. On the other hand, Cys2 in WD4 is buried (Figure 4.5C), and interacts with the conserved Ser and Phe (Figure 4.4C, D), Val20 in helix $\alpha_1$, Val38 in strand $\beta_2$ and Leu40. The Cys-Cys distance distribution in WD4 consists of two peaks centered at ~5.5 and ~7 Å, which differs significantly from that found for WD1, WD2, WD5 and WD6 (Figure 4.4E). Notably, the Cys-Cys distances found here for WD4 are similar to those previously observed for apo-Atox1.

On the other hand, WD5’s final structure differs significantly from the initial one (Table 4.1). It also differs more than 2 Å in rmsd with the rest of the structures, with WD4 being the closest match and WD1 most different (Table 4.2, Figure 4.3). Helix $\alpha_1$ has enhanced flexibility in WD5: it loses secondary structure and becomes distorted (Figure 4.3). As opposed to WD4, in WD5 helix $\alpha_1$ unwinds and does not return to the original conformation during the entire 140 ns run. Because of the unwinding and
bending of this helix, interactions with strand β2 are lost, and this interface is therefore significantly weakened (Figure 4.3). The orientation of helices α1 and α2 with respect to the antiparallel β-sheet changes because (i) strand β1 is extended and as a result the Cu loop is shorter and the conserved Met is no longer part of the loop and (ii) the conserved Phe in strand β4 in this domain, is completely buried in the core (Figure 4.3).

In WD5 the Cu loop is particularly short (only four residues, Figure 4.3). The Met sidechain interacts most of the time with Val37 and Leu39 in strand β1, and with the conserved Phe. However, the destabilization of the interface between helix α1 and strand β2 allows for interactions between the Met and the solvent (Figure 4.5A). As in the other domains, Cys1 interacts mainly with the Ser (Figure 4.4A, B). Because of the unusual proximity between the Cu and α2-β4 loops in WD5, the conserved Thr interacts with Gly66 in the α2-β4 loop. Cys2 is significantly more buried than Cys1 (Figure 4.5C), and adopts two major conformations: it points to the core and interacts with Phe or it interacts with Cys1 (Figure 4.4D, E). As a result, the Cys-Cys distance distribution in WD5 consists of two peaks, both centered at short distances of ~4 and ~5 Å.

4.4.2 Holo-forms of WD2, WD4 and WD6

WD2c, WD4c and WD6c were simulated for more than 100 ns, and the backbone rmsd of the total simulation indicates that WD4c undergoes the largest conformational change (Table 4.1, Figure 4.6A). In the last 100 ns of the simulation, all the domains are stable and exhibit small fluctuations with respect to the corresponding apo-forms (Table 4.1). Backbone fluctuations of WD2c and WD6c are similar (Figure 4.6B) and
comparable to those observed previously for holo-Atoxl. Interestingly, in WD4c the Cu loop is more flexible than the rest of the protein, even more floppy than observed in the same region in apo-WD1, WD2 and WD6. Nevertheless, in all cases including WD4c, Cu(I) coordination reduces the overall backbone fluctuations (Figure 4.6C).

![Graphs showing rmsd and rmsf](image)

**Figure 4.6.** A: Rmsd (in Å, with respect to the first structure) of the backbone heavy atoms as a function of simulation time for the holo-domains. B: Average fluctuations (rmsf in Å) of backbone heavy atoms per residue (with respect to the average structure of the last 100 ns) for the holo-domains. The secondary structure elements are shown (reference is WD1 final structure). Red, WD2c; blue, WD4c; cyan, WD6c. C: Comparison of rmsf between apo- (black) and holo- (blue) forms of WD2, WD4 and WD6.

To validate our generated holo-structures, we compared them with the corresponding holo-MKs, for which NMR solution structures are available (the structure of MK4 was solved in the presence of Ag(I), which has similar coordination chemistry to
Cu(I)) (Banci et al., 2004; Banci et al., 2005b; Gitschier et al., 1998). Our holo-structures are between 1.1 and 1.8 Å from the corresponding NMR structures of MKs (Table 4.3), validating our structures. The geometry of the active sites of our QM-MM optimized structures, which we used for the Cu(I) site parameterization (section 4.3.2), and the corresponding NMR structures, is similar in all cases. WD2c optimized with a S-Cu-S angle of 156.1° versus 140 ± 40° found for the solution structure of MK2c (Banci et al., 2004). WD4c optimized with a geometry in which the S atoms of the two Cys are separated by 4.31 Å, similar to the 4.8 ± 0.2 Å in Ag(I) loaded MK4 (Gitschier et al., 1998). In the case of WD6c, it also optimized with a S-S distance of 4.31 Å, which is close to an average of 4.27 ± 0.03 Å in the solution structure of MK6c (Banci et al., 2005b).

Table 4.3: Backbone rmsd (in Å) between the holo-WD (initial and final) and the corresponding holo-MK structures. MBD2c: WD2c versus 1S6U.pdb, (Banci et al., 2004). MBD4c: WD4c versus 2AW0.pdb, (Gitschier et al., 1998). MBD6c: WD6c versus 1YJV.pdb, (Banci et al., 2005b).

<table>
<thead>
<tr>
<th></th>
<th>initial</th>
<th>final</th>
</tr>
</thead>
<tbody>
<tr>
<td>MBD2c</td>
<td>1.5</td>
<td>1.6</td>
</tr>
<tr>
<td>MBD4c</td>
<td>1.7</td>
<td>1.8</td>
</tr>
<tr>
<td>MBD6c</td>
<td>1.2</td>
<td>1.1</td>
</tr>
</tbody>
</table>

The final structures of WD2c and WD6c are very similar to each other, with backbone rmsd of only 1.0 Å (Table 4.2, Figure 4.7). However, the WD4c final structure differs 1.7 Å with the other holo-structures (Table 4.2, Figure 4.7). The conformation of key residues around the Cu loop are very similar in WD2c and WD6c (Figure 4.7), and this conformation matches that found for holo-Atox1. The structure of WD4c differs from the other holo-proteins mainly in the Cu loop and the beginning of helix α1 (Figure
In WD4c, the Cu loop is longer: eight residues (including Cys2) in WD4c versus four and five residues in WD2c and WD6c, respectively (Figure 4.7). This means that the N-terminal end of helix α1 unwinds and both Cys are located in the Cu loop in WD4c, which may explain the observed higher fluctuations in this region (Figure 4.6B). This structural change exposes the Cu(I) bis-thiolate center in WD4c to solvent as opposed to in the other holo-domains.

**Figure 4.7.** Top: Final structures of WD2c (red), WD4c (blue) and WD6c (cyan) in the holo-forms. For each domain, the Cu(I) atom and residues M, X₁, C₁, X₃ and C₂ from the conserved MX₁C₁X₂X₃C₂ motif and the conserved Phe are shown. Bottom: Superposition of 50 representative conformations of the Cu loop in the production run for the holo-domains.

The interactions around the Cu center are similar between WD2c and WD6c (Figures 4.7 and 4.8). The conserved Thr and Ser sidechains interact extensively with Cys1 (Figure 4.8A, B), which provides electrostatic stabilization to the Cu center. The Phe sidechain remains close to the Cu loop and interacts with Cys2 (Figure 4.8D), which
further stabilizes the Cu site. Moreover, Cys1 and Cys2 form a HB with the backbone amide of Ser and Thr, respectively. The Met sidechain is completely buried in WD2c and WD6c holo-domains (Figure 4.9A), like in the apo-forms, and interacts with residues in the β2-β3 sheet and with Phe and Cys2. Altogether, this extensive network of interactions appears to maintain stable and rigid structures of these two holo-domains. In both domains, Cys1 is exposed to the same extent (Figure 4.9B), because it has a similar conformation in both structures. However, Cys2 is more buried in WD6c (Figure 4.9C), because in contrast to in WD2c, it points towards the loop backbone.

**Figure 4.8.** Histograms (arbitrary units) of distance distribution (in Å) between heavy atoms of residues X, Thr (OG1) and Cys1 (SG) (A), Cys1 (SG) and residue X3, Ser (OG) (B), residue X3, Ser (OG) and Cys2 (SG) (C), and between Cys2 (SG) and Phe (CZ) (D), for the holo-domains. Red, WD2c; blue, WD4c; cyan, WD6c.
Figure 4.9. Protein-solvent $g(r)$ of Met(SE) (A), Cys1(SG) (B) and Cys2(SG) (C) for the holo-domains. Red, WD2c; blue, WD4c; cyan, WD6c.

In contrast, most of the mentioned interactions around the Cu loop are absent in WD4c (Figures 4.7 and 4.8). The dynamics of Met in WD4c changes throughout the production simulation: during the first ~80 ns, it interacts with Ile11 in strand β1, Leu40 and Val38 in β2, and with the Phe ring, but in the last ~20 ns this sidechain points to the Cu loop where it becomes exposed to solvent (Figures 4.7 and 4.9A). This may also contribute to an increase in the fluctuations of the Cu loop. Furthermore, the Phe sidechain does not interact with Cys2 (Figures 4.7 and 4.8D) but with the hydrophobic core. This is mainly due to the conformation of the loop, which is protruding from the surface of the protein and too far away for Phe to reach it (Figures 4.7). Therefore, the
Cu(I) bis-thiolate center is more exposed in WD4c as testified by a marked increase in solvent exposure of Cys2 (Figure 4.9C). The Cu site is only stabilized by Ser, which interacts with both Cys (Figure 4.8B, C) but the Thr remains far from the Cu center (Figures 4.7 and 4.8A). Overall, it emerges that the Cu(I) atom is more exposed and less protected in WD4c as compared to in the other holo-domains studied.

4.4.3 Surface electrostatic potential analysis of WDs

To address the involvement of electrostatic attraction in Atoxl-WD and WD-WD interactions, we calculated the surface electrostatic potentials (as explained in Chapter 2) at the end of the simulations for all the domains (apo- and holo-forms) (Figure 4.10). For comparison, we also performed this calculation for holo-Atoxl (using the structure at the end of our previous simulation of Chapter 3). The structures in Figure 4.10 have the same orientation as that in Figures 4.3 and 4.7, which shows the protein side in which the metallochaperone and target domains are believed to interact (Achila et al., 2006; Banci et al., 2007; Banci et al., 2008; Banci et al., 2009a; Banci et al., 2009b).

The surface charge distribution in Atoxl is mainly positive (Figure 4.10), because of the presence of one Arg and Lys in helix α1 and three Lys in helix α2. On the other hand, WD2 (apo and holo), WD4 (apo and holo) and WD5 have significant patches of negative surfaces, complementary to the positive ones of holo-Atoxl. Whereas the WD2 and WD5 structures are mainly negative in helix α2 and the α2-β4 loop, most of the WD4 surface is negative, including helices α1 and α2 and the α2-β4 loop. These are the regions believed to form the metallochaperone-target adduct interface (Achila et al., 2006; Banci et al., 2007; Banci et al., 2008; Banci et al., 2009a; Banci et al., 2009b).
WD1 has also patches of negative surface, although not as much as WD2, WD4 and WD5. On the other hand, WD3 and WD6 (apo and holo) have mainly positive charge distribution on this side.

**Figure 4.10.** Comparison of the surface electrostatic potential of the apo- (top) and holo- (bottom) domains, using the final structures from the simulation as reference. The same analysis is shown for the human Cu chaperone Atoxl in its holo-form (Atoxl). Red and blue indicate negative [-2 kcal/(mol e)] and positive [+2 kcal/(mol e)] potentials, respectively.
4.4.4 Biological relevance

To address the role of intrinsic biophysical properties of individual domains, we have compared the in silico behavior of the six WDs of ATP7B. We hypothesize that selective Cu(I) delivery by Atox1 to WDs in ATP7B is governed by differential conformational flexibility and/or Cys solvent exposure among the WDs. For example, one may propose that some domains are more poised than others to form partially folded intermediates that are involved in adduct formation. We have determined that Atox1 is a rigid and stable protein in comparison to the bacterial Cu chaperone CopZ, and backbone fluctuations are significantly reduced upon Cu(I) binding (Chapter 3). It is tempting to speculate that because of this, Atox1 will form more stable complexes with WDs that are more floppy and can adapt. Moreover, we have seen in Chapter 3 that key residues near or in the Cu loop have evolved differently in Cu chaperones from different species to allow fine-tuning of the flexibility of the loop.

As predicted, our simulations suggest that the individual WDs are not equivalent at the molecular level. First, due to the presence of a unique Pro, WD3 behaves differently from the rest: the Cu loop is rigid (and adopts a different conformation than in the other domains) whereas the overall fold is more floppy. WD5 is also an outlier because its final structure is different from the starting structure and helix $\alpha$1 becomes distorted. WD1, WD2 and WD6 display many similar trends: their folds are stable and their Cu loops are more well-defined and less floppy than in WD4 and WD5. Of all the domains, WD6 is the least flexible. WD4 retains the initial structure but exhibits enhanced dynamics during the simulation. In WD4, half of helix $\alpha$1 completely unwinds and rewinds during the simulation whereas the N-terminus of this helix only partially
unwinds in WD1, WD2 and WD6. No helix $\alpha_1$ unwinding motion is found in WD3 or WD5. Therefore, we propose that unwinding of helix $\alpha_1$ may be important for adduct formation between partner proteins during Cu(I) transfer, for example, between Atox1 and WD1, WD2 and WD4, and between WD4 and WD6. The many experimental implications of favorable Atox1 interactions with WD4 (Achila et al., 2006; Banci et al., 2008; Banci et al., 2009b; van Dongen et al., 2004) can thus be explained by a requirement of helix $\alpha_1$ unwinding in order to form a stable complex with Atox1.

The Cys-Cys distance in the different WD domains also vary: from short (in WD1, WD2, WD5, WD6) to intermediate (in WD4) to long (in WD3). The Cys-Cys distance distribution in WD4 is similar to the one found in Atox1, and Cys1, which is proposed to be the first Cys to bind Cu(I) during transfer (Banci et al., 2006b), is most exposed in WD4. These features may facilitate the formation of stable Cu-intermediates during the Cu(I) exchange process between Atox1 and WD4. In addition, subtle differences among the domains have been noted regarding the network of interactions around the Cu site and the $\alpha_2$-$\beta_4$ loop, regions believed to be involved in metallochaperone-target adduct formation (Achila et al., 2006; Banci et al., 2007; Banci et al., 2008; Banci et al., 2009a; Banci et al., 2009b), which may further affect the stability of the Atox1-WD adduct.

Inspection of the surface electrostatic potentials predict favorable interactions between the highly positive holo-Atox1 and both negatively charged WD2 and WD4, as also noted experimentally (Achila et al., 2006; Banci et al., 2009b; van Dongen et al., 2004; Walker et al., 2004). However, WD4's surface appears better optimized for interactions with Atox1, which may explain why Atox1 interacts more strongly with
WD4 than WD2 in vitro (Achila et al., 2006). Thus, surface electrostatic complementarity appears as an important component for Atox1-WD adduct formation. Moreover, WD4's surface becomes more negative upon Cu(I) binding, especially on the Cu loop, which complements the positive patches in apo-WD6 and not the negative ones in apo-WD5. Electrostatic attraction may thus explain why WD4c preferentially delivers Cu(I) to WD6 in a WD56 construct (Achila et al., 2006). In addition, we found that WD4c is the most flexible of all holo-domains simulated here, and its Cu site protrudes from the surface of the protein. It is possible that WD4c's unique flexibility is a good match for interaction with WD6, which is the most rigid WD, as also observed experimentally (Achila et al., 2006).

Several missense mutations found in Wilson disease patients have been located to the WDs: Gly85Val (Loudianos et al., 1998), Asp196Glu (Ye et al., 2007), Leu492Ser (Loudianos et al., 1998), Tyr532His (Cox et al., 2005), Val536Ala (Davies et al., 2008), Gly591Asp (Cox et al., 2005), Ala604Pro (Cox et al., 2005), Arg616Trp (Caca et al., 2001), Arg616Gln (Gromadzka et al., 2005; Todorov et al., 2005) and Gly626Ala (Figus et al., 1995). Inspection of the behaviors of residues Gly85 (Gly30 in WD1), Asp196 (Asp57 in WD2), Leu492, Tyr532 and Val536 (Leu8, Tyr48 and Val52 respectively, in WD5), and Gly591, Ala604, Arg616 and Gly626 (Gly28, Ala41, Arg53 and Gly63, respectively, in WD6) during the simulations, may reveal why and how mutations at these positions cause disease. Gly30/28 in WD 1/6 is located in the α1-β2 loop and is completely conserved among all the WDs (Figure 4.1). Mutation of this small residue to a bigger one, such as Val or Asp, will likely be detrimental to the fold, since there is no space for a bulkier sidechain in this position. Asp57 in WD2 is located at the N-terminus
of helix α2, and HBs with Gln30 in α1-β2 loop and Gln54 in β3-α2 loop throughout the simulation. Although replacement of this residue by a Glu appears as a conservative substitution, the longer sidechain of Glu may not correctly preserve the contacts made by Asp. Leu8 in WD5 is located in strand β1 and is conserved as a hydrophobic residue among the WDs (Figure 4.1). Throughout the entire simulation, Leu8 forms an extensive network of hydrophobic interactions with residues in helix α2 (Ile58, Ala59 and Ile62): mutation of this position to a polar residue (like Ser as in the disease variant) will likely disrupt the stability of the core. Tyr48 is located at the end of strand β3 in WD5 and is also conserved among the WDs, replaced by a Phe only in WD6 (Figure 4.1). During the simulation, its aromatic ring forms a stable interaction with the conserved Pro50 and Pro55 in the β3-α2 loop and helix α2, respectively, stabilizing this loop. Mutation to a polar His (as found in a disease variant) will likely perturb these interactions. Val52 in WD5, located in the β3-α2 loop, is conserved in WD1 and WD4 and replaced by hydrophobic residues in the rest of the WDs (Figure 4.1). Its hydrophobic sidechain interacts with Gly31 in α1-β2 loop and Asp49, Glu51 and Ile53 in β3-α2 loop. Mutation of this residue to a small Ala will likely disrupt these hydrophobic interactions. Ala41 is located in strand β3 in WD6; it is fully conserved among all the WDs (Figure 4.1) and interacts with the protein core. Mutation of this residue to a Pro (as found in a disease variant) will probably disrupt the conformation of this β strand. Arg53 is located in helix α2 in WD6, and is facing the protein surface. This residue may therefore be involved in inter-domain interactions, as we will later see in this Chapter. Gly63 in WD6 is conserved in all WDs but WD3 (Figure 4.1) and it is located in the α2-β4 loop, one residue
upstream of the conserved Phe, and thus in the region of metallochaperone-domain
interactions. Changing this small side chain may reduce or alter loop dynamics and
thereby affect adduct formation. Taken together, we can predict that many disease-
causing mutations will result in domain unfolding/instability but some may instead
disturb protein-protein interactions. Although WD5 is not found to be of particular
importance in Atox1 interactions or in inter-domain interactions, destabilizing mutations
in this domain can cause Wilson disease. This clearly suggests that the domains affect
each other, directly or via the linkers, as we will see in the next section.

4.5 Inter-domain interactions modulate collective dynamics of ATP7B metal-
binding domains

Although individual domains are similar in size, the two-domain constructs
WD12, WD34 and WD56 are very different in size, because of differences in the linker
regions (Figure 4.1). WD1 and WD2 are separated by 12 residues, and the total construct
is of 156 residues. The two-domain construct WD56 consists of a total of 149 residues,
with 6 residues between WD5 and WD6. In the case of WD34, the construct used for
NMR consists of additional segments apart from the 27-residue inter-domain linker: a N-
terminal 17-residue linker (before WD3) and a C-terminal 11-residue linker (after WD4).
Therefore, the total size of this construct is 202 residues.

All constructs were simulated for more than 150 ns (Table 4.4). WD12 was
simulated longer times, because instead of starting from an experimental-determined
structure, as in the rest, we started from a built model. Figure 4.11A shows the backbone
rmsd of the entire two-domain constructs (including linkers) as a function of the total
simulation time. Upon inspection, we can see that the WD34, which contains the longest linker regions, is the protein with higher deviations, as expected. However, since we are interested in protein dynamics of each individual domain, we simulated until the backbone rmsd of each domain was stable for at least 100 ns (Figure 4.11B, C).

Table 4.4: Total simulation time (ns), backbone rmsd (Å) for the entire simulation (rmsd\textsubscript{total}, with respect to the first structure) and for the 100 ns of the production simulations (rmsd\textsubscript{100}, with respect to the average structure). The relative deviation from the mean value of the rmsd\textsubscript{100} (rmsd\textsubscript{100} rel. dev., in %) is also shown.

<table>
<thead>
<tr>
<th></th>
<th>time</th>
<th>rmsd\textsubscript{total}</th>
<th>rmsd\textsubscript{100}</th>
<th>rmsd\textsubscript{100} rel. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>WD1.12</td>
<td>239</td>
<td>2.0 ± 0.3</td>
<td>1.4 ± 0.2</td>
<td>14</td>
</tr>
<tr>
<td>WD2.12</td>
<td>239</td>
<td>2.3 ± 0.4</td>
<td>1.0 ± 0.2</td>
<td>20</td>
</tr>
<tr>
<td>WD1.12c</td>
<td>169</td>
<td>1.8 ± 0.3</td>
<td>0.9 ± 0.2</td>
<td>22</td>
</tr>
<tr>
<td>WD2c.12c</td>
<td>169</td>
<td>2.2 ± 0.6</td>
<td>0.8 ± 0.2</td>
<td>25</td>
</tr>
<tr>
<td>WD3.34</td>
<td>160</td>
<td>2.4 ± 0.1</td>
<td>1.1 ± 0.1</td>
<td>9</td>
</tr>
<tr>
<td>WD4.34</td>
<td>160</td>
<td>2.3 ± 0.6</td>
<td>1.3 ± 0.2</td>
<td>15</td>
</tr>
<tr>
<td>WD3.34c</td>
<td>161</td>
<td>1.5 ± 0.3</td>
<td>1.0 ± 0.2</td>
<td>20</td>
</tr>
<tr>
<td>WD4c.34c</td>
<td>161</td>
<td>2.4 ± 0.4</td>
<td>1.1 ± 0.3</td>
<td>27</td>
</tr>
<tr>
<td>WD5.56</td>
<td>161</td>
<td>2.2 ± 0.1</td>
<td>0.8 ± 0.1</td>
<td>13</td>
</tr>
<tr>
<td>WD6.56</td>
<td>161</td>
<td>2.1 ± 0.4</td>
<td>1.3 ± 0.3</td>
<td>23</td>
</tr>
<tr>
<td>WD5.56c</td>
<td>159</td>
<td>1.1 ± 0.2</td>
<td>0.9 ± 0.2</td>
<td>22</td>
</tr>
<tr>
<td>WD6c.56c</td>
<td>159</td>
<td>2.1 ± 0.7</td>
<td>1.0 ± 0.2</td>
<td>20</td>
</tr>
</tbody>
</table>

In the entire simulations, WD3 in the two-domain construct WD34 (WD3.34) and holo-WD4 in the two-domain construct WD34c (WD4c.34c) were the domains that underwent the greatest conformational changes, and the one with highest fluctuations was WD6c in WD56c (WD6c.56c) (Table 4.4). In the last 100 ns, WD4c.34c exhibited the highest fluctuations, whereas WD3.34 was the one with smallest fluctuations. In any case, all domains have stable rmsd in the production simulation time, which suggests that none of them underwent significant conformational changes.
4.5.1 Protein dynamics

4.5.1.1 WD12

As opposed to what we previously observed for the individual domains, backbone fluctuations of WD1 and WD2 in the WD12 construct are not the same: WD1’s Cu loop is much more flexible than WD2 (Figure 4.12A). In the case of WD2, Cu loop fluctuations do not significantly change by the presence of the other domain, although the C-terminus of helix α1 and α2-β4 loop are more flexible in the double construct. In particular, the end of helix α1 loses some secondary structure.
Figure 4.12. Average fluctuations (rmsf in Å) of backbone heavy atoms per residue (of the production run) for each domain in the apo- and holo-two-domain constructs: WD12 (A), WD34 (B) and WD56 (C). The secondary structure elements are indicated (α1, α2; β1-β4), using the first apo-domain as a reference (i.e. WD1, WD3 and WD5). Black, WD1.12; red, WD2.12; brown, WD1.12c; violet, WD2c.12c; light green, WD3.34; blue, WD4.34; yellow, WD3.34c; magenta, WD4c.34c; orange, WD5.56; cyan, WD6.56; maroon, WD5.56c; dark green, WD6c.56c.
On the other hand, the Cu loop and N-terminus of helix α1 of WD1 are much more flexible in WD12 than in the isolated domain. As we observed for the individual domains, the N-terminus of helix α1 loosens up (unwinds) and returns to the original conformation several times during the production time for both domains. Interestingly, Cu(I) binding to WD2 in the WD12c construct not only reduces the overall flexibility of WD2 but also of WD1, maintaining the same tendencies.

The starting model of WD12 was built via homology threading calculations using the structure of WD56 as a model. In WD12’s initial structure, whose conformation is maintained for the first ~200 ns of our total simulation, the domains are interacting in a tight unit, and the linker lays between and interacts with the two domains, mediating many of the interactions (Figure 4.13A). In particular, WD1’s helix α2 and the β3-α2 loop interact with WD2’s strands β1, β2 and β3. This inter-domain network includes vdW, electrostatic and HB interactions that include, but are not limited to WD1’s Gln55, Gln56 and His59 in helix α2 and WD2’s Glu3 in strand β1, Arg34 in strand β2 and Thr47 in strand β3. However, in our production simulation, the Rg increases from a mean of 17.3 ± 0.2 Å in the first ~67 ns to a final value of 21 Å (Figure 4.14). In this final structure, the linker is extended, the interface is weakened (Figure 4.13A), and both domains are able to move more freely.

To quantify the strength and nature of the domain-domain interacting energy, the ΔG and its different contributions were estimated for each of the two-domain constructs (Table 4.5). Upon inspection, we can clearly see how the interacting energy between WD1 and WD2 significantly weakens from the first 65 ns to the last 35 ns of the
production run. In particular, the intermolecular vdW and change in non-polar solvation energy (proportional to the SASA that becomes buried upon hetero-complex formation) significantly decreased, which indicates that the domain-domain interface falls apart.

**Figure 4.13.** Structure of WD12 from different snapshots of the total MD simulation of apo- (A) and holo- (B) forms. Snapshots are shown in order: first, after 68 ns and last for WD12; and first and last for WD12c. The two Cys residues are shown in sticks.
Figure 4.14. Rg (in Å) of the backbone heavy atoms of the two-domain constructs (including both individual domains and linker regions) as a function of the total simulation time. Black: WD12; red: WD12c; green: WD34; blue: WD34c; orange: WD56; cyan: WD56c.

On the other hand, in the holo-form, the Rg is constant throughout the production simulation, with a mean of 16.6 ± 0.2 Å (Figure 4.14), indicating that the domains remain in a tight unit interacting with each other. The domain-domain interface, and consequently the inter-domain energy, in the holo-form is similar to the one in the apo-form (before they separate, compare ΔG of WD12c to ΔG of WD12 in the first 65 ns, Table 4.5), although it also covers WD1’s strand β4 (Figure 4.13B). For example, Glu67 in WD1’s strand β4 makes a strong electrostatic contact with Lys36 in WD2’s strand β2, which was absent in the apo-form. This additional interaction found in WD12c and not in WD12 is evidenced by an increased of the intermolecular electrostatic energy in the former (Table 4.5). Glu67 is highly conserved among eukaryotes, often replaced by an Asp, whereas Lys36 is conserved as such in domain 2 of some eukaryotes, including MK2 (Arnesano et al., 2002). Conservation of these residues suggests that this interaction may potentially be important for the function of the protein.
Table 4.5. Inter-domain interacting free energy (ΔG) of each two-domain construct in apo- and holo-forms (calculated with sietraj as explained in Chapter 2). The non-polar and polar energetic contributions are indicated: intermolecular van der Waals interactions (vdW), change in non-polar solvation energy (cavity), intermolecular Coulomb interactions (electrostatic) and change in reaction field energy (field). All energy values are in kcal/mol; errors correspond to the standard deviation. Reported energy values correspond to the average of 1000 calculations based on the last 100 ns (production run) of MD simulations of each two-domain construct. As recommended in the sietraj webpage\(^8\), the ΔG should be calculated over a stable part of the trajectory. In the case of WD12, WD34c and WD56c, however, large changes in construct conformation and relative orientations are found in the production simulations. Therefore, the reported values for these constructs correspond to the first 65 and last 35 ns, first 15 and last 85 ns, first 30 and last 70 ns, respectively, of the production runs.

<table>
<thead>
<tr>
<th></th>
<th>ΔG</th>
<th>non-polar contributions</th>
<th>polar contributions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>vDW</td>
<td>cavity</td>
</tr>
<tr>
<td>WD12</td>
<td>1st 65 ns</td>
<td>-5.6 ± 0.6</td>
<td>-24 ± 5</td>
</tr>
<tr>
<td></td>
<td>last 35 ns</td>
<td>-3.5 ± 0.4</td>
<td>-5 ± 4</td>
</tr>
<tr>
<td>WD12c</td>
<td>100 ns</td>
<td>-5.6 ± 0.4</td>
<td>-22 ± 3</td>
</tr>
<tr>
<td>WD34</td>
<td>100 ns</td>
<td>-2.8 ± 0.4</td>
<td>-4 ± 3</td>
</tr>
<tr>
<td>WD34c</td>
<td>1st 15 ns</td>
<td>-4.1 ± 0.6</td>
<td>-16 ± 7</td>
</tr>
<tr>
<td></td>
<td>last 85 ns</td>
<td>-7.4 ± 0.7</td>
<td>-45 ± 6</td>
</tr>
<tr>
<td>WD56</td>
<td>100 ns</td>
<td>-3.7 ± 0.5</td>
<td>-5 ± 3</td>
</tr>
<tr>
<td>WD56c</td>
<td>1st 30 ns</td>
<td>-3.7 ± 0.3</td>
<td>-4 ± 3</td>
</tr>
<tr>
<td></td>
<td>last 70 ns</td>
<td>-6.1 ± 0.6</td>
<td>-28 ± 5</td>
</tr>
</tbody>
</table>

In both apo- and holo-forms, the active sites of both domains remain always facing opposite sides, so that there is no direct interaction between the Cu loops. Also, by looking at the protein structures, we can see that the relative orientation between domains is close to parallel, meaning that both Cu loops point to similar directions (Figure 4.13), as opposed to an antiparallel conformation, in which one Cu loop would be up and the other one down. As a qualitatively measure of the relative orientation between the domains, we calculated the angle between the domains (or inter-domain angle) as a function of the simulation time (Figure 4.15).

\(^8\) http://www2.bri.nrc.ca/ccb/pub/sietraj_main.php
Figure 4.15. Inter-domain angle, defined as the dihedral angle between the vectors formed by the backbone heavy atoms of the conserved Met and Pro in the Cu and β3-α2 loops of each domain, respectively (A), of the two-domain constructs WD12 (B), WD34 (C) and WD56 (D) in apo- (black) and holo- (blue) forms as a function of the total simulation time.

The inter-domain angle was defined as the dihedral angle between the vectors formed by the backbone heavy atoms of the conserved Met and Pro in the Cu and β3-α2 loops of each domain, respectively (Figure 4.15A). We picked these residues because they are conserved in all six domains (Figure 4.1) and also because they are located in loops that are at opposite ends of each domain (Figure 4.15A). In the apo-form, although in the first ~67 ns of the production run the two domains are interacting, the relative orientation between them varies slightly, meaning that the domains are able to rotate with
respect to each other (Figure 4.15B). However, they never reach an antiparallel conformation (Figure 4.13A). On the other hand, the relative orientation between WD1 and WD2 is almost fixed in the holo-form (Figure 4.15B), probably due to their reduced fluctuations.

4.5.1.2 WD34

In the case of WD3 and WD4, covalent attachment also affects backbone fluctuations. As a monomer, WD3 behaved in a unique way, because its Cu loop was rigid in comparison with the rest to the protein. In the two-domain construct WD34, WD3’s behavior is closer to the rest of the domains (Figure 4.12B). Its Cu loop is more flexible, whereas helices α1 and α2, and strands β3 and β4 are less flexible. The generalized reduction in backbone fluctuations in WD3 may be due to interactions of this domain with linker 2 (between WD3 and WD4).

WD4’s behavior also changes in the presence of WD3; overall, backbone fluctuations are reduced, except the Cu loop that is more flexible (Figure 4.12B). Helices α1 and α2 and the α2-β4 loop are less flexible, due to interactions with linker 2. Although helix α1 does not completely unwind and rewinds, as previously seen in the individual domain, it loosens up and returns to its original conformation several times. Moreover, the Cu loop adopts a unique conformation, in which both Cys are sticking out from the surface of the protein.

Backbone fluctuations of WD3 are affected by Cu(I) binding to WD4. The α1-β2 and β3-α2 loops, as well as helix α2 and α2-β4 loops all become less flexible, because these regions interact with WD4 and linker 3 (post WD4) only when Cu(I) is bound to
WD4, and this restricts backbone mobility in that region. On the other hand, the β2-β3 sheet has greater fluctuations. As opposed to what seen in the individual domain, WD4 fluctuations increase upon Cu(I) binding in the two-domain construct (Figure 4.12B). The change is focus on the Cu loop and helices α1 and α2. This holo-domain exhibits significantly higher fluctuations than any of the other two holo-domains, as also observed in the individual domains. Interestingly, the N-terminus of helix α1 unwinds and rewinds several times, which greatly exposes the Cu site, allowing it to protrude out from the surface of the protein, as also observed in the individual domain.

The linker between WD3 and WD4 (linker 2) is long enough to possibly allow free motion of both domains, as if they were independent beads on a string. However, this is not necessarily the case for WD34. In both apo- and holo-forms, linker 2 interacts extensively with both monomers. These interactions maintain the domains not far from each other, because the linker is not extended. However, there is a drastic difference between the inter-domain interacting surface and energy between the apo- and holo-forms during the production simulation (in Figure 4.16 the last snapshots of both apo- and holo-forms correspond to the conformation adopted during most of the production run). In the apo-form, the interface is weak, and the domains do not directly interact between each other, except for some few contacts that include vdW interactions between WD3’s α1-β2 loop and WD4’s β3-α2 loop. Instead, most of the interactions are mediated by the linker regions (linker 2). This is evidenced by a weak inter-domain energy (the weakest among all two-domain constructs, Table 4.5), with a very poor change in non-polar solvation energy, indicating that WD3 and WD4 in WD34 are not forming a complex by themselves.
Figure 4.16. Structure of WD34 from different snapshots of the total MD simulation of apo- (A) and holo- (B) forms. Snapshots are shown in order: first, after 15 ns and last for WD34; and first, after 45 ns, after 62 ns and last for WD34c. The two Cys residues are shown in sticks.
On the other hand, in the holo-form WD3 and WD4 are packed against each other, the interacting surface is extensive and covers a large surface area (the largest among all two-domain constructs, compare the cavity contributions in Table 4.5), formed by helices α1 and α2 from WD3 and the β1-β2-β3 sheet from WD4, as well as loops and linker regions. This is evidenced by the interacting energy between WD3 and WD4c, which is the strongest among all two-domain constructs (during the last 85 ns of the production simulation), and presents the largest change in non-polar solvation energy (Table 4.5), indicating that WD34c forms a compact and stable complex. The interactions involve an intricate network of HB and vdW interactions, which include WD3’s Asn20, Glu23 and Asn24 in helix α1 and Ala63 and Leu64 in helix α2 and WD4’s Thr6 in strand β1, Gln34, Gln35 and Ser37 in strand β2, and Thr46 and Leu48 in strand β3. Some of these residues are conserved among eukaryotic ATPases, or replaced by conservative mutations (Arnesano et al., 2002).

In apo-WD34, the Rg varies homogeneously during the simulation, with a mean value of 20.2 ± 0.5 Å in the production run (Figure 4.14), because although the domains remain at a relative constant distance, the linker regions are very flexible. Extensive dynamics were also previously observed for the linker regions by NMR (Banci et al., 2008). However, the relative orientation of the domains varies along the simulation. Initially, their orientation is close to being parallel (Figures 4.15C and 4.16A). However, between the first 4 and 15 ns of the total simulation, their arrangement changes (they rotate) and they adopt an antiparallel conformation, in which they remain the rest of the simulation (Figures 4.15C and 4.16A).
In the holo-form, the Rg is initially close to the one in the apo-form, with a mean value of 19.7 ± 0.8 Å in the first 10 ns of the production run, after which the domains collapse against each other, to reach a mean value of 17.6 ± 0.3 Å in the last 90 ns (Figure 4.14). In this conformation, WD3 and WD4 remain in a tight and compact unit, interacting between each other and with the linker regions (Figure 4.16B). In this case, in the initial structure the domains are in an extended and antiparallel orientation (Figure 4.16B), in which they are able to rotate between each other (Figure 4.15C). However, they finally rotate to adopt a parallel orientation, prior to the collapse, in which they remain for the rest of the simulation (Figures 4.15C and 4.16B). In any case, the active sites of each domain are far away and never face each other.

4.5.1.3 WD56

Previously, we found that WD5 exhibits increased conformational flexibility as an individual domain, and helix α1 was very disordered and distorted. On the other hand, the entire backbone flexibility of WD5 is significantly reduced in the two-domain construct (Figure 4.12C). In particular, the Cu loop and helix α1 are more structured. We also found that WD6 was the most rigid of all individual six domains. In the two-domain construct, however, WD6 is more flexible, and it is now WD5 the most rigid one. As an individual domain, only a subtle movement of the N-terminus of helix α1 in WD6 could be observed. In the two-domain construct this unwinding/rewinding is intensified. When Cu(I) is bound to WD6 in WD56c, the C- and N-terminus of helices α1 and α2 in WD5, respectively, become more flexible. These regions define an interface at the surface of the protein, which may be important for partner interactions in the cell. On the other hand,
WD6’s flexibility is reduced upon Cu(I) binding in the two-domain construct. However, the individual holo-domain is much more rigid than holo-WD6 in WD56c.

In the initial (NMR) structure of apo-WD56, both domains are in a tight unit interacting between each other (Rg of 17.5 Å), and their orientation is parallel (Figure 4.17A). However, when the simulation starts, WD5 and WD6 separate (after ~4 ns), evidenced by an increase in the Rg up to ~23 Å (Figure 4.14) and an extension of the linker that separates them (Figure 4.17A). After ~55 ns the domains come closer again and remain together during the production run (Rg of 18.4 ± 0.5 Å, Figure 4.14), however, they adopt a more extensive conformation than the initial one, because of a different conformation of the linker (Figure 4.17A). In particular, the interacting energy for these domains is relatively weak, comparable for example to the one between WD1 and WD2 at the end of the simulation (after they separate) (Table 4.5). The conformational change takes place in two steps. First the domains separate (see Rg plot at ~4 ns, Figure 4.14) and rotate so that now their orientation is such that they are antiparallel (note how the inter-domain angle smoothly starts to increase in magnitude at ~4 ns until it approaches ~|180| degrees at ~12 ns, Figure 4.15D, and Figure 4.17A). Finally, they rotate again to the initial orientation (after it reaches ~|180| degrees, the inter-domain angle smoothly starts to become closer to zero again, Figure 4.15D) and come closer together (see Rg plot at ~55 ns, Figure 4.14, and Figure 4.17A). Although some variation in the inter-domain angle suggests that the construct is not rigid, during the production run, WD5 and WD6 remain in a similar relative orientation (Figure 4.15D). This is consistent with the rotational correlation time of the two-domain construct being twice the one for a monomer in NMR experiments (Achila et al., 2006).
Figure 4.17. Structure of WD56 from different snapshots of the total MD simulation of apo- (A) and holo- (B) forms. Snapshots are shown in order: first, after 12 ns, after 30 ns and last for WD56; and first, after 30 ns and last for WD56c. The two Cys residues are shown in sticks.
The final conformation adopted by WD5 and WD6 in the apo-form is different to the one observed in the initial (NMR) structure (Achila et al., 2006), although both conformations likely represent minima in the phase space of the system. In the initial structure, WD6’s strands β1, β2 and β3 are facing WD5’s helix α2 and strand β4 (Figure 4.17A). During our simulations, the domains flipped, so that now WD5’s strands β1, β3 and β4 interact with residues in WD6’s N-terminus of strand β1 and helix α2 and C-terminus of strand β4. The interactions involve a network of vdW, electrostatic and HB interactions; the most stable is between WD6’s Arg53 at the N-terminus of helix α2 with WD5’s Phe7 in strand β1 and Glu45 in strand β3, which are pretty conserved residues among the corresponding domains in eukaryotes (Arnesano et al., 2002). Arg53 in WD6 aligns with Gln55 in WD1 (Figure 4.1), pointing to a similar conformation in which these domains interact in the two-domain constructs. Most importantly, mutation of Arg53 to a Trp (Caca et al., 2001) or a Gln (Gromadzka et al., 2005; Todorov et al., 2005) results in Wilson disease. This change will likely disrupt a strong salt bridge between WD5’s Glu45 and WD6’s Arg53 that may be relevant in vivo. In the NMR structure, Arg53 is pointing to the solvent and makes no contacts with residues in WD5 (Achila et al., 2006). The fact that Arg53 is mutated in Wilson disease patients supports the relevance of the WD56 conformation observed in our MD simulations.

On the other hand, in the holo-form, WD5 and WD6 are initially further apart with a Rg of 19.2 ± 0.7 Å in the first ~30 ns of the production run, after which they collapse into a tight unit with a Rg of 16.9 ± 0.2 Å in the last ~70 ns (Figures 4.14 and 4.17B). In this case, the domains are able to slightly move between each other only in the
first ~30 ns of the production run (Figure 4.15D), as in the apo-form. The similarity between the behavior of WD56 and the first ~30 ns (of the production run) of WD56c is also evidenced by a similar interacting energy (Table 4.5). However, after the first ~30 ns, WD5 and WD6c remain in a constant orientation, in which they interact extensively (Table 4.5). The conformation adopted by the domains is similar to the one in the apo-form (the nature of the interactions found in the apo-form are conserved in the holo-form, although the strength significantly changes, Table 4.5). This points to the importance of Arg53 in mediating protein-protein interactions also in the presence of Cu(I). However, these domains are more tightly packed against each other in the holo-form (Table 4.5, compare the change in non-polar solvation energy between WD56 and WD56c), and so the interacting surface covers a bigger surface area (Figure 4.17B), including WD5’s strand β2 and the β2-β3 loop and WD6’s entire helix α2, α2-β4 loop and strand β4. This inter-domain surface is very stable, thanks to an extensive network of vdW, electrostatic and HB interactions, which maintain the domains very close to each other, as part of a functional unit.

4.5.2 Cross-correlated motions

To gain insight into the cooperative dynamics and long-range communication between individual domains in the two-domain constructs, as well as the effect of Cu(I) binding, the cross-correlation matrix for the Cα atoms was calculated (as explained in Chapter 2) for all protein forms (see Figure 4.18 for individual domains and Figure 4.19 for two-domain constructs).
Figure 4.18. Cross-correlation matrices of fluctuations of Cα atoms from their average values during the last 100 ns of the simulation for the six individual apo-domains (WD1-WD6) and three holo-domains (WD2c, WD4c and WD6c). Degree of correlation is color-coded: red, strong positive (0.6 ≤ C_{ij} ≤ 1); yellow, moderate positive (0.4 ≤ C_{ij} < 0.6); black, strong negative (0.6 ≤ |C_{ij}| ≤ 1.0); cyan, moderate negative (0.4 ≤ |C_{ij}| < 0.6); white, weak or no correlation (-0.4 < C_{ij} < 0.4).
Figure 4.19. Cross-correlation matrices of fluctuations of Cα atoms from their average values during the production simulation for WD12, WD12c, WD34, WD34c, WD56 and WD56c. Degree of correlation is color-coded: red, strong positive (0.6 ≤ C_{ij} ≤ 1); yellow, moderate positive (0.4 ≤ C_{ij} < 0.6); black, strong negative (0.6 ≤ |C_{ij}| ≤ 1.0); cyan, moderate negative (0.4 ≤ |C_{ij}| < 0.6); white, weak or no correlation (-0.4 < C_{ij} < 0.4). The limits of each domain are indicated.
4.5.2.1 Intra-domain correlations

In all cases, we find mostly moderate cross-correlated regions in the individual proteins (Figure 4.18). On the other hand, correlations of the three two-domain constructs show that both domains are strongly correlated and the presence of the other domain significantly enhances both in-phase and out-of-phase correlations within each individual domain (intra-domain correlations) (Figure 4.19). This suggests that presence of the other domain is needed to induce specific motions within each individual domain. This also agrees with previous simulations of multidomain proteins, in which intra-domain correlations were also significantly weaker in the individual domains (Cansu and Doruker, 2008; Falconi et al., 2003). All six domains present an overall similar correlation pattern, which is expected since all have similar fold and overall similar dynamics (e.g. in all cases the Cu loop is the most flexible region).

In all cases, positive correlation is observed between residues forming secondary structure elements, which hold the fold together: helices α1 and α2; sheets β1-β3, β2-β3 and β1-β4; helix α1 with strand β2; strand β1 with helix α2, strand β3 and helix α2 with strand β4. In-phase motions are also found between loops and secondary structure elements: Cu loop with helix α1; strand β1 with α1-β2 and β2-β3 loops; α1-β2 loop with β3-α2 loop; Cu loop with α2-β4 loop.

On the other hand, negative correlations are observed between residues in the Cu loop and helix α1 and residues distant in both sequence and tertiary structure, including the β3-α2 loop and strand β4. Similarly, the α2-β4 loop presents out-of-phase motions with distant residues including strand β1 and the β3-α2 loop. Besides the Cu loop, the
α2-β4 loop may also have functional relevance since the conserved Phe residue is located here, and this loop is believed to form part of the metallochaperone-target interaction interface (Achila et al., 2006; Banci et al., 2007; Banci et al., 2008; Banci et al., 2009a; Banci et al., 2009b). Therefore, these concerted motions may altogether help modulate the flexibility of two key loops in these domains.

Cu(I) coordination to WD2, WD4 and WD6 in the two-domain constructs not only changes the correlation pattern within the holo-domain, but also within the consecutive apo-domain (Figure 4.19). In all cases, intra-domain out-of-phase motions are reduced or absent. However, WD5 and WD6 in WD56c still present significant negative correlations.

4.5.2.2 Inter-domain correlations

In all three constructs, inter-domain correlations are significant, and involve both residues that are close to each other, which form the domain-domain interacting interface, and residues that are distant both in sequence and space. This coupling between domains, which is possible because of the linker regions that connect them, emphasize a long-range domain-domain interaction that might be important for the function and regulation of ATP7B in vivo.

Some similarities are found between inter-domain correlations in WD12, WD34 and WD56 (Figure 4.19). In all cases, Cu(I) binding to one domain significantly affects the inter-domain collective motions. And, whereas in the apo-form we found both in-phase and out-of-phase correlations, upon Cu(I) binding inter-domain correlations are only or mainly negative. These changes may provide individual domains a way to
“communicate” among each other. For example, Cu(I) binding to one domain may alter the interplay between the consecutive domain, which at the same time may affect the tertiary structure of the full-length N-terminal domain, as previously observed experimentally (DiDonato et al., 2000). This conformational change of the entire N-terminal domain may be crucial for the ATPase regulation in vivo (DiDonato et al., 2000).

Interestingly, many of the inter-domain motions, both in the apo- and holo-forms, involve the Cu loop, helix α1 and the α2-β4 loop of each domain with each other and several other distant residues located in the other domain. These shared collective movements may be functionally important, since all these regions contain conserved residues with functional relevance (e.g. for interaction with the metallochaperone). In the apo-forms, the type of correlation involving these residues varies among the two-domain constructs. In the holo-form, however, both Cu (including the two Cys residues) and α2-β4 (including the conserved Phe) loops are anti-correlated between each other. Moreover, some regions are similarly correlated in all constructs. For example, in the apo-forms, the C-terminus of helix α1, the α1-β2 loop and N-terminus of strand β2 of each domain are always positively correlated with the C-terminus of strand β1, Cu loop and the N-terminus of helix α1.

4.5.2.2.1 WD12

In WD12, the linker region and beginning of WD2’s β1 strand is strongly anti-correlated with several residues in WD1, including: strands β1, β2 and β3; helices α1 and
and the \( \alpha_1-\beta_2 \) and \( \beta_3-\alpha_2 \) loops. Some of these regions define the domain-domain interface during the first \( \sim 67 \) ns of the production simulation (after which the interface weakens). Interestingly, WD1’s Cu loop is negatively correlated with residues in WD2 and the opposite is true for WD2’s Cu loop; however, both Cu loops are negatively correlated.

Although in WD12c the individual domains make more contacts than in the apo-form, inter-domain correlations are weaker than in WD12, probably because backbone fluctuations are significantly reduced upon Cu(I) binding. Only few inter-domain out-of-phase motions are found in WD12c, and these include residues that are close between each other in space, and are part of the domain-domain interface: WD1’s N-terminus of helix \( \alpha_2 \) with the linker and WD2’s N-terminus of strand \( \beta_1 \); and the linker with both WD1’s strand \( \beta_4 \) and WD2’s strand \( \beta_1 \). Although the domains are in close contact, the decreased in overall flexibility may account for the weaker correlations found in the presence of Cu(I). However, both Cu loops are still flexible, especially WD1’s Cu loop, which may explain why the majority of the out-of-phase movements involve these regions.

4.5.2.2.2 WD34

In WD34, correlations are weaker to the ones found in WD12. This may be due to the fact that the linker between WD1 and WD2 is much shorter than the one between WD3 and WD4, so domain-domain correlations are more easily propagated through a shorter linker. As mentioned before, in apo-WD34 the domain-domain interface is small. Cross-correlations are found between residues that interact directly and between nearby
regions. In-phase correlations are found between residues forming the interacting interface, which include WD3's C-terminus of helix α2 and α2-β4 loop with WD4's β3-α2 loop. On the other hand, out-of-phase movements are found between WD3's α1-β2 loop and the same region in WD4; WD3's α1-β2 loop, strand β2 and β2-β3 loop and WD4's strand β3 and β3-α2 loop; WD3's strand β3 and β3-α2 loop and the same residues in WD4; and between WD3's β3-α2 loop and N-terminus of helix α2 and WD4's β3-α2 loop, helix α2, α2-β4 loop and strand β4. Moreover, as opposed to in WD12, we found the Cu loop and helix α1 of each domain to be positively correlated.

As opposed to WD12c, correlations are accentuated upon Cu(I) binding to WD34c. This may be explained by two reasons. First, the interacting interface between WD3 and WD4 significantly increases upon Cu(I) binding, whereas it increases to a lesser extent in WD12c. Second, the backbone fluctuations of WD3 and WD4c are much larger than those of WD1 and WD2c. To recall, WD4c in WD34c was found to be the most flexible holo-domain. In fact, WD4c fluctuations increased upon Cu(I) binding, instead of decreasing as in the rest of the domains. For these reasons, inter-domain correlations are strong and negative. Many of these correlations involve residues forming or near the interface between the two domains: WD3's helices α1 and α2, strand β4 and the Cu and α2-β4 loops with WD4's strands β1, β2, β3, the Cu and β2-β3 loops; WD3's helix α2, α2-β4 loop and strand β4 with WD4's strands β1 and β4, helix α2 and the Cu and α2-β4 loops. This strong negative coupling between the motion of WD3 and WD4c may be relevant from a functional point of view, for example, to promote inter-domain Cu(I) transfer from WD4c to a more C-terminal domain (e.g. WD6).
4.5.2.2.3 WD56

Inter-domain correlations in WD56 are similar to the ones found in WD12. This may be because the linker between WD5 and WD6 is comparable in length to the one in WD12. Therefore, correlations are strong instead of moderate as found in WD34, probably because backbone motions can be more easily propagated through a shorter linker. However, despite the short linker that separates them, the interacting interface between WD5 and WD6 is small. The motion of residues near the domain-domain interface is negatively coupled: WD5’s N-terminus of strand β1 with WD6’s C-terminus of strand β3 and the β3-α2 loop. As in WD12, the motion of the Cu loop and N-terminus of helix α1 of each domain is positively correlated.

The effect of Cu(I) binding to WD56c is similar to the one found in WD34c: Cu(I) binding intensifies the degree of cross-correlations motions instead of decreasing it, as in WD12c. The reasoning is similar to the one for WD34: WD56c interacting surface is extensive and the flexibility of WD56c upon Cu(I) binding is not reduced as in WD12c. Negative inter-domain correlations are found between residues that form or are near the domain-domain interface: WD5’s strand β1, Cu loop, helix α1 and α1-β2 loop with WD6’s strands β1 and β4, helices α1 and α2, Cu and α2-β4 loops; WD6’s helix α2, α2-β4 loop and strand β4 with WD5’s strand β4, helix α2 and β2-β3, β3-α2 and α2-β4 loops; WD5’s C-terminus of strand β3, β3-α2 loop and N-terminus of helix α2 with the same residues in WD6. Again, as suggested for WD34c, this strong interplay between WD5 and WD6c found only in the holo-form may be key to promote Cu(I) transfer to the transmembrane sites of ATP7B.
4.5.3 Biological relevance

We have explored the conformational dynamics of the six WDs of ATP7B as pairs (WD12, WD34, WD56) with and without one Cu(I) added to each pair, to shed light into inter-domain interactions that may regulate Cu transport. In all cases, protein dynamics significantly changed when each domain was covalently linked to its consecutive domain, in comparison to the individual domains. Moreover, Cu(I) binding to one domain modulated not only the behavior of that same domain, but also the one of the apo-domain that was covalently linked to. This suggests that structure and dynamic changes occurring in one domain upon Cu(I) binding are being translated or propagated to the next domain through the linker as a way of “communicating” between each other. This “communication” is possible because of the presence of inter-domain correlations. Interestingly, Cu(I) binding to one domain induced not only changes in the individual domains, but also tertiary structure rearrangements of the entire construct. Secondary and tertiary structure changes were also observed experimentally upon Cu(I) binding to the entire N-terminal domain of ATP7B (DiDonato et al., 2000; Tsay et al., 2004). In particular, in all our MD simulations, Cu(I) binding induced a more compact tertiary structure, in which the domains interact more stably and with a larger interacting surface area. This is in agreement with a “structural tightening” observed in vitro upon Cu(I) binding to the N-terminus domain of ATP7B (Bartee et al., 2009). Moreover, the transition between apo- and holo-forms was further accompanied by conformational changes of the linker regions. This observation is consistent with the limited proteolytic digestion pattern of the N-terminal domain being affected by the presence of Cu(I) (Bartee et al., 2009; Walker et al., 2004). These conformational changes upon Cu(I)
binding amplified by linkers may alter the overall arrangement of the domains in the full-length N-terminal domain, which may be important for the function and regulation of the protein in vivo (Lutsenko et al., 2007b). For example, there is evidence that the N-terminal domain of ATP7B is able to interact with the ATPBD but only in the absence of Cu(I) (Tsivkovskii et al., 2001). Conformational changes of the entire N-terminal domain upon Cu(I) binding, similar to the ones observed here in the two-domain constructs, may reverse its interaction with the ATPBD and trigger Cu(I) transport across the membrane. Additionally, these conformational changes may trigger ATP7B translocation from the Golgi membrane to a vesicular compartment when Cu levels are elevated (Guo et al., 2005; Schaefer et al., 1999).

Whereas Cu(I) binding to WD12 weakened the inter-domain correlations, the opposite was true for Cu(I) binding to WD34 and WD56. This difference in the coupling between the domains may highlight different individual functions of these proteins in the context of the entire ATPase. For example, strong anti-correlations between WD3-WD4 and WD5-WD6 upon Cu(I) binding to WD4 and WD6, respectively, may promote vectorial Cu(I) transfer to WD6 (Achila et al., 2006) and the transmembrane Cu sites (Lutsenko et al., 2007b), respectively.

Interestingly, the interacting surface in all domains covered similar protein regions. In all our apo- and holo-two-domain constructs the Cu loops of each domain remain at opposite sides and far from each other. This suggests that inter-domain Cu(I) transfer between these consecutive domains (e.g., from WD2c to WD1, WD4c to WD3 or from WD6c to WD5) is not likely to occur. Instead, it is more likely that inter-domain Cu(I) transfer will occur between domains that are further apart, for example between
WD2 and WD4, and between WD4 and WD6. Although the linker between WD3 and WD4 is long enough to allow domain rearrangements that may allow inter-domain Cu(I) transfer, Cu loops were never close enough to suggest this as a possibility.

In a recent study based on 6 ns MD simulations of two-domain MBD constructs from different organisms, it was proposed that each domain exhibited increasing freedom of reorientation and weaker interacting energy with increasing linker length (Sharma and Rosato, 2009). However, based on more than 150 ns of MD simulation for three different two-domain constructs, our results suggest that the interplay between linker length, interacting energy and conformational freedom is less trivial. For example, even though the linker between WD1 and WD2 is two times longer than the one between WD5 and WD6, individual domains were able to reorient more freely in WD56 than in WD12 with respect to the initial structures. Starting from a parallel orientation, whereas WD56 was able to adopt an antiparallel orientation to finally return to the original conformation in 55 ns, individual domains in WD12 were unable to reach such orientation in 239 ns. Also, the interacting energy between WD1 and WD2 (during the first ~65 ns of production run) was stronger than the one between WD5 and WD6 in the apo-forms. Moreover, Cu(I) binding can significantly alter the relationship between linker length and interacting energy. For example, because apo-WD34 has the weakest inter-domain energy, one might conclude that this is due to the long inter-domain linker. However, this may not be the case, since WD34c exhibited the strongest inter-domain energy.

On the other hand, previous NMR experiments on the same WD34 construct suggested that these domains behave independently from each other, as beads on a string (Banci et al., 2008). Although we do see independent tumbling of these domains, most of
the simulated time they remain at a relative constant orientation making contacts between each other. Moreover, the fact that elements from both domains are strongly correlated between each other suggests that their behavior is not independent but instead strongly coupled. This also highlights the intricate relationship between linker length and relative conformational freedom between domains. Similar NMR experiments on the same construct revealed that holo-Atox1 is only able to form a stable Cu-dependent adduct with WD4, although Cu(I) is transferred efficiently to both domains (Banci et al., 2008). Two possible scenarios could explain this observation: Atox1-Cu-WD3 adduct is too weak for NMR detection or, once WD4 receives Cu(I) from Atox1, Cu(I) is transferred inter-molecularly to WD3 (Banci et al., 2008). Because at no time of our simulations WD3’s and WD4’s active site were close enough to allow interactions, we propose the first scenario as more likely to occur. We propose two reasons as to why Atox1-Cu-WD3 adduct was not detected by NMR. First, WD4’s Cu loop and helix α1 are more flexible than WD3’s, and this may facilitate interactions with the metallochaperone. Second, WD4 is more accessible than WD3 within the two-domain construct, which may increase the probability of interactions between WD4 and Atox1. Whereas WD3 interacts with the linker regions and its Cu loop faces most of the time the domain-domain interface, WD4’s Cu loop is always facing the surface. These differences observed between WD3 and WD4 may result in a more transient and less stable (as we will later see in Chapter 5) complex between Atox1 and WD3.

Because of the longest and shortest linker separating WD4 from WD5, and WD5 from WD6, respectively, and based on the WD56 NMR solution structure, it has been proposed that these two last domains behave as a functional and compact unit (Achila et
al., 2006). A recent 6 ns MD simulation on the same construct showed that WD56 also remain in a tight unit (Sharma and Rosato, 2009). Because WD6 is the most C-terminal domain, i.e., closest to the membrane (and because bacterial ATPases contain only one or two MBDs), it has been proposed to be the one that, after picking up Cu(I) from other domain, e.g. WD4 (Achila et al., 2006), or from Atox1 (Banci et al., 2009b), delivers Cu(I) to the transmembrane sites for subsequent Cu(I) translocation (Lutsenko et al., 2007b). In order to perform the aforementioned tasks, WD56 may still need to have certain degree of conformational plasticity, rather than remaining as a tight and rigid structure. Here, thanks to our extensive simulation time, which allowed us to explore a vast range of conformational space, we can conclude that, although WD56 remains in a close conformation in most of the production simulation time, the individual domains still have enough conformational flexibility to allow for movements with respect to each other, suggesting that they do not behave as a rigid entity.

4.6 Conclusions

Taken together, our computational data on the WD domains provides many structural and dynamic details not easily obtainable via experiments, which can be used as a basis for interpretation and explanation of experimental in vivo and in vitro data on these proteins. We conclude that the six WDs from ATP7B are not equivalent at the molecular level. Future experiments could be performed to test our hypothesis, for example, by assessing the protein stabilities (via unfolding experiments) and dynamics (via NMR experiments) of individual WDs and two-domain constructs. Based on the molecular properties found for individual WDs and two-domain constructs we propose
that (i) overall-fold and Cu loop dynamics (floppy favoring Atoxl interaction), (ii) Cys-Cys distance and positioning (rather long distance and solvent exposure favor Atoxl interaction), (iii) helix α1 conformation (unwinding/rewinding motion favors Atoxl interaction), and (iv) electrostatic attraction are properties that vary distinctly among the six WD domains and can act as selectivity filters to guide Atoxl interactions and inter-domain Cu(I) transfer. We propose that WD4 appears as the WD best optimized for Atoxl interactions. This agrees with \textit{in vivo} (van Dongen et al., 2004) and \textit{in vitro} (Achila et al., 2006; Banci et al., 2008; Banci et al., 2009b) data, which show that Atoxl preferentially interacts with WD4.

Moreover, regardless of the length of the linker that separates them, the WDs cannot be considered as independent beads on a string. Our simulations suggest that these domains are strongly correlated and interact with each other, which may likely be important for the function of the ATPase \textit{in vivo}. Other factors, such as quaternary structure of the full-length six-domain segment and the surrounding environment may further limit these reactions \textit{in vivo}. Our computational findings agree with experiments, which show conformational changes of the entire N-terminal domain (Bartee et al., 2009; DiDonato et al., 2000; Tsay et al., 2004; Walker et al., 2004) and alterations in the interplay between the N-terminal domain and the ATPBD (Tsivkovskii et al., 2001) upon Cu(I) binding. To corroborate our predictions, future efforts should focus on the determination of the solution or crystal structures of the two-domain construct WD12 in apo- and holo-forms, the holo-forms of WD34 and WD56, and constructs containing more than two domains.
Chapter 5:  

Inter-protein interactions and copper transfer mechanism between Atox1 and ATP7B metal-binding domains

5.1 Requirement of protein-protein specificity

Given the different array of metals within cells, how do proteins acquire the correct metal to perform their biological functions? It is believed that the relative affinities of proteins to different metals is not sufficient for selectively acquiring the correct metal, as some proteins can bind stronger metals in vitro than the cognate metal in vivo (Bertini et al., 2007). For example, Cu(I) (Frausto da Silva and Williams, 2001) and Cu(II) (Irving and Williams, 1948) are the highest competitor monovalent and divalent ions, respectively, so this metal will strongly bind to many protein ligands. A minimum requirement for a protein to bind its cognate metal in vivo is that the metal should be available (Bertini et al., 2007). However, given the overcapacity of metal chelation, the free concentration of transition metals inside cells is almost non-existent (Bertini et al., 2007). For example, it is believed that both prokaryotic and eukaryotic organisms have less than one free Cu atom per cell (Changela et al., 2003; Rae et al., 1999). Therefore, to make metals kinetically and thermodynamically accessible for proteins (Bertini et al., 2007), cells have evolved highly regulated pathways, in which trafficking proteins called metallochaperones guide the metal from the plasma membrane to the final cellular target (Figure 1.1).

Taking Cu as an example, which tends to bind to undesirable sites and has a high redox potential, it is reasonable that these metal-homeostasis pathways require high protein-protein specificity, to avoid the mishandling of the metal and irreversible
macromolecular oxidation. In the case of Cu transport to the secretory pathway, in which Atoxl delivers Cu(I) to the MBDs of ATP7A/B, this specificity appears to arise from the fact that the metallochaperone and target domains share the same fold and metal-binding motif (Figure 1.3). However, as pointed out in Chapter 4, presence of a ferredoxin-like fold, although required, is not sufficient for a target WD to interact with Atoxl, as not all of the six WDs interact with the metallochaperone in vitro or in vivo. In Chapter 4, based on MD simulations of individual and two-domain constructs WDs, we proposed that subtle structural, dynamical and electrostatic differences among individual WDs might act as selectivity filters for Atoxl recognition. In this Chapter, we went a step further and assessed the interaction between Atoxl and the six different WDs, to test our previous predictions.

5.2 Protein-protein interactions

Cu(I) transfer between the metallochaperone and target domains is believed to occur via direct protein-protein interactions and ligand-exchange reactions, with the formation of an Atoxl-Cu-MBD adduct in which Cu(I) is shared by both proteins (Banci et al., 2006b; Banci et al., 2009a; Benitez et al., 2008; Wernimont et al., 2000). In vivo, it is important for this adduct to be transient to assure that the metal is efficiently transferred to the target domain. How is then this transient metallochaperone-target complex assured?

Several bioinformatic approaches have been performed to identify the propensity of residues to appear in protein-protein interfaces, based on reported PDB structures (Glaser et al., 2001; Jones and Thornton, 1996; Ofran and Rost, 2003). In general, it emerges that large hydrophobic residues (like Trp, Tyr and Met) are overrepresented
(Glaser et al., 2001; Ofran and Rost, 2003), suggesting that complex formation is driven by the hydrophobic effect (Glaser et al., 2001). However, hydrophobic interactions appear to be more common in permanent and not transient protein-protein interactions (Jones and Thornton, 1995). On the other hand, there is more variability among the charged residues: whereas Lys is underrepresented, Arg is overrepresented (Ofran and Rost, 2003). However, hydrophobic-charge (like Trp-Arg), polar-polar and ionic charge-charge (like Arg-Glu) interactions are also common (Glaser et al., 2001; Jones and Thornton, 1996; Ofran and Rost, 2003). Smaller residues, such as Ser, Ala, Gly, Thr, Leu and Val are underrepresented (Ofran and Rost, 2003). In any case, it is believed that interacting proteins have both a high degree of surface and electrostatic complementarity (Glaser et al., 2001).

To date, the PDB structure of two Cu-bridged metallochaperone-target adducts, and a metallochaperone homodimer are available: Atox1 with MK1 (Banci et al., 2009a), the yeast Cu chaperone Atx1 with one of its target MBD (Ccc2a) (Banci et al., 2006b) and Atox1 homodimer (Wernimont et al., 2000). These structures show that the interacting interface between the metallochaperone and target domain is not a typical protein-protein interface as the ones described above. In particular, for the Cu-bridged Atox1-MK1 complex, the interacting surface is small and few hydrophobic but more charged residues are present (Banci et al., 2009a). Moreover, contrary to the residues propensities described above, Gly and Ala are overrepresented (Banci et al., 2009a). This "weak" interface appears to be a strategy for these proteins to interact only in the presence of Cu, and prevent the formation of stable complexes in the absence of Cu (Banci et al., 2009a).
To date, the adduct between Atoxl and any of the six WDs was not structurally characterized. In this Chapter we will study the nature, strength and dynamics of protein-protein interactions in the six Atoxl-WDs hetero-complexes by MD simulations. We showed in Chapter 3 that residues in (Met10 and Thr11) or near (Lys60) the Cu loop tune the flexibility of this loop in apo-Atoxl, whereas stabilize the Cu site in holo-Atoxl. In this Chapter, we will further assess if these residues play also a role in the formation of the Atoxl-WD hetero-complex. Since the protein-protein interface seems to be weak to begin with, we predict that mutations of residues that are directly involved in inter-protein interactions will have a large impact in adduct stability.

5.3 Copper transfer mechanism

Atoxl and WDs bind Cu(I) with essentially the same affinity (absolute values vary depending on the methods employed): 1-10 μM according to (Wernimont et al., 2004) and ~0.1 nM according to (Yatsunyk and Rosenzweig, 2007). Although the interacting interface in the metallochaperone-MBD adduct is not typical, it is stable enough for NMR (Banci et al., 2006b; Banci et al., 2009a) and gel-filtration (Hussain et al., 2009) isolation. This suggests the apparent absence of a thermodynamic gradient for Cu(I) transfer (Huffman and O'Halloran, 2000). On the other hand, because of the strength of the Cu(I) and Cu(II) ligand bonds, the Cu exchange rates in general are believed to be slow (Frausto da Silva and Williams, 2001). However, the presence of the Cu binding motif in a surface exposed loop, instead of hidden inside the core, may be a means of increasing Cu exchange rates. It has been proposed that the yeast homolog Atxl functions as an enzyme, in the sense that it lowers the kinetic barrier of Cu(I) transfer.
(Huffman and O'Halloran, 2000). How is then Cu(I) transferred from Atox1 to a target domain?

The Cu-transfer reaction can be thought of as having a starting point or reactants (holo-Atox1 and apo-WD), an end point or products (apo-Atox1 and holo-WD), and one or more intermediates. The nature of such intermediates involves an Atox1-Cu-WD ternary complex (intermediates are considered ternary complexes, since the Cu(I) atom is shared between the two proteins). All these species represent minima in the PES of the transfer process. In principle, the Cu-transfer reaction could potentially involve 2-, 3- and/or 4-coordinate Cu-intermediates (Banci et al., 2009a; Holt and Merz, 2007; Wernimont et al., 2000), through single or multiple reaction steps. However, the precise mechanism, thermodynamics and kinetics of this reaction are unknown. Because the transfer reaction involves bond breaking and forming, this process cannot be studied by MM methods, as we have done in almost all our previous simulations. Therefore, in this Chapter, we will probe the existence of all possible intermediates and reaction pathways of Cu(I) transfer from holo-Atox1 to one WD, via QM-MM calculations. The nature of the intermediates will be studied assuming that once Atox1 and WD come to close contact, Cu(I) transfer will take place.

5.4 Computational Methods

5.4.1 Initial structures

The initial structures for the six different reactant Atox1c-WD hetero-complexes (consisting in holo-Atox1 and each apo-WD), and WD4c-WD5 and WD4c-WD6 hetero-complexes (consisting in holo-WD4 and apo-WD5 and WD6, respectively) were
generated using the final structures obtained from the MD simulations of the corresponding WDs (see Chapter 4) and holo-Atox1 (see Chapter 3). The initial structure for the product Atox1-WD4c hetero-complex (consisting in apo-Atox1 and holo-WD4) is the same that the reactant, except that the Cu(I) atom coordinates the Cys in WD4. At the time we started this analysis, the structure of Atox1-Cu-MK1 was not available (Banci et al., 2009a). Therefore, the relative orientation of the proteins in the Atox1-WD hetero-complex was obtained by homology threading calculations (as explained in Chapter 2) using the holo-Atox1 homodimer crystal structure (1FEE.pdb) (Wernimont et al., 2000) as reference. Based on NMR data, including the recently published Atox1-Cu-MK1 structure (Achila et al., 2006; Banci et al., 2007; Banci et al., 2008; Banci et al., 2009a; Banci et al., 2009b), the contact regions in our hetero-complexes match the contact regions in the metallochaperone-target adduct. The following mutations were generated in silico in the Atox1c-WD4 hetero-complex structure: Met10Ala, Thr11Ala, Thr11Ser, Lys60Ala and Lys60Tyr in Atox1.

5.4.2 MM-MD simulations

The initial structures were solvated as explained in Chapter 2; water molecules extended at least 10 Å from the surface of the proteins. The parameters of the coordinating Cys and Cu(I) in holo-Atox1 (for the reactant Atox1c-WDs hetero-complexes) and holo-WD4 (for WD4c-WD5, WD4c-WD6 and the product Atox1-WD4c hetero-complexes) were the same to those reported in Chapter 3 and Chapter 4, respectively. All MM-MD simulations were performed using AMBER, as explained in Chapter 2. The total simulation time for the different hetero-complexes varied: ~30-40
ns for Atox1c-WD1,2,3,5,6, WD4c-WD5 and WD4c-WD6 hetero-complexes; -50 ns for the mutant Atox1c-WD4 hetero-complexes; and 120 ns for the reactant (Atox1c-WD4) and product (Atox1c-WD4c) hetero-complexes. Because there is only PDB-reported structure for WD4 (Banci et al., 2008) (and not of WD2), and Atox1c-WD4 was the strongest complex, as we will see later, we decided to further explore its conformational space by simulating this hetero-complex longer time (in comparison to the other hetero-complexes). However, when comparisons are made between the hetero-complexes, the same simulation time frame is analyzed.

5.4.3 Initial structure for QM-MM calculations

A snapshot from the MM-MD simulation of the reactant Atox1c-WD4 hetero-complex, in which WD4’s Cys were close to Atox1’s active site (S-Cu < 5Å) was selected as the initial structure of the reactant. Starting from this structure we simulated 1 ns of MM-MD constraining WD4’s Cys geometries by applying a harmonic restraint to both SWD4-CuAtox1 distances (equilibrium distance of 3.5 Å) and to both WD4’s Cys dihedral angles along the axe of the CA-CB bond, so that SG points towards Atox1’s Cu(I) atom, followed by a 200 ps SA calculation in which the temperature was slowly decreased to 0 K, applying the same geometrical constraints. We then performed a free MM geometry optimization of the whole system followed by a QM-MM optimization in which the Cu(I) atom plus the four CH3S- groups are treated as the QM subsystem (see simulation details below).
5.4.4 QM-MM calculations

To investigate the reaction mechanisms of Cu(I) transfer from holo-Atoxl to WD4, we performed QM-MM restrained energy optimizations along selected reaction coordinates, using SIESTA (Chapter 2). For this purpose, an additional term is added to the potential energy according to equation 2.39 (Chapter 2). In all cases, the force constant was 100 kcal/(mol Å²) and minimizations were performed at each value of $\zeta_0$ until the gradient was 0.04 eV/Å. Each reactant, intermediate and product was reoptimized without any constraints until the gradient was 0.03 eV/Å. For reactions A through S, the Cu(I) atom plus the four CH$_3$S$^-$ groups were selected as the QM-subsystem, which comprises 21 atoms. For reactions A’, B’ and R’ the Cu(I) atom, three CH$_3$S$^-$ groups plus one methylthiol (CH$_3$SH) group (WD4’s Cys2, WD4’s Cys1 and Atoxl’s Cys2 for reactions A’, B’ and R’, respectively) were selected as the QM-subsystem, comprising 22 atoms. The rest of the protein, counter ions and water molecules were treated classically.

5.5 Investigating the stability of Atoxl-WD and WD-WD hetero-complexes

5.5.1 Choosing the best reactant: Atoxl-WD hetero-complexes

One of the main goals of this Chapter is to elucidate the reaction mechanism of Cu(I) transfer from Atoxl to an apo-WD. To do this, we first need to build the reactant. In the actual reaction, there are two reactants: holo-Atoxl and apo-WD, as individual and separated proteins. However, as a first approximation to address the problem, we will assume that holo-Atoxl and the apo-WD are already forming a complex. With this assumption, the number of reactants reduces to one: holo-Atoxl-WD hetero-complex.
This is not a ternary complex (and that is why we called it hetero-complex), since Cu(I) is only coordinated to Atox1. Next, since there are six different WDs, we need to chose the best reactant for our Cu-transfer reaction, i.e., we will pick the most stable reactant. To do so, we generated the six different possible hetero-complexes between holo-Atox1 and WDs (Atox1c-WDs) and subjected them to up to ~40 ns MD simulations (Figure 5.1A). Figure 5.2A shows Atox1c-WD4 as an example.

All Atox1c-WD hetero-complexes except for Atox1c-WD3 were stable during the simulation time, maintained a similar conformation (see Figure 5.2A for Atox1c-WD4 structure at ~40 ns) and exhibited similar backbone fluctuations (Figure 5.1B). On the other hand, Atox1c-WD3 hetero-complex was unstable and floppy; on average, fluctuations are three times greater than in the other hetero-complexes. After ~10 ns the interacting surface between Atox1c and WD3 weakens and the complex adopts a different conformation, in which both proteins are further apart from each other, as testified by a marked increase in the backbone rmsd time evolution (Figure 5.1A). Finally, after ~25 ns, the hetero-complex falls apart completely. Because of this, data analysis of Atox1c-WD3 was performed for the first ~20 ns only.

To quantify the strength and nature of the protein-protein interactions, we calculated the ΔG of hetero-complex formation (Table 5.1). This analysis shows that WD4 is the WD that binds strongest to Atox1c with ΔG ~ -9.5 kcal/mol, whereas WD3 binds the weakest with ΔG ~ -4.2 kcal/mol (value calculated before the complex falls apart). The second best Atox1c binder is WD5, followed by WD2, WD1 and WD6, which all have similar ΔG values (Table 5.1). The intermolecular vdW interactions that contribute to the calculated ΔG are in the range of -45 to -50 kcal/mol in the different
hetero-complexes, except in Atox1c-WD3 complex in which it is -12 kcal/mol (Table 5.1).

**Figure 5.1.** A: Rmsd (in Å, with respect to the first structure) of backbone heavy atoms as a function of simulation time for each hetero-complexes. For Atox1c-WD4, only the first ~40 ns of the total ~120 MD simulation are shown. B: Average fluctuations (rmsf in Å) of backbone heavy atoms per residue (of the last 10 ns of MD simulations) for each hetero-complexes. For Atox1c-WD4, the analysis was done for the first 40 ns of the total ~120 MD simulation, since the rest of the hetero-complexes were simulated ~30-40 ns. For Atox1c-WD3 and WD4c-WD5, the analysis was done for the first ~20 and ~13 ns, respectively, before the complex falls apart. The approximate limits between the apo- (WD1-6 and WD5/6 in Atox1c-WDs and WD4c-WDs hetero-complexes, respectively) and holo- (Atox1c and WD4c in Atox1c-WDs and WD4c-WDs hetero-complexes, respectively) proteins are indicated. Black, Atox1c-WD1; red, Atox1c-WD2; blue, Atox1c-WD3; green, Atox1c-WD4; orange, Atox1c-WD5; cyan, Atox1c-WD6; brown, WD4c-WD5; purple, WD4c-WD6.
Figure 5.2. Holo-Atox1-WD4 hetero-complex. A: Structure corresponding to the ~40 ns snapshot of the MD simulation; the Cu(I) atom and residues C₁ and C₂ from the conserved MX₁C₁X₂X₃C₂ motif for each protein are shown (top), and surface representation of the same structure (bottom). B: Zoom-in of holo-Atox1 and WD4 active sites (Cu and α₂-β₄ loops) showing key intermolecular interactions. C: Intermolecular interactions between helices α₁, α₂ and the α₂-β₄ loop. In all cases, holo-Atox1 is in magenta and WD4 in green. Distances between atoms are in Å.
Table 5.1: Free energy ($\Delta G$) of Atoxic-WDs, WD4c-WD5 and WD4c-WD6 hetero-complex formation (calculated with *sietraj* as explained in Chapter 2), including its different energetic contributions: intermolecular van der Waals interactions (vdW), change in nonpolar solvation energy (cavity), intermolecular Coulomb interactions (electrostatic) and change in reaction field energy (field). All energy values are in kcal/mol; errors correspond to standard deviation. Reported energy values correspond to the average of 100 calculations based on the last 10 ns of MD simulations of each hetero-complex. For Atoxic-WD4, the analysis was done for the first 40 ns of the total ~120 MD simulation, since the rest of the hetero-complexes were simulated ~30-40 ns. For Atoxic-WD3 and WD4c-WD5, the analysis was done for the first ~20 and ~13 ns, respectively, before the complex falls apart.

<table>
<thead>
<tr>
<th></th>
<th>$\Delta G$</th>
<th>vdw</th>
<th>cavity</th>
<th>electrostatic</th>
<th>field</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atoxic-WD1</td>
<td>-8.1 ± 0.6</td>
<td>-47 ± 5</td>
<td>-8.4 ± 0.6</td>
<td>-74 ± 11</td>
<td>79 ± 10</td>
</tr>
<tr>
<td>Atoxic-WD2</td>
<td>-8.2 ± 0.7</td>
<td>-48 ± 6</td>
<td>-9.0 ± 0.6</td>
<td>-99 ± 11</td>
<td>106 ± 9</td>
</tr>
<tr>
<td>Atoxic-WD3</td>
<td>-4.2 ± 0.6</td>
<td>-12 ± 5</td>
<td>-2 ± 1</td>
<td>-38 ± 12</td>
<td>40 ± 11</td>
</tr>
<tr>
<td>Atoxic-WD4</td>
<td>-9.5 ± 0.4</td>
<td>-50 ± 4</td>
<td>-9.7 ± 0.4</td>
<td>-172 ± 12</td>
<td>169 ± 10</td>
</tr>
<tr>
<td>Atoxic-WD5</td>
<td>-9.1 ± 0.5</td>
<td>-50 ± 5</td>
<td>-9.4 ± 0.8</td>
<td>-120 ± 11</td>
<td>119 ± 9</td>
</tr>
<tr>
<td>Atoxic-WD6</td>
<td>-8.0 ± 0.5</td>
<td>-45 ± 4</td>
<td>-8.5 ± 0.7</td>
<td>-50 ± 9</td>
<td>54 ± 8</td>
</tr>
<tr>
<td>WD4c-WD5</td>
<td>-4.1 ± 0.4</td>
<td>-18 ± 4</td>
<td>-2.8 ± 0.7</td>
<td>40 ± 15</td>
<td>-31 ± 15</td>
</tr>
<tr>
<td>WD4c-WD6</td>
<td>-6.3 ± 0.6</td>
<td>-29 ± 5</td>
<td>-6 ± 1</td>
<td>-30 ± 8</td>
<td>32 ± 8</td>
</tr>
</tbody>
</table>

All electrostatic interactions are favorable, consistent with all WDs having a certain degree of charge complementarity with Atoxic (Figure 4.10). However, the contributions differ among the hetero-complexes (Table 5.1). Atoxic-WD4 hetero-complex has the greatest intermolecular electrostatic interaction, which is also accompanied with the greatest penalty of charge burial (change in reaction field energy, Table 5.1). The second strongest intermolecular Coulomb interactions are found in the Atoxic-WD5 hetero-complex, followed by Atoxic-WD2, -WD1, -WD6 and -WD3 (Table 5.1). Taking into account the polar contributions to the $\Delta G$ (electrostatic + reaction field), we can see that this contribution is only favorable in Atoxic-WD4 and -WD5 hetero-complexes. These results are in complete agreement with and provide a quantitative interpretation of the surface electrostatic potential calculations of the individual domains (Figure 4.10). Finally, the last contribution to the calculated $\Delta G$ is
the change in nonpolar solvation energy, which is proportional to the SASA that becomes buried upon hetero-complex formation. The Atox1c-WD4 hetero-complex has the greatest nonpolar solvation energy contribution (cavity, Table 5.1) and is evidenced in the surface rendering of the complex (Figure 5.2A). The cavity contribution in Atox1c-WD2 and -WD5 is similar to the one in Atox1c-WD4, in accord with these complexes having a similar interacting surface. On the contrary, in Atox1c-WD3, the cavity contribution is very small, in agreement with these proteins being further apart, and not in a tight unit as in the rest of the hetero-complexes.

The energy contributions to the calculated ΔG are the result of a number of specific intermolecular contacts between Atox1c and the WDs that form the interaction interface. In Atox1c-WD4, an extensive and stable electrostatic network of interactions keeps the active sites in close contact (Figure 5.2B). Thr11 in Atox1c forms alternatively an intermolecular HB with Thr13 (both residues correspond to X1 of the conserved motif) and Cys14 (1st Cys) in WD4, whereas Lys60 in Atox1c forms alternatively an intermolecular HB with WD4’s Thr13 and Ser16 (residue X3 of the conserved motif), and interacts with both WD4’s Cys. As was previously proposed (Wernimont et al., 2000), Thr11 in Atox1 HBs with the first Cu-coordinating Cys of the target domain. At the same time, Thr13 in WD4 interacts with Cys15 (2nd Cys) in Atox1c. This extensive HB and electrostatic network appears to stabilize the hetero-complex. On the other hand, Phe66 in WD4 points to Atox1c’s α2-β4 loop, where it interacts with Thr58, Gly59 and Lys60, as well as with Ala18 in Atox1c’s helix α1.

Close to the active sites, interactions between helices α1 and α2 and the α2-β4 loop also contribute to the binding interface (Figure 5.2C). Gly59 in Atox1c interacts
with Ser16 and Ser20 in WD4 helix α1. Also, Met64 in WD4 helix α2 (only present in WD1, WD2 and WD4, but replaced by another hydrophobic residue in the rest, Figure 4.1) interacts with Lys57 and Thr58 backbones in Atox1c helix α2. Finally, three strong intermolecular salt bridges stabilize the hetero-complex (Figure 5.2C): Lys57 in Atox1c with Asp63 in WD4 helix α2 (conserved among the WDs except in WD3 and replaced by a conservative Glu in WD6, Figure 4.1), Arg21 in Atox1c helix α1 with Glu67 in WD4 strand β4 (present in WD1, WD2, WD4 and WD5, Figure 4.1), and Lys25 in Atox1c helix α1 with Glu62 in WD4 helix α2 (only present in WD3, WD4 and WD6, Figure 4.1).

In Atox1c-WD2 and -WD6 hetero-complexes, the active sites interact in a similar way and with a similar orientation to Atox1c-WD4, although the strength of the interactions is not the same. In Atox1c-WD1 and -WD5, the active sites are further apart and are orientated in a different way: the corresponding X1 residue (Thr) in each protein does not always interact with the partner’s active site. This difference may affect the efficiency of formation of a ternary complex in vitro or in vivo. It appears that interactions between residue X1 and the target’s active site in both metallochaperone and WDs are important to assure a proper orientation of the proteins in the hetero-complex. In Atox1c-WD2 and -WD5, similar salt bridges to the ones found in Atox1c-WD4 stabilize the Atox1c-WDs hetero-complexes. In Atox1c-WD1 and -WD6, however, only two of them are present, which agrees with the intermolecular Coulombic contribution to the ΔG (Table 5.1). On the other hand, the Atox1c-WD3 hetero-complex is weak; before the hetero-complex falls apart, the proteins make only few intermolecular vdW contacts and a single salt bridge.
5.5.2 Probing inter-domain interactions: WD4-WD5/6 hetero-complexes

Because Cu-ATPases from higher organisms contain multiple MBDs separated by linkers of different length, it has been proposed that Cu(I) could be transferred intramolecularly from one domain to the other. To date, there are two studies that report the existence of such type of interactions (Achila et al., 2006; Bunce et al., 2006). In one NMR study, holo-WD4 preferentially delivered Cu(I) first to WD6 and then to WD5 in a WD56 construct, without the formation of detectable adducts (Achila et al., 2006). In Chapter 4, based on the surface electrostatic potential and Cu loop flexibility and exposure of individual domains, we suggested possible reasons for why WD4c preferentially delivers Cu(I) to WD6 in a WD56 construct. Now, to quantify the nature and strength of interactions between WD4 and WD5/6, we generated the WD4c-WD5 and WD4c-WD6 hetero-complexes, subjected them to ~40 ns MD simulations (Figure 5.1A), and calculated the ΔG of hetero-complex formation (Table 5.1).

As expected based on our previous predictions, a strong difference between these hetero-complexes is observed. WD4c-WD6 hetero-complex is relatively stable, although it exhibits greater fluctuations than those found in Atox1c-WDs hetero-complexes (Figure 5.1). On the other hand, WD4c-WD5 hetero-complex is very unstable and floppy (Figure 5.1). In fact, after ~13 ns WD4c and WD5 interacting surface is compromised, and the proteins interact non-specifically, as evidenced by a marked increase in the backbone rmsd time evolution data (Figure 5.1A). This difference is also evidenced by the ΔG of interaction (Table 5.1): WD4c-WD6 hetero-complex is more than 2 kcal/mol more stable than WD4c-WD5. All energy contributions are reduced in WD4c-WD5 hetero-complex. In fact, Coulombic interactions are unfavorable between these two
proteins, as opposed to all other hetero-complexes investigated. This is consistent with
the surface electrostatic potential of these proteins (Figure 4.10), which shows that both
WD4c and WD5 surfaces are largely negative, whereas WD6 surface is more positive.
Finally, the change in nonpolar solvation energy is more than 3 kcal/mol more stable in
the WD4c-WD6 hetero-complex (Table 5.1), indicating that the complex is more
compact and involves a bigger surface area. These results all-together may explain why
WD4c preferentially interacted with and delivered Cu(I) to WD6 when presented to a
WD56 two-domain construct (Achila et al., 2006).

5.5.3 Biological relevance

In our structural models of Atox1-WDs hetero-complexes, the orientation of
Atox1 and the target domains was obtained using the holo-Atox1 homodimer crystal
structure (Wernimont et al., 2000) as reference, because no structure of an Atox1-WD or
Atox1-MK adduct was available at that time. Recently, the solution structure of a Cu-
bridged Atox1-MK1 adduct was determined (Banci et al., 2009a). A structural
comparison between our Atox1c-WD1 model and the Atox1-Cu-MK1 structure reveals a
backbone rmsd of 1.7 Å (Figure 5.3), although MK1 and WD1 share only ~42 %
sequence identity. Both structures show the same orientation of the metallochaperone and
target domain, and similar intermolecular interactions. This analysis validates not only
our Atox1c-WD1 structure, but also all our Atox1c-WDs structural models.
Figure 5.3. Structural comparison of Atoxl-Cu-MK1 adduct (up, 2K1R.pdb) and our holo-Atoxl-WD1 hetero-complex model (bottom). The Cu(I) atom and residues X1, C1, C2 and Lys60 of Atoxl, and residues X1, C1, X3 and C2 from the conserved MX1C1X2X3C2 motif of the target domain are shown. Atoxl is in magenta, MK1 in blue and WD1 in silver.

Simulations of the six different Atoxlc-WD hetero-complexes provide insights into the nature and strength of intermolecular interactions. We found that from the six WDs, WD4 binds most strongly to holo-Atoxl, with the electrostatic interactions having the greatest contribution. This was expected based on the surface electrostatic potential of the individual domains (Figure 4.10), which predicted that the complementary charged surface would be maximized in the Atoxlc-WD4 hetero-complex. Although the Atoxlc-WD2 hetero-complex is also energetically favorable, we found a difference of ~1.3
kcal/mol in favor of the Atox1c-WD4 complex, which explains why, when presented as individual domains in vitro, WD4 interacts stronger than WD2 with holo-Atox1 (Achila et al., 2006). Moreover, our results are consistent with Atox1 and WD4 interacting in a stable way in the presence of Cu(I), being this complex able to be isolated by gel filtration (Hussain et al., 2009).

Holo-Atox1 forms an hetero-complex with WD1 with similar ΔG than the one formed with WD2, which explains their interaction in vitro (Banci et al., 2009b). Although WD6 also exhibits a similar, yet smaller, ΔG with Atox1c, the intermolecular vdw and electrostatic interactions are weaker, indicating that this complex is not as energetically favorable as the rest. On the other hand, WD3 binds very poorly to holo-Atox1, with a difference in ΔG of over 5 kcal/mol as compared to the strongest Atox1 binder, WD4. This explains why holo-Atox1 preferentially delivered Cu(I) to WD4 in a WD34 construct in vitro (Banci et al., 2008). To note, electrostatic interactions appear to play an important role in Atox1c-WD adduct formation. These interactions are the component of ΔG that varies the most among the six hetero-complexes and strongly correlate with the stability of the hetero-complexes (Table 5.1).

In Chapter 4, we suggested that because holo-Atox1 was rigid, it would preferentially bind to more floppy domains. As predicted, the most flexible domains, WD4 and WD5 (Chapter 4), form the most stable hetero-complexes with the rigid chaperone holo-Atox1. However, no metallochaperone-target adduct formation was detected in vitro for WD5 (Achila et al., 2006; Banci et al., 2009b). We propose two possible explanations for this discrepancy. First, our classical MD simulations do not allow for bond breaking and forming, so formation of a ternary complex, in which Atox1
and WD share the Cu(I), is not possible. However, in vitro, the complex would likely survive only when the ternary complex is formed, facilitated by the correct positioning of the two active sites. We found that although Atox1c and WD5 form a stable hetero-complex, the corresponding active sites are not properly oriented, which may likely affect the stability of the complex and its detection in vitro. Second, we are studying the binding of individual domains to Atox1c, as opposed to experiments in which two-domain (Achila et al., 2006) or six-domain (Banci et al., 2009b) constructs were used. As seen in Chapter 4, presence of other domain/s may affect the dynamics and structure of the isolated domain. In fact, whereas WD5’s flexibility is significantly reduced in the WD56 construct, WD4 is still flexible in the WD34 construct (Chapter 4). Thus, backbone rigidity (in a two-domain construct) and/or its relative location in the full N-terminus of ATP7B (in a six-domain construct) may explain lack of detectable interactions between WD5 and Atox1 in vitro.

Overall, we can conclude that WD4 appears to be the domain best optimized for interaction with Atox1. The dynamic and thermodynamic differences found here among the different Atox1c-WDs hetero-complexes, together with the relative location of each WD in the entire N-terminus of ATP7B and the surrounding media may affect the stability and detection of the ternary complexes in vitro and in vivo.

Interestingly, we found that inter-domain interactions are much weaker than Atox1-WDs interactions. This finding agrees with the inability to detect a Cu-dependent WD-WD adduct by NMR, as opposed to Atox1-WD adducts, although WD-WD Cu(I) transfer was demonstrated (Achila et al., 2006). This suggests that the metallochaperone, although sharing a similar fold and metal-binding properties with the WDs, has evolved
to interact in an efficient way with its partner domains. Because of the presence of multiple WDs separated by linker of different lengths, these domains may in principle form strong and stable adducts, blocking the access of the metallochaperone to the target sites and the transfer of Cu(I) across the membrane. However, if intramolecular Cu(I) transfer from WD4 to WD6 occurs in vivo, it is possible that a lower WD4c-WD6 hetero-complex stability assures rapid Cu(I) transfer to WD6, promoting vectorial Cu(I) transfer across the membrane. At the same time, in this model and based on the stability of the complexes, holo-Atox1 would easily displace WD6 from the WD4-WD6 adduct, contributing to the net irreversible Cu(I) transfer across the membrane.

5.6 Role of conserved residues in Atox1-WD4 interactions

We have seen in Chapter 3 that residues in (Met10 and Thr11) and near (Lys60) the Cu loop are important for the stability and flexibility of Atox1 in both apo- and holo-forms. Now is the time to assess if these same residues are important for interactions between Atox1 and its targets. In section 5.5.1 we have anticipated that Thr11 and Lys60 interact directly with the target’s active site. However, the exact role of these residues in hetero-complex stability and/or Cu(I) transfer is not clear. Because Atox1 appears to form the strongest complex with WD4, both in vitro (Achila et al., 2006) and in silico, and because there is PDB-reported structure for WD4 (Banci et al., 2008) and not for WD2, we decided to continue our analysis on this hetero-complex only. Therefore, to explore the molecular roles of Met10, Thr11 and Lys60 in holo-Atox1-WD4 adduct stability, we performed MD simulations of the Met10Ala, Thr11Ala, Thr11Ser, Lys60Ala and Lys60Tyr Atox1 variants hetero-complexes.
All hetero-complexes are stable during each MD simulation of ~50 ns, except that involving the Lys60Ala mutant (Figure 5.4A). Lys60Ala(Atox1c)-WD4 hetero-complex is only stable for the first ~20 ns, followed by a conformational change in which the proteins interact non-specifically, as testified by an increase in the backbone rmsd (Figure 5.4A). Instability of the Lys60Ala(Atox1c)-WD4 complex is also evidenced by a more than two-fold increase in fluctuations with respect to WT(Atox1c)-WD4 (Figure 5.4B).

As explained in section 5.5.1, in the WT(Atox1c)-WD4 hetero-complex, a stable interaction network keeps the two Cu-binding sites in close contact (Figures 5.2B and 5.5). This intermolecular network is disrupted in all mutants (at different extents), causing WD4’s Cys to be more distant from holo-Atox1’s Cu site (Figure 5.5). In Met10Ala, Thr11Ala and Thr11Ser Atoxic complexes, WD4’s Cys are most of the time facing Atox1 surface, whereas in the Lys60 Atoxic mutants, these Cys primarily face the opposite side. Met10Ala(Atoxic)-WD4 shows the behavior closest to WT(Atoxic)-WD4, a result consistent with the distance of this residue from the interacting surface (Figure 5.5).

Lack of a polar residue in position 11 of Atox1, as in Thr11Ala, disrupts an electrostatic interaction with WD4’s active site (Figure 5.5). As a result, WD4’s active site moves further away, although the two proteins remain in a tight and compact complex. In this complex, the active sites orientation of the metallochaperone and target domain seems not to be optimal for Cu(I) transfer. On the other hand, presence of a polar residue in position 11 of Atox1, as in Thr11Ser, appears to preserve a proper orientation
between the active sites, although the proteins are further apart from each other in comparison to WT.

Figure 5.4. A: Rmsd (in Å, with respect to the first structure) of backbone heavy atoms as a function of simulation time for WT and each Atox1 variant in Atox1c-WD4 hetero-complexes. For WT, only the first ~50 ns of the total ~120 MD simulation are shown. B: Average fluctuations (rmsf in Å) of backbone heavy atoms per residue (of the last 20 ns of MD simulations) for each Atox1c-WD4 hetero-complexes. For WT, the analysis was done for the first 50 ns of the total ~120 MD simulation, since the rest of the hetero-complexes were simulated ~50 ns. For K60A, the analysis was done for the first ~22 ns, before the complex falls apart. The limits that comprise each individual protein in the hetero-complex are indicated. Magenta, WT; blue, M10A; yellow, T11A; orange, T11S; red, K60A, and cyan, K60Y.
Figure 5.5. Zoom-in of holo-Atoxl and WD4 active sites (both Cu loops and Atoxl’s α2-β4 loop) in each hetero-complex; the structure after ~50 ns (except K60A) of each MD simulation is shown. For K60A(Atoxlc)-WD4, the structure corresponds to ~20 ns of the MD simulation, before the complex falls apart. Distances between atoms are in Å.

In the Lys60 Atoxl substitutions, the proteins lose considerable both intermolecular sidechain and backbone interactions, as important electrostatic and HB
interactions provided by Lys60 are disrupted (Figure 5.5). However, during at least 50 ns of the simulation, replacement of Lys60 by a HB donor (Tyr) prevents rupture of the hetero-complex, as opposed to the impact of an Ala substitution. Although a HB donor like Tyr is more tolerated than Ala, it seems that the prokaryotic residue is too big to allow proper docking of the target domain, as the proteins are still much further away in the Lys60Tyr Atoxl hetero-complex than in WT.

To quantify the strength of the inter-protein interactions, we estimated the ΔG of hetero-complex formation (Table 5.2).

Table 5.2. Free energy (ΔG) of holo-Atoxl-WD4 hetero-complex formation for Atoxl variants and WT (calculated with sietraj as explained in Chapter 2), including its different energetic contributions: intermolecular van der Waals interactions (vdW), change in nonpolar solvation energy (cavity), intermolecular Coulomb interactions (electrostatic) and change in reaction field energy (field). All energy values are in kcal/mol; errors correspond to standard deviation. Reported energy values correspond to the average of 100 calculations based on the last 10 ns of the MD simulations of each hetero-complex. For WT(Atoxlc)-WD4, the analysis was done for the first 50 ns of the total ~120 MD simulation, since the rest of the hetero-complexes were simulated ~50 ns only. For K60A(Atoxlc)-WD4, the analysis was done for the first ~20 ns of the MD simulation, before the complex falls apart.

<table>
<thead>
<tr>
<th></th>
<th>ΔG</th>
<th>vDW</th>
<th>cavity</th>
<th>electrostatic</th>
<th>field</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT(Atoxlc)-WD4</td>
<td>-9.5 ± 0.6</td>
<td>-53 ± 4</td>
<td>-9.9 ± 0.4</td>
<td>-144 ± 18</td>
<td>144 ± 15</td>
</tr>
<tr>
<td>M10A(Atoxlc)-WD4</td>
<td>-9.4 ± 0.7</td>
<td>-51 ± 6</td>
<td>-9.3 ± 0.6</td>
<td>-136 ± 14</td>
<td>134 ± 12</td>
</tr>
<tr>
<td>T11A(Atoxlc)-WD4</td>
<td>-9.7 ± 0.5</td>
<td>-57 ± 5</td>
<td>-9.7 ± 0.6</td>
<td>-119 ± 15</td>
<td>121 ± 13</td>
</tr>
<tr>
<td>T11S(Atoxlc)-WD4</td>
<td>-7.2 ± 0.6</td>
<td>-36 ± 5</td>
<td>-7 ± 1</td>
<td>-138 ± 16</td>
<td>140 ± 15</td>
</tr>
<tr>
<td>K60A(Atoxlc)-WD4</td>
<td>-6.1 ± 0.8</td>
<td>-28 ± 6</td>
<td>-5 ± 1</td>
<td>-65 ± 18</td>
<td>68 ± 16</td>
</tr>
<tr>
<td>K60Y(Atoxlc)-WD4</td>
<td>-7.9 ± 0.5</td>
<td>-43 ± 4</td>
<td>-8.4 ± 0.6</td>
<td>-95 ± 18</td>
<td>98 ± 15</td>
</tr>
</tbody>
</table>

Whereas Met10Ala and Thr11Ala Atoxl mutations do not alter overall hetero-complex stability, Thr11Ser, Lys60Ala and Lys60Tyr Atoxl mutations have detrimental effects on adduct stability (Table 5.2). Intermolecular vdw interactions correlate with stability: they are not altered in Met10Ala and Thr11Ala variants, but they are reduced in Thr11Ser and Lys60 variants. On the other hand, the intermolecular Coulombic
interaction is greatly reduced in both Lys60 mutant complexes, more than two fold in
Lys60Ala (Table 5.2). Consistent with Ser11 being able to engage in HB interactions as
opposed to Ala11, Thr11Ser hetero-complex has stronger electrostatic interactions than
Thr11Ala, although its overall stability is smaller. Moreover, the nonpolar solvation
energy is reduced only in the Thr11Ser and the Lys60 mutants (and again the greatest
change is found for Lys60Ala), evidencing a reduction in the packing of these complexes.

5.6.2 Biological relevance

To date, only few Atoxl mutants were probed for interactions with its partner
proteins. In a recent NMR study, the ability of Arg21Gln, Val22Ala and Thr58Ala Atoxl
variants to interact with MK1 was probed, and the results suggest that all three mutants
are able to interact with MK1 to a similar extent than WT (Banci et al., 2009a). Arg21
and Val22 are located in helix α1, whereas Thr58 is located at the end of helix α2. We
have seen in section 5.5.1 that Arg21 is involved in an intermolecular salt bridge with
WD4's Glu67, and the backbone of Thr58 interacts with WD4's Met64. In any case,
these three residues in Atoxl are located further away from the Cu loop, and therefore
also away from the partner's Cu loop in the hetero-complex. It appears then that the
interactions between the metallochaperone and target domain away from the Cu loops are
not essential for complex stability.

Our results provide insights into the role that conserved residues in (Met10 and
Thr11) or around (Lys60) the Cu loop of Atoxl play in the interaction with one of its
partners WD. Furthermore, our results also agree with experimental results on the same
complexes (except Thr11Ser, which was not tested experimentally) (Hussain et al.,
2009). For example, *in vitro*, from all Atoxl variants, Met10Ala exhibited the most WT-like behavior (Hussain et al., 2009), which agrees with our *in silico* results. The amount of complex formation and Cu(I) transfer from Met10Ala(Atoxlc) to WD4 was reduced ~22% and 30%, respectively, with respect to WT (Hussain et al., 2009). These differences may probably be due to the fact that, as seen in Chapter 3, Met10Ala lowers stability of Atoxl. It thus appears that Met10 is not conserved in this class of proteins for protein-partner recognition, as a Met10Ala mutation does not significantly compromises Atoxlc-WD4 adduct stability. Met10 may therefore be important for preserving the integrity of the ferredoxin-like fold, as seen in Chapter 3.

On the other hand, Thr11 appears to be important for placing the metallochaperone in a correct position for interaction with its target domain. In WT Atoxl, Thr11 HBs to both Thr13 and Cys14 of WD4, and this HB network contributes to the proper docking of WD4. In agreement, based on the Atoxl Cu-bridged homodimer, Thr11 was proposed to HB with the first Cu-coordinating Cys of the partner MBD during Cu(I) transfer (Wernimont et al., 2000). Mutation of this residue to Ala does not affect adduct stability (in fact, we observe an increase of stability), but instead alters the docking of the two molecules, in a way that the active sites are not facing each other. This finding agrees with the higher proportion of Cu-containing Thr11Ala(Atoxlc)-WD4 adduct detected by gel-filtration (~17% more than WT) but decreased amount of Cu(I) transfer to WD4 (~70% less than WT) (Hussain et al., 2009). However, it appears that a HB donor at position 11 in Atoxl is not enough to allow efficient Cu(I) transfer to the partner domain. Mutation of Thr11 to the residue found in prokaryotic Cu chaperones, Ser, preserves a proper docking orientation between the two molecules, but nevertheless
has detrimental effects in adduct stability (more than 2 kcal/mol less stable than WT). This result suggests that a HB donor is necessary at position 11 of Atoxl to allow proper docking of the incoming target molecule, but Ser cannot replace Thr. This subtle difference between the prokaryotic and eukaryotic systems suggests that each metallochaperone is best designed to interact with their own cellular targets.

As opposed to any other Atoxl variant, in the Lys60(Atoxlc)-WD4 mutants, WD4’s Cys are not facing Atoxl surface but primarily face the opposite side. This may explain the much lower extent of Cu(I) transfer to WD4 found experimentally for these mutants (~98 % and 85 % lower than WT, for Lys60Ala and Lys60Tyr variants, respectively) (Hussain et al., 2009). Moreover, we observed strong detrimental effects on adduct stability of Lys60Ala and Lys60Tyr Atoxl mutants, in accord with the absence of detectable complexes in experiments (Hussain et al., 2009). By participating in HBs and electrostatic interactions with WD4’s Thr13, Cys14, Ser16 and Cys17, Atoxl’s Lys60 appears to provide strong electrostatic stabilization of the hetero-complex and proper orientation of WD4’s Cys for efficient Cu(I) transfer. Whereas the Lys60Ala hetero-complex fell apart during our MD simulation, the Lys60Tyr hetero-complex, although having ~2 kcal/mol less stability than WT, did not fall apart. It appears then that the 50 ns simulated here were not enough to observe the rupture of the complex.

In summary, the interacting surface between Atoxl and WD4 appears to be the result of the sum of small interactions, in which Lys60-mediated electrostatic interaction plays a dominant role. This interface can be disrupted with at least one substitution at position 60 in Atoxl, which points to the subtle balance of forces that contribute to hetero-complex stability, as also recently suggested for the similar Atoxl-MK1 adduct.
based on NMR data (Banci et al., 2009a). Thus, metallochaperone-target electrostatic complementarity (provided by Lys60) appears essential to assure sufficient hetero-complex stability and thereby also efficient Cu(I) transfer.

5.7 Copper transfer reaction pathway search: from holo-Atox1 to WD4

Next, we investigated the reaction pathway of Cu(I) transfer from holo-Atox1 to WD4 using QM-MM calculations (Figure 5.6). To do so, we performed restrained energy minimizations (as explained in Chapter 2) along selected reaction coordinates defined in Table 5.3. In this way, we probed the energetics and kinetics of all possible reaction pathways, involving all possible 2-, 3- and/or 4-coordinate Cu-intermediates (Table 5.4).

![Figure 5.6. Left: QM-MM optimized structure of the reactant Atox1c-WD4 hetero-complex. The QM subsystem (Cu(I) plus four CH3S' groups) is shown in sticks; the rest of the system (protein plus solvent) corresponds to the MM subsystem. Right: Zoom-in of holo-Atox1 and WD4 active sites. The QM subsystem is shown in sticks, and relevant MM residues (Atox1's Thr11 and Lys60, plus three water molecules) are shown in CPK.](image-url)
Table 5.3. Definition (ζ) of reaction coordinate in equation 2.39 for restrained energy minimizations (Chapter 2), initial $\xi_0^{\text{initial}}$ and final $\xi_0^{\text{final}}$ values (in Å) of the reference reaction coordinate, for each reaction A through S, A', B' and R'. ζ is always defined as a linear combination of the distance between Atoxl’s and WD4’s (SG)Cys1 and (SG)Cys2 (S1-S4) and Cu(I). S1 and S2, Atoxl’s (SG)Cys1 and (SG)Cys2, respectively; S3 and S4, WD4’s (SG)Cys1 and (SG)Cys2, respectively.

<table>
<thead>
<tr>
<th>reaction</th>
<th>reaction coordinate (ζ)</th>
<th>$\xi_0^{\text{initial}}$</th>
<th>$\xi_0^{\text{final}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>(S3-Cu)</td>
<td>3.5</td>
<td>2.3</td>
</tr>
<tr>
<td>B</td>
<td>(S4-Cu)</td>
<td>3.8</td>
<td>2.3</td>
</tr>
<tr>
<td>C</td>
<td>(S4-Cu)</td>
<td>3.8</td>
<td>2.3</td>
</tr>
<tr>
<td>D</td>
<td>(S3-Cu) fixed</td>
<td>3.5</td>
<td>3.5</td>
</tr>
<tr>
<td>E</td>
<td>(S3-Cu) - (S3-Cu)</td>
<td>0.7</td>
<td>-1.3</td>
</tr>
<tr>
<td>F</td>
<td>(S3-Cu) - (S3-Cu)</td>
<td>-1.4</td>
<td>1.4</td>
</tr>
<tr>
<td>G</td>
<td>(S3-Cu) - (S3-Cu)</td>
<td>-1.4</td>
<td>1.4</td>
</tr>
<tr>
<td>H</td>
<td>(S3-Cu) - (S3-Cu)</td>
<td>-1.6</td>
<td>1.4</td>
</tr>
<tr>
<td>I</td>
<td>(S3-Cu) + (S4-Cu) - (S1-Cu)</td>
<td>5.1</td>
<td>0.8</td>
</tr>
<tr>
<td>J</td>
<td>(S3-Cu) + (S4-Cu) - (S2-Cu)</td>
<td>5.1</td>
<td>0.8</td>
</tr>
<tr>
<td>K</td>
<td>(S3-Cu) + (S4-Cu)</td>
<td>7.3</td>
<td>4.6</td>
</tr>
<tr>
<td>L</td>
<td>(S1-Cu)</td>
<td>2.3</td>
<td>3.6</td>
</tr>
<tr>
<td>M</td>
<td>(S2-Cu)</td>
<td>2.3</td>
<td>3.6</td>
</tr>
<tr>
<td>N</td>
<td>(S4-Cu)</td>
<td>3.6</td>
<td>2.3</td>
</tr>
<tr>
<td>O</td>
<td>(S1-Cu) - (S4-Cu)</td>
<td>-1.3</td>
<td>1.3</td>
</tr>
<tr>
<td>P</td>
<td>(S2-Cu) + (S4-Cu)</td>
<td>-1.3</td>
<td>1.3</td>
</tr>
<tr>
<td>Q</td>
<td>(S1-Cu) + (S2-Cu) - (S4-Cu)</td>
<td>1.0</td>
<td>5.0</td>
</tr>
<tr>
<td>R</td>
<td>(S1-Cu)</td>
<td>2.3</td>
<td>4.0</td>
</tr>
<tr>
<td>S</td>
<td>(S3-Cu) + (S4-Cu) - (S1-Cu) - (S2-Cu)</td>
<td>2.9</td>
<td>-2.8</td>
</tr>
<tr>
<td>A'</td>
<td>(S3-Cu)</td>
<td>3.5</td>
<td>2.3</td>
</tr>
<tr>
<td>B'</td>
<td>(S4-Cu)</td>
<td>3.7</td>
<td>2.3</td>
</tr>
<tr>
<td>R'</td>
<td>(S1-Cu)</td>
<td>2.3</td>
<td>3.7</td>
</tr>
</tbody>
</table>
Table 5.4. List of reactant, product and all possible intermediates in the Cu-transfer reaction from holo-Atoxl to apo-WD4, indicating the corresponding Cu(I) center geometry. S1 and S2, Atoxl’s (SG)Cys1 and (SG)Cys2, respectively; S3 and S4, WD4’s (SG)Cys1 and (SG)Cys2, respectively.

<table>
<thead>
<tr>
<th>reactant</th>
<th>2-coordinated intermediates</th>
<th>3-coordinated intermediates</th>
<th>4-coordinated intermediate</th>
<th>product</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><img src="image1" alt="Diagram" /></td>
<td><img src="image2" alt="Diagram" /></td>
<td><img src="image3" alt="Diagram" /></td>
<td><img src="image4" alt="Diagram" /></td>
</tr>
</tbody>
</table>

5.7.1 Finding the first intermediate

Starting from the optimized structure of holo-Atoxl and apo-WD4, the reactant (Figure 5.6), we first probed the energetics of binding either WD4’s Cys (reactions A and B, Table 5.3, Figure 5.7). Coordination of WD4’s Cys1 occurs without activation energy ($E_a$), and this intermediate, V, is more energetically stable than the reactant (Table 5.5). Coordination of WD4’s Cys2 occurs with a non-zero $E_a$ (Table 5.5) and yields a different intermediate than expected: VII instead of VI (Figure 5.7). This suggests that the 3-coordinate intermediate in which Cu(I) is coordinated by Cys1 from
both proteins (VII) is more stable than the one in which only Cys1 from Atoxl coordinates Cu(I) (VI).

\[ \text{Figure 5.7. Potential energy profile (kcal/mol) along the corresponding reaction coordinate (Å) defined in Table 5.3 for reactions A (blue) and B (red). The schematic of the reactant and product of each reaction is shown.} \]

Since we wanted to assess the energetics of binding of WD4’s Cys2, we performed two different reactions: (i) same as reaction B but fixing WD4’s Cys1 (reaction C, Table 5.3, Figure 5.8), to avoid formation of intermediate VII; and (ii) starting from intermediate VII to obtain intermediate VI (reaction D, Table 5.3, Figure 5.8). Both reactions yield the desired intermediate VI, which has higher energy than the reactant (Table 5.5) and it is not at a minimum in the PES (Figure 5.8). From these reactions, it appears that WD4’s Cys1 will bind first to Cu-Atoxl to form a stable 3-
coordinated intermediate (V). Binding of WD4’s Cys2 alone (without subsequent binding of WD4’s Cys1) is not energetically or kinetically favorable.

Table 5.5. Activation energy (E<sub>a</sub>), difference in potential energy (ΔE), and type of intermediate found for each reaction A through S, A’, B’ and R’, for the corresponding 1<sup>st</sup>, 2<sup>nd</sup> or 3<sup>rd</sup> steps.

<table>
<thead>
<tr>
<th>reaction</th>
<th>1&lt;sup&gt;st&lt;/sup&gt; E&lt;sub&gt;a&lt;/sub&gt; (kcal/mol)</th>
<th>2&lt;sup&gt;nd&lt;/sup&gt; ΔE (kcal/mol)</th>
<th>3&lt;sup&gt;rd&lt;/sup&gt; intermediates-like</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1st</td>
<td>2nd</td>
<td>3rd</td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>B</td>
<td>2.9</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>C</td>
<td>6.0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>D</td>
<td>4.5</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>E</td>
<td>0</td>
<td>17.3</td>
<td>-</td>
</tr>
<tr>
<td>F</td>
<td>0</td>
<td>10.8</td>
<td>-</td>
</tr>
<tr>
<td>G</td>
<td>2.2</td>
<td>1.2</td>
<td>-</td>
</tr>
<tr>
<td>H</td>
<td>4.0</td>
<td>7.8</td>
<td>-</td>
</tr>
<tr>
<td>I</td>
<td>0</td>
<td>8.4</td>
<td>-</td>
</tr>
<tr>
<td>J</td>
<td>1.3</td>
<td>8.9</td>
<td>-</td>
</tr>
<tr>
<td>K</td>
<td>0</td>
<td>5.6</td>
<td>-</td>
</tr>
<tr>
<td>L</td>
<td>17.5</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>M</td>
<td>12.6</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>N</td>
<td>7.5</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>O</td>
<td>11.5</td>
<td>3.8</td>
<td>-</td>
</tr>
<tr>
<td>P</td>
<td>9.5</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Q</td>
<td>10.1</td>
<td>2.1</td>
<td>3.7</td>
</tr>
<tr>
<td>R</td>
<td>8.1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>S</td>
<td>2.8</td>
<td>5.3</td>
<td>7.1</td>
</tr>
<tr>
<td>A’</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>B’</td>
<td>6.1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>R’</td>
<td>10.2</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

To assess the possibility of a 2-coordinated specie being the first intermediate of the Cu-transfer reaction, we performed reactions E-H (Table 5.3, Figure 5.9). None of these reactions occur in a concerted fashion, instead they all occur via formation of a 3-coordinated intermediate, which is always more stable than the product. This suggests that 2-coordinated intermediates are not likely to occur during the Cu-transfer process from Atox1 to WD4.
To form 2-coordinated species in which WD4’s Cys1 is bound to Cu(I), both reactions E and F start by binding WD4’s Cys1 to form the stable 3-coordinated intermediate V with no $E_a$ (Table 5.5, Figure 5.9). Following formation of intermediate V, the reactions continue to obtain the 2-coordinated meta-stable species (I and II). In both cases, de-coordination of Atox1’s either Cys occurs with a large kinetic barrier.
However, specie II is more stable than I, suggesting again that intermediates in which the Cys1 residues from both Atox1 and WD4 are bound to Cu(I) are the most stable.

Figure 5.9. Potential energy profile (kcal/mol) along the corresponding reaction coordinate (Å) defined in Table 5.3 for reactions E (black), F (green), G (red) and H (blue). The schematic of the reactant and product of each reaction is shown.

On the other hand, neither reaction G nor H yield the desirable intermediate (Figure 5.9), suggesting that these two di-coordinated intermediates (III and IV) are not
stable species in the Cu-transfer reaction, and therefore, not likely to be populated in vivo or in vitro. Both reactions begin by coordination of both WD4’s Cys and de-coordination of Atox1’s Cys2, to form the stable 3-coordinated intermediate VII (Table 5.5, Figure 5.9). As in reaction B, this step occurs with $E_a \sim 2$-4 kcal/mol. In the case of G, the reaction continues to form a similar 3-coordinated intermediate VII, which only differs by $\sim$1 kcal/mol. In the case of H, the reaction continues to form the final product of the global reaction, which is higher in energy by $\sim$8 kcal/mol. Overall, based on reactions E-H we propose that (i) 3-coordinated intermediates are more stable than 2-coordinate intermediates in the Cu-transfer reaction and (ii) the 3-coordinated intermediate V (more reactant-like) is more stable than the 3-coordinated intermediate VII (more product-like).

We next explored the possibility of obtaining the 3-coordinated intermediates VII or VIII as the first intermediates of the reaction (reactions I and J, Table 5.3, Figure 5.10) starting from the reactant geometry. As for reactions E-H, both reactions I and J occur in a non-concerted way, starting by the initial binding of WD4’s Cys1 to form intermediate V, with no (reaction I) or small (reaction J) kinetic barrier (Table 5.5). However, as opposed to reaction J, reaction I did not yield the desired intermediate VIII, suggesting that intermediate VII and not VIII is likely an intermediate of the global Cu-transfer reaction. In the case of I, the reaction proceeds to form the final product of the global reaction, with a kinetic barrier of $\sim$8 kcal/mol. However, in J, the reaction continues from intermediate V to finally form specie VII, after overcoming an $E_a$ of $\sim$9 kcal/mol.

Finally, we calculated the energy profile of forming the 4-coordinated specie (intermediate IX), in which Cu(I) is coordinated by both Cys of Atox1 and WD4
(reaction K, **Table 5.3, Figure 5.11**). In agreement with all reactions above, the initial step involves binding of WD4’s Cys1 without kinetic barrier, to form intermediate V (**Table 5.5**). Finally, WD4’s Cys2 binds to form the meta-stable tetrahedral intermediate IX with $E_a \sim 6$ kcal/mol. Intermediate IX is not a minimum in the PES (**Figure 5.11**), suggesting that a tetrahedral species is not likely to be a probable intermediate in the biological Cu-transfer reaction between these proteins.

**Figure 5.10.** Potential energy profile (kcal/mol) along the corresponding reaction coordinate (Å) defined in **Table 5.3** for reactions I (blue) and J (red). The schematic of the reactant and product of each reaction is shown.

Overall, these calculations suggest that the initial step in the global Cu-transfer reaction from Atox1 to WD4 is coordination of WD4’s Cys1 to form 3-coordinated intermediate V. This reaction occurs with no or small activation barrier and is
energetically favorable. Also, intermediate VII appears to be another 3-coordinated intermediate likely to occur during the global reaction.

![Figure 5.11. Potential energy profile (kcal/mol) along the corresponding reaction coordinate (Å) defined in Table 5.3 for reaction K. The schematic of the reactant and product is shown.](image)

5.7.2 Finding the second intermediate

Once the first intermediate was identified, we continued exploring the reaction mechanism by assessing the existence of a second intermediate. First, we probed the possibility of a 2-coordinated intermediate occurring after intermediate V (reactions L and M, Table 5.3, Figure 5.12). Consistent with reactions E and F, 2-coordinated intermediates I and II are not stable species in the potential energy profile, and both are formed with large kinetic barriers (Table 5.5). Next, we assessed the formation of a tetrahedral intermediate taking place after intermediate V (reaction N, Table 5.3, Figure 5.13). This reaction occurs with $E_a \sim 8$ kcal/mol and yields a meta-stable intermediate IX (Table 5.5), supporting our previous conclusions based on reaction K.
Figure 5.12. Potential energy profile (kcal/mol) along the corresponding reaction coordinate (Å) defined in Table 5.3 for reactions L (red) and M (blue). The schematic of the reactant and product of each reaction is shown.

Figure 5.13. Potential energy profile (kcal/mol) along the corresponding reaction coordinate (Å) defined in Table 5.3 for reaction N. The schematic of the reactant and product is shown.
Once, based on our findings, 2- and 4-coordinated species were ruled out from the reaction mechanism, we investigated the potential energy profiles of binding WD4’s Cys2 and de-coordinating either Atox1’s Cys1 or Cys2 to form 3-coordinated intermediates VII and VIII from intermediate V (reactions O and P, Table 5.3, Figure 5.14). Reaction O proceeds with high $E_a$ and the formation of one intermediate, which is similar to the product of this reaction (intermediate VIII). On the other hand, reaction P proceeds in a concerted fashion, yielding the stable intermediate VII after overcoming an activation barrier of 9.5 kcal/mol (Table 5.5). Consistent with our previous results based on reactions E-H, intermediate V is more stable than intermediate VII by ~6 kcal/mol; and is kinetically more feasible to form VII than VIII.

Figure 5.14. Potential energy profile (kcal/mol) along the corresponding reaction coordinate (Å) defined in Table 5.3 for reactions O (blue) and P (red). The schematic of the reactant and product of each reaction is shown.
We also probed the possibility of a concerted reaction towards the overall reaction product from intermediate V (reaction Q, Table 5.3, Figure 5.15). As expected, this reaction did not occur in a concerted fashion, but instead proceed with the existence of two VII-like intermediates, which are similar in energy, after overcoming an activation barrier of \(\sim 10\) kcal/mol (Table 5.5). All together, this group of reactions supports the existence of the 3-coordinated species VII as the second intermediate of the global Cu-transfer reaction from Atoxl to WD4. These results again emphasize that 3-coordinated species are more stable than 2- or 4-coordinated species.

![Figure 5.15](image)

**Figure 5.15.** Potential energy profile (kcal/mol) along the corresponding reaction coordinate (Å) defined in Table 5.3 for reaction Q. The schematic of the reactant and product is shown.

We finally studied the last step of the reaction, in which starting from intermediate VII, Atoxl’s Cys1 is de-coordinated from Cu(I) (reaction R, Table 5.3, Figure 5.16). This reaction occurs with \(E_a \sim 8\) kcal/mol to yield the product of the reaction, which is \(\sim 6\) kcal/mol less stable than the second intermediate VII (Table 5.5).
5.7.3 Overview of the reaction mechanism

Based on all the above reactions, we can suggest the most probable reaction mechanism of Cu(I) transfer from holo-Atoxl to apo-WD4 (Figure 5.17). The proposed reaction involves the two 3-coordinated intermediates V and VII, in which the Cys1 residues of both Atoxl and WD4 are always coordinated to the Cu(I) ion.

Table 5.6 lists relevant geometrical parameters of the reactant, the two intermediates and product of the overall reaction. In all cases, the Cu-center is surrounded by three water molecules that HB with Atoxl’s Cys1, WD4’s Cys1 and Atoxl’s Thr11 (see Figure 5.6 for reactant). Neither 2- nor 4-coordinated species appear to be intermediates of this reaction. This finding agrees with NMR data on Atxl-Ccc2a (Banci et al., 2006b) and Atoxl-MK1 (Banci et al., 2009a) Cu-dependent adducts in which only three of the four functional Cys residues were required for adduct formation, and among
these, the Cys1 residues of both the metallochaperone and target domain were strictly required. Also, the intermediates V and VII in our proposed reaction mechanism agree with Cys1 being the first Cys to bind Cu(I) during Cu(I) transfer (Banci et al., 2006b).

Figure 5.17. Top: Proposed reaction mechanism of the Cu-transfer reaction, showing a schematic of the Cu(I) center geometry of the reactant, two intermediates and product. The name of the reaction step and species involved, the activation energy (Ea, kcal/mol) and difference in potential energy (ΔE, kcal/mol) of each step are shown. Bottom: schematic of the potential energy (kcal/mol) as a function of the generalized reaction coordinate, showing the relative energies of the reactant (R), intermediates V and VII and product (P), as well as the two transition states (TS).
Table 5.6. Relevant distances (in Å) and angles (in degrees) from the four species of the Cu-transfer reaction: reactant, product and intermediates V and VII. A, Atox1; W, WD4.

<table>
<thead>
<tr>
<th></th>
<th>reactant</th>
<th>intermediate V</th>
<th>intermediate VII</th>
<th>product</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1-Cu</td>
<td>2.18</td>
<td>2.29</td>
<td>2.35</td>
<td>3.97</td>
</tr>
<tr>
<td>S2-Cu</td>
<td>2.19</td>
<td>2.33</td>
<td>3.26</td>
<td>3.70</td>
</tr>
<tr>
<td>S3-Cu</td>
<td>3.54</td>
<td>2.37</td>
<td>2.32</td>
<td>2.16</td>
</tr>
<tr>
<td>S4-Cu</td>
<td>3.77</td>
<td>3.60</td>
<td>2.43</td>
<td>2.15</td>
</tr>
<tr>
<td>S1-Cu-S2</td>
<td>158.4</td>
<td>134.8</td>
<td>101.9</td>
<td>79.7</td>
</tr>
<tr>
<td>S2-Cu-S4</td>
<td>101.6</td>
<td>101.6</td>
<td>106.8</td>
<td>104.9</td>
</tr>
<tr>
<td>S4-Cu-S3</td>
<td>70.5</td>
<td>77.2</td>
<td>106.7</td>
<td>158.2</td>
</tr>
<tr>
<td>S3-Cu-S1</td>
<td>90.8</td>
<td>104.5</td>
<td>109.3</td>
<td>83.5</td>
</tr>
<tr>
<td>T11A(OG1)-Cu</td>
<td>3.61</td>
<td>3.49</td>
<td>4.34</td>
<td>4.71</td>
</tr>
<tr>
<td>T11A(OG1)-S3</td>
<td>3.94</td>
<td>3.75</td>
<td>3.83</td>
<td>3.78</td>
</tr>
<tr>
<td>T13W(OG1)-Cu</td>
<td>6.10</td>
<td>6.27</td>
<td>5.57</td>
<td>6.02</td>
</tr>
<tr>
<td>T13W(OG1)-S1</td>
<td>4.91</td>
<td>4.97</td>
<td>4.89</td>
<td>4.84</td>
</tr>
<tr>
<td>K60A(NZ)-Cu</td>
<td>3.79</td>
<td>3.75</td>
<td>3.51</td>
<td>3.17</td>
</tr>
<tr>
<td>K60A(NZ)-S2</td>
<td>3.13</td>
<td>3.09</td>
<td>3.05</td>
<td>3.06</td>
</tr>
<tr>
<td>K60A(NZ)-S4</td>
<td>2.96</td>
<td>2.96</td>
<td>3.01</td>
<td>3.08</td>
</tr>
</tbody>
</table>

Our proposed reaction mechanism was further validated by performing a final calculation in which the reaction coordinate was defined to go from the reactant all the way to the product (reaction S, Table 5.3, Figure 5.18). In this calculation, we can observe how the global reaction occurs, step by step, with sequential formation of the two intermediates V and VII, confirming all our above predictions that are based on assessing individual steps. Moreover, upon inspection of reaction S we note that the product of the reaction is less stable than the reactant.

The first step of the overall reaction (reaction A) proceeds without kinetic barrier to obtain the first intermediate V, which is the most stable specie of all (Figure 5.17). This result is in agreement with the observation of a stable Cu-dependent hetero-complex that Atox1 and WD4 form in vitro (Hussain et al., 2009), with the ability to detect a similar Atox1-MBD Cu-dependent adduct by NMR (Banci et al., 2009a) and with our previous in silico calculations of Atox1c-WD4 hetero-complex stability. After this initial
step, the system has to overcome the rate-limiting step, which is formation of the second intermediate VII (reaction P), to finally form the product (reaction R).

![Potential energy profile](image)

**Figure 5.18.** Potential energy profile (kcal/mol) along the corresponding reaction coordinate (Å) defined in Table 5.3 for reaction S. The schematic of the reactant and product is shown.

The overall reaction A-P-R (Figure 5.17) has a relatively low activation barrier, 9.5 kcal/mol, which suggests that the overall reaction is kinetically feasible. However, our results suggest that this reaction is not favored energetically, with an overall ΔE of 7.7 kcal/mol (Figure 5.17). In the absence of a favorable thermodynamic gradient, a low activation barrier may still allow for rapid Cu(I) transfer from the metallochaperone to the target domain if this process is coupled to subsequent Cu(I) translocation into the Golgi lumen and ATP hydrolysis, as previously suggested (Huffman and O'Halloran, 2000). On the other hand, it is tempting to speculate that in the absence of a driving force, *i.e.* ATP hydrolysis, formation of the Cu-bridged hetero-complex (specie V) might be favored with respect to Cu(I) transfer. However, we note that our QM-MM calculations
do not take into account entropic effects, which will likely drive separation of the hetero-
complex into individual species.

To explore the effect of Cys protonation in our proposed reaction mechanism of
Cu(I) transfer from Atox1 to WD4 we repeated some of the key reactions (A’, B’ and R’,
Table 5.3, Figures 5.19-5.21, compare reaction B' with C) with the Cys residues that
were free both at the initial and end of the calculation protonated (for this reason,
reactions O and P could not be performed). These results are in agreement with our
previous findings (Table 5.5) and also suggest that our proposed reaction pathway shown
in Figure 5.17 and our proposed reaction intermediates are independent of whether the
Cys that are not coordinating the Cu(I) atom are protonated or not. This further suggests
that deprotonation/protonation of those Cys may occur before or after Cu(I) binding to
the other Cys residues.

Figure 5.19. Potential energy profile (kcal/mol) along the corresponding reaction
coordinate (Å) defined in Table 5.3 for reaction A’. The schematic of the reactant and
product is shown.
Figure 5.20. Potential energy profile (kcal/mol) along the corresponding reaction coordinate (Å) defined in Table 5.3 for reaction B'. The schematic of the reactant and product is shown.

Figure 5.21. Potential energy profile (kcal/mol) along the corresponding reaction coordinate (Å) defined in Table 5.3 for reaction R’. The schematic of the reactant and product is shown.
Since our computed overall Cu-transfer reaction appears to be endothermic, we searched for molecular determinants that could account for the difference in energy between the reactant and product. Close to the active sites, the most important difference between Atoxl and WD4 is the presence of Lys60 in Atoxl, which is a Phe in WD4. This Lys is believed to stabilize the Cu-bound state by neutralizing the overall negative charge of the Cu(I)-Cys$_2$ center. Initially, we thought that presence of this Lys in Atoxl would account for the lower energy of the reactant hetero-complex (as compared to of the product hetero-complex). However, inspection of the active sites of the QM-MM optimized structures of the reactant and product reveals that the position of Lys60 is such that it is close to both active sites (Table 5.6), being able to stabilize both reactant and product states to a similar extent. On the other hand, although both proteins contain a Thr in position $X_1$ of the MX$_1$CX$_2$X$_3$C motif, Thr is only close to the Cu center in Atoxl (Thr11), whereas in WD4, it is facing the opposite side (Thr13) (Table 5.6). In the reactant, Atoxl’s Thr11 is close to the Cu(I) atom, stabilizing in this way the Atoxl’s Cu-bound state. Instead, this same residue is further away from Cu(I) in the product, which may account for the reduced stabilization of WD4’s Cu-bound state. The distance of Atoxl’s Thr11 to Cu(I) correlates with the relative energy in all species: reactant, intermediates X and XII, and product (Table 5.6).

5.7.4 Dynamics of interactions in reactant and product hetero-complexes

To gain insight into structural and dynamic differences between the reactant and product, we performed ~120 ns MD simulations of the product hetero-complex, and extended our MD simulations of the reactant (Figure 5.22A).
Interestingly, the $\Delta G$ of hetero-complex formation of the product is $\sim$2 kcal/mol less stable than for the reactant hetero-complex (Table 5.7), which is consistent with the observed endothermic nature of the overall Cu-transfer reaction. Although Coulombic interactions increase in the product with respect to the reactant, the change in reaction field energy increases even more, to the point that the combined contribution is favorable for the reactant but unfavorable for the product. Also, contributions from vdw
interactions and changes in nonpolar solvation energy are smaller in the product hetero-complex, which indicates less compactness of this complex. It appears that the less favorable $\Delta G$ of the product hetero-complex may arise in part from an increase in fluctuations of holo-WD4 in this complex (Figure 5.22B). It is tempting to speculate that the increase in flexibility of holo-WD4 (as discussed in Chapter 4) is a way to assure that the complex will dissociate once Cu(I) is transferred to WD4, contributing in this way to promote directional Cu(I) transfer into the Golgi lumen in vivo.

Table 5.7. Free energy ($\Delta G$) of reactant (Atox1c-WD4) and product (Atox1-WD4c) hetero-complex formation (calculated with sietraj as explained in Chapter 2), including its different energetic contributions: intermolecular van der Waals interactions (vdW), change in nonpolar solvation energy (cavity), intermolecular Coulomb interactions (electrostatic) and change in reaction field energy (field). All energy values are in kcal/mol; errors correspond to the standard deviation. Reported energy values correspond to the average of 500 calculations based on the last 50 ns of the ~120 ns MD simulations of each hetero-complex.

<table>
<thead>
<tr>
<th></th>
<th>$\Delta G$</th>
<th>vdW</th>
<th>cavity</th>
<th>electrostatic</th>
<th>field</th>
</tr>
</thead>
<tbody>
<tr>
<td>reactant</td>
<td>-9.2 ± 0.7</td>
<td>-50 ± 5</td>
<td>-9.1 ± 0.6</td>
<td>-137 ± 21</td>
<td>136 ± 19</td>
</tr>
<tr>
<td>product</td>
<td>-7.1 ± 0.8</td>
<td>-35 ± 7</td>
<td>-7 ± 1</td>
<td>-205 ± 17</td>
<td>206 ± 16</td>
</tr>
</tbody>
</table>

5.8 Conclusions

Overall, the results presented in this Chapter provide unique insights into the molecular mechanism of Cu(I) transfer from Atox1 to a WD, and the nature and strength of interactions that participate in formation of the transient complex. The details observed here with our computational approaches are difficult to obtain with experimental techniques and also provide explanations for in vitro observations. We have shown that, from all six WDs, WD4 forms the strongest complex with holo-Atox1, as was also observed in vivo (van Dongen et al., 2004) and in vitro (Achila et al., 2006; Banci et al., 2008; Banci et al., 2009b). It appears that Cu loop and helix $\alpha$1 flexibility, as well as
electrostatic complementarity correlate with Atox1-WD hetero-complex stability. Future experiments aimed to calculate the affinity of holo-Atox1 for each of the individual WDs could be performed to corroborate our *in silico* predictions.

We investigated the role of key conserved residues in eukaryotic Cu chaperones in formation of the Atox1-WD4 hetero-complex. Our findings are in complete agreement with *in vitro* experiments on the same Atox1 mutants (Hussain et al., 2009). We conclude that Met10 is not essential for adduct formation, Thr11 is important for proper docking of the incoming target domain, and thus, efficient Cu(I) transfer, and Lys60 is essential for adduct formation. Atox1 interacts with WD4 via a "weak" interface that can be disrupted by a single mutation in Atox1, Lys60. Therefore, it appears that metallochaperone-target electrostatic interactions (mediated by Lys60) are essential for providing an optimal hetero-complex stability that can assure efficient Cu(I) transfer to the target.

Finally, we have used a hybrid QM-MM approach to investigate all possible reaction schemes for how Cu(I) is transferred from the two Cys residues in Atox1 to the corresponding two Cys in WD4. Our results suggest that 3-coordinate Cu-intermediates are more stable than 2- or 4-coordinated intermediates. Furthermore, we propose that two intermediates exist on the reaction pathway from Cu-Atox1 and apo-WD4 to apo-Atox1 and Cu-WD4 (species V and VII, Table 5.4). First, Cys1 of WD4 coordinates Cu(I) together with the two Cys from Atox1 (V). The next step, which is rate-limiting for the overall reaction, involves a concerted switch to another 3-coordinated species, in which Cys1 of Atox1 and both Cys in WD4 bind Cu(I) (VII). Our findings suggest that the overall reaction, although kinetically accessible, is endothermic. Although entropic effects may contribute to hetero-complex separation, this suggests that other factors (e.g.,
ATP hydrolysis) must contribute to assure that Cu(I) transfer is directional in vivo. The reaction mechanism suggested here may be also extended to the other WDs that receive Cu(I) from Atoxl (for example, WD2) or to the MKs of ATP7A. Future experiments to corroborate our in silico findings should focus on the determination of the rates and thermodynamics of Cu(I) transfer from Atoxl to the WDs. Although NMR structures of the Atx1-Ccc2a (Banci et al., 2006b) and Atoxl-MK1 (Banci et al., 2009a) Cu-bridged adducts support the existence of trigonal intermediates in the Cu-transfer reaction, as also suggested here with our computations, the determination of the high resolution (crystal) structure of Atoxl in complex with WD4 or with any of the WDs will further help elucidate the Cu(I) coordination in these metal-mediated adducts.
Chapter 6:

Overview, Conclusions and Perspectives

6.1 Present understanding of copper transport pathways

With the identification of the genetic basis for Menkes and Wilson diseases (Bull et al., 1993; Mercer et al., 1993; Vulpe et al., 1993), and the discovery of the first Cu chaperones, the yeast Atx1 (Dancis et al., 1994b) and the human Atox1 (Klomp et al., 1997), much effort has been placed in understanding the structure, function and regulation of Cu transport pathways in all organisms. However, many questions still remain open. First, Cu chaperones from prokaryotes and eukaryotes vary in primary sequence although they share the same topology. The role of conserved and divergent residues in the vicinity of the Cu loop on Cu chaperone structure, dynamics and interaction with their protein partners has not yet been investigated. Second, whereas mammalian Cu-ATPases contain up to six N-terminal MBDs, including the six MKs and WDs from ATP7A and ATP7B, respectively, prokaryotes only have between one and two. Although several studies have been aimed to explore the roles of these extra MBDs in mammals, it is still not clear if individual MBDs in higher organisms are equivalent or have intrinsic molecular differences that ultimately dictate their function and interaction with the metallochaperone. Moreover, the interplay between these multiple MBDs separated by linkers of different lengths has not yet been explored. Third, the detailed reaction mechanism from the metallochaperone to a target domain is poorly understood. Based on structural data, only qualitative intermediaries have been proposed, but the kinetics, energetics, nature of intermediaries and the reaction mechanism are still unknown.
In this thesis, we have performed a thorough study of the structural dynamics, Cu-binding properties, protein-protein interactions and Cu-transfer mechanisms of Cu transport proteins, and how conserved residues modulate these properties. As a complementary approach to experimental methods, we took a computational approach using a wide array of state-of-the-art computational schemes, which employ different levels of theory, such QM, MM and hybrid QM-MM methodologies, in combination with MD simulations sampling techniques. These methodologies allowed us to obtain molecular and mechanistic details, with residue-specific resolution, hardly attainable via experiments. Our computational results are in agreement with available in vivo and in vitro data, and also provide molecular explanations for experimental observations. To our knowledge, our work is pioneering in nature; to date, only few computational studies have been published on this group of Cu proteins (Arnesano et al., 2004; Dalosto, 2007; Fuchs et al., 2006; Holt and Merz, 2007; Poger et al., 2005; Sharma and Rosato, 2009).

Following, we summarize the most important findings of this thesis work.

6.2 Residues in and near the copper loop are conserved differently in prokaryotic and eukaryotic copper chaperones to fine tune their flexibility

We have shown that prokaryotic CopZ and human Atox1 Cu chaperones have distinct Cu(I) geometries and overall protein dynamics (Rodriguez-Granillo and Wittung-Stafshede, 2008). In particular, CopZ is a much floppier protein than Atox1, and point mutations significantly affect the stability and dynamics of CopZ (Rodriguez-Granillo and Wittung-Stafshede, 2009a) whereas subtler effects were found in Atox1 (Rodriguez-Granillo and Wittung-Stafshede, 2009b). This finding suggests that although sharing the
same ferredoxin-like fold, the human chaperone has evolved into a more robust protein, probably to handle the more complex environment and higher-order level of regulation present in the human system.

We have explored the role of three key residues that, although not directly involved in Cu(I) coordination, are conserved differently in CopZ and Atoxl to provide a unique environment around the Cu loop (Rodriguez-Granillo and Wittung-Stafshede, 2009a, b). Met10/11 is highly conserved among Cu transporting proteins, including Cu chaperones, Cu-ATPases and other Cu proteins, from archaea, bacteria and eukaryotes (Arnesano et al., 2002). We found that although sequentially conserved, the location of Met sidechain in the protein structure varies between CopZ and Atoxl, and its location correlates with the backbone flexibility of the protein. Thus, in the holo-forms, Met is shielded inside the core and appears to provide key hydrophobic interactions, which are essential for protein stability. This finding correlates with some bacteria having similar hydrophobic residues at this position (Arnesano et al., 2002). On the other hand, in the apo-forms, Met10 is completely buried in Atoxl whereas it is exposed in CopZ. This subtle difference appears to be one of the reasons as to why CopZ is a much more floppy and unstable protein than Atoxl (Hussain and Wittung-Stafshede, 2007; Rodriguez-Granillo and Wittung-Stafshede, 2008). This key structural difference might have evolved differently in each system to fulfill the functional requirements of each protein given the distinct cellular environments.

On the other hand, although distant in sequence, two key residues near the Cu loop appear to have similar roles in Cu chaperone structure and dynamics. Residue X1 of the conserved MX1C1X2X3C2 motif is almost always a residue capable of forming a HB
(mainly Thr, Ser, His or Asn) in Cu transporting proteins from archaea, bacteria and eukaryotes (Arnesano et al., 2002), corresponding to Ser12 and Thr11 in CopZ and Atoxl, respectively. Residue 60 (Atoxl numbering) is located in the α2-β4 loop, which is structurally close to the Cu loop. This position is mainly conserved as Lys, Tyr and Phe in Cu transporting proteins from archaea, bacteria and eukaryotes (Arnesano et al., 2002), corresponding to Tyr65 and Lys60 in CopZ and Atoxl, respectively. Of particular interest is the latter residue, because despite the variability, its sidechain is always found to be close to the Cu loop. Thr11 in Atoxl was proposed to HB with Cys1 of the target domain during Cu(I) transfer, and Lys60 was proposed to neutralize the overall negative charge of the Cys-Cu(I)-Cys center (Wernimont et al., 2000). Our results provide further insights into the role of these residues in Atoxl and CopZ structural dynamics. We found that both residues tune the flexibility of the Cu loop in both apo- and holo-forms, and they provide a pivotal stabilization force for the Cu-bound state through key electrostatic and HB interactions. The latter finding is consistent with a polar residue found at position X1 of other organisms (Arnesano et al., 2002). Importantly, prokaryotic residues in the eukaryotic chaperone do not behave like the WT, and vice versa for the prokaryotic chaperone. This suggests that residues in Cu chaperones have evolved differently to maintain a delicate balance of Cu loop flexibility and stability, which may likely be important for Cu(I) uptake and release, as found experimentally for Atoxl (Hussain et al., 2008). Similar experiments to the ones performed for the different Atoxl variants (Hussain et al., 2008) could be conducted in the future for the CopZ variants, to corroborate our in silico findings. The structural changes observed here in silico for the different Atoxl and CopZ mutants could be corroborated by the determination of the
corresponding high-resolution structures, and protein dynamics could be studied through NMR spin relaxation experiments on the same protein variants.

6.3 ATP7B metal-binding domains are unique, interact differently with Atox1, and “communicate” with each other

To date, only the solution structure of the two-domain constructs WD34 (Banci et al., 2008) and WD56 (Achila et al., 2006) from Wilson disease protein have been published. Although all the domains appear to have the same fold, only a subset of them interact with Atox1. This suggests that subtle differences in the structure and dynamics of each domain may be important for their unique functions. We have explored the conformational dynamics of the six WDs (Rodriguez-Granillo et al., 2009e), including WD1 and WD2, whose structures were not previously published. Our in silico findings on WD1 and WD2 could be experimentally validated in the future by the determination of the solution or crystal structures of the corresponding domains. We have found unique molecular differences among the six WDs, including overall-fold and Cu loop dynamics, Cys-Cys distance and positioning, helix α1 conformation and dynamics, and surface electrostatic potential, which have not previously been observed experimentally. We conclude that the six WDs from ATP7B are not equivalent at the molecular level, and propose that these differences are important to guide Atox1 interactions and inter-domain Cu(I) transfer. To corroborate our computational findings, protein dynamics of individual WDs could be studied by NMR experiments. Alternatively, subtle differences in the structural dynamics as well as interacting networks could potentially yield to differences in protein stabilities, as found for example for WD2 and WD4 (Hussain and Wittung-
We found that WD4 stands out from all the rest, because of its unique features in both apo- and holo-forms. In the apo-form, WD4 is highly flexible, has a unique Cu loop conformation and a highly negative surface electrostatic potential, which appear to be good complement for the rigid and highly positively charged Atoxl. In agreement, we have shown that WD4 forms the strongest complex with holo-Atoxl (Rodriguez-Granillo et al., 2009c), as was also observed in vivo (van Dongen et al., 2004) and in vitro (Achila et al., 2006; Banci et al., 2008). Therefore, WD4 appears as the WD best optimized for Atoxl interactions. To validate our in silico estimations on relative stabilities among the different Atrox1-WD hetero-complexes, future efforts should be aimed to calculate the experimental absolute affinity of holo-Atoxl for each of the individual WDs.

At the time we started the Atoxlc-WD hetero-complexes analysis, the structure of Atoxl in complex with any of the six WDs or MKs was not available. Therefore, the orientation of Atoxl and the target domains in the six Atoxlc-WD hetero-complexes was obtained using the holo-Atoxl homodimer crystal structure (Wernimont et al., 2000) as reference. Our models (Rodriguez-Granillo et al., 2009c) have been now validated, since the solution structure of a Cu-bridged Atoxl-MK1 adduct was recently determined (Banci et al., 2009a), showing the same orientation of the metallochaperone and target domain and similar interactions between them.

Furthermore, in the holo-form, WD4 was also distinctive with respect to the other holo-domains studied, because it remains significantly flexible despite the bound metal, and its Cu site protrudes from the surface of the protein. This unique features of this holo-domain may be important for inter-domain Cu(I) transfer, as this holo-domain has been
proposed to be able to transfer Cu(I) to other WDs (Achila et al., 2006; Bunce et al., 2006). In particular, holo-WD4 was found to preferentially deliver Cu(I) to WD6 in the two-domain construct WD56 (Achila et al., 2006). In agreement, we found that holo-WD4C interacts ~2 kcal/mol more strongly with WD6 than with WD5.

Interestingly, metallochaperone-WD hetero-complexes were found to be more stable than WD-WD hetero-complexes. This suggests that these proteins are designed to allow efficient interactions between the metallochaperone and its partner domains, and weak and transient inter-domain interactions that would otherwise block the access of the chaperone. In this model, the incoming Atox1 molecule would always displace possible WD-WD inter-domain interactions, facilitating irreversible Cu(I) transport across the membrane.

Simulation of the two-domain constructs WD12, WD34 and WD56 provided unique insights into domain-domain interactions that may be important for regulation of the ATPase activity. Based on our findings, we propose that intramolecular Cu(I) transfer between WD1-WD2, WD3-WD4 and WD5-WD6 is not likely to occur, as the Cu sites always remain far from each other and facing opposite protein sides. On the other hand, the WDs strongly interact between each other, regardless of the length of the linker that separates them (Rodriguez-Granillo et al., 2009b). To note, these interactions will not likely hinder interactions with Atox1, because they do not involve the Cu loops. This suggests that the WDs in the N-terminus of the Wilson domain protein are not independent beads on a string, but rather an intricate “communication” holds them together, as also stressed recently (Barry et al., 2009). Structural and dynamic changes upon Cu(I) binding to one domain propagated through the inter-domain linker and
induced changes in the consecutive domain. Our in silico findings agree with secondary and tertiary structure changes occurring in the N-terminus domain of ATP7B upon Cu(I) binding detected by CD and limited proteolysis experiments (Bartee et al., 2009; DiDonato et al., 2000; Tsay et al., 2004; Walker et al., 2004). These conformational changes may alter the interplay between the N-terminal domain and other domains, such as the ATPBD (Tsivkovskii et al., 2001), and induce Cu(I) translocation across the membrane or trafficking of the ATPase to vesicles (Guo et al., 2005; Schaefer et al., 1999).

Moreover, in agreement with our simulations on individual WDs, the individual WDs in the two-domain constructs as well as the two-domain constructs as a whole were not equivalent. It is tempting to speculate that because the linkers that separate WD2 from WD3 and WD4 from WD5 are the longest within the N-terminal domain, WD12, WD34 and WD56 behave as relatively independent units. In support of this idea, we found significant interactions between individual domains in the two-domain constructs. These different structural units may also result in different functional units in vivo.

The structural changes observed here upon Cu(I) binding to WD34 and WD56 two-domain constructs could be further corroborated by the structure determination of the holo-forms and comparison to the corresponding apo-forms (Achila et al., 2006; Banci et al., 2008). Our findings regarding the orientation and interactions between WD1 and WD2 in the WD12 construct could be validated in the future by the determination of the high-resolution structure of this construct in both apo- and holo-forms. Also, future efforts should focus on the determination of the structure of constructs containing more than two domains, or even the entire N-terminal soluble domain of ATP7B or ATP7A,
since the quaternary structure of the full-length six-domain segment and the effect of Cu(I) coordination will likely shed light on protein-protein interactions that might be key for the in vivo function of the ATPases.

6.4 The interaction interface between Atoxl and a target domain is “weak”

We found that holo-Atoxl and the six WDs do not interact with a strong interface; only few contacts hold the two proteins together in the hetero-complex (Rodriguez-Granillo et al., 2009c). This was also recently suggested based on the solution structure of an Atoxl-MK1 adduct (Banci et al., 2009a), which validates our findings based on a structural model. In particular, we were able to dissect the specific roles of Atoxl’s Met10, Thr11 and Lys60 in hetero-complex formation with WD4, which provide molecular explanations and are in agreement with experimental findings on the same variants (Hussain et al., 2009). We found that Met10 is not essential for adduct formation, since mutation of Met10 to an Ala only slightly decreases the hetero-complex affinity. This suggests that Met10 is more likely conserved in Cu chaperones for the integrity of the ferredoxin-like fold, as discussed above.

On the other hand, Thr11 appears to be important for proper docking of the incoming target domain, and thus, for efficient Cu(I) transfer, by providing key HBs to residues X1 and Cys1 of the conserved motif of the acceptor domain. However, replacement of Thr11 by the prokaryotic residue, Ser, although it preserves a correct orientation of the proteins in the hetero-complex, it is still detrimental for adduct stability. Finally, we found Lys60 to be essential for adduct stability and therefore also for Cu(I) transfer, by providing key electrostatic and HB interactions to both Cu loops in the
hetero-complex. In particular, Lys60 forms HBs with residues $X_1$ and $X_3$ and interacts with Cys1 and Cys2 of the conserved motif of the acceptor domain, and this in turn appears to aid in the correct position of the acceptor’s Cys residues. As in the case of Thr11, replacement of Lys60 by the prokaryotic residue, Tyr, did not result in a behavior like WT. This suggests that prokaryotic and eukaryotic chaperones have evolved differently to optimize the interaction with their own partner domains. Altogether, we conclude that Atoxl interacts with WD4, and possibly with other target MBDs, via a “weak” interface, in which residues Thr11 and Lys60 play key roles. This “weak” interface can be disrupted with at least one substitution at position 60 in Atoxl, which points to the subtle balance of forces that contribute to adduct stability. This “weak” interface may be a strategy of these proteins to interact only in the presence of Cu(I), to assure vectorial Cu(I) transfer across the membrane. Because we have shown that Atoxl interacts with different affinities with the different WDs, it would be interesting to test whether the effects in adduct stability and Cu(I) transfer between the different Atoxl variants and WD4 are similar for the other WDs.

6.5 Copper transfer from Atoxl to a target domain appears to be energetically unfavorable but kinetically feasible

Finally, we have used a hybrid QM-MM approach to investigate the reaction mechanism of Cu(I) transfer from holo-Atoxl to WD4. To our knowledge, this is the first study that addressed the Cu(I) exchange mechanism in this group of proteins. Our results suggest that Cu(I) transfer occurs in three steps, with the existence of two stable 3-coordinate Cu-intermediates, in which Cys1 from both the metallochaperone and the
target domain are always coordinated to the Cu(I) atom (Rodriguez-Granillo et al., 2009a). The trigonal intermediates proposed here are supported by NMR experiments of Atx1-Ccc2a (Banci et al., 2006b) and Atoxl-MK1 (Banci et al., 2009a) interactions. Furthermore, the study of the Cu-transfer reaction by time-resolved X-ray crystallography, for example, may help to elucidate the transfer mechanism by isolation of the different intermediates and corroborate our in silico predictions.

We were able to obtain both the energetics and kinetic barriers of each individual step. Our findings suggest that the overall reaction appears kinetically feasible (E_a of rate limiting step \( \sim 9 \) kcal/mol), despite being endothermic (\( \Delta E \sim 8 \) kcal/mol). Although entropic effects may likely drive the separation of the hetero-complex into individual species, this suggests the absence of a thermodynamic gradient for Cu(I) transfer, as previously suggested (Huffman and O'Halloran, 2000). We propose ATP hydrolysis as an important driving force to assure that Cu(I) transfer is directional in vivo. Importantly, the reaction mechanism suggested here for Atoxl and WD4 may also apply to the other WDs and MKs that receive Cu(I) from Atoxl. Future experiments to corroborate our computational predictions should focus on the determination of the rates and thermodynamics of Cu(I) transfer from Atoxl to the WDs.

6.6 Future perspectives

With the increasing reports of Cu involvement in human pathologies, such as Wilson (Gitlin, 2003) and Menkes (Kaler, 1998) diseases, as well as neurodegenerative disorders (Gaggelli et al., 2006) and cancer (Brewer, 2001; Campbell et al., 1981; Tapiero et al., 2003; Yoshida et al., 1995), it is critical to obtain molecular and
mechanistic details of the proteins that participate in cellular Cu homeostasis. These Cu transport pathways are complex, with the participation of several players, and different levels of regulation. In this thesis, we have taken an unconventional yet essential computational approach to broaden our understanding on the structure-function relationships of Cu chaperones, MBDs, and their interactions. We strongly believe that we have set the basis for the study of these proteins from a computational point of view, obtaining structural, dynamic, thermodynamic and mechanistic details that are difficult to obtain via experiments, and can be used complementary to experimental results to provide functional insights.

With the constant advance in the accuracy and speed of computational methodologies, it is easy to predict that computational methods will increasingly become a fundamental tool for the biophysicist or biochemist to access chemical mechanisms and microscopic detailed information that cannot be obtained by any other method. The platform we have established, in conjunction with more efficient computers, will serve to delineate more complex systems in the future. In essence, future studies should focus on the computational modeling of the entire N-terminal domain of ATP7B or ATP7A and the simultaneous interaction of Atox1 with multiple MBDs and multiple Cu-transfer events.

Future experimental studies could focus on testing the molecular and mechanistic insights about these proteins obtained here via in silico simulations. For example, in vitro studies should focus on the biophysical properties (i.e., protein stability and dynamics) of individual WDs and combined WDs (two- and multiple-domain constructs). Also, efforts should concentrate on the determination of the absolute affinity of Atox1 for each
individual WD, as well as how is this affinity affected in multiple-domain constructs. Our proposed mechanism of Cu(I) transfer from Atox1 to a WD should be validated by *in vitro* mechanistic studies aimed to identify the sequential intermediates with the corresponding Cu(I) coordination geometry, as well as the kinetics and thermodynamics of the transfer reaction. Finally, the determination of the high-resolution structure of the full-length N-terminal soluble domain of ATP7A or ATP7B, or even of the entire protein, will significantly broaden our understanding of the function and regulation of these ATPases.
References


