BIRZEIT UNIVERSITY

FACULTY OF GRADUATE STUDIES

Ferric Uptake Regulation Protein (Fur): A Study Using Molecular Dynamics (MD) and Molecular Modeling

By

Salih J. Al-Jabour

Submitted in Partial Fulfillment of the Requirements for the Master Degree in Scientific Computing From the Graduate Faculty at Birzeit University

Supervisor

Dr. Mazen Y. Hamed
Chemistry Department

Birzeit, Palestine
Spring, 2005

Ferric Uptake Regulation Protein (Fur): A study using molecular dynamics (MD) and molecular modeling

By

Salih J. Al-jabour

This thesis was defined on 16/6/2005 and approved by:

Committee Members	Signature
1- Dr. Mazen Y. Hamed	_____
2- Dr. Hani Awad	_____
3- Dr. Mohammad Abu Alhaj	_____
4- Dr. Wael Qara'en	_____

ABSTRACT

Iron is an essential element in many living cells. It must be regulated in the living system because its accumulation is toxic. In bacteria the regulation of gene expression is performed by Ferric Uptake Regulation repressor protein (Fur). Fur has been proposed to bind iron as a co-repressor and to it act as a negative regulator of genes. It binds to DNA at specific sequence in Escherichia Coli (A.T rich region called the iron box **5'GATAATGATAATCATTATC'3)**. In this work the structure and conformational changes of Fur E.coli were studied using computational methods, to uncover its structure-function relationship.

The comparative protein modeling was used to model the structure of Fur. Fur consists of three domains: N-terminal, central and C-terminal domains. The N-terminal contains the helix turn helix motif which binds the DNA. The central domain is responsible for dimerization of Fur. Thed C-terminal contains the metal ion binding enclaves. Fur structure carries some resemblance with DtxR especially their DNA binding domains. Visualization of the Fur-DNA complex showed that the N-terminal domain of Fur interacts directly with the major groove of the iron box (which was built using consense 19bp plaindromic DNA sequence). Extensive computations using molecular dynamics proved that metal-binding and DNA-binding induces conformational changes in the Fur dimer. The N-terminal domain of Fur binds directly to the major groove of the iron box. The calculations of the distances between the two monomer subunits of Fur showed that the domain consisting

of residues 40-65 near the N-terminal is responsible for dimerization of Fur. Iron (II) binding sites of fur are discussed. Two major sites were seen on the C-terminal. A site 1 involves Cys92, Cys95, His71, Ile50, Asn72, Gly97 and Ala109. Site 2 involves His145, His143, Asp137, Asp141, Arg139 and Glu140 and iron II is present in distorted octahedral environment. This study shows that metal ion binding to the C-terminal induces conformational changes in the N-terminal. This enhanced the binding affinity of the Fur protein to the DNA. The Fur binding to DNA resulted in DNA tilting and a change in its conformation, Fe^{+2} was found to associate with DNA at high concentrations and mediate the Fur dimer binding to DNA.

Acknowledgments

I would like to express my sincere gratitude to my supervisor Dr. Mazen Hamed for all the good he has bestowed upon me. I thank him for encouraging me to practice good science and to explore new research areas. Thanks for his patience and his continuous help. Without his encouraging attitude and motivation words, things might have been impossible to peruse. Many thanks for offering me the opportunity to work in the field of my interest. My thanks are also dew to him for immense and fruitful knowledge I gained through the courses he taught me.

I would also like to express my gratitude to: Dr. Zaki Abdel Majeed for his continuous support and encouragement, Dr. Hijazi Abu Ali for his warm feelings, encouragement and invaluable help. In addition I would like to thank all the chemistry department staff at Birzeit University.

Great thanks are dew to my family for their continuous support and endeavour to help me leading fabulous life. Special thanks to my great father and mother in this world. Thanks father, thanks mother. Thank God that I have been born to this family.

The researcher

Contents

List of Figures

List of tables

Chapter 1

Introduction

1.1 Proteins

Studying the protein folding problem and protein dynamics is necessary in order to understand the protein structure, function and dynamics. There are two major ways to present the protein: top-down, or bottom-up. The top-down can be achieved by looking through the microscope and gradually increasing the magnification factor. The second option is to start at the individual atoms, and diminish the details as we decrease the power magnification.

In the following section the bottom-up approach is used to clarify the description of the protein at decreasing levels of detail.

1.1.1 A physicist view of the protein

From the physics point of view, the protein is described as a set of soft spheres with different colors (white, red, blue, or yellow) with varying radii, and masses as indication of the different atoms (table 1.1). Each sphere is connected to four other spheres by short rigid rods [1], representing the chemical bonds as can be seen in fig. 1.1.

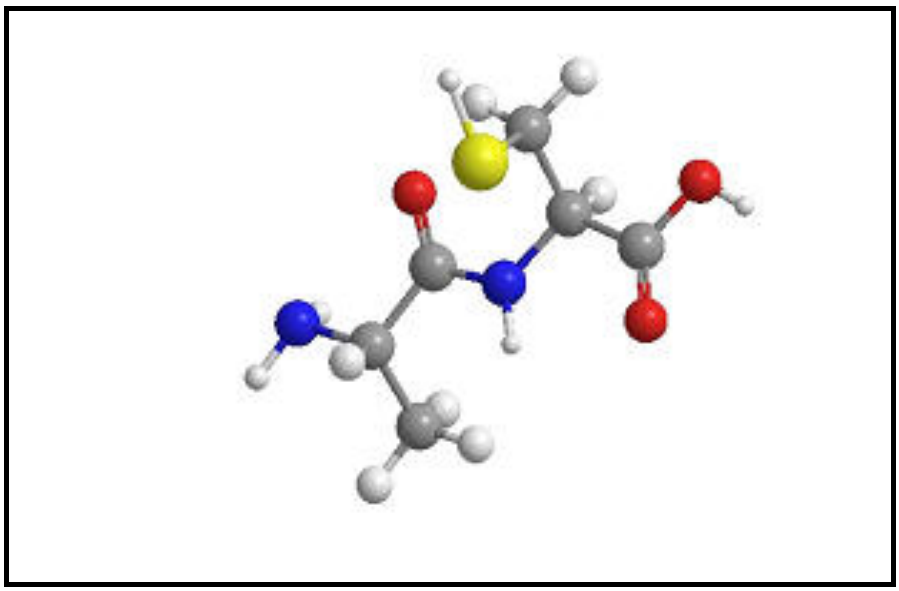


Figure 1. 1: A short stretch of amino acids connected through peptide bonds. Drawn using Chem Draw[†] package.

By defining the colors of spheres and the number of the short rigid rods (topology), a unique description of the protein can be produced. Since the protein is a subset of all possible molecules, the above description applies to all molecules.

Table 1. 1: Simplified overview of the atoms occurring in proteins with their color.

Atom	Mass (amu)	Radius (nm)	Color
Carbon	12	0.2	Black[‡]
Hydrogen	1	0.1	White
Oxygen	16	0.23	Red
Nitrogen	14	0.22	Blue
Sulphur	32	0.3	Yellow

[†] Cambridge software.
[‡] In some software it will be gray.

1.1.2 A chemist view of the protein

A protein can be compared to a string of beads. There are twenty one types of beads, each with different shape and size. The shapes of beads may vary, each bead is connected to one or two other beads. It represents a residue which consists of two parts. The first part is the backbone, which contains covalent bonds between residues. The backbone consists of the peptidic amide units and the alpha carbons. The second part is the side chains, which contains the remaining atoms in the molecule (i.e. the "R" groups of each amino acid) (figure 1.2) [2].

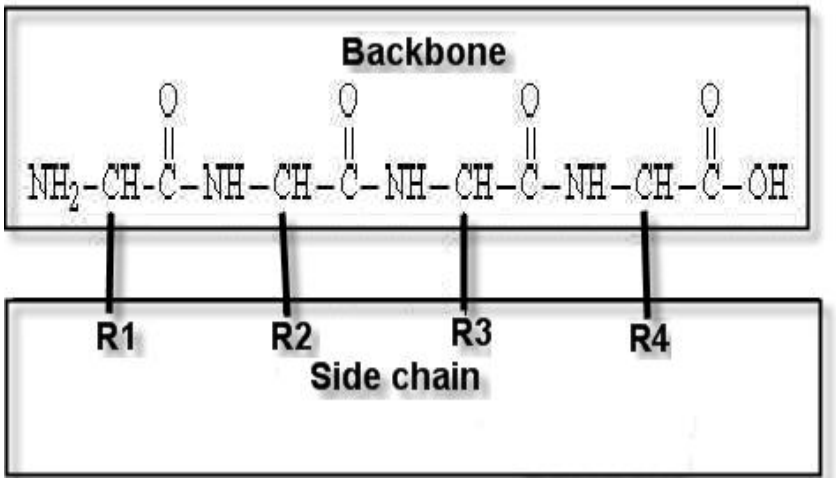


Figure 1. 2: The backbone and R1, R2, R3 and R4 are the side chains for a sequence ofAmino acids.

Residues are classified into hydrophilic, hydrophobic and polar according to their side chains. Amino acids are connected by peptide bonds which can be described conveniently by a residual sequence i.e.: Alanine-Tyrosine-Valine-Serine (full name presentation), or by three letter representation: Ala- Tyr-Val-Ser, or using one letter representation: ATVS (figure 1.3).

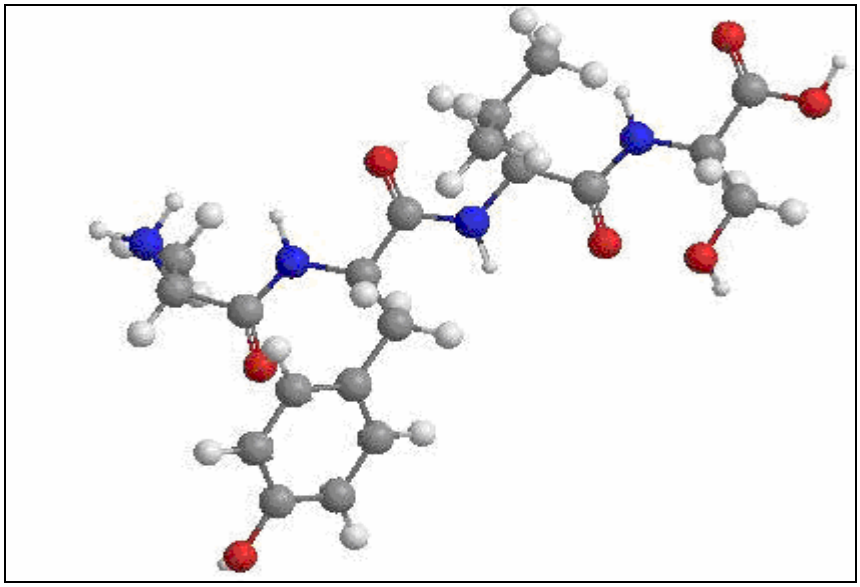


Figure 1. 3: The peptide bonds (in blue color) between the sequences (Ala-Tyr-Val- Ser).

1.1.3 A biologist view of the protein

Biologists think of a protein as serving a certain function in biological systems. Proteins have an enormous range of complex functions. Some proteins have a structural function like Ferric Uptake Regulation Protein (Fur). Others are located in the cell to perform a specific job such as to facilitate transportation of small particles such as ions. Other proteins can serve as enzymes in catalytic reactions, others have a regulatory function.

DNA binding proteins have a vital function in the living system. Biologists work has uncovered many of the relations between the proteins and other parts of the cell, and between the proteins themselves. Attention is paid to understanding certain processes that are regulated by protein molecules. For biologist explaining the protein function does not always require knowledge of its three-dimensional structure (3D). Instead, they

pay more attention to other properties such as function and binding properties in order to explain the role of proteins [3].

1.2 Protein structure

Most known protein structures are determined by either X-ray crystallography, or Nuclear Magnetic Resonance (NMR) spectroscopy. These structures are collected in data banks that contain well defined protein structures.

The BrookHaven data bank and Cambridge structural data bank provide most of the well determined structures by the X-ray crystallography and/or NMR spectroscopy. These structures were obtained using a crystal in the case of X-ray diffraction, or solvent in the case of the NMR spectroscopy. Ligand or inhibitor binding to the protein helpes in protein structure determination (figure 1.4).

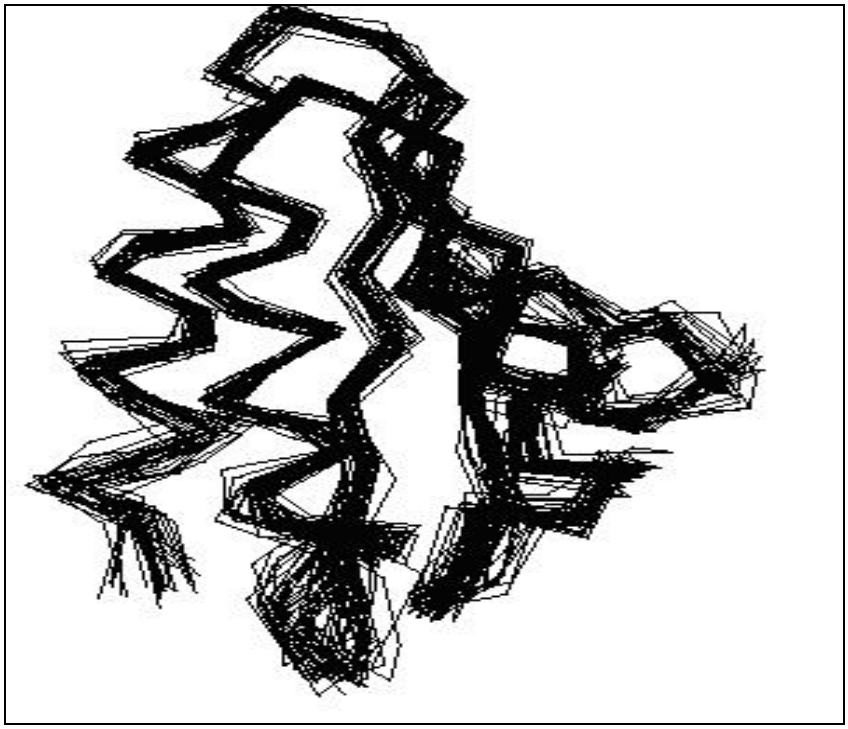


Figure 1. 4: Ensemble of 30 structures determined by NMR spectroscopy (PDB entry 1HDN [4]).

1.2.1 3D structure of protein

The complex three-dimensional (3D) structure of proteins can be described by six levels of structural organization:

- Primary structure: The linear amino acid sequence of the polypeptide chain including post-translational modifications and disulfide bonds.
- Secondary structure: Defined by the phi (φ) and psi (ψ) angles of the backbone atoms of the amino acid residues, and the hydrogen bonds between main chain atoms (figure 1.5).

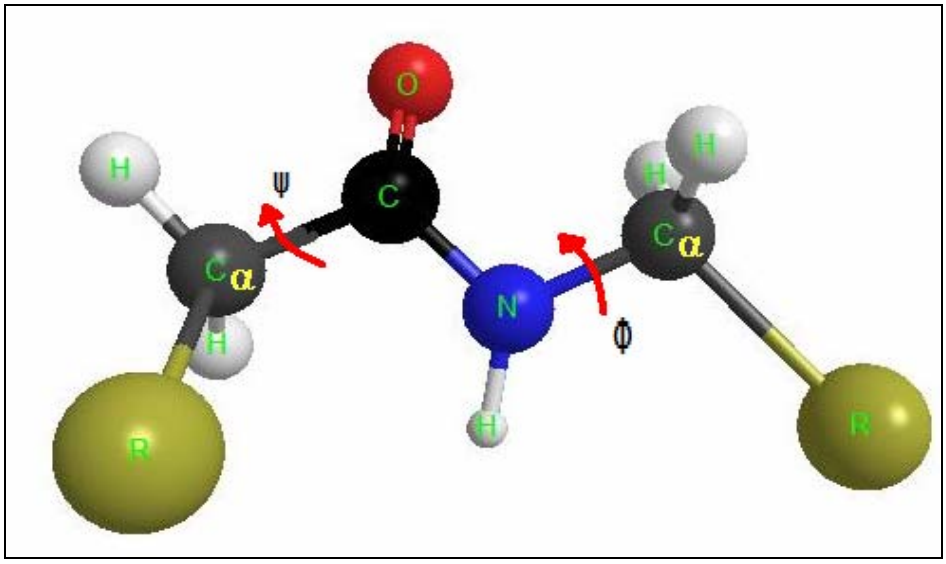


Figure 1. 5: The phi (Φ) and psi (ψ) angles of the backbone atoms.

- Super-secondary structure: include the alpha-alpha unit (two antiparallel alpha-helices joined by a 'hairpin' bend changing the chain direction by 180°); the beta-beta unit (two antiparallel strands connected by a hairpin); and the beta-alpha-beta unit (two parallel strands, separated by an alpha-helix antiparallel to them, with 2 hairpins separating the three secondary structures) (figure 1.6).

- Domain structure (folds): combinations of the super-secondary structural motifs described above are observed in proteins. For example, there are a considerable number of proteins with a four-helix bundle, consisting of two alpha-alpha units connected by a loop (figure 1.6).

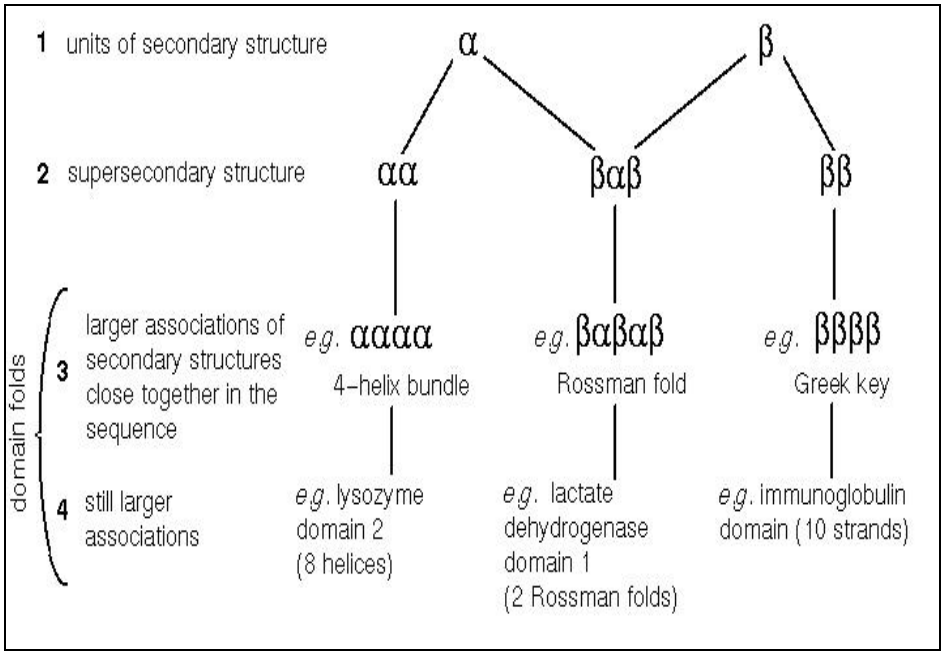


Figure 1. 6: Levels 2, 3 and 4 come under the umbrella of 'tertiary structure', but tertiary structure can also describe how domains pack together. Possibly level 3 can be considered constitute supersecondary structure as well as 2. Not all domain folds consist of motifs of supersecondary structure.

A common motif is the beta-alpha-beta-alpha-beta unit- alias the Rossman fold (effectively two consecutive beta-alpha-beta units sharing a strand). Arguably such units can be thought of as more complex super-secondary structural motifs (figure 1.7).

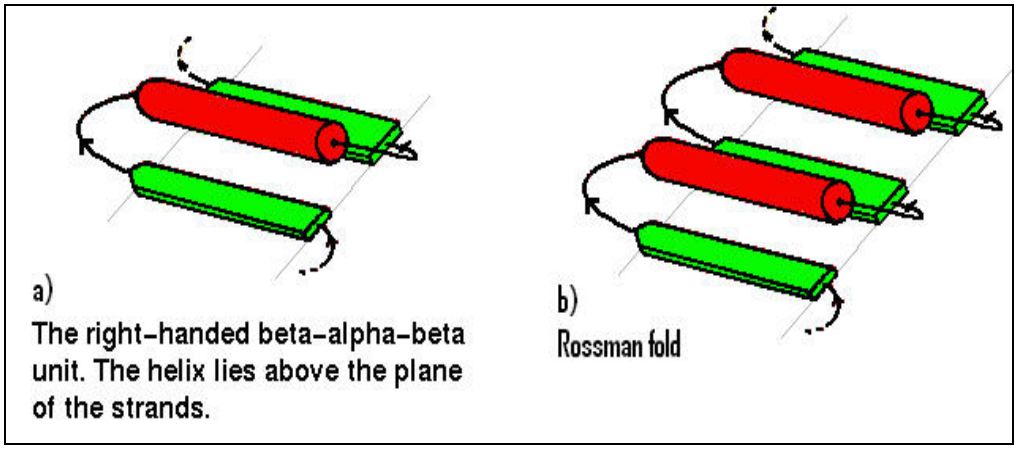


Figure 1. 7: a) The right handed Beta-Alpha-Beta unit. b) Rossman folding.

- Tertiary structure: Explains how the secondary structure units associate within a single polypeptide chain to give a three-dimensional structure.

- Quaternary structure: The quaternary structure is that level of form in which units of tertiary structure (separate polypeptide chains) aggregate to form homo- or hetero-multimers. This is found to be remarkably common, especially in the case of enzymes (figure 1.8).

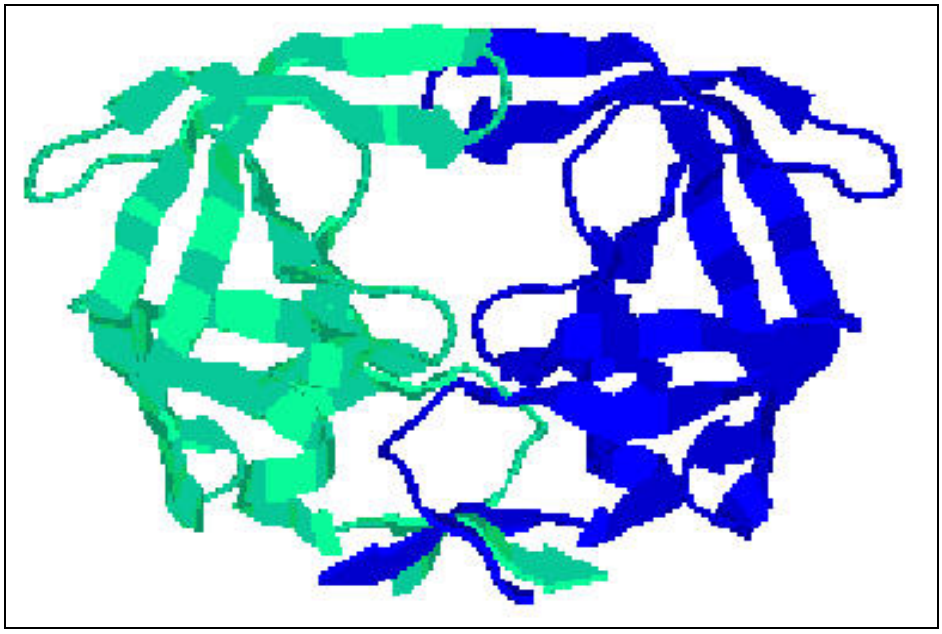


Figure 1. 8: Proteins are composed of identical subunits (chains). A simple example is the dimer HIV protease.

The first three-dimensional structure of the protein myoglobin (at 6Å resolution) was reported by John C. Kendrew and co-workers in 1957 [5]. During the 1930's, W. T. Astbury at the University of Leeds showed that a human hair gave a characteristic X-ray diffraction pattern, which changed dramatically upon stretching the hair. He gave the term α (alpha) for one and β (Beta) for the other form. Analysis of the diffraction pattern of the stretched hair led him to propose a model for the stretched or beta form corresponding to a nearly fully extended polypeptide chain with hydrogen bonds between the adjacent antiparallel chains. It followed that the unstretched or alpha form must adopt some kind of a "folded" conformation [6].

Later it was demonstrated that nearly all naturally occurring protein fibers could be grouped into one of three classes according to their X-ray diffraction pattern:

1. The alpha-type (α) contains (among others) the proteins of unstretched hair, fingernail, horn and bacterial flagella.
2. Beta-type (β) characterized by stretched hair and silk fibroin.
3. Gamma (γ) containing the protein collagen.

Meanwhile, researchers at the California Institute of Technology set out on another, albeit indirect path, recognizing the limitations of the X-ray technique as applied to fibers. They turned their attention to crystals of amino acids and simple polypeptides with the hope of learning enough about the covalent geometry of the polypeptide chain in order to permit a guess on how the folded polypeptide chain might look [7]. By the 1950s, they had produced the structures of a few amino acids, simple polypeptides and related molecules at atomic resolution. From these data, the general characteristics of the polypeptide chain were revealed. Bond distances and bond angles were measured with an accuracy of 0.02Å. From these data it was determined that the atoms comprising the peptide bond were in a trans-configuration [8].

The geometry of the polypeptide hydrogen bonds involving the peptide backbone atoms were also analyzed and shown to be fairly independent of side chain influence. This geometry analyzing was based on the following assumptions: The geometry of the polypeptide backbone is the same as that found in the X-ray crystal structure of amino acids and related compounds, all residues are present in equivalent positions and each residue participates in at least one hydrogen bond [9].

Linus Pauling and Robert B. Corey (Pauling et al., 1951) proposed the structures of two helical conformations called alpha and gamma [10]. Similar reasoning has led to the proposal of the double helical DNA structure by Watson and Crick [25,26]. Closely following these developments, the Danish researcher K. Linderstrøm-Lang proposed that there should be, at least, four levels of structural organization present in the protein structure [11].

In Linderstrøm-Lang's hierarchy of protein structure, each level was characterized by a particular type of organizing force and the higher levels of organization were composed of elements described by the previous level. It was proved later on that this type of organization was an oversimplification, but the organization of structure into levels is still useful from a pedagogical viewpoint. Linderstrøm-Lang proposal of structural hierarchy was based on the following:

A primary structure: in which the chemical structure of the polypeptide chain or chains in a given protein i.e. the number and sequence of amino acid residue is linked together by peptide bonds.

A secondary structure: the folding is brought about by linking the carbonyl and amide groups of the backbone together by means of hydrogen bonds (figure 1.9).

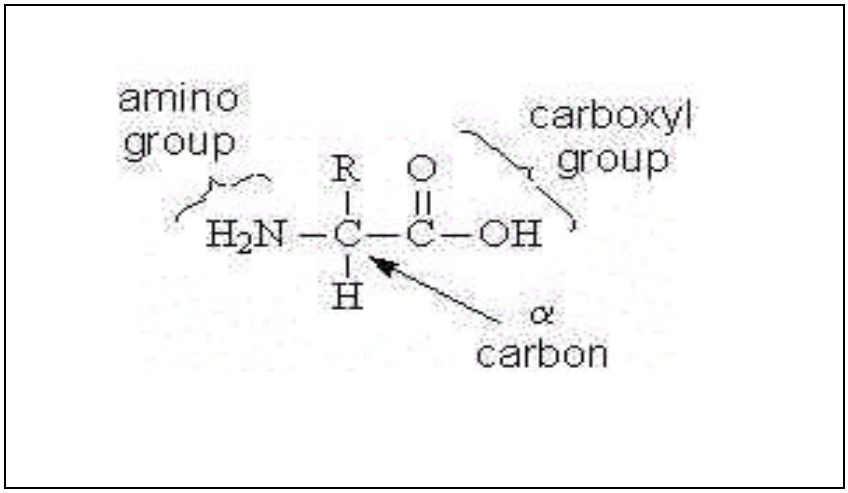


Figure 1. 9: Amino group and carboxyl group.

A tertiary structure: which is an organization of secondary structures linked by "looser segments" of the polypeptide chain, and this is stabilized (primarily) by side chain interactions. Disulfide bonds are included in this level.

A quaternary structure: The aggregation of separate polypeptide chains into the functional protein.

1.2.2 Peptide's structure

The definition of protein does not vary from the definition of the peptide structure. Both of them consist of amino acid chains linked by peptide bonds that are present between an amino group and carboxyl group with capped or charged end group (figure 1.9). In addition they have the ability (probability) to contain the disulfide group. The protein can be described as a peptide if it consists of a shortest amino acid (10 or 20 residues). If the protein contains more residues it will be called a polypeptide chains (range up to 50 or 60 residues).

1.3 Aspects of protein interaction

This work studies protein folding and conformation, taking into consideration all views of interaction that are present in the protein. Protein interaction can be classified into: interactions between atoms, between atoms and residues, and between the protein and its environment.

Most of these interactions are electrostatic in nature, i.e. interaction between charged sites (Arg, Lys, His, Glu, Asp, side chain, N- and C-termini, ions), dipole-dipole interaction (NH, NH2, C=O, OH group, water), and quadropole-quadropole interaction (Tyr, Phe, Trp, side chains)[12].

Burly and Pestsko have given an excellent overview of electrostatic interactions in aromatic peptides. They propose a classification of electrostatics interaction based on the physical nature of the groups involved, rather than the customary division in salt bridge, hydrogen bonds and the van der waal interaction [13].

According to their study, the net electrostatic interaction VC can be quantified using Coulombs law for all the atoms involved [14]:

$$Vc = \sum_{i<j} \frac{q_i * q_j}{4\pi\varepsilon_0 r_{ij}} \qquad (1.1)$$

where q is the partial charge on the atom and r_{ij} is the distance between the atoms i and j.

Apart from the Columbic interaction, London dispersion and the electron repulsion should be considered. These are electronic interactions.

The sum of these two terms is referred to as the Lennard-Jones interaction VLJ [15]:

$$V_{LJ} = \sum \frac{A_{ij}}{r^{12}_{ij}} - \frac{B_{ij}}{r^{6}_{ij}} \qquad (1.2)$$

where A_{ij} is the repulsion constant and B_{ij} is the dispersion constant term and both must be positive. According to this the repulsion term is always repulsive, and the dispersion term always represent attractive.

The sum of the net electrostatic interaction VC and Lennard-Jones interaction VLJ gives the total energy V:

$$V = V_C + V_{LJ} \qquad (1.3)$$

Using this equation we can calculate the energy of a protein in its environment, if we know particle positions and solvent atoms position. However, this description has some problems. Firstly, the partial charges on the atoms are not static, a molecule maybe polarized due to its environment, which will give rise to a different charge distribution on the atoms. Secondly, there are interactions between nuclei and inner shell electrons which are quantum-mechanical in nature. This should also be included in calculating the repulsion between atoms, so the term r^{-12} functional form of the repulsion can not be derived from these interactions.

A more realistic functional form, containing an exponential repulsion term, is known as the Buckingham potential (VB):

$$V_B = A_{ij} e^{-C_{ij} r_{ij}} - \frac{B_{ij}}{r^6_{ij}} \qquad (1.4)$$

The current form of the function was introduced since it's computationally efficient. Third, the terms q_i, A_{ij} and B_{ij} in equation 1.1 and equation 1.2, cannot be determined in a vigorous way (the A_{ij} and B_{ij} parameters are usually fitted to reproduce an experimental observable. In principle the charges q_i can be derived with reasonable accuracy from quantum mechanical calculations). It should be also noted that more terms in the energy expression (equation 1.3) are necessary to describe covalent bonds, angles and dihedrals, which will describe the protein [16].

These terms need parameters and the complete set of the parameters is called the force fields. These fields play an important role in scientific computational softwares, such as: AMBER7 [17], CHARMM [18], GROMOS [19], OPLS [20] and NAMD [21].

Neglecting these problems, and using a force field from the scientific literature (e.g. AMBER, GROMOS…etc), makes it possible to calculate the energy of a protein in a solution. The force on each atom can be calculated from the derivative of the energy with respect to the atomic position:

$$F_i = -\frac{\partial V}{\partial t_i} \qquad (1.5)$$

This can be used numerically to integrate the equation of motion for all atoms given by Newton second law:

$$F_i = m_i a_i \tag{1.6}$$

where m_i is the atom mass and $a_i = \frac{d^2 r_i}{dt^2{}_i}$ is the acceleration.

If this is done repeatedly then the molecular dynamics (MD) can be simulated. From molecular dynamics calculations a trajectory can be generated, which describes the atom position as a function of a time [19].

1.4 Protein folding problem

A protein consisting of N amino acids, in principle, can be in about 3N conformations, since there are three possible combinations of (**α/γ**) angles [2]. It almost folds into a single conformation under physiological condition (native state).

The unfolded protein in vitro can refold into its native conformations. It is generally accepted that the "information" necessary to find this conformation is retained in the residue sequence [22]. If the protein would search all the available conformations randomly, it would take longer than the life time of the universe. This paradox is known as Levinthals paradox [6], and it defines the protein folding problem: how does the protein find its native conformation?

To solve the problem, we must first understand the three dimensional structure of a protein given its sequence, and then have a physical model which can be used to predict the folding pathway. The most prominent model is the diffusion-collision model, which assumes that the secondary structure elements are formed much faster than the final structure. The secondary structure elements move in a diffusive way until they collide with other pieces of protein, and form their final structure [23].

1.4.1 Three dimensional structure of protein

To understand the protein folding, and how a protein finds its native conformation, it is necessary to understand the three-dimensional structure of the protein. The three dimensional structure of a protein is basically determined by the localization of a rigid α and β structural segments of the protein chain **(**figure 1.10**),** establishing their connection in space (protein topology) and by the mutual orientations of the connecting segments (protein packing) [24].

Figure 1. 10: Three dimensional structure of hamster and mouse prion proteins. The beta sheet and helix are shown.

Proteins can be divided to four classes according to their predominant type of secondary structure:

- all -α proteins with predominant α structure.
- all -β proteins with the predominant β structure.
- α/β proteins with alternating α- and β- segments.
- α + β proteins with α- and β- structures segregated along the chain.

Predicting the protein structure involves building the structural class of a given protein from its primary structure, establishing which type

of topology and packing are allowed for the proteins of such class, and how the protein chooses between allowed topologies and packing [27].

The approach to solve this problem is based on the assumption that for each protein there is a definite folding path way, which ensures a rapid and correct folding, This means that protein folding passes through several stages forming sequence, i.e. one structure is followed by the other, in such a way that the outcome of each stage determines the subsequent stage of folding [28].

Only the most stable intermediate structure formed at each stage survives, which gives the necessary time for their growth or rearrangement at the subsequent stage, while the less stable intermediate is rapidly destroyed by thermal motion. The stability of the intermediate structure, not intermolecular diffusion, is the factor which determines the kinetics parameters for protein folding as diffusion time (10^{-7} sec), which is very small compared to the time of folding (~ 10^{-2} sec). There fore, the folding process must pass through most stable intermediate structure for each stage in the self-organization process, this leads to the 3D structure of native protein [29].

1.4.2 Experimental approach to protein folding problems

The protein folding can be represented as a reaction where three stages are present: unfolded structure, native protein structure and the intermediate structure of the protein. The reaction then can be written as [30]:

$$U \leftrightarrow I \leftrightarrow N \tag{1.7}$$

where U represents the unfolded protein, I the intermediate stage, and N the native structure of the protein.

From that it appears that the unfolded protein can have many conformations: If this were not the case it should be possible to detect a conformation experimentally. Many attempts to predict the folding intermediate structure and some candidates have been reported formed. These proteins with partially folded core have been named "molten globules" [31, 32].

The protein that refolds without detectable intermediates and within milliseconds has grown considerably [33]. This comes from fact that protein fold so fast implies there are no larger free energy barriers along their folding pathway, for example, Cytochrome C can be trapped in a molten globule state at low pH, but under a different solvent condition it will folds fast without detectable intermediates [34]. So the stability of the intermediate is an artifact of possibility that the same intermediate is on the "true" folding pathway. Most of the experiment in the protein folding field study protein in isolated form in a test tube "vitro", but its very complex techniques, since it will not have the same conditions as a living cells [35]. Such as in the living cell's there are a presence of the ribosome, on which the protein is synthesized, chaperone-proteins which may aid in folding and the environment of the cells for all kinds of organic compounds, which all of them are absent in the vitro environment [36].

1.4.3 Theoretical approach to protein folding

There are many theoretical approaches (models) to determine whether protein folding is possible These models have a requirements that include:

- A polypeptide chain should have a unique free energy minimum, corresponding to the native state.
- The polypeptide chain should fold in finite time, there by solving the levinthal paradox [37]. The protein folding is generally accepted to be in a cooperative process, to describe the folding dynamics properly [38].

karplus and shakhnovich [23] consider possible models for a protein folding simulation, with the residue-based models at the one end and all-atom model in explicit solvent at the other end. Dill and co-workers have built protein models using two kinds of residue on a cubic lattice and performed Monte Carlo simulation of protein folding [39, 40]. Due to the time scale (millisecond) all-atom simulations of folding are currently not feasible; therefore many scientists have tried to find simpler models. However, protein and peptide unfolding simulations were performed using all atom based simulation.

For intermediate levels there are a number of models. One of the earliest example is a simulation of bovine pancreatic trypsin inhibitor by levitt and warshell using a simplified residue [32]. In this work the side chain and backbone were both represented by a single particle. Skolnick and co-workers used a lattice model to describe the first stage of folding of

a four helix bundle, and an all-atom model to refine the end structure of the lattice simulations [41].

For all cases, the interaction between particles to be its residue or atoms, are important to be described in a proper way. If the lattice simulation dose not yield a structure that is sufficiently close to the native structure refinement of such structure is useless. If for a moment, we can devise that all atom model that fulfils requirement 1 (see above) then we can describe the interaction between two residues as a sum of pair-interactions between atoms. So the net interaction depends on the distance between the residue, the relative orientation and the internal conformation of the residue.

It should be kept in mind that a residue like lysine which is usually typed as hydrophilic has **$5CH_2$** groups in its side chain, which are definitely hydrophobic. We allow for averaging over local details of similar nature, but retain important differences. A description of all residues in similar way has been used for folding studies using the ***LINUS*** program [42].

The conclusion of this subsection is that lattice simulations are too crude and all-atom simulations are too slow to predict the folding pathway of a protein.

1.5 Molecular dynamics software used in this work

1.5.1 *Amber7*

Amber is a suite program that allows users to carry out molecular dynamics calculations. ***Amber*** is also an empirical force field that is implemented in ***AMBER7*** and other molecular dynamics software such as ***NAMD*** and ***GROMOS*** [43].

AMBER7 came out in 2002 and represented a significant change from the previous version, amber6, which was released in 1999. The between two versions differ as follows:

1. Several force fields are included in ***AMBER7*** for the proteins and nucleic acids, which include a version with polarizable dipoles on atoms, and off-center charges. Also, several models of water are provided such as, **TIP3P**, **TIP4P**, **TIP5P**, **SPCIE** and **POL3** [44].
2. ***AMBER7*** includes Antechamber program, which converts three-dimensional models into files appropriate for molecular mechanics calculations [44].
3. As a result of implementation of the three variants of the generalized Born (**GB**) code, it provides a better energy balance between surface-exposed and buried atoms [45].
4. More efficient **PME** simulation, with a better performance on both single-processor and parallel machines [46].

5. The updating script for **MM-PBSA**, making the input easier to create and providing more options for analysis of the result [44].
6. ***AMBER7*** includes the Sander program, which can carry out free energy minimization using a thermodynamics integration method [45].
7. New types of restraints forces can be defined, that are based on **RMS** super position to a reference structure.

Flowing into *AMBER7*: In order to use ***AMBER7***, it is important to understand where to begin in ***AMBER7*** and how it flows. It is important to first understand, what information is needed by the simulation programs, and to know where it comes from and how it gets into the form that the energy package needs.

The most common information needed in all simulation programs is:

- The Cartesian coordinates of each atom, which comes from the **X-ray** crystallography, **NMR** spectroscopy, or model-building. They are usually in **PDB** (protein data bank) formats.
- Amber7/dat/leap/prep directory provides the standard topology files format for the amino acids, **N-** and **C-**terminal charged amino acids, **RNA** and **DNA**.
- Force fields parameters for all bonds, angles, dihedral, and atom types in the system, these force fields found in the

amber7/dat/leap/parm directory. The users specify the procedural options and state parameters desired [43, 45].

1.5.2 Docking Software

The **AUTODOCK** program (version 2.4) was developed to provide an automated procedure for predicting the interaction of ligands with macromolecule targets [47]. In any docking scheme two conflicting requirement must be balanced: The desire for a robust, to keep the computational demands at reasonable level. Accurate procedure, the ideal one would find the global minimum in the interaction energy between the substrate and the target protein, exploring all available degrees of freedom (**DOF**) for the system.

The original procedure developed for ***AUTODOCK*** used a Monte Carlo (**MC**) simulated annealing (**SA**) technique for configurationally exploration with a rapid energy evaluation using grid based molecular affinity potentials. This leads to the combined advantages of exploring a large search space and a robust energy evaluation [48]. Using this procedure give a powerfull approach to the problem of docking a flexible substrate into the binding site of a state protein.

Overview of the method: The energy evaluation for each atom described by the Goodford, which depend on precalculating the atomic affinity potential for each atom. A p`rotein embedded into three dimensional grid map and placing the active site in the center using the ***AUTOGRID*** program. The interaction energy is assigned to each grid point. Then the affinity energy is calculated, as well as a grid of

electrostatic potential, either using Poisson-Boltzmann finite difference method or using point charge of +1 as a probe. Using tri-linear interpolation of affinity values of eight grid point that surround the atom to calculate the energy of this atom. Interpolating the values of the electrostatic potential and multiplying by the charge of the atom to calculate the electrostatic interaction. The docking simulation is carried out using Monte Carlo simulated annealing (Metropolis method). The substrate performs a random walk around the static protein. At each step, a small displacement is applied to each of the degrees of freedom of the substrate: rotation its center of gravity; translation; orientation around its flexible internal dihedral angeles. The energy for new displacement is evaluated using the grid interpolation procedure. This new energy is compared to the energy of the preceding step. The new configuration is accepted if the new energy is lower. If the new energy is higher, then the configuration is accepted or rejected according to the probability expression, which depend on the temperature T. The probability acceptance is given by [47, 48]:

$$P(\Delta E) = e^{\frac{\Delta E}{K_B \times T}} \qquad (1.8)$$

Where $\mathbf{K_B}$ is the boltzmann constant and ΔE is the energy difference.

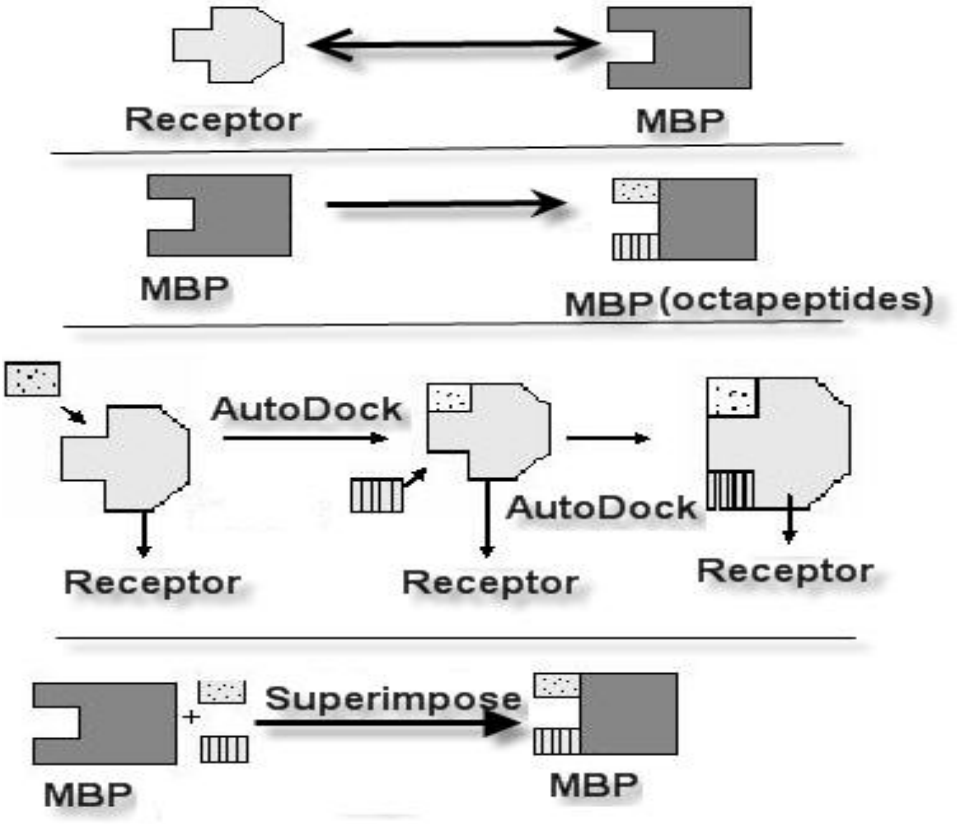
All steps are accepted at high temperature, and fewer at low. This simulation proceeds as a series of cycles, with a specified temperature for each cycle. Based upon the current temperature, each cycle will contain a large number of acceptances and rejections. The temperature will be

lowered for the next cycle after a specified number of acceptances or rejections. The temperature will be lowered by a specific schedule such as:

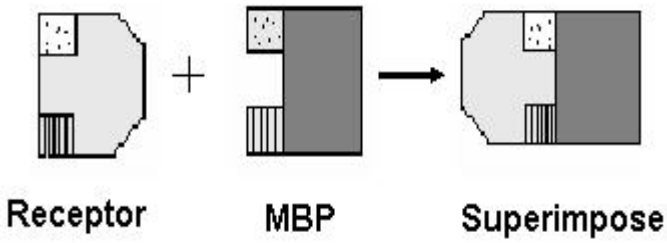
$$T_i = gT_{i-1} \tag{1.9}$$

Where T_i is the temperature at cycle i, g is a constant between 0 and 1.

AUTODOCK application: Many investigations of the ***AUTODOCK*** were done [47]. The investigations of unknown structure of maltose-binding protein (**MBP**) and the ligands domains of aspartate receptors to predict the structure of the receptor-protein complex. The investigators used knowledge from mutational studies of **MBP** to select two octapeptides on the protein known to be involved in the binding to the receptors, which they docked independently to the model of the receptor using automated docking code, where the backbones of the peptide were fixed, but the side chain conformation and overall orientation were unrestrained [47].



The distance and orientation of the two peptides as docked to the receptor corresponded to that in the intact MBP, thus enabling a reasonable prediction of the protein-receptor complex. This technique could be useful in situations where there are data on multi-site interactions.



1.6 Ferric Uptake Regulation Protein

Iron is required by most biological systems due to its diverse role in the biochemistry of the cell. Although iron is the most abundant transition metal on Earth, the solubility of iron is very low at physiological **pH** in aerobic environments and therefore, the bioavailability of iron is poor. In order for iron to be used by the bacterium, it must be host be specialized uptake mechanisms or extracted from the environment [49]. The major role of iron inside bacterial cell is its involvement in enzymatic redox reaction. On the other hand, iron plays a structural role in protein and change the reactivity of active side residues. Although the reactivity of the iron atom makes it useful in many biological applications, undesirable side reactions can occur. Through Fenton-type chemistry will, which arises from the spontaneous combination of superoxide anions created by oxidative metabolism in cells. The resultant radicals induce the formation of unsaturated bonds in lipids, decreasing membrane fluidity and causing cell lysis. They also react with thiol groups in proteins, causing cross-linking and inactivation. Hydroxyl radicals can also extract hydrogen atoms from **DNA** and **RNA**, causing mutations or cleavage of the phosphodiester backbone [50].

A collection of different mechanisms have evolved in order to deal with toxic radicals. A number of enzymes and cofactors function in prokaryotes to detoxify oxygen radicals. However, a simpler method to reduce radical formation by iron is to limit the availability of the iron atom itself; by sensing adequate iron levels, limiting uptake and sequestering excess iron in storage proteins [50].

A variety of high-affinity iron uptake systems exist in bacteria to extract iron from the environment or an animal host, by directly binding iron or iron-binding proteins. Most pathogenic bacteria can procure iron directly from the host iron binding proteins or indirectly through heme from hemoglobin (**Hb**) [51]. The ferric uptake regulation proteins regulate the iron concentration in the bacterial cells to avoid the side effect reactions. The E. coli Fur protein (17 kDa [52,53]) is one of these regulation genes works in the Eschesichia Coli and has homologs in both Gram-negative (Salmonella , Yersinia, Vibrio, Neisseria, Pseudomonas, Campylo-bacter, Legionella, Bordetella, Haemo-philus, Helicobacter pylori, Acinetobacter baumannii, and Erwinia chrysanthemi) and Gram-positive bacteria (Staphylococcus and Bacillus subtilus)[54].

Structurally, very little is known about Fur. In solution, Fur seems to be a dimer in the presence or absence of $\mathbf{Fe^{2+}}$ [55] and it has the ability to multimerize [56]. Experimental approaches like **X-ray** [57], **NMR** [58] and **CD** [59] spectra were used in order to solve the structure of the Fur. Unfortunately, these experiments give very little information about its structure. On the other hand, it was proposed that metal binding may affect the conformation of Fur since increased rates of proteolysis occur in the presence of excess metal [60]. The protein is proposed to have three domains, with the N-domain for **DNA** binding, central domain for dimerization and the C-domain for metal binding [60, 61]. However, specific amino acids involved in the function of Fur have yet to be determined even though several mutant Fur proteins have been discovered

[60, 62]. Likely possibilities for iron coordination are the 12 His and 4 Cys residues present in the amino acid sequence [60, 63 and 64]. Yet, results from a variety of spectroscopic techniques, including paramagnetic **NMR** experiments [58], spin-label studies and electronic absorption spectra [60] are inconclusive and disagree about what type of coordination occurs. Fur can also bind other divalent metal ions such as $\mathbf{Mn}^{+2}$, $\mathbf{Co}^{+2}$, $\mathbf{Cu}^{+2}$, $\mathbf{Zn}^{+2}$ and $\mathbf{Cd}^{+2}$ in vitro when binding to **DNA** [65, 66 and 68]. As such, previous work to identify the coordinating ligands of $\mathbf{Fe}^{2+}$ may have actually been examining two different metal binding sites [67]. There appears to be some conservation of residues in the putative metal binding sites (His and Cys clusters) in Fur [66] with the homologous **Zur** protein, which regulates zinc uptake in E. coli [65].

In light of the work being done, the structure of protein will be established using the comparative protein modeling and the conformational changes during binding process will be also investigated. On other hand, the type of ligands coordination of the Fe^{2+} with metal binding sites will also be examined in this work. Solving the **3D** structure of the Fur and the following which will be done in this work will help us to understand how the Fur regulates iron metals in the E.Coli.

Chapter 2

Methods and Calculations

2.1 Comparative protein modeling of ferric uptake regulation protein

The experimental elucidation of **3D**-structure of proteins is often hampered by difficulties. Therefore the number of solved **3D**-structures increases slowly compared to the rate of the sequencing registered in the protein data base. The **3D**-structure of proteins is of great assistance when planning an experiment aimed at the understanding of the protein function and conformation.

Comparative modeling is a way to avoid the experimental elucidation for getting the **3D**-structure of a protein; comparative modeling depends on the fact that proteins from different sources and diverse biological functions can have similar sequences. It is also accepted that if the protein sequences have high similarity, this reflects a distinct structural similarity. This comes from finding the relative mean square deviation (**RMSD**) of **Cα** coordinates for protein cores sharing 50% residue identity was expected to be around 1Å. The comparative modeling provides the molecular biologist with low-resolution models which hold essential information about the spatial arrangement of important residues and this will guide to a design of the experiments [69].

2.1.1 Comparative modeling procedure

Several steps are involved in the comparative modeling technique:

1- Identification of modeling templates

Homology modeling requires, at least, one sequence of a known 3D-structure with significant similarity to the target unknown protein. The sequence of the target protein is compared with database in the Brookhaven Protein Data Bank (**PDB**) using certain programs, such as ***SWISS-MODEL*** [70], ***BLAST*** [71] and ***FastA*** [72]. This step will produce several templates for the target sequence and the one with the highest sequence similarity will serve as a reference. Maximizing **C*α*** in the common core while minimizing their **RMSD** leads to the optimization of the superposition. Then every residue in the reference is aligned with a residue in the templates if their **RMSD** falls within a 3.0Å range. This procedure leads to structurally corrected multiple alignments.

2- Aligning the target sequence with the template sequence

The target sequence which needs to be aligned with several templates will be selected according to the corrected multiple sequence alignment. The residue located in the conserved loops will not be selected. Thus, at least one template structure is built after being completely defined by the loops and the common core of the target protein [73].

3- Building the model: In building the model the following technique is implemented:

- Framework construction and building the non-conserved loops

The construction of a framework is computed by averaging the position of each atom in the target protein. This process is based on the location of each atom in the template. Since there is more than one likely template available then the relative contribution of each structure can be determined according to the local degree of sequence identity with target sequence. Because the template does not give enough structural information, loops must be constructed. Most of the known 3D-structures may have similarities with their loop regions regardless of the unavailable share in overall similarity with templates. Most of the homology modeling programs uses the spare part algorithm. This algorithm search for fragments which could be accommodated onto the framework of the Brookhaven Protein Data Bank (**PDB**) entries with resolution better than 2.5 Å. Each loop is defined by its alpha carbon (**Cα**) atom coordinates for the four residues following and preceding the loop. The fragments will then be retained and sorted in the model if they don't overlap with neighboring [74].

- Addition of the side chains and completing the backbone

Building the non-conserved loops relies only on **Cα**. Nitrogen atoms and carbonyl must be completed in these regions. The library of the pent-peptide backbone fragment that is derived from the Brookhaven Protein Data Bank (**PDB**) will be performed in order to complete the missing fragment. This method will give resolution better than 2.0Å, and these fragments are fitted to overlap within the **Cα** atoms. The coordinates of central backbone atoms (**C**, **O** and **N**) are averaged and added to the target model. According to this process the **RMSD** of the resulting

modeled backbone will differ from the experimental coordinates by approximately 0.2 Å [74, 75].

A structure lacking the side chains will result. The number of the side chains is added according to the degree of sequence identity between the model and the template sequence. The allowed rotamers of each residue from the table of rotamers are analyzed to check if they are acceptable by the Van Der Waals exclusion test. The most favored rotamer is then added to the model. The atoms defining the α1 and α2 angles of the incomplete side chains can be used to restrict the choice of rotamers that fit these angels.

- Model refinement

Idealization of bond geometry and removal of unfavorable non-bonded contacts can be performed by energy minimization using the force fields such as ***CHARMM***, ***AMBER*** or ***GROMOS***. Refinement of the primary model should be performed by no more than 100 steps of steepest descent, followed by 200-300 steps of conjugate gradient energy minimization. The number of steps used in the energy minimization plays an important role in the model optimization. Constraining the positions of selected atoms or using a **B-factor** based functioṇ for each residue, generally, helps to avoid excessive structural drift during force field computations [74, 75].

2.1.2 Modeling of the Ferric Uptake Regulation Protein Structure

- **The web servers used**

The ***SWISS-MODEL*** server was the first web server to automate the comparative modeling process. ***SWISS-MODEL*** was initiated in 1993 by Manuel Peitsch, and is now being developed within the Swiss Institute of Bioinformatics (**SIB**) in collaboration between Torsten Schwede at the Structural Bioinformatics Group, Biozentrum (University of Basel) and Nicolas Guex at GlaxoSmithKline [70]. In addition to the ***SWISS-MODEL*** [70] server there are other servers such as ***CPH Models*** [76], ***SDSC1*** [77], ***VADAR*** [78] and ***ModWeb*** [79]. A sequence is submitted by the user to any of these servers and an all atom comparative is returned when possible. The ***SWISS-MODEL*** server and ***VADAR*** program were used to model the structure of the Ferric Uptake Regulation Protein from E.Coli (**Fur**). These servers accept the sequence in ***FastA*** format and returned the analysis and the templates that agree with the Fur structure according to the above procedure.

- Alignment of ferric uptake regulation protein sequence

The structure of the ferric uptake regulation protein is still unresolved, although some **NMR** [58] and **X-ray** [57] work has been performed but the structure function relationship is still unclear. In this work the homology modeling was used to predict the structure of the protein. The sequence of the Ferric Uptake Regulation protein (**Fur**) extracted from E.Coli in **FastA** format is:

> fur

**MTDNNTALKKAGLKVTLPRLKILEVLQEPDNHHVSAEDLYKRLIDMGEEIG
LATVYRVLNQFDDAGIVTRHNFEGGKSVFELTQQHHHDHLICLDCGKVIEFS
DDSIEARQREIAAKHGIRLTNHSLYLYGHCAEGDCREDEHAHEGK.**

This sequence was submitted to the ***SWISS-MODEL*** web-server by selecting the first approach mode and submitting the sequence. The ***SWISS-MODEL*** server searches in the data base and aligns it to the similar known proteins. The server also sends some properties of the protein, template groups and Global alignment. The predicted structure and the best alignment with known proteins were sent as Brookhaven Protein Data Bank (**PDB**) file.

2.2 Energy minimization of the ferric uptake regulation protein

Energy minimization is a basic modeling technique used to characterize molecular conformations and refinement of protein structure. The number and type of molecular conformations and their relations to minimum potential energy function (**U**) is a topic related to energy minimization. The total potential energy function is expressed in the following equation [45]:

$$U(R) = \sum_{bonds} K_r (r - r_{eq})^2 \qquad \text{bond}$$

$$+ \sum_{angels} K\theta(\theta - \theta eq)^2 \qquad \text{angel}$$

$$+ \sum_{dihedrals} \frac{Vn}{2}(1 + \cos[n\phi - \gamma]) \qquad \text{dihedral}$$

$$+ \sum_{i<j}^{atoms} \frac{A_{ij}}{R^{12}_{ij}} - \frac{B_{ij}}{R^{6}_{ij}} \qquad \text{van der Waals}$$

$$+ \sum_{i<j}^{atoms} \frac{q_i q_j}{\varepsilon R_{ij}} \qquad \text{electronics}$$

Non-additive force fields based on atom-centered dipole polarization can be also used. The term of polarization can be added to the following term:

$$E_{pol} = -\frac{1}{2}\sum_{i}^{atom} \mu_i \bullet E_i^{(0)} \qquad \text{polarization}$$

The kinetic energy at time t is obtained by considering the kinetic energy at time $t - \frac{\mathrm{dt}}{2}$ and $t + \frac{dt}{2}$:

$$KE(t) \approx \frac{[KE(t - \frac{dt}{t}) + KE(t + \frac{dt}{2})]}{2}$$

The kinetic energy at time **t** is obtained technically by using ***AMBER7***. A better way to calculate the kinetic energy is by using the half step time to find the velocity at time **t** from $t - \frac{dt}{2}$. The commonly

available minimization algorithms used in ***AMBER7*** are Steepest Descent and Conjugate Gradient [43].

Minimization algorithms are designed to head down-hill toward the nearest minimum. Remote minima (separated from the initial conformation by an energy barrier) are not detected by energy minimization because this would require a period of up-hill movement. Each starting conformation may result in the detection of one minimum energy conformation (barring pathological problems). Different starting conformations should lead to the same minimum.

Energy minimization locates a minimum energy conformation mathematically on the energy function minima. During the energy minimization process, the nearest minimum energy conformation will be found using the smallest number in calculations. The degrees of freedom of the molecule (**3n-1**) are also included in calculation.

2.2.1 Preparing the coordination file for ferric uptake regulation protein

During the simulation process, ***Linux*** (***Red hat 9***) was installed on Pentium ***III PC's***. ***AMBER7*** was used for energy minimization computation, molecular dynamics simulation and analysis. The ***AMBER7*** suite consists of 60+ programs. However, to carry out a traditional molecular dynamics simulation two programs are used: ***xleap*** (graphical interface) or ***tleap*** (text-based interface) and ***sander***. In addition to these two programs, ***carnal*** and ***ptraj*** are used. These two programs are the data analysis workhorse of the ***AMBER7*** suite [45].

The comparative modeling of ferric uptake regulation protein leads to the coordination protein file (**PDB**). To remove bad connections in the structure and make the structure in a minimum energy state, energy minimization was performed.

Before starting ***xleap***, the environment of the ***AMBER7*** must be set up. After installing ***AMBER7*** in the local directory, all parts of the ***AMBER7*** are installed correctly. The terminal command line in the Linux opened and the following lines written to setup the environment of ***AMBER7***:

>cd Path… /amber7
>csh
>setenv AMBERHOME /usr/local/amber7
>set path= ($path $AMBERHOME/exe)

At this point the suite programs of ***AMBER7*** are ready to be used. The following command line was used to start the ***xleap***:

>$AMBERHOME/exe/xleap –f leaprc.ff99

AMBER7 will load the necessary library files included in the **dat** directory. These library files contain the parameters for each atom and the residues in the protein:

```
Welcome to LEaP!
(no leaprc in search path)
Sourcing: /usr/local/amber7/dat/leap/cmd/leaprc.ff99
Log file: ./leap.log
Loading parameters: /usr/local/amber7/dat/leap/parm/parm99.dat
Loading library: /usr/local/amber7/dat/leap/lib/all_nucleic94.lib
Loading library: /usr/local/amber7/dat/leap/lib/all_amino94.lib
Loading library: /usr/local/amber7/dat/leap/lib/all_aminoct94.lib
Loading library: /usr/local/amber7/dat/leap/lib/all_aminont94.lib
Loading library: /usr/local/amber7/dat/leap/lib/ions94.lib
Loading library: /usr/local/amber7/dat/leap/lib/solvents.lib
> ^
```

After this step ***xleap*** is ready to load the protein data bank file, after assigning it to a variable called "**fo**":

```
Unknown residue:    number: 149   type: Terminal/last
..relaxing end constraints to try for a dbase match
  -no luck
  Added missing heavy atom: .R<CMET 149>.A<CA 3>
  Added missing heavy atom: .R<CMET 149>.A<CB 5>
  Added missing heavy atom: .R<CMET 149>.A<C 16>
  Added missing heavy atom: .R<CMET 149>.A<CG 8>
  Added missing heavy atom: .R<CMET 149>.A<O 17>
  Added missing heavy atom: .R<CMET 149>.A<OXT 18>
  Added missing heavy atom: .R<CMET 149>.A<SD 11>
  Added missing heavy atom: .R<CMET 149>.A<CE 12>
Creating new UNIT for residue:  sequence: 150
Created a new atom named:  within residue: .R< 150>
  Added missing heavy atom: .R<NMET 151>.A<N 1>
  total atoms in file: 4645
  Leap added 21 missing atoms according to residue templates:
       9 Heavy
       12 H / lone pairs
  The file contained 1 atoms not in residue templates
> saveOff  "/salih1"
saveOff: Improper number of arguments!
usage:  saveOff <object> <filename>
> ^
```

The original file of the **Fur** protein data bank does not contain hydrogen atoms. ***Xleap*** will add them directly after loading the protein data bank file. 1191 hydrogen atoms are added to the **Fur** proteins.

If ***xleap*** did not give error messages, then the process is going correctly. In **Fur** file ***xleap*** complained that there are unknown atoms in the protein:

```
FATAL:  Atom .R<A3 306>.A<C3' 8> does not have a type.
FATAL:  Atom .R<A3 306>.A<C4' 9> does not have a type.
FATAL:  Atom .R<A3 306>.A<C5' 10> does not have a type.
FATAL:  Atom .R<A3 306>.A<O4' 11> does not have a type.
FATAL:  Atom .R<A3 306>.A<N6 12> does not have a type.
FATAL:  Atom .R<A3 306>.A<H61 13> does not have a type.
FATAL:  Atom .R<A3 306>.A<H62 14> does not have a type.
FATAL:  Atom .R<A3 306>.A<C6 15> does not have a type.
FATAL:  Atom .R<A3 306>.A<C5 16> does not have a type.
FATAL:  Atom .R<A3 306>.A<N7 17> does not have a type.
FATAL:  Atom .R<A3 306>.A<C8 18> does not have a type.
FATAL:  Atom .R<A3 306>.A<N9 19> does not have a type.
FATAL:  Atom .R<A3 306>.A<C4 20> does not have a type.
FATAL:  Atom .R<A3 306>.A<N3 21> does not have a type.
FATAL:  Atom .R<A3 306>.A<C2 22> does not have a type.
FATAL:  Atom .R<A3 306>.A<N1 23> does not have a type.
FATAL:  Atom .R<A3 306>.A<H3T 24> does not have a type.
FATAL:  Atom .R<T5 307>.A<H5T 1> does not have a type.
FATAL:  Atom .R<T5 307>.A<O3' 2> does not have a type.
FATAL:  Atom .R<T5 307>.A<O5' 3> does not have a type.
FATAL:  Atom .R<T5 307>.A<C1' 4> does not have a type.
FATAL:  Atom .R<T5 307>.A<C2' 5> does not have a type.
FATAL:  Atom .R<T5 307>.A<C3' 6> does not have a type.
FATAL:  Atom .R<T5 307>.A<C4' 7> does not have a type.
FATAL:  Atom .R<T5 307>.A<C5' 8> does not have a type.
```

This problem arises from uncorrected loading of parameter files. To solve this problem, loading needed parameter files correctly and edits the missing atom was done using ***xleap*** (graphical user interface).

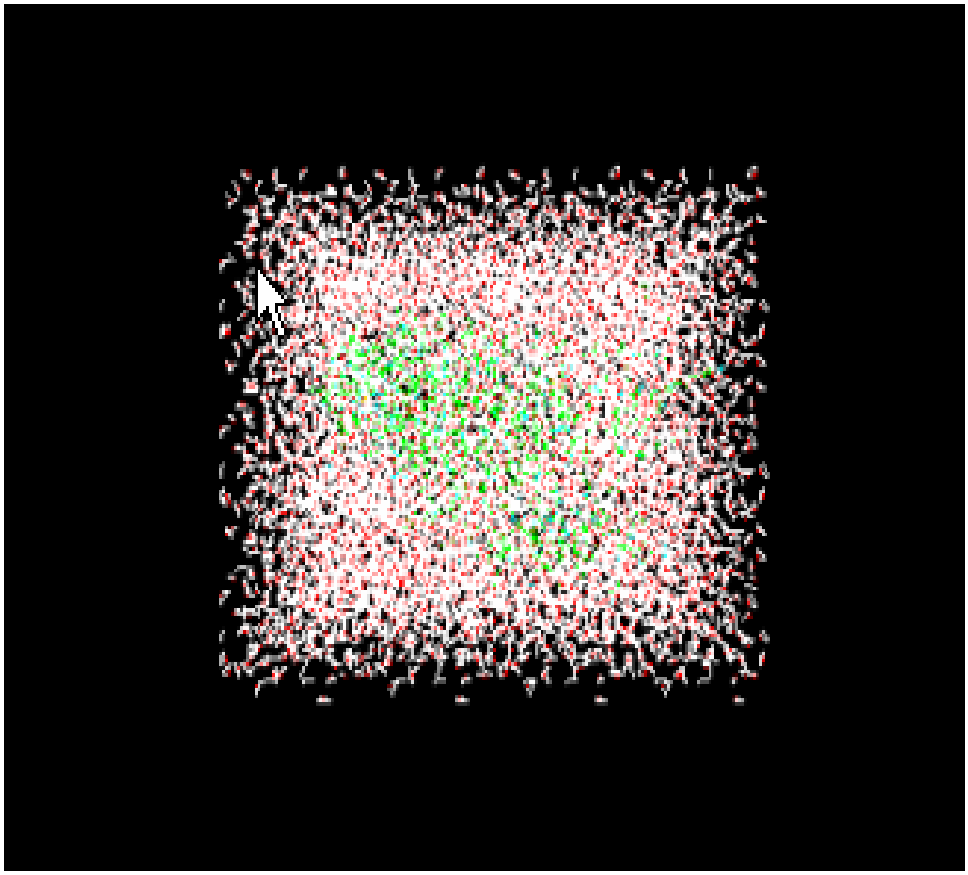
At this point, a **PDB** file will be loaded, which will add hydrogen's, other missing heavy atoms and assign parameters to all atoms. Then minimization of structure is run computationally intensive steps implementing in solution.

2.2.2 Minimization in solution

Minimization in solution requires setting a solution (water box) around the protein. ***AMBER7*** has this facility and it can be performed easily. ***xleap*** was used to generate a water box around the protein in order

to complete the model system. The command line used to generate the box is called "**solvateBox**", and there are several ways to generate the water box. The most straightforward way is by using the **WATBOX216** which is a pre-equilibrated box of **TIP3P** water. The buffering distance between the edges of the box and the protein are determined by the user according to size of the protein. In our work, it was determined and found to be 14.0 angstroms. A larger dimension of the water box will result in an unnecessary computing time. Also, using a smaller water box during the simulation will result in the protein undergoing conformational changes and as a consequence part of the protein may stick outside the box.

```
> solvateBox fo WATBOX216 14.0

Solute vdw bounding box:    49.995    53.684    37.063

Total bounding box for atom centers: 69.995   73.684    57.063

Solvent unit box:        18.774    18.774        18.774

Total vdw box size: 73.439    76.851    60.149 angstroms.

Volume:   339472.593  A^3

Total mass 162173.797 amu,  Density 0.793 g/cc

Added 8064 residues.

>
```



The next step was adding counter ions to the system. Before adding ions, the charge of the system must be figured out. If the system is positively charged, negative counter ions like **Cl^-** must be added and if it's negatively charged, positive counter ions like **Na^+** must be added. The following command was used to figure the charge of the system:

>charge fo

```
Total unperturbed charge : -9.00

Total perturbed charge: -9.0
```

Since the system was found to be negatively charged, then positive counter ions (**Na^+**) were added. The following command used to add

counter ions to the system which will make a net charge zero for the system. A total of 9 counter ions were added:

>addIons fo Na+ 0

Finally, the system is ready to generate the needed files for sander in order to start the minimization process. The topology and trajectory files for the protein were generated before starting the minimization process.

```
>·saveAmberParm·mol·wt1mg.parm7·wt1mg.crd·↵
Checking·Unit.↵
Building·topology.↵
Building·atom·parameters.↵
Building·bond·parameters.↵
Building·angle·parameters.↵
Building·proper·torsion·parameters.↵
Building·improper·torsion·parameters.↵
total·468·improper·torsions·applied↵
Building·H-Bond·parameters.↵
Marking·per-residue·atom·chain·types.↵
(Residues·lacking·connect0/connect1·--·↵
these·don't·have·chain·types·marked:↵
res·total·affected↵
CGLN·1↵
NCYS·1↵
WAT·8064↵
)↵
(no·restraints)↵
>·↵
```

The molecular dynamics simulation is ready to startup after generating the topology and coordination files for Fur. The energy minimization was performed in two stages. The first stage was to perform

the energy minimization in order to minimize the water molecules while holding the protein static. The following is the input control file which was used to run this stage of calculation:

```
minimization
$cntrl
imin =1, ntmin = 1, maxcyc= 200,
irest =0,ntx =1,
ntb =1, ntp =0,
ntpr =10,
nsnb =5,
cut =9.0,
scee =1.2,
ntf =2,
ntc =2,
```

The above file command can be explained as follows:

imin: flag to run minimization

=0 perform molecular dynamics (no minimization).

=1 performs energy minimization.

=5 this will read in a trajectory file which is needed for analysis.

ntmin: flag for the method of the minimization, since there are many of the minimization algorithms, we must inforce sander the method of Minimization to be used.

=0 this flag will perform full conjugate gradiant minimization. The first 10 cycle it will perform steepest descent and after every nonbonded pair list update.

=1 it will perform the steepest descent minimization for NCYC

steps, then it will switch to gradiant minimization.

=2 in this option the steepest descent will be used.maxcyc:

It's indicating the maximum number of the minimization steps.

irest: flag to start run.

=0 no effect (default value).

=1 restarts calculation. Requires velocities in coordinates input file.

ntb: periodic boundary condition. The value of the **NTB** specifies weather using constant volume or constant pressure during the simulation. The constant pressure is the only way to equilibrate density if the starting state is not correct. The default value using of the **NTB** = 1 which using constant volume.

ntp: Flag for constant. When the NTB set to be 2, the NTP must be set to 1 or 2.

=0 used when the value of NTB dose not equal 2, no pressure boundaries condition are used.

=1 used with molecular dynamics and isotropic pressure scaling.

=2 used with molecular dynamics and anisotropic (x-, y-, z-) pressure scaling: is used only when the orthogonal boxes are used.

This generated file was saved as "**min.in**" and fed to ***sander*** in order to start the simulation process. The following command describes how to run the sander and the important files need to be loaded:

$ sander -O -i min.in -p fur.parm7 -c fur.crd -r fur_water.rst -o fur_water.out

The flags that were used in the above command are also necessary.

-O: overwrite the output file and restart file if its really excite.

-i: flag for the input control parameters file.

-p: flag to load the topology file of the model.

-c: flag for the coordination file of the model.

-r: flag of the restart file.

-o: flag for the output file.

On a Pentium **III** computer, sander finishes the calculation in 3 hours, and generates an output file in addition to the trajectory and topology files. These are necessary to generate a new **PDB** file for the protein and needed for analysis of the output using the ***AMBER7*** suite analysis program.

The previous minimization resulted in relaxing the water box around the protein. To be more meticulous, we must minimize the solute while keeping the water molecules fixed, and then relax the whole system together. Since we want to remove the bad connection in the protein only,

we proceed by minimizing the entire system as a whole. The following is the input control file used to perfume this step:

```
Minimization second stage
&cntrl
imin =1, maxcyc=300,
ntpr=5,
&end
```

The above input control file was saved and fed to sander to start calculations. The following command will start the next stage of calculations:

$sander –O –i min1.in –p fur.parm7 -c fur.crd -r fur2_all.rst -o fur2_all.out

Minimization of the entire system took two hours on a Pentium **III** machine. The **PDB** file was generated from the output topology and restart files after we removed the bad connection by performing the energy minimization. The suite program ***ambpdb*** did this job. To do this job the topology and restart files were fed to ***ambpdb*** program and the following command was used to start the job:

$ambpdb –p fur.parm7 <fur2_all.rst> fur_minimized.pdb

Pentium **III** took 20 minute to generate the refined **PDB** file. ***SWISS-DEEP*** viewer is the most appropriate program to view the

generated protein. It can be downloaded freely from the ***SWISS-DEEP*** viewer site.

2.3 Preparation of the dimer protein

2.3.1 Installation and running the Autodock program

The dimer ferric uptake regulation protein was prepared after performing energy minimization. ***AutoDock*** version **3.0.5** (Automated docking of flexible ligands to receptors) was used to carry out the docking process of the protein. The program was obtained after Academic Software License Agreement from the molecular design institute, University of California, San Francisco. The program was sent via ftp file, which is compatible with the Linux operating system after preparing the machine file. The machine file contains all the operating and computer hardware which can be used to run the ***Autodock.*** Before installing the program, the machine file was prepared to run under ***Linux*** operating system. To install the program, the command prompt was used to add the following two lines to the (**.cshrc**) directory.

> ***> setenv AUTODOCK_UTI /path/to/the/directory/share***
> ***> set path= ($path $AUTODOCK_UTI)***

In any docking scheme two conflicting requirements must be handled: the desire to keep the computational demand at reasonable level and the desire for the robust and accurate procedure.

2.3.2 Setting up AutoGrid and AutoDock Jobs

The macromolecule first needs polar hydrogen's to be added and then partial atomic charges to be assigned. This can be done efficiently in ***SYBYL***, ***AMBER7*** and ***SWISS-DEEP*** viewer. The next step was to assign the atomic solvation parameters file ”**PDBQS**“ for the macromolecule. The ***addsol*** included program in the ***AUTODOCK*** has been used to get the input “**fur.pdbq**” and gave the output a **PDBQS** file, “**fur.pdbqs**”, using following command:

> addsol Fur.pdbq fur.pdbqs

Mkgpf3 and ***mkdpf3*** have been used to generate the grid parameters file “**macro.gpf**” and dock parameters file “**lig.macro.dpf**” for both ligand and macromolecular, as shown below:

> mkgpf3 lig.pdbq macro.pdbqs
> mkdpf3 lig.pdbq macro.pdbqs

The resultant files fed to the ***autogrid3*** (executable file) to start the grid calculation of the macromolecular “**macro.glg**”. The terminal command prompt was used to run this command:

>autogrid3 -p macro.gpf -l macro.glg

Autodock3 was fed with the generated docking parameters and the following command line was performed to start the docking process:

>autodock3 -p lig.macro.dpf -l lig.macro.dlg

The docking process on Pentium **III PC'**s took 45 minute before termination. The resultant **lig.macro.dlg** file was used to generate the docking **PDB** file of the dimer protein, as follows:

>get-docked lig.macro.dlg

2.4 Building the iron box and using it for energy minimization process

The program ***nucgen*** which comes with ***AMBER7*** builds cartesian coordinate canonical A- and B- models for standard **DNA:DNA**, **RNA:RNA, RNA:DNA** duplex. ***Nucgen*** requires an input file which can be prepared using the ***nukit*** program included in the ***AMBER7.*** The output **PDB** file contains all atoms named according to ***AMBER7*** naming conventions. The missing hydrogen atoms in the ***PDB*** file were added directly after loading the model into ***xleap***.

The study of Fur-binding sites on the **DNA** allowed the early recognition of a **19-bp** consensus called iron-box [53]. The sequence alignment of the iron-box **5'GATAATGATAATCATTATC'3** confirmed that this sequence is the functional target of the Fur protein. In order to study the DNA-binding of Fur protein and conformational changes, an iron-box must be prepared.

2.4.1 Starting nukit

Nukit is the sample session for building a decamer **DNA** duplex in the ***Amott*** canonical **B** geometry with ***nukit*** and **parm94.dat** naming conventions. Setting ***AMBER7*** environment, path and loading all parameter files was a pre-request to start ***nukit***. When ***nukit*** was started it asked for the following input parameters file that is needed to build the iron box:

Q1: Residue naming convention?(O = pre-94, N= 94) O/N:

A1: N

Q2: JOB NAME?

A2: Iron-box

The following line will be printed on the screen:

--------- (from here on, use Capitals)----

---------- e.g. CGCCATAT ----------

Q3: ENETER SEQUENCE {5-prime to 3-prime}:

A3: GATAATGATAATCATTATC

Q4: DNA OR RNA?(D/R):

A4: D

Q5: ENETER SEQUENCE {5-prime to 3-prime}:

A5: GATAATGATAATCATTATC

Q6: DNA OR RNA? (D/R):

A6: D

Q7: CONFORMATION:

ARNA right handed A rna (arnott)

APRNA right handed A-prime rna (arnott)

LBDNA right handed B dna (langridge)

ABDNA right handed B dna (arnott)

SBDNA left handed B dna (sasisekharan)

ADNA right handed A dna (arnott)

NIXON none of above- nucgen.pdb user-defined

A-forms may need work to place H1-primes properly.

CONFORMATION?

A7: ABDNA

- The first question asks about the kind of force field to be used. ***AMBER7*** includes the old force field, and the earliest force field which was generated in ***1994*** and ***1999***. ***AMBER7*** uses two force field conventions. The oldest one uses three letter names. The other uses one or two letter symbols for each residue.

- The second question asks for the title of the job. It is answered and defined according to the user.
- The third question that asks for the sequence of the Iron-box was followed by two warning messages, indicating only capital letters must be used (small letters will not be successful in the following steps). The sequence used started from 5' and ended at 3' (**5' GATAATGATAATCATTATC 3'**).
- The fourth question asks about the sequence that should add either **DNA** or **RNA** duplex. Since we deal with Iron-box, which is a part of **DNA**, the **DNA** convention was used.
- The fifth question asks again for the sequence of the second strand of the Iron-box. The sequence starts with 5' end with 3'.
- The sixth question asks about the sequence that should be added if it's a part of the **DNA** or **RNA**.
- The seventh question asks about the kind of conformation of the duplex to be chosen from the list shown. **ABDNA** was selected as the best conformation of the Iron-box.

The program was terminated with the generation of two files **Iron-box.in** and **lin.in**. These two files contain all the needed parameters to

generate the **PDB** file that is going to be used in the ***nucgen*** at a later stage.

The generated files Iron-box and **lin.in** are necessary for building the Brookhaven Data Bank file. The Iron-box file looks exactly as shown below:

```
Iron-box 1
D
A5 G A T A A T G A T A A T C A T T A T C A3

Iron-box 2
D
T5 G A T A A T G A T A A T C A T T A T C T3

END
$ABDNA
```

2.4.2 Preparing the Brookhaven data bank file for iron-box

The ***nucgen*** program was used to generate the **PDB** file using the output files of the ***nukit***. Before starting this program, the ***nucgen*** data base file "**nucgen.dat**" was transferred to the work directory. The generated files from ***nukit*** program were used as input file for ***nucgen***. The following command line was used to start generating the **PDB** file for the iron-box.

>nucgen –O –i Iron-box.in –o Iron-box.out –d Iron-box.dat –p Iron-box.pdb

It is possible that nucgen gives an error massage if the appropriate flags were not used. The flags can be explained as follows:

-O: this flag is known in many ***AMBER7*** programs; if the out put file does exit in the current directory, overwrite it.

-i: flag for the input file.
-o: flag for the output file.
-d: flag for the data base generation file.
-p: flag for the PDB file.

Generated **PDB** file from ***nucgen*** must be slightly modified, so as to let other programs figure out where one strand terminates and the next begins. The **PDB TER** card was added manually to the **iron-box.pdb** file (placed between the two strands).

2.4.3 Building the topology (prmtop) and coordination (inpcrd) files

The structure of the iron-box is not at its minimum energy and this may have produced bad connections between atoms or residues. Further minimization for the iron-box removed these bad connections and relaxed the structure which now was became at minimum energy. According to what was discussed before, the topology and coordination files were

needed to start the minimization. These files can be generated using: ***xleap***, ***prep***, ***link***, and ***parm***. The best method to generate these files is the graphical model building program, ***xleap***.

When loading the iron-box **PDB** file, the blank screen appears as a result of unloading the nucleic acid incorrectly. To avoid this problem, the residue was rebuilt into the ***AMBER7*** database. This was executed using the following command:

> ***loadAmberPrep "$AMBERHOEM/dat/leap/prep/all_nuc94.in"***

AMBER7 must define all residues before adding the water box around the model. The check command was used to determine if the structure was correct. The solvate box using **WATBOX216** model was applied and **SaveAmberParm** was used to generate and save the topology and coordination files. These files were fed into ***AMBER7*** to start the minimization process in solution. The following control input file was used to start the minimization.

```
Minimization second stage
&cntrl
imin =1, maxcyc=300,
ntpr=5,
&end
```

2.5 Docking the dimer protein to the iron-box

The most important step in studying the ferric uptake regulation protein conformational changes and the relation to its function was to

study the Fur protein interaction with the iron-box. The complex compound of the dimer protein and iron-box must be prepared. Due to the difficulty of implementing a docking command line that is needed to generate the **Fur-DNA** complex, the ***AUTODOCK*** interface (**mgl** tools) was used instead. This package is provided by scripps research institute at the molecular graphic laboratory, **USA**, but it could not carry out the grid and docking calculations if the executable **auodock3** and **autogrid3** source files were not installed in the directory of the ***AUTODOCK*** program. The **mgl** tools interface is compatible with the Linux operating system; therefore the docking process was generated under Linux platform.

2.5.1 Editing the PDB files

The **PDB** file needs to be corrected since it may have a variety of potential problems such as missing atoms, added water molecules, chain breaks and alternate locations. These problems have been corrected using the edit and repair commands in the **autodock** tools (**ADT**) which was built on the python molecule viewer (PMV) and has an evolving set of tools designed to solve such problems.

2.5.2 Preparing a ligand file for autodock

AUTODOCK distinguishes between aromatic and aliphatic carbons, the names for aromatic start with "**A**" instead of "**C**". Also ligands have partial charges for each atom. The ligands are written using special keywords in the **autodock**. The root keywords define the rigid set of atoms, while the branches define the rotatable groups of atoms

connected to the rigid root. The following steps were implemented to prepare the ligand files:

ADT detects the charges of the ligand. It also searches for the peptide in the ligand after loading the lignads. If all the residue is peptide the kollman charges will be added, Otherwise it will compute the Gasteiger charges.

- Defining the best root and this comes into two ways: the user can define it or let **ADT** determine it automatically by searching for atoms in the ligand with the smallest subtree. In the case of a tie if either the atom in cycle will be picked as a root. If neither atom in a cycle, the first one was picked.

- Defining the rotatable bonds and setting the number of active torsions. The included tools can recognize the rotatable bonds and the number of active torsions which are necessary to write the **PDBQ** file. The **PDBQ** is an ***AUTODOCK*** specific file format, **PDB** augmented by '**Q**', a charge.

2.5.3 Preparing the macromolecule file

Macromolecule file must be in **PDBQS** format, where the first three letters refer to the protein data bank, '**Q**' refers to the charges and 'S' to the solvation parameters. The macromolecule (dimer protein) was loaded using the following command:

Grid → macromolecule → choose macromolecule

ADT checks the charge of the macromolecule after being loaded; if the molecule appeared to not have a charge then **ADT** will check the peptide. If it has a peptide it will add the Kollman charges to it, otherwise it will add the Gasteiger charges. After that **ADT** will add the solvation parameters: **AtVol**, the atomic fragmental volume, and **AtSolPar**, the atomic salvation parameters which were used to calculate the energy contribution of the desolvation of the receptor by ligand binding. The resultant PDBQS file was saved in order to be used later for calculating the grid and docking parameters.

2.5.4 Preparing the grid parameters

The grid map was calculated for each element in the ligands. The location and pair-wise potential energy were computed by the autogrid function. For each atom in the ligands one map and electrostatic were calculated. In each map the following parameters were specified:

- Self consistent 12-6 Lennard-Jones energy parameters (R_{ij}).
- Equilibrium internuclear separation (epsij).
- Energy well depth.

In order to calculate the above parameters the following steps were followed:

1. Setting the map type which depends on atom type in the ligand. This was performed using the command:

Grid→ set map type

2. The ligand widget was allowed to modify the type of the map to be calculated, and whether to model the hydrogen model bonding.

3. Setting the grid parameters using the Grid Option Widget, this will display the current total grid points per map. The grid map size is determined by the following relation :

$$(n_x+1)(n_y+1)(n_z+1)$$

where n_x , n_y ,and n_z define the number of grid points in three dimension.

Grid → set Grid

The box lines generated around the protein have three thumbwheel widgets which were used to set the grid points in the x, y, and z direction to cover the entire model. The dimensions were set to be 60, 120, 80 which gave a total of 597861 grid points per map.

4. Writing the grid parameters file (**gpf**) by choosing the following command:

Grid → write GPF

The file was then saved and the convention (**gpf**) was used as an extension for the file.

2.5.5 Starting the autogrid

The ligands and parameters' files were placed in the same directory of ***autogrid3*** before starting the ***autogrid*** for the macromolecule. In

alternative manner the above files can be moved to the current directory of autogrid3. The following command was used instead of selecting the starting autogrid3 from the menu.

% autogrid3 –p dimerprotein.gpf –l dimerprotein.glg &

2.5.6 Preparing the docking parameters and starting the docking

The parameter files are needed in order to carry out the docking process. Each docking parameter file contains the necessary information needed to start the docking, which are:

- Define which map files have been generated before use.
- Define the center, number of torsions, where to start the ligand and number of iterations.
- Specify the docking algorithm to be used; Usually there are four different algorithms supplied in ***AUTODOCK***:
    1. The original Monte Carlo (MC) Simulated Annealing (**SA**).
    2. The traditional Darwinian genetic algorithm (**GA**).
    3. The local search algorithm (**LS**).
    4. The larmarckian genetic algorithm (**GA-LS**).

In order to start the ***AUTODOCK*** job, the dimer protein **PDBQS** was loaded and the current ligand and its parameters were chosen. The number of torsional degrees of freedom, atom types, atom center and the

number of active torsions were also specified. The algorithm parameters have been specified by opening the search parameters in the panels menu:

Docking → set search parameters→ algorithms parameters

The genetic algorithm with default parameters was used. The number of iterations to be used in calculations was defined by opening the following panel:

Docking →set docking run parameters

The parameters needed for the docking process were set as:

- Random number generator =500.
- Random number generator seeds =1000.
- The energy outside the grid =300.
- Step size parameters = 0.1.
- Other parameters are set as default.

The Docking parameters and instructions were written and saved as "**fur.dpf**" file. All previously built files were used to start the docking process using the following command:

%autodock3 –p fur.dpf –l fur.dlg &

The resulting **PDB** file generated by ***AUTODOCK*** was picked from the ***autodock3*** directory and used to derive the topology and trajectory file to start the minimization of the resulting model.

2.6 Preparing the parameter file for the ferric ion

AMBER7 has built in parameter files for several ions like **Na^+**, **Li^+**, **Cl^-** …etc. But unfortunately it does not contain parameters for iron. In order to add iron to the protein, the parameter file for iron must be prepared. This procedure was achieved easily by using the same procedure used for other ions. The parameter library files in the directory **/amber7/dat/leap/lib**, include the parameters for amino acid, **DNA**, **RNA**, and all ions supplied in ***AMBER7***. Some ion parameters are found in both versions **91** and **94** library files.

The following script was written to add the ferric ion to the protein using ***xleap.*** The default name of ferric was set to be **QF**; it will appear in the atom type query when loading the file into ***xleap***. The following script was added to **ions91**.cmd and **ions94.cmd** files:

```
i = createAtom  Fe+2  QF  2.0

set i   element Fe

set i   position { 0 0 0 }

r = createResidue Fe+2

add r i

Fe+2 = createUnit Fe+2

add Fe+2 r

saveOff Fe+2 ./ions91.lib
```

The following parameters for ions could figure out in the two parameter library files (**ions91.lib** and **ions94.lib**): name of atoms, atom types, atom charges, atomic numbers and flags. In order to define the Fe ion parameters for ***AMBER7***, the following script has been added to the library files of the ions.

```
!entry.Fe+2.unit.velocities table  dbl x  dbl y  dbl z
 0.0 0.0 0.0
!entry.Fe+2.unit.atoms table  str name  str type  int typex  int resx  int flags  int seq
int elmnt  dbl chg
 "Fe+2" "QF" 0 1 131072 1 26 2.000000
```

```
!entry.Fe+2.unit.atomspertinfo table str pname str ptype int ptypex int pelmnt
dbl pchg
 "Fe+2" "QF" 0 -1 0.0
!entry.Fe+2.unit.boundbox array dbl
 -1.000000
 0.0
 0.0
 0.0
 0.0
!entry.Fe+2.unit.childsequence single int
 2
!entry.Fe+2.unit.connect array int
 0
 0
!entry.Fe+2.unit.hierarchy table str abovetype int abovex str belowtype int
belowx
 "U" 0 "R" 1
 "R" 1 "A" 1
!entry.Fe+2.unit.name single str
 "Fe+2"
!entry.Fe+2.unit.positions table dbl x dbl y dbl z
 0.0 0 0 0.0
```

```
!entry.Fe+2.unit.residueconnect table  int c1x  int c2x  int c3x  int c4x  int c5x  int
c6x
 0 0 0 0 0 0
!entry.Fe+2.unit.residues  table   str  name   int  seq   int  childseq   int  startatomx   str
restype  int imagingx
 "Fe+2" 1 2 1 "?" 0
!entry.Fe+2.unit.residuesPdbSequenceNumber array int
 0
!entry.Fe+2.unit.solventcap array dbl
 -1.000000
 0.0
 0.0
```

2.7 Molecular dynamics simulation of ferric uptake regulation protein

The dynamic simulation is a key for understanding the structure and function of the ferric uptake regulation protein function. In order to get an accurate simulation of the model, the following steps were followed to prepare the model for computer simulation.

2.7.1 Preparing the input files

Sander needs three input files: topology, coordination files which were generated using sander and the input control file. The model was loaded using **loadPDB** command in ***xleap***; ***xleap*** complains that there was an unknown residue (**FE**), this was a result of un- loading the parameter

file of the ion. This problem was solved during loading the parameters library for the iron using the following command:

> loadOff fe.off

The xleap program has the capability to recognize that the **C**-terminal is missing an **OXT** atom, so it automatically added **OXT** atom to the C-terminal. At this point, the following have been prepared: loading the **PDB** file, adding hydrogen and missing heavy atoms and assigning parameters to all atoms. ***xleap*** was used to generate the topology and coordination files for the model in a vacuum using the following command:

>saveAmberParm fo furricproteinvacum.top furricproteinvacum.crd

Also it's reasonable to run the simulation in solution. So the water box was added around the model using the following command:

> solvateBox fo WATBOX216 2.0

The whole system is negatively charged, and must be neutralized. There are two algorithms in ***AMBER7*** to add ions, **addIons2** and **addIons**. The first algorithm is simply draws a grid around the solute and solvent and places the ions at the point where the energies are lowest. This approach was implemented to add ions to the system in a manner that ensures that the $\mathbf{Na^+}$ ions are at some distance from the molecule and the ion charges will not affect the system. The second approach draws a grid point around the solute, ignoring the water molecules when locating the position of added ions.

The following command used the "**addIons2**" algorithm to add counter ions (**Na^+**). This algorithm computes the total charge for the system and adds the proper number of sodium ions to neutralize it.

>addIons2 fo Na+ 0

SaveAmberParm command was used to generate the topology and trajectory files after neutralizing the system.

> saveAmberParm fo bothfur.parm7 bothfur.crd

2.7.2 Energy Minimization

Before starting the molecular dynamics simulation process the bad connections in the system were removed in order to relax the system. The energy minimization was performed in two stages: The first was to minimize the energy of the water molecules, while the protein and the ferric ions were held fixed. The following input control file was used to perform this stage of minimization:

```
Minimization
$cntrl
imin =1, ntmin =2, maxcyc= 200,
irest =0,ntx =1,
ntb =1, ntp =0,
ntpr =10,
nsnb =5,
cut =9.0,
scee =1.2,
ntf =2,
ntc =2,
Group input for restrained atoms
100.0
RES 1 155
$end
```

In the second stage the whole system was minimized including the solvent molecules. The input file for this stage is:

```
Minimization
$cntrl
imin =1, ntmin =2, maxcyc= 500,
irest =0,ntx =1, ntb =1, ntp =0,
ntpr =10, nsnb =5, cut =9.0,scee =1.2,
ntf =2,ntc =2,
$end
```

This input file was saved as "**min.in**", and fed into sander using the following command:

>sander –O –i min.in –p model.parm7 –c model.crd –o min.out –r min.rst

2.7.3 Equilibration

The atoms in the macromolecules and in the surrounding solvent undergo a relaxation that usually lasts for tens or hundreds of picoseconds before the system reaches a stationary state in the molecular dynamics simulations. The initial non-stationary segment of the simulated trajectory is typically discarded in the calculation of the equilibrium properties. This stage of MD simulation is called the equilibration stage.

There are different equilibration protocols set according to what results are needed from the calculation. Calls for an elaborate procedure to gradually increase the temperature in a step-wise fashion involves the one equilibration protocol. Some other protocols use a linear temperature gradient and heat the system up to a desired temperature. ***AMBER7*** has its own protocol for equilibration which uses two stages: the first one starts at low temperature (100k) and gradually heat up to 300k using 10 picoseconds intervals. This protocol was used in the present model equilibration. The suggested parameters for this protocol were as follows:

```
Equilibration stage 1

 &cntrl
 nstlim=5000, dt=0.002, ntx=1, irest=0, ntpr=500, ntwr=5000, ntwx=5000,

 tempi =100.0, temp0=300.0, ntt=1, tautp=2.0, ig=209858,
 ntb=1, ntp=0,
 ntc=2, ntf=2,
 nrespa=2,
&end
```

Where **nstlim** specifies the run steps.
dt specifies the time step.
ntx, irest the previous velocity information will not be saved for the next run.

Ntwr specifies the number of steps to be run before printing the energy output.
ntwx the saving coordinate after the number of step specified.
tempi initial temperature.
temp0 the reference temperature using the Berendsen coupling algorithm.
tautp the time constant for temperature coupling.
ig random seed number for initial velocity.
ntb periodic boundary with constant volume.
ntp pressure control (=0 no pressure control).
ntc for using the shake algorithm.
ntf omit force evaluation.
nrespa evaluating the slow-varying terms in the force field.

To start the equilibration process the above input file was used as an input control file for sander and the parameter and coordination files fed to sander also:

Commands:

> sander –O eq_TEMP.in –p model.parm7 –c model.rst –r modeleq_TEMP.rst –x modeleq_TEMP.crd –o modeleq_TEMP.out

After the calculation has terminated, the temperature information from the output file from of last stage were collected. The following command line was used to extract the temperature:

>grep TEMP modeleq_TEMP.out | awk'{print $6,$9}'> temperature.dat

This command line was used to extract the temperature information and save it as **temperature.dat** file. Excel was used to plot temperature values against time.

The next step was to bring the density of water box close to the experimental value. The equilibration was run out at constant temperature and pressure. The following input control file (**eq_p.in**) was fed into sander to start the job:

```
Equilibration fo the density of the water box
&cntrl
nstlim=5000, dt=0.002, ntx=5,  ntpr=500, ntwr=5000, ntwx=5000,
temp0=300.0, ntt=1, tautp=2.0, irest=1, ntp=1,
ntb=2,
ntc=2, nrespa=1, ntf=2,
&end
```

2.7.4 Production

The production stage occurs while we continue running the simulation process for the protein which has been prepared in previous calculation. This lasts until the structure became as close as possible to real environment. This step is performed at constant temperature and pressure. The following input file was used to run the simulation:

```
Production ( last step)
Nstlim= 20000, dt=0.002, ntx =5,
 irest =1, ntpr =200, ntwx=4000, ntwr=4000,
temp0 =300.0, ntt=1, tautp=2.0,
ntb =2, ntp=1,
ntc=2, ntf=2,
nrespa=1,
&end
```

This file was saved as **pro.in** and fed into ***sander*** using the same topology and trajectory file previously used in the equilibration process:

>sander –O –pro.in –p model.parm7 –c modellast.crd –r modelpro.rst –x modelpro.crd –o modelpro.out

This step took 523 hours on a Pentium III machine with Linux platforms. The **PDB** file has been generated using the restart and topology files after the production process has finished.

Commands:

> ambpdb –p model.parm7 < modelpro.rst> lastmodel.pdb

2.8 Analysis and Visualization

The ***pitraj*** suite program included in ***AMBER7*** was used as a workhorse to run the analysis steps. Also a prepared script file was used to extract the necessary variables that were needed to follow the analysis (see appendix IV). ***EXCEL*** was used to plot the data and Swiss deep view and Pymol were used for visualization.

Chapter 3

Results and Discussion

3.1 Prediction of the ferric uptake regulation protein 3D structure using homology modeling

The three dimensional structure of **Fur** protein from E.Coli was generated from the atomic coordinates of **Fur.** The predicted structure was based on sequence alignment between **Fur** and known proteins (figure 3.1). Although the ferric uptake regulation proteins primary sequences are 25% - 45% homologous with that of the E.Coli ferric uptake regulation protein, alignment was facilitated by comparison with other iron recognition proteins. The resultant comparison of the Fur protein with known protein structures presented in fig. 3.2. It is noticeable that most residues are preserved in the family with few exceptions. The main feature is the preserved hydrophobic residues (AGLIV) on 17 positions on the N-terminal and to a less extent (4 major positions) on the C-terminal domain. Hydrophobic basic K and R (residues K9, K10, K14, R19-X-K21, K41, K41, R42, R52, R70 and K77) repeated 7 times on the N-terminal domain, and once on the C-terminal (R110-X-K112). All proteins in the **Fur** family are histidine rich His32, His33, His71, His86, His87, His88, His90, His118, His125, His132, His143 and His145 (figure 3.1). Also the unit Cys93-X-Cys95 in coil folding is highly preserved and Cys133 in β-sheet. These residues play an important role in the Fur function; as in sensing of

metal ion dimerization and reversible binding to the metal [68, 69, 73 and 81].

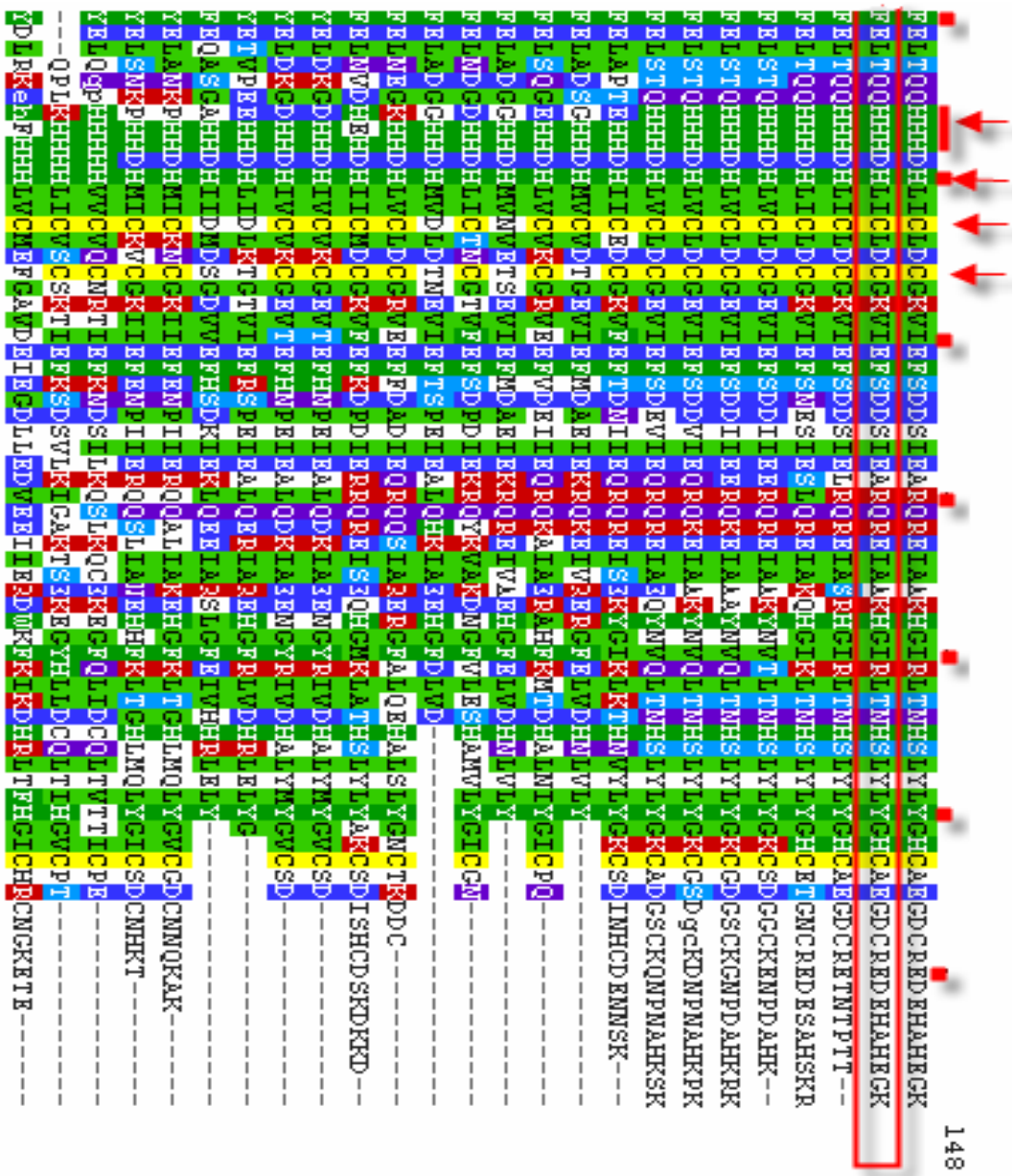


Figure 3. 1: An alignment of iron acquisition subfamily and that of Fur E.Coli protein domain predicted using SWISS-MODEL server.

The SWISS-MODEL predicted solvent accessibility composition (core/surface ratio) for the ferric uptake regulation protein of the E.Coli. The calculated value shows that 66.89% of the **Fur** residues are exposed with more than 16% of their surface. This is especially clear for residues forming the loops and residues at both C- and N-terminal. While 32.43% of **Fur** protein residues were buried (Figure 3.2). Table 1 shows the predicted bonding state of the Cysteine residues in the **Fur** protein. The predicted properties of **Fur** shows that Cys93, Cys96 and Cys133 did not favor formation of S-S bridging while it is more likely to happen in Cys138 (41% favored).

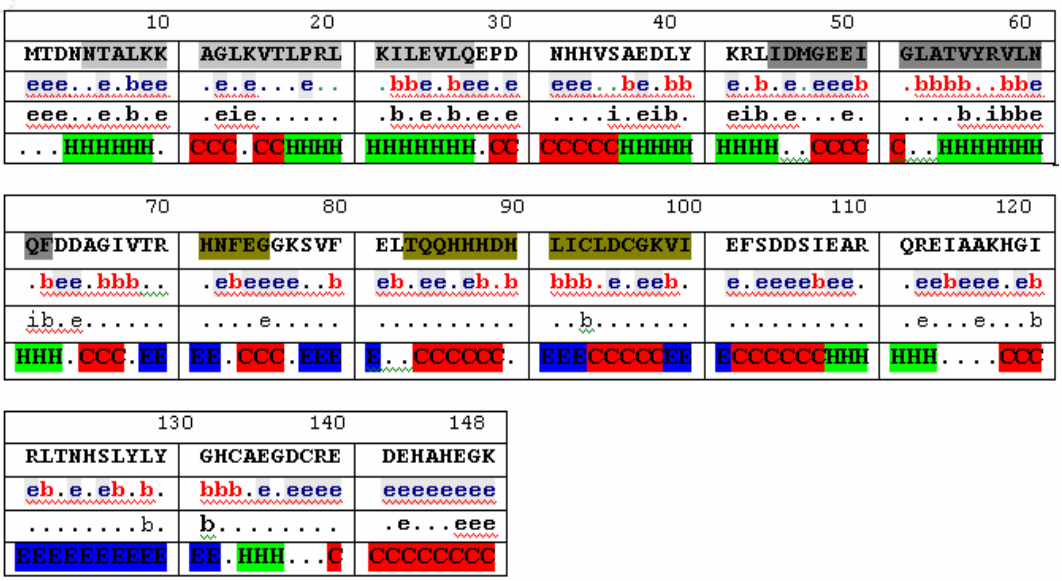
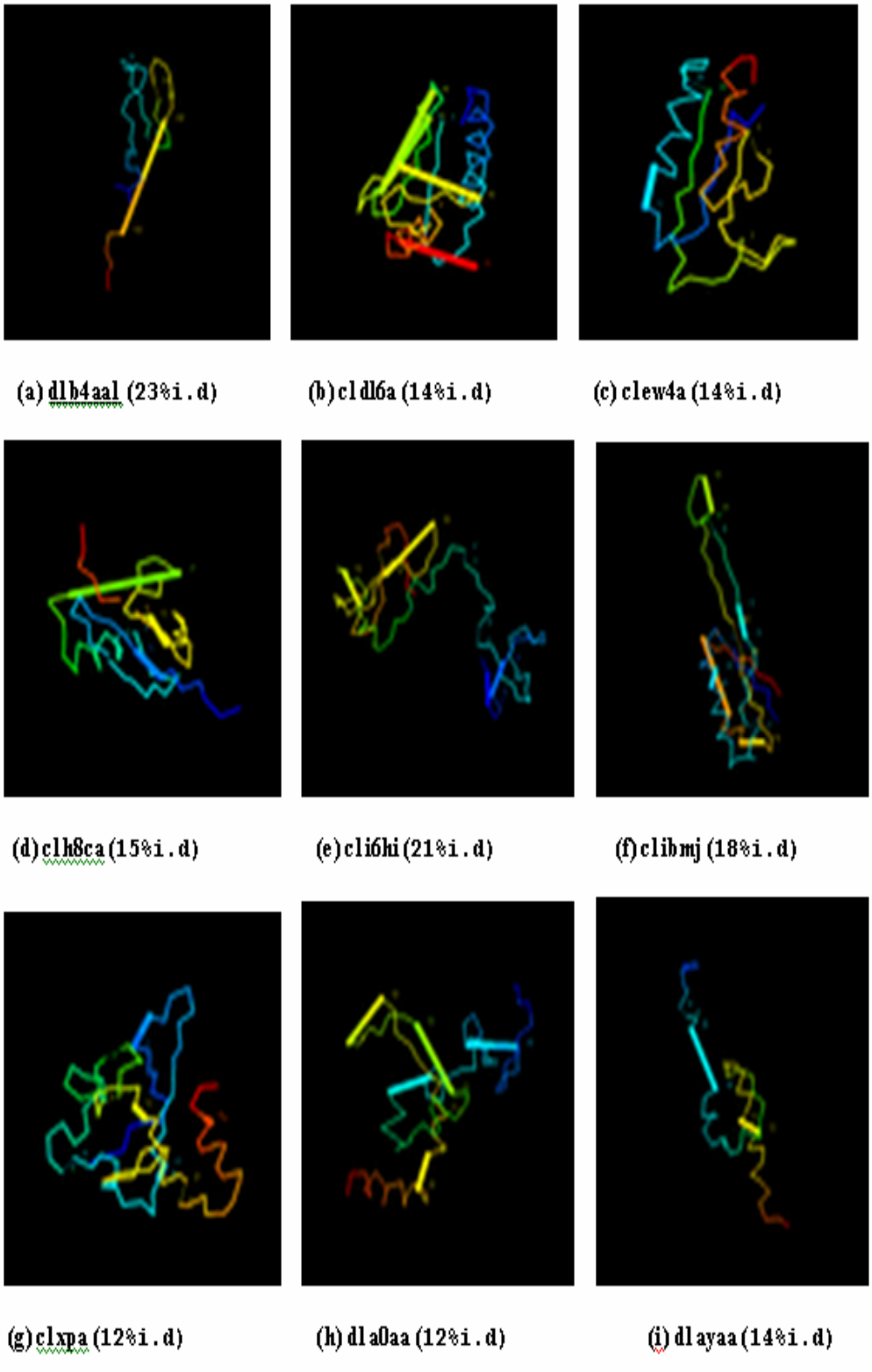


Figure 3. 2: The predicted secondary structure of Fur. The second row contains the Fur amino acids sequence. Third row contains the predicted solvent accessibility composition (core/surface ratio) for Fur protein: e: residues exposed with more than16% of their surface, b: all other residues. The fourth row contains the observed relative accessibility, where b = 0-9%, i= 9-36%, e=36-100%. Predicted solvent accessibility composition and observed relative solvent accessibility calculated by PROF server [94]. The fifth row contains the predicted secondary structure of the Fur.

Table 3. 1: Prediction of the bonding state of the Cysteine residues in the Fur E.Coli protein using PHD server [96].

Bonding state of cysteins residues			
N.cys	**Bonded**	**Non-bonded**	**Disulfide**
93	0.165	0.708	NO
96	0.098	0.804	NO
133	0.175	0.692	NO
138	0.414	0.313	YES

Upon homology modeling using the 3D ***PSSM*** server, parts of the **Fur** sequence resembled (tertiary structure) the winged helix protein family. The tertiary structure of **Fur** was aligned with similar proteins belonging to the winged helix protein family. The alignment was color coded ranging from high similarity (color coded red) to poor similarity (color coded blue) as seen in fig. 3.3. The parts of **Fur** which did not align were omitted (not shown).



(a) d1b4aa1 (23%i.d) (b) c1dl6a (14%i.d) (c) c1ew4a (14%i.d)

(d) c1h8ca (15%i.d) (e) c1i6hi (21%i.d) (f) c1ibmj (18%i.d)

(g) c1xpa (12%i.d) (h) d1a0aa (12%i.d) (i) d1ayaa (14%i.d)

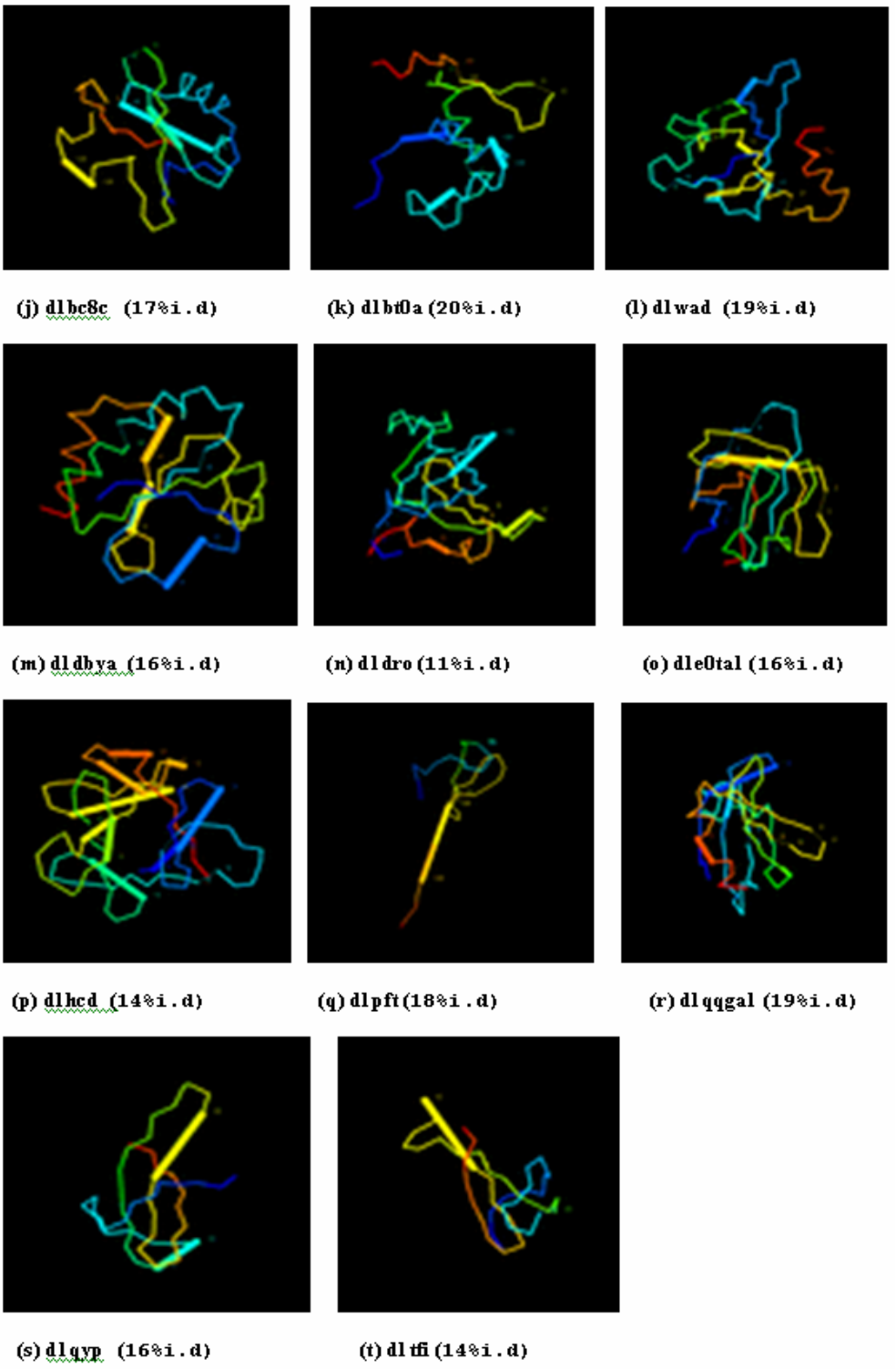


Figure 3. 3: Parts of the known proteins which fit the Fur protein. The red color shows the parts of the protein with high similarity with Fur and the blue color the low ones, according to the following color code.

During the comparative protein modeling the secondary structure of the ferric uptake regulation protein was predicted as shown in fig. 3.2 and fig. 3.4. The best fit was selected using different servers; the parts with the highest confidence level were selected from each server result. The predicted secondary structure (Figure 3.5), especially the conserved region, was compared with those predicted by NMR [58]. The **Fur** E.coli sequence was aligned with the **Fur** sequence from *Pseudomonas aerugnosa* with a known crystal structure [97]. This was found to bind zinc (II) in two different binding sites and did not have sequence similarity with **Fur** E.coli. The result of alignment gave P(N)= $9e^{-40}$ and 62.9% sequence identity. Similarity with high confidence level was for residues Lys10-Pro19, Gly48-Thr54, Arg121-Gly136, His71-Ser79 and Thr84-Ala110. **Fur** was found to have considerable similarity with Dtxr. Both proteins are iron-dependent repressor proteins but differ in their DNA specific binding [85] although both of these proteins regulate iron uptake. The fur monomer resembles a great deal the determined structure of Dtxr which contains two clearly defined domains; the N-terminal domain consists of 70 residues and contains three helices, two antiparallel β strands plus the first half of α_4. The second domain (70 residues) contains α_4, α_5 and α_6. The structure contains the helix-helix interactions α_1 with α_4 and α_5 , α_2 -α_4 , and α_1- α_5 thought to be crucial for protein function.

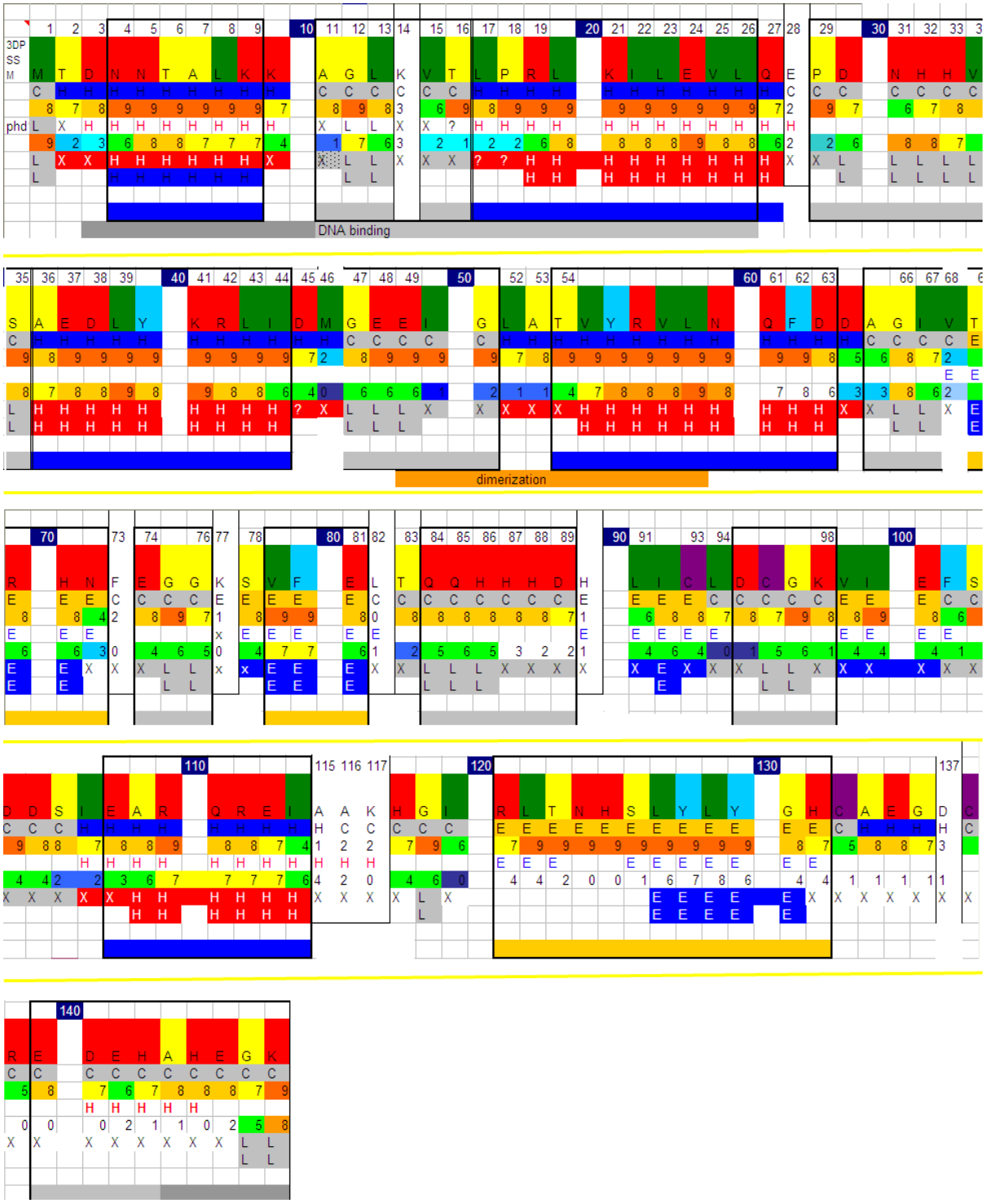


Figure 3. 4: The secondary structure prediction of the Fur (ferric uptake regulation protein), 9: shows the best prediction and as the prediction goes down to poor prediction 2, 1. The helix region (blue), coil (gray) and yellow for β sheet.

Residues near N-terminal		Folding predicted by NMR ( )	folding	Confidence level
4 to 6	**DNA binding HTH motif**		1 helix α H1	9
11 to 16			Coil T1	8 to 9
17 to 27			α **2** helix H2	9
29 to 35		coil	**Coil T2**	9
36- 44	**wing**	α helix	α **3** helix H3	9
47-51 (**G**)	Dimerization Region	44-48 coil	Coil T3	9
52-63	Y55-F61	49-59 helix 60-64 coil	4 helix α H4	9
65,66,67	suggested DNA binding domain(REF)	65-74 helix	Coil T4	8 to 9

Residues near **C**-terminal		NMR predicted	folding	Confidence level
69-72			1 sheet β	6 to8
74-76 **(LYS)**	**Metal ion binding sites**		Coil T5	8 -9
78-81			2sheet β	9
83-89 contain **His**			Coil T6	8
90-93 contain **His** and **Cys 92**			3sheet β	8
94-98 **Cys 95**			Coil T7	8-9
99-101			4 sheet β	8 -9
102-107			Coil T8	8-9
108-113		107-117 α helix	5 helix α H5	7, 8 to9
118-120			Coil T9	7,9,6
121-132			5 sheet β	9
134-136			6 helix α H6	8,8,7
140-148			Coil T10	7, most 9

Figure 3. 5: Results of homology modeling of fur from different sources compared to that elucidated by NMR study by Williams et. al [58]. Column 2 shows the suggested role previously reported for each domain.

The result of homology modeling from different servers coincided with each other to a great extent and this allowed us to propose a three dimensional structure for the fur monomer. The fit was in good homology with ***winged helix*** proteins with an RMSD value of 1.3 Å which falls within the accepted value for protein alignment (1-2Å). The final 3D structure of **Fur** agrees with the suggested function; the N-terminal domain contains the HTH motif, a helix in the central domain which was reported to be responsible for **Fur** Dimerization. The C-terminal which was reported to be the metal binding domain contains two helices separated by a β-strand and a coil (Figure 3.6).

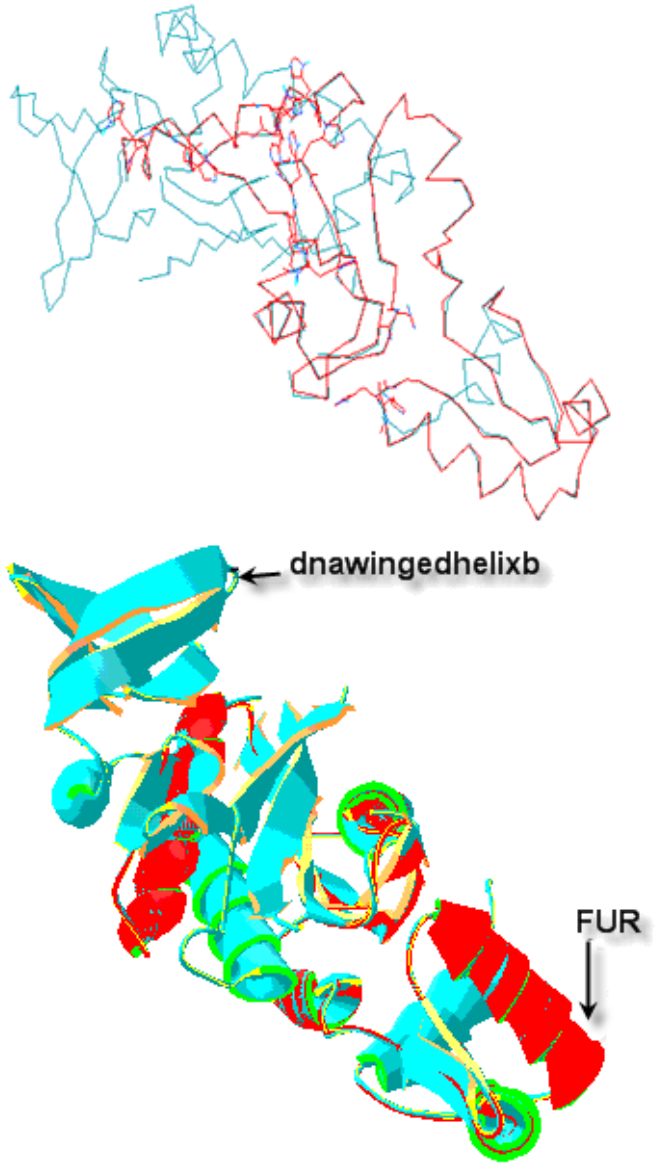


Figure 3. 6: The fitting between the predicted ferric uptake regulation protein (red) and dnawingedhelixb (green). The dnawingedhelixb shows 75% similarity to the ferric uptake regulation protein (Drawn using Swiss deep viewer (spdv)).

3.1 Fur structure modeling using Amber7

In order to validate the predicted structure of the **Fur** an energy minimization plot was used. The resultant minimized structure was predicted after running the energy minimization using ***AMBER7*** (see 2.2 for more information). The energy reached a minimum value after **~10** ps as seen in figure 3.7.1. The total **RMSD** was calculated between the starting structure of Fur and the final minimized structure was 1.18Å indicating an acceptable simulation. Energy minimization idealized the geometry of bonds and removed unfavorable connections. The ***VADAR*** server was used to calculate the Ramachandran Plot (In a polypeptide the main chain **N-Cα** and **Cα-C** bonds are relatively free to rotate. These rotations are represented by the torsion angles **Φ** and **ψ,** respectively) in the Fur protein. The sterically allowed and forbidden values of **Φ** and **ψ** conformation were plotted in fig. 3.7.2. The green line indicates the sterically allowed **ψ** and **Φ** angles for all residues except **Gly** and **Pro**. While the red color represents the conformational angels of several secondary structures (Figure 3.7.2). The Ramachandran indicates that all the calculated **Φ** and **ψ** angles fall within the allowed regions for α helices indicating a reasonable determined structure.

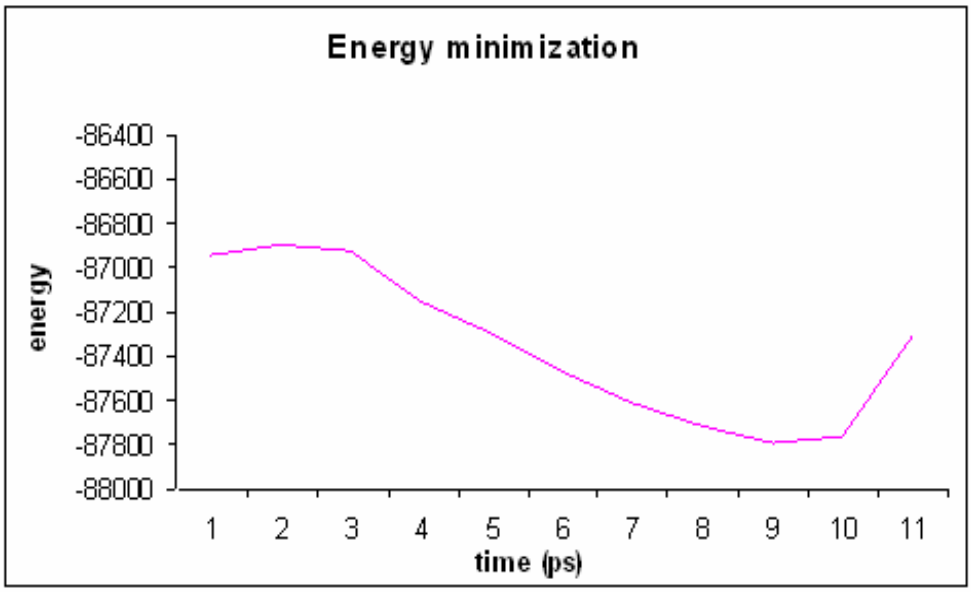


Figure 3. 7: Energy minimization of the ferric uptake regulation protein (Fur).

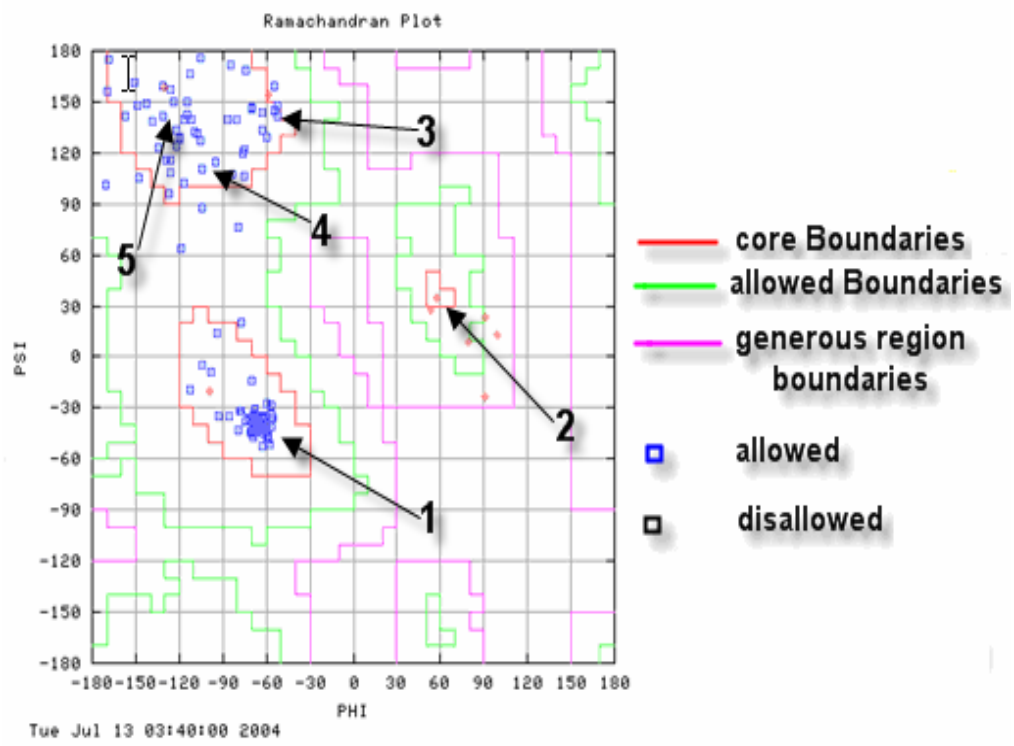


Figure 3.7.2: Ramachandran plot of the ferric uptake regulation protein. 1, the right handed α helix; 2, left-handed α helix; 3, collagen helix; 4, parallel β sheet; 5, antiparallel β sheet.

The three dimensional structure of the **Fur** dimer in the Escherichia coli reveals that each monomer consists of three domains (Figure 3.8). The N-terminal domain which is responsible for **DNA** binding, the central domain which play a role in dimerization, and the C-terminal (residues 70–148) which has the metal binding sites. The C-terminal contains the Cyestine residues. These residue play an important role in the regulation function (metal ion binding sites). This architecture helps to understand the Fur function. The DNA-binding domain contains helix-turn-helix (HTH) motif which was observed in other dimeric DNA-binding proteins [85, 55, 96 and 88]. The Fur protein in Escherichia coli like other Fur proteins in Vibro cholerae, Yersinia pestis, Pesudomonas aeruginosa, related Gram-negative bacteria and DtxR is a dimeric protein activated by divalent cations like $\mathbf{Fe^{+2}}$, $\mathbf{Mn^{+2}}$,$\mathbf{Ni^{+2}}$ and $\mathbf{Co^{+2}}$. These divalent cations bind in the C-terminal and specifically in the proximity of the interface area of the two domains in each subunit [88, 89 and 93]. Metal ion binding tunes the orientation of the N-terminal DNA-binding domain. This consequently determines the distance between the DNA-binding sites on the N-terminal in the two subunits, and hence the DNA-binding affinity of Fur. Indeed, Saito and Williams [58] suggested a structure based on NMR data which bares a great similarity with slight differences in the region between aa106-aa178. They proposed that the region (residues106-108) and (residues114-aa117) are not present in the helix domain.

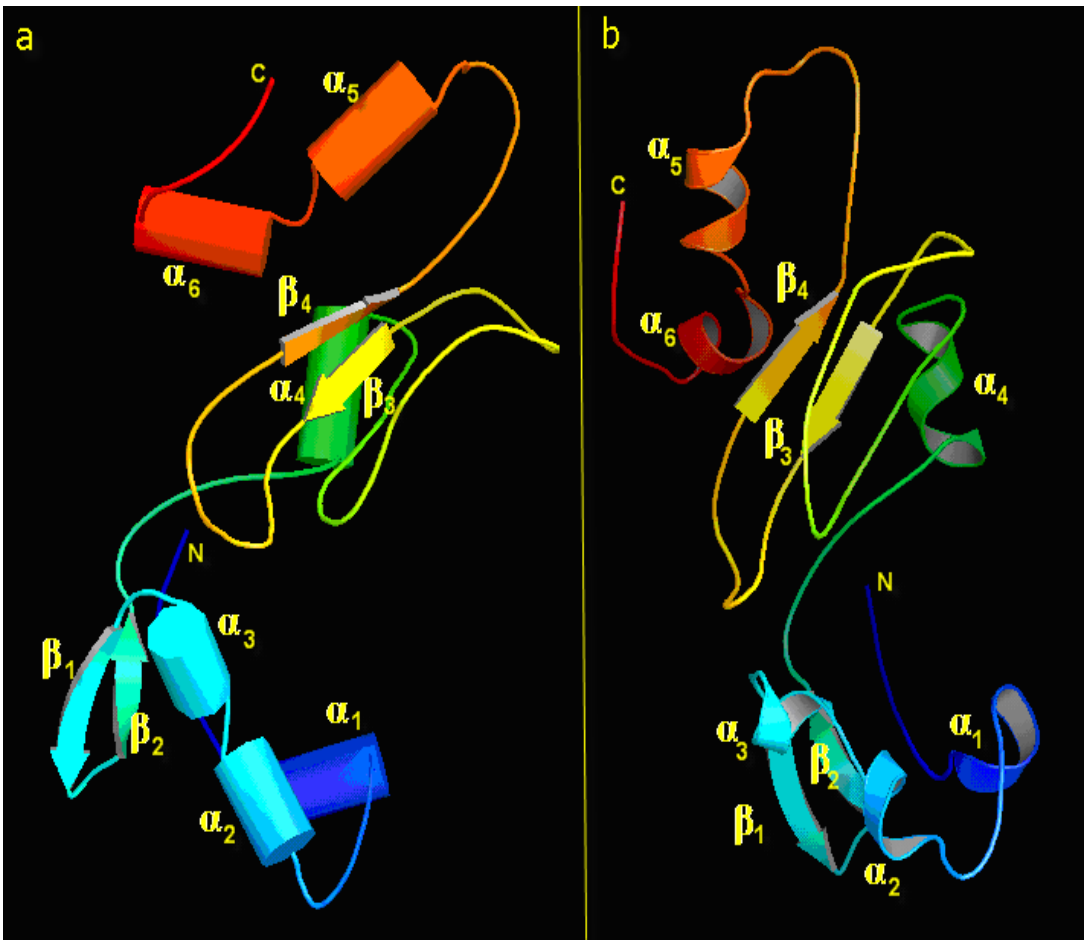


Figure 3. 8: The three dimensional structure of the ferric uptake regulation protein from E.Coli at minimum energy calculated using AMBER7 program. a) Three dimensional structure of Fur presented using cartoon. b) Ribbon display Fur 3D structure.

3.2 Fur dimer

The Fur dimer was built using Autodock (see 2.3). It is evident from the dimer structure (Figure 3.9) that dimerization takes place in central domain which lies in the region aa45-aa70. This finding agrees with other experimentally results about the dimerization site on the Fur [60, 84, 87 and 92]. The calculated distances between residues in the dimer are shown in Table 3.2. It is clear that most interactions occur between the residues Val 55, Leu 53, Gln 52, Glu 49 and Tyr 56 in one monomer with the closest contact occurring at residues 49-56 (Figure 3.10). These residues are present in the α-helix region near the N-terminus which agrees with the experimental findings of Coy et al. [60], and Kolade et al. [57] which concluded that H-H interactions occurs at the central domain of the N-terminal. Figure 3.9 shows the helix-helix interaction between the two Fur monomers protein. The negatively charged and highly polar Glumatic acid residue (Glu49) seems to aid the establishment of a hydrogen bond across Fur monomers. Extensive hydrophobic interactions occur between the surface of Fur protein and DNA due to the hydrophobic properties of valine (Val 55) and leucine (Leu 53). The aromatic cycle of tyrosine (Tyr 56) also helps to establish a hydrogen bond between the two monomers. All residues together take part in bringing the two monomers together in an antifriction process which produces the Fur dimer. It is evident that the dimerization process is mainly due to highly polar and negatively charged residues.

Table 3. 2: The calculated distances between the two monomers of the ferric uptake regulation protein before binding to the iron box.

Residues	Distance (Å)
N-terminal – N-terminal	20.4
VAL25-VAL25	3.2
LEU52-LEU82	0.7
GLN51 ─ GLN85	0.02
GLU49 ─ GLU 82	0.02
THR54-THR84	0.5
GLU49-GLU49	28.2
THR69-THR69	12.1
GLN85-GLN85	32.4
ALA53 ─ ILE 107	8.6
THR54 ─ GLU 108	9.5
ARG112-ARG112	12.7
C-terminal ─ C-terminal	34.9

The N-terminal domain which is approximately 70 residues (1-70) is suggested as a candidate responsible for the DNA binding and dimerization [84, 87 and 89]. The canonical HTH motif of the N-terminal of Fur appears to be unique since there is no significant structural homology between Dtxr or IdeR and Fur [85, 90]. On the other hand, the large C-terminal domain was thought to be responsible for metal sensing

and binding [60, 91]. It was reported that residues which associate with metal ions, (Cys and His) are present near the C-terminal domain [83, 84]. Also, the C-terminal packing lead to the stability of Fur structure which contribute to the stability of the Fur dimer.

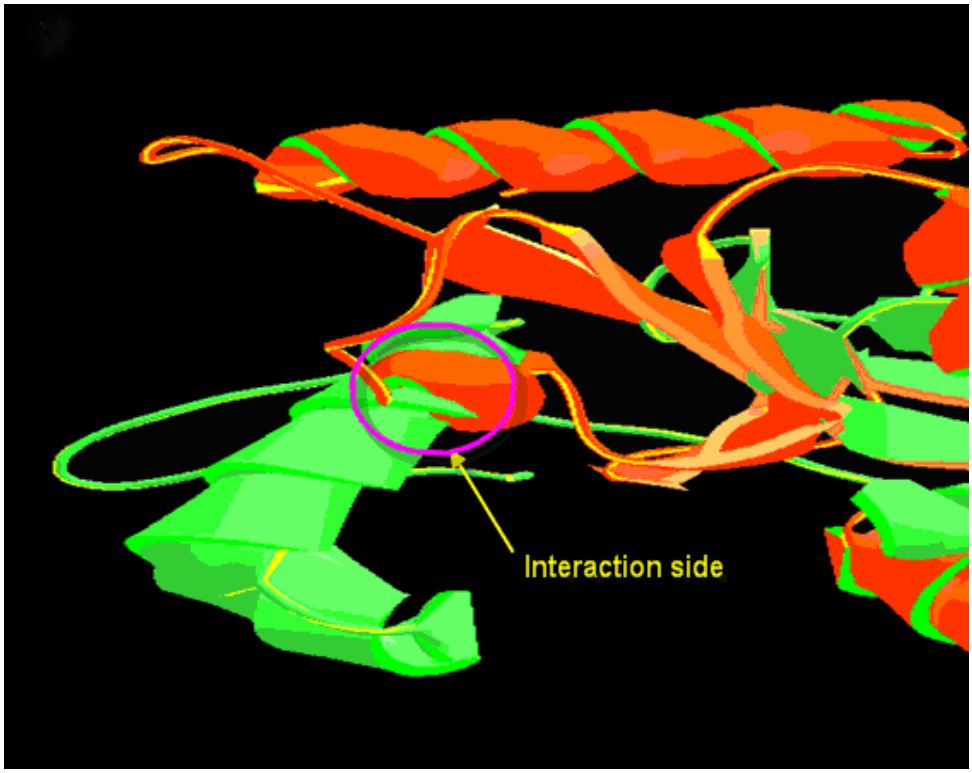


Figure 3. 9: The helix-helix interaction between two monomers. Red and green colors represent the two monomers.

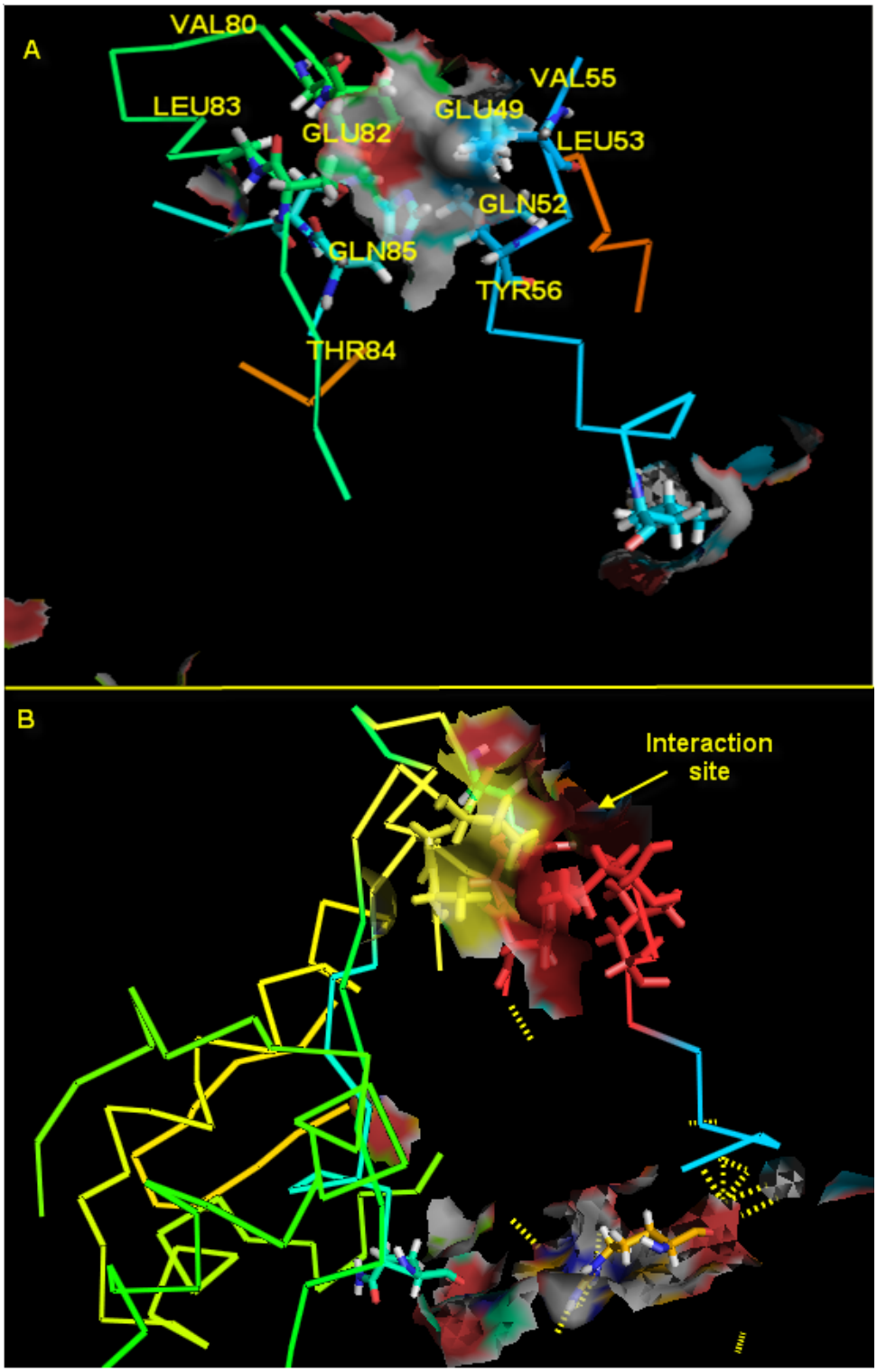


Figure 3. 10: A) Shows the residues that take part in dimerization. According to the picture the dimerization take place between these residues Glu, Val, Leu, Tyr and Gln. B) Interaction sites in (A) shown using surface bonded display.

3.3 Building the iron box

In order to study the interaction of Fur dimer with DNA, 19bp inverted repeat (GATAAT) was built and used to represent the 3D structure of the iron box (Figure 3.11). The docking process was used to dock the Fur dimer on DNA in the absence and presence of iron (II). The result shown in fig. 3.12 confirms previous reports [55, 58] that the GATAAT binds the Fur dimer by its symmetric core ATAT. The type of interaction between Fur dimer and DNA is discussed in the next section.

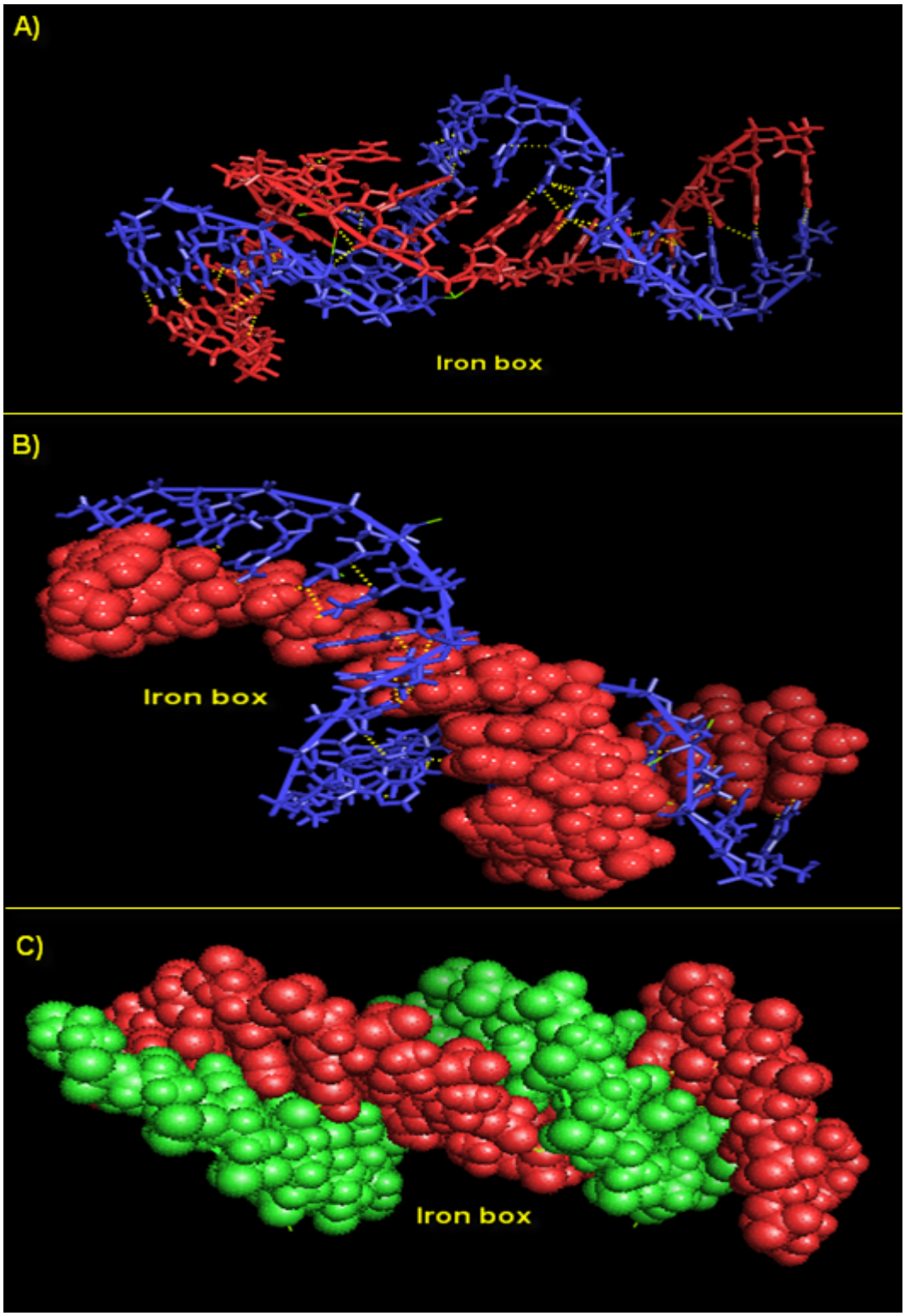


Figure 3. 11: The tertiary structure of the iron box generated using AMBER7. A) Line display of the helices (red, blue) the discontinuous yellow line shows the sulphide bond. B) One of the two helices is represented using sphere display while another presented using line mode. C) The two helices represented using sphere mode.

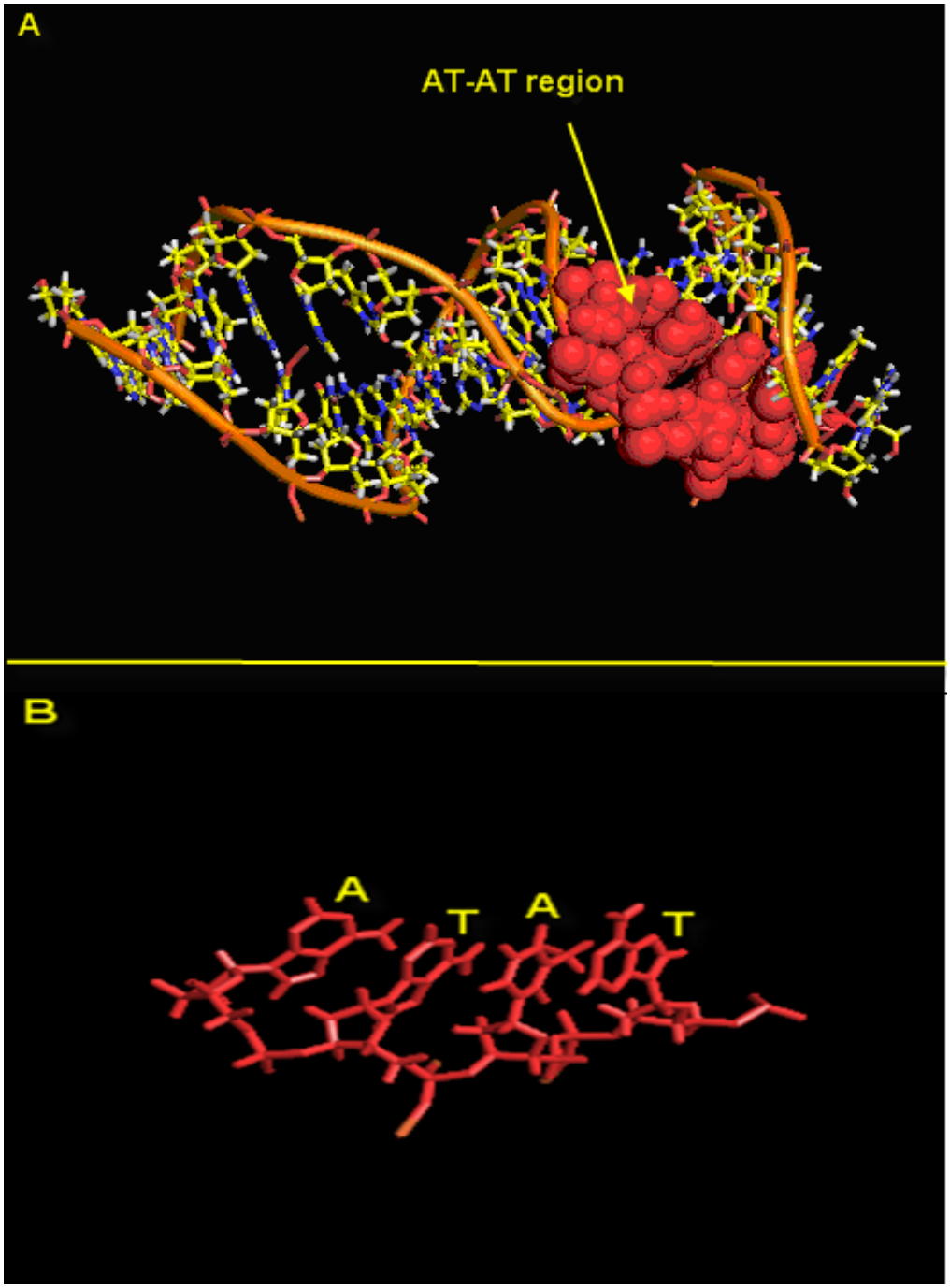


Figure 3. 12: The repeated AT-AT presentation. A) Shows the site where the repeated AT-AT occurs. B) AT-AT drawn alone which shows how they appear in the 3D structure.

3.4 Fur-DNA interaction

Specific Fur-DNA interaction with the promoter region of aerobactin has been visualized using ***Pymol*** [95]. Fur uses Fe^{+2} as a co-repressor to bind to a consensus sequence (iron box), which was found in a promoter region of several iron-regulated genes [55, 58]. It was previously reported that Fur does not interact with **DNA** (iron box) in the absence of divalent metal ions. Metal ions like $\mathbf{Mn^{+2}}$, $\mathbf{CO^{+2}}$, $\mathbf{Cu^{+2}}$…etc were found to act as co-repressors for fur protein [83, 86 and 89]. The resultant model shows that Fur-DNA recognition which maybe attributed to specific interactions between the amino acid residues of the recognition helix of the Fur and the exposed major groove of the DNA (iron box) (Figure 3.13). It is evident that the N-terminal region of the Fur interacts with iron box using helix-turn-helix (HTH). This model shows that dimeric HTH containing recognition sequence which attaches adjacent grooves on the iron box. According to this model, Fur dimer interacts with DNA through a symmetric AT-AT unit. It was proposed in several previous works that the AT-rich region is the most probable binding site that binds the Fur dimer [55, 87]. Escolar et al. [87] proposed that 19-bp iron box can be viewed as a head-to-head-to-tail repeat of a simple hexamar GATAAT, which contains a rich AT repeated unit.

The model of DNA binding apo-fur dimer structure clearly suggest that putative DNA-binding helices α_2 and α'_2 fit into the major groove well and α_2-α'_2 overlap is presented by conformational changes in fur dimer. Recognition and binding is the result of direct interaction between the base

pairs in the major groove of DNA and the amino acid side chains of α-helix. The most likely residues on the fur which represent the binding site to DNA were: valine (Val15), leucine (Leu13), proline (Pro18) and alanine (Ala11). It can be seen in Table 3.3 that these residues moved closer to the DNA. The extensive hydrophobic properties of valine, leucine and alanine residues show hydrophobic interaction between the surfaces of the Fur and edges of the bases of suger-phosphate backbone on the iron box groove. The aromatic cycle on proline performs hydrogen and salt bridges with sugar-phosphate backbone. These interactions induce an affect on the DNA by overwinding the middle four base pairs and compression of the minor groove in the center of the operator, such that the phosphate to phosphate distance is reduced from 11.4 Å for canonical B-DNA to 9.3 Å (figure 3.14).

Table 3. 3: The calculated distances between the amino acid residues on Fur dimer and AT region on the DNA binding domain (before adding iron (II)).

Residue	Distance (Å) from AT groove
N-terminal	7.8
Leu 13	0.7
Pro 18	1.3
Ala 11	0.9
Gly 12	0.8
Arg19	7.4
Asn8	6.9
Leu20	0.8
C-terminal	24.5

The result indicates that in the presence of activating metal Fur undergoes conformational changes. These conformational changes render to N-terminal domain of the Fur, which was proposed as a result of DNA binding [81, 87 and 91]. Kolade et al. suggest that these conformational changes occur through a rearrangement of the packing of the N-terminal domain (HTH) motifs to reveal two competent DNA binding motifs [57]. Table 3.4 shows the calculated distance between the two monomers of the Fur bind to iron box. The comparison between the data in table 3.4 shows that the two monomers become close to each other as a result of DNA binding. This leads to the conclusion that Fur undergoes conformational changes during DNA binding. The calculated RMSD between two model shows 2.5Å difference. This value indicates the conformational changes during the Fur-DNA interaction. The considered operator sites of the Fur structure show a similarity with DtxR. Both bind with operator 19bp inverted repeat sequence of iron box. Noel Baichoo and John D. Helmann reported that Fur-DNA complex maybe structurally similar to the DtxR complex [55]. This suggestion and the model they obtained from DtxR–DNA complex agree with our result. DtxR and Fur bind to DNA using N-terminal domain due to the helix-turn-helix (HTH) DNA-binding motif.

Table 3. 4: The calculated distances between the two monomers on the Fur dimmer before and after binding to the iron box, the residues indicated by arrows moved closer to each other upon DNA binding (before adding the iron metals).

Residues	apo Fur dimer	Fur dimer/DNA
N-terminal – N-terminal →	20.4	15.4
VAL25 – VAL25	3.2	5.60
PRO29 – PRO29	0.5	1.7
GLU49-GLU49 →	28.2	12.8
THR69-THR69 →	12.1	9.5
GLN85-GLN85	32.4	34.5
ARG112-ARG112 →	12.7	4.2
C-terminal – C-terminal	34.9	32.7

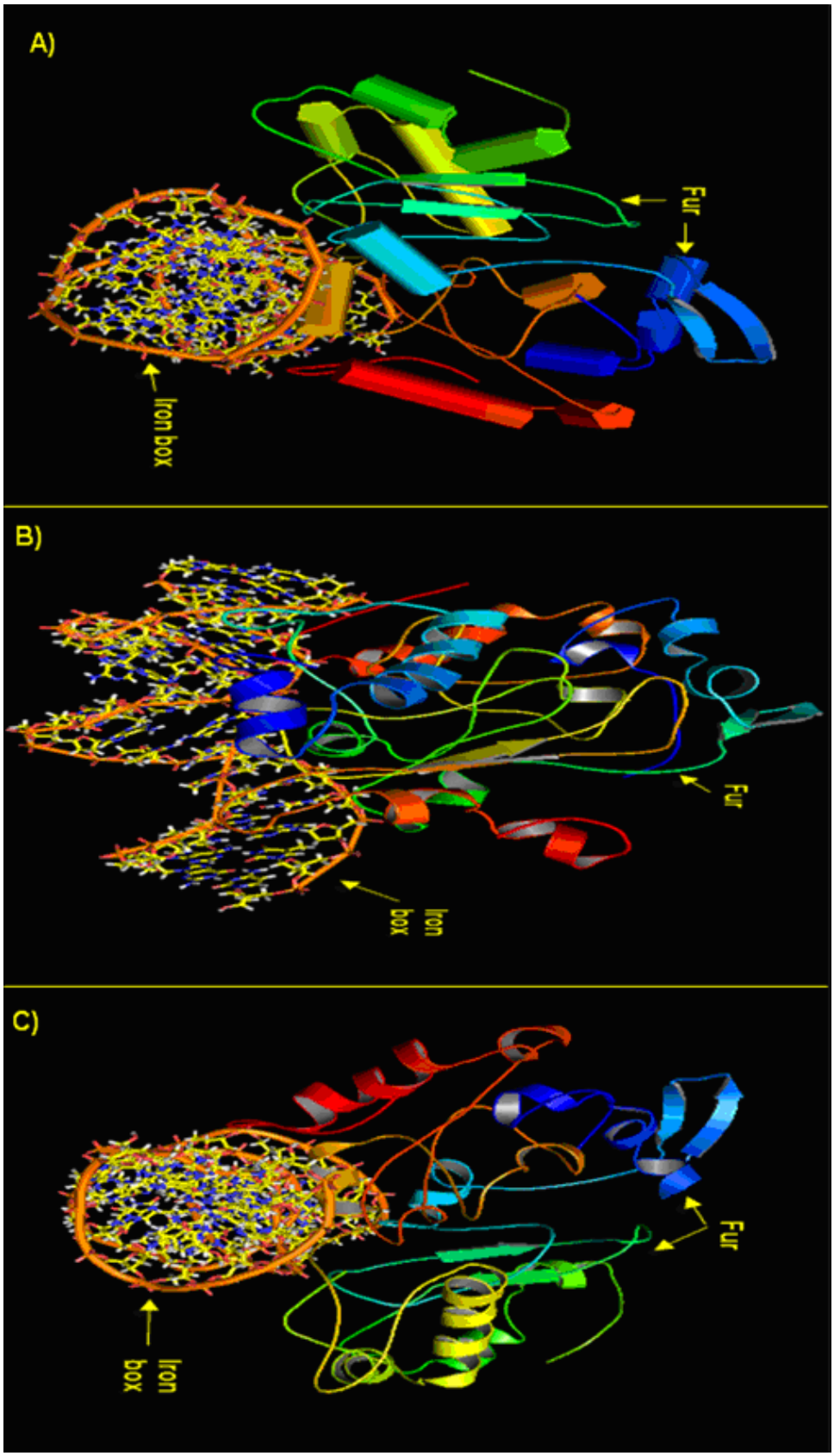


Figure 3. 13: The DNA-Fur models. The two structures show the binding sites of the Fur protein to the iron box. The red color in the second structure shows the binding region of the Fur to the AT-AT unit. A, B and C are drawn from different site view.

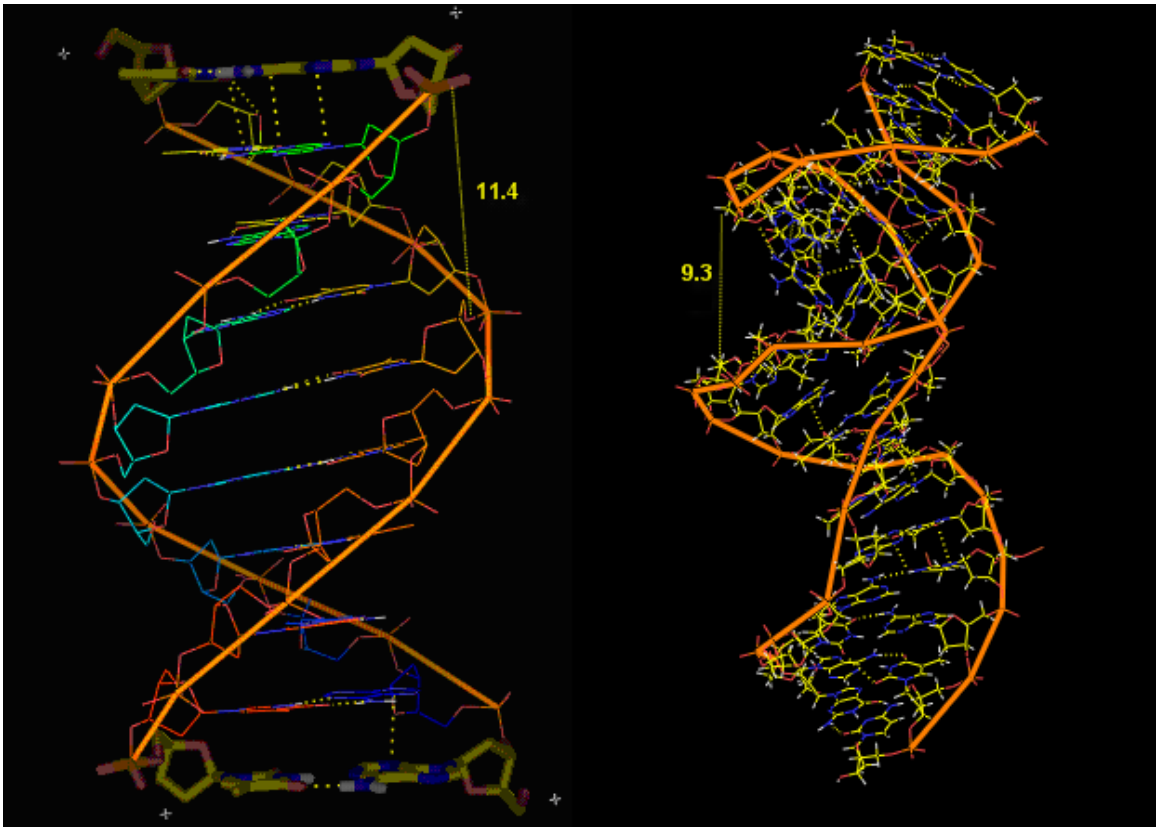


Figure 3. 14: DNA tilting, the three dimensional structure of the canonical B-DNA, before (left) and after (right) binding to the fur dimer. The calculated distances between phosphate atoms in the first major grove of the two models are shown

3.5 Validation Of the dynamics simulation

Different properties were calculated from the simulation output to validate the results. The **B-factor**, **RMSD** and energy of the simulation data were calculated. These results gave a view of the direction of the simulation and conformational changes occuring in the model. Figure 3.15 and fig. 3.16 show the calculated **B-factor** and **RMSD** extracted from the simulation output. The **B-factor** shows the correlation of the residues during the simulation. From the **B-factor** plot, it appears that residues which are close to the dimer region are protected from fluctuation as a result of strong binding. The C-terminal however, has the highest correlation since there is no restriction of fluctuation. On the other hand the dimer domain and DNA binding domain in the N-terminal have the lowest fluctuation due to the fact that the binding affinity with **DNA** is strong like dimerization. As there is no resolved X-ray structure the experimental **B-factor** is still unknown. To avoid this elucidation the **B-factor** is calculated using two time windows: 10 **ps** and 25 **ps**. The two plots are fitted and showed that the correlation paths have the same order of magnitude. The **RMSD** value is less than 2**Å** (~ 1.5**Å**) which indicates a good simulation and acceptable structure. After the first 3 **ps** the production gave high values indicating that the system is still in equilibration phase. A good simulation must be performed under real conditions, such as: constant temperature, pressure and volume. Fig. 3.17 shows the heating of the system up to 300 **K** and then keeping it constant at this value during the simulation. Fig. 3.18 shows the effect of raising the

density of the water box up to the real value prior to starting the simulation. All previously listed conditions lead to a good simulation process and hence a reliable simulated structure.

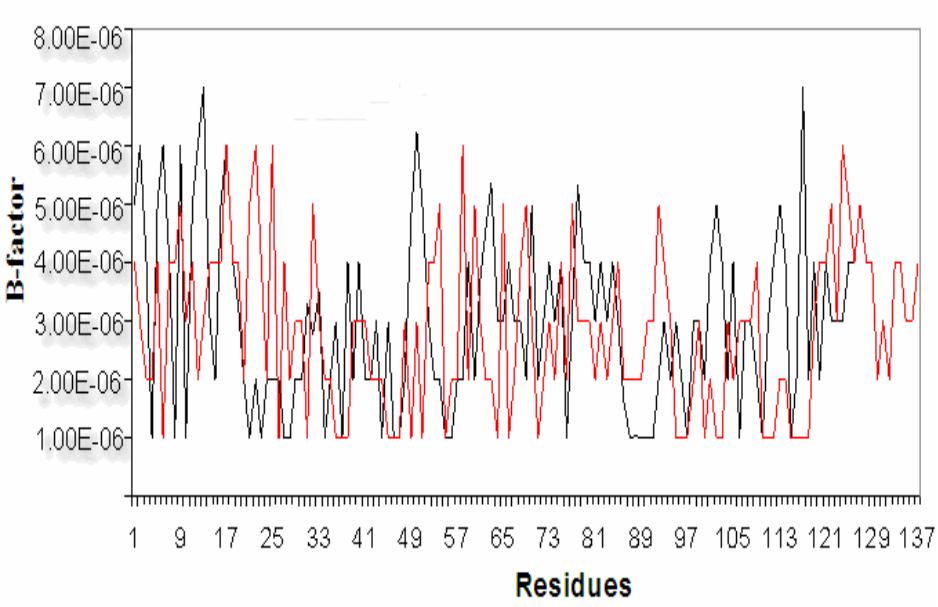


Figure 3. 15: The calculated B-factor from the simulation (Red; 25 ps, black; 10 ps).

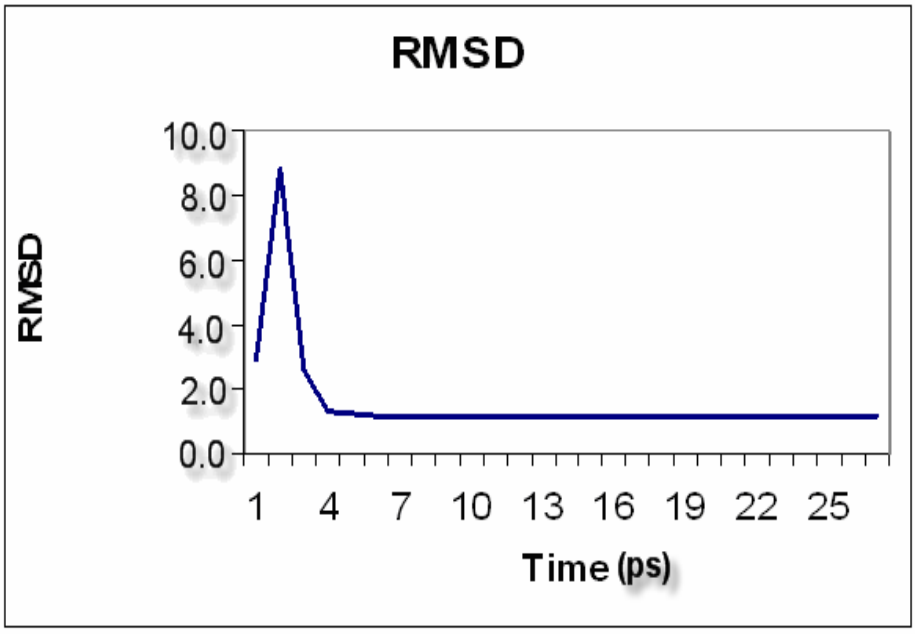


Figure 3. 16: Show the calculated RMSD of the molecular dynamics simulation.

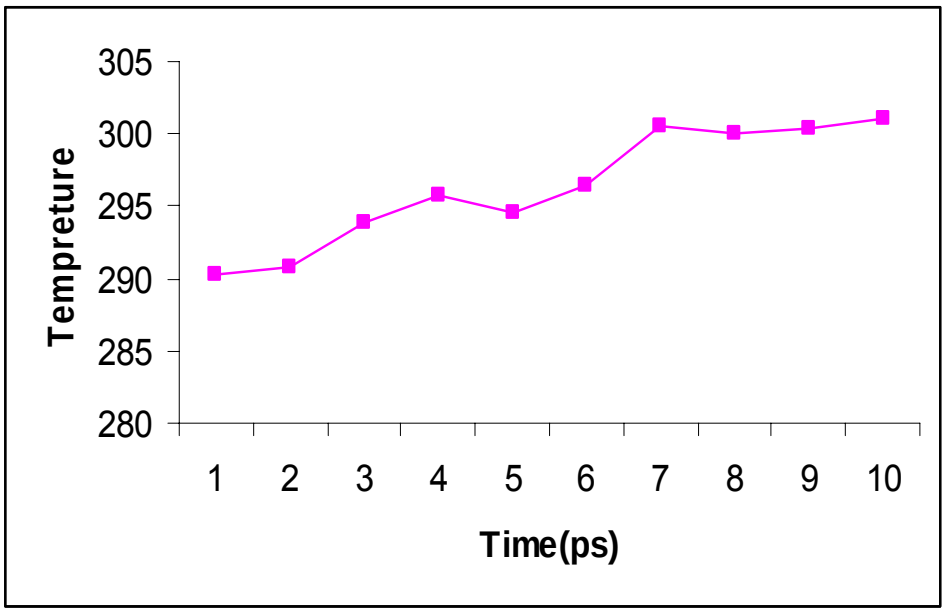


Figure 3. 17: Heating the water box and protein before starting the molecular dynamics simulation.

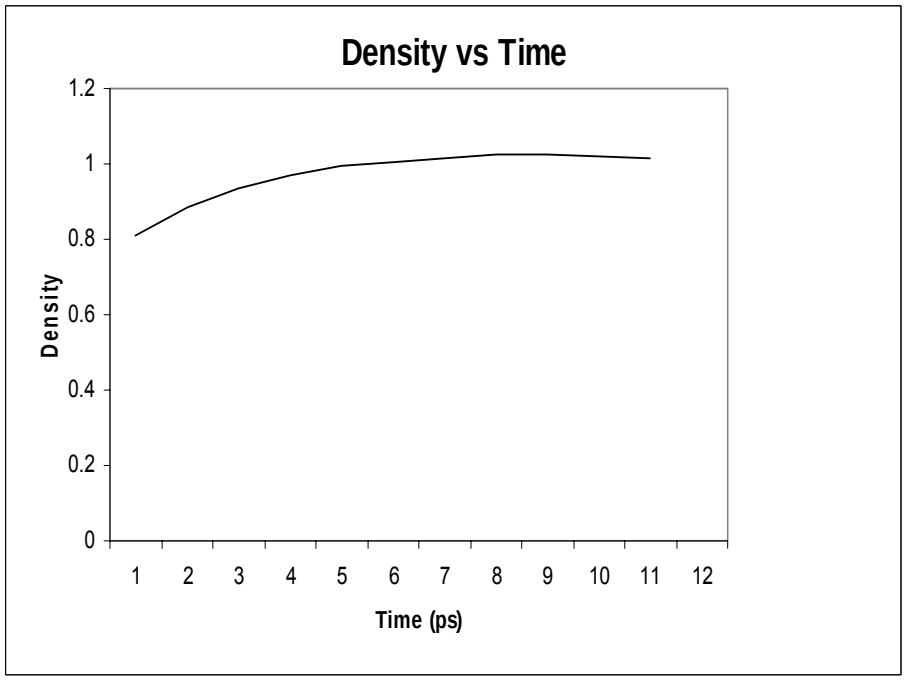


Figure 3. 18: Raising the density of the water box before starting the molecular dynamics simulation.

3.5 Metal ion binding to the DNA-Fur model

The monomer of Fur protein contains 12 Histidine residues and 4 Cystine residues. Some of these residues participate in the metal binding sites. The resultant model of DNA-Fur Fe^{+2} complex shows that Fe^{+2} mainly bind to residues close to the C-terminal. The nature of ligands provided by the Fur dimer to metal ion, and the number of metal ion sites, were always a matter of debate and it is worth the attention as it plays a key role in the whole process. There are two major sites provided by the fur dimer to Fe^{+2}: site 1, which involves Cys92 and Cys95 and other residues with N and O ligands (Table 3.5). Cys92 and Cys95 were always reported to play a crucial role in metal ion binding and fur function. An EXAFS analysis result [98] suggested a metal environment consisting of a total of 5 oxygen and nitrogen atoms at an average distance of 2.13 Å (either 2O at 2.05 Å/3N at 2.17 Å or 3O at 2.08 Å/2N at 2.19 Å). This structure indicates that Cys92 and Cys95 are present in close vicinity to the metal ion which makes them part of the binding site (Figure 3.19). Cys92 and Cys95 residues do not seem to participate directly in binding Fe^{+2} ions. The calculated distance between the Fe^{+2} and Cys92-Cys95 were 2.2 Å for Cys92, and 1.6 Å for Cys95. Probably bound through H-bonded H_2O intermediate or a protonated SH as indicated by the weak binding evident in the Mossbauer parameters for Fe^{+2} and the reported dissociation constant which ruled out a strong sulfur–Fe^{+2} [84]. The carboxylate group

of Asp108 participates directly in Fe (II) ions. It is found another Fe^{+2} is coordinate by the side-chains of residues His71 (end of β strands), Asp 105 (coil), Ala 109 (α helix), Asn72 (β starnd) and Ile50 (coil). The coordinate sphere of this site can be described as distorted octahedral (figure 3.20). Table 3.5 shows the calculated distance between the donor atom of these residues and Fe^{+2} ions, and it appears that His 71 plays an important binding role to the Fe^{+2} [60, 66]. Recent experiments suggest that apo fur possesses at least one Zn^{+2} ion in each monomer which coordinates with Cys92 and Cys95 residues while another metal sites contains the iron binding. We suggest that metal binding site one probably is the Zn^{+2} binding site while another site is coordinate of Fe^{+2} ions. Another suggestion for the Zn^{+2} binding site thought to involve the C-terminal Cys132 and Cys137 was not found to binds metal ions in our study [81]. The second site was found to involve His143, His145, Glu140, Arg139, Asp141, Asp137 and iron(II) present in distorted octahedral environment. Table 3.5 shows the calculated distances between the donors atom of these residues and iron(II). It also shows the position of residue in structure.

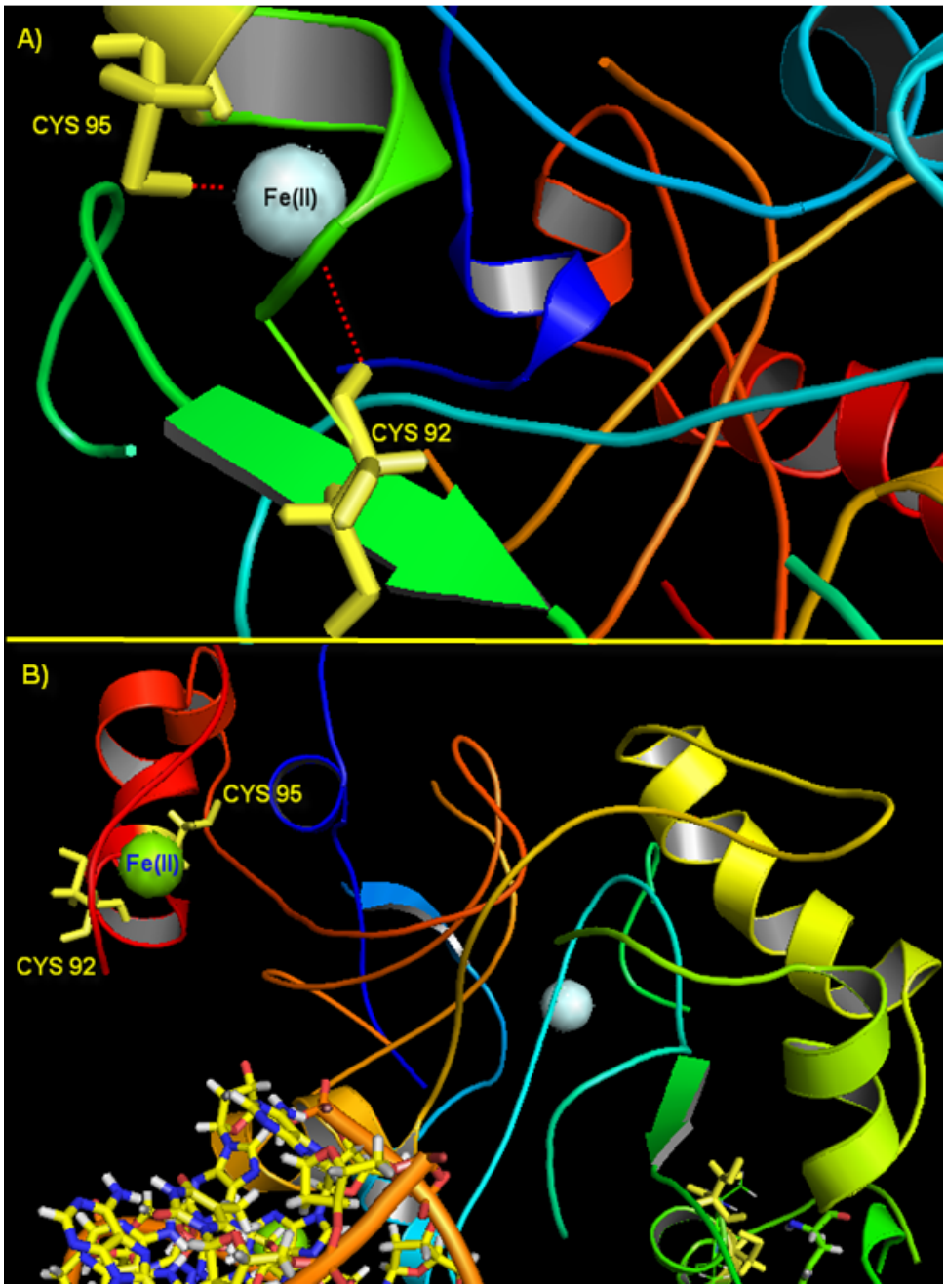


Figure 3. 19: Cys92 and Cys95 are also found in the metal binding pocket. A) Shows that Cys92 and Cys95 may provide by two of six ligands for. B) Shows that residue which lies between Cys92 and Cys95 play a role in the iron bind binding.

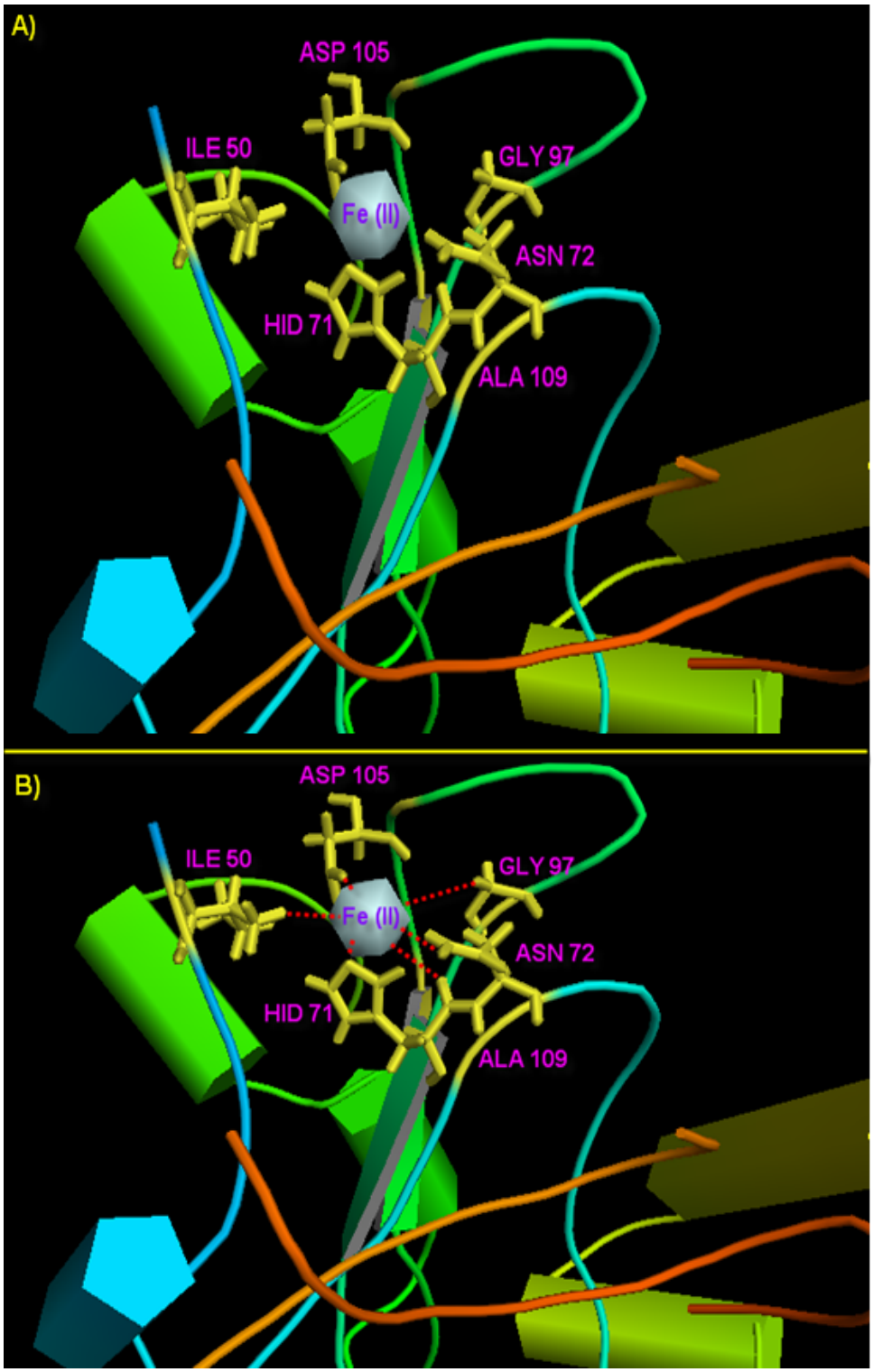


Figure 3. 20: One of the 12 Histadine found to be included in the iron ions binding site. A) Shows that His 71 (Pymole software used HID prefix) included in the metal binding domain. B) Ala109, Asn72, Gly97, Ile50 and Hid 71 performing the six coordinate binding

Table 3. 5: Calculated distances between Fe (II) and closest residues on the fur for the first two iron ions added.

Residue	Position of residue in structure	Donor atom	Residue-Fe(II) Distance (Å)
Fe2-His 71	strand β End of	N	1.3
Fe2-Ile 50	Coil	hydrophobic	2.3
Fe2-Asn 72	strand β	N	1.5
Fe2-Gly 97	Coil	polar	2.3
Fe2-Asp 105	Coil	O	1.4
Fe2-Ala 109	helix α	hydrophobic	2.1
Fe1-Asp 137	coil	O	1.3
Fe1-Asp 141	coil	O	1.5
Fe1-Arg 139	coil	N	1.7
Fe1-Glu 140	coil	O	1.3
Fe1 –His 145	coil	N	1.2
Fe1-His 143	coil	N	1.5
Fe1-Cys 92	coil	H-bonded H_2O	2.2
Fe1-Cys 95	coil	H-bonded H_2O	1.6
Fe-His 32	helix α	N	3.6
Fe-His 33	helix α	N	4.2
Fe –Arg 57	helix α		5.1
Fe –Gln 61	coil		4.9
Fe –Phe 62	coil		7.9
Fe –Ile 67	coil		8.3
Fe –Arg 70	coil		3.4
Fe –Phe 73	coil		3.1
Fe –Ile 114	coil		4.9
Fe –Ile 120	coil		6.2
Fe –His 132	helix α		5.4

Since the Fur dimer was proposed to change conformation upon metal binding, this was suggested to facilitate the DNA binding process. The Simulation was repeated in the presence of higher concentration of Fe (II) (up to 8 Fe (II)/dimer), in order to understand the effect of metal ion binding on producing Fur conformation changes. The calculated distances between residues of two monomers showed that Fur has undergone conformational changes of iron (II) binding (Table 3.6). It appears that the two monomers became closer to each other upon increasing iron as evident in (Figure 3.21). This finding confirms the role of mutation in initiating conformational changes on fur dimer which triggers the DNA binding process and this is the basis for the sensing repressing process of Fur. It is also evident that at high concentrations of Fe^{+2}, the N-terminal moved closer to the iron box as seen in Table 3.7 and Fig. 3.22 [87, 91]. Figure 3.23 shows a comparing of the two model of Fur-DNA at low (up to 4 Fe(II)/dimer) and high (up to 8 Fe(II)/dimer) concentration of metals ions. The two models show a different conformation with an increase in the affinity binding of the C-terminal and the N-terminal. This result agrees with what was proposed before [10, 55, 85]. The calculated RMSD between the two models is 2.1 Å, which indicates the occurrence of conformational changes. These results give insight as to how the Fur regulates the iron uptake process. N-terminal and C-terminal binding affinity to the DNA will increase with increasing iron concentration which will prevent further iron binding to the Fur dimer.

Table 3. 6: The calculated distances between residues on each monomer of the Fur dimer. The first apo-fur dimmer, 2nd column for apo-fur dimmer with DNA, the last two columns show the distances after adding Fe (II).

Residue	ApoFur dimer	ApoFur dimer/DNA	Fur dimer/DNA +4 Fe^{+2}	Fur dimer/DNA +8 Fe^{+2}
N-terminal-N-terminal	20.4	15.4	10.6	10.3
$\alpha_1 \rightarrow \alpha_1$	18.3	16.4	10.9	9.2
$\alpha_2 \rightarrow \alpha_2$	10.8	6.5	4.8	2.4
Val 25- Val 25	3.2	5.60	3.10	3.9
Pro29-Pro29	0.5	1.7	1.5	1.5
$\alpha_3 \rightarrow \alpha_3$	27.5	15.6	17.3	11.0
$\alpha_4 \rightarrow \alpha_4$	13.2	12.7	10.1	8.9
Leu52-Leu82	0.7	1.2	1.6	1.2
Gly51-Gln85	0.02	0.02	0.5	0.5
Glu49-Glu81	0.02	0.4	0.09	0.09
Thr54-Thr83	0.5	1.2	0.9	0.7
Glu49-Glu49	18.2	12.8	10.2	8.2
Thr69-Thr69	12.1	9.5	8.4	8.0
Gln85-Gln85	32.4	34.5	20.4	19.2
Ala53-Ile107	8.60	13.7	12.7	12.5
Thr54-Glu108	9.50	11.8	9.3	8.9
$\alpha_5 \rightarrow \alpha_5$	33.5	34.6	20.8	19.2
Arg112-Arg112	12.7	10.2	8.5	7.6
$\alpha_6 \rightarrow \alpha_6$	34.9	32.1	15.7	14.8
C-terminal-C-terminal	34.9	32.7	15.6	14.2

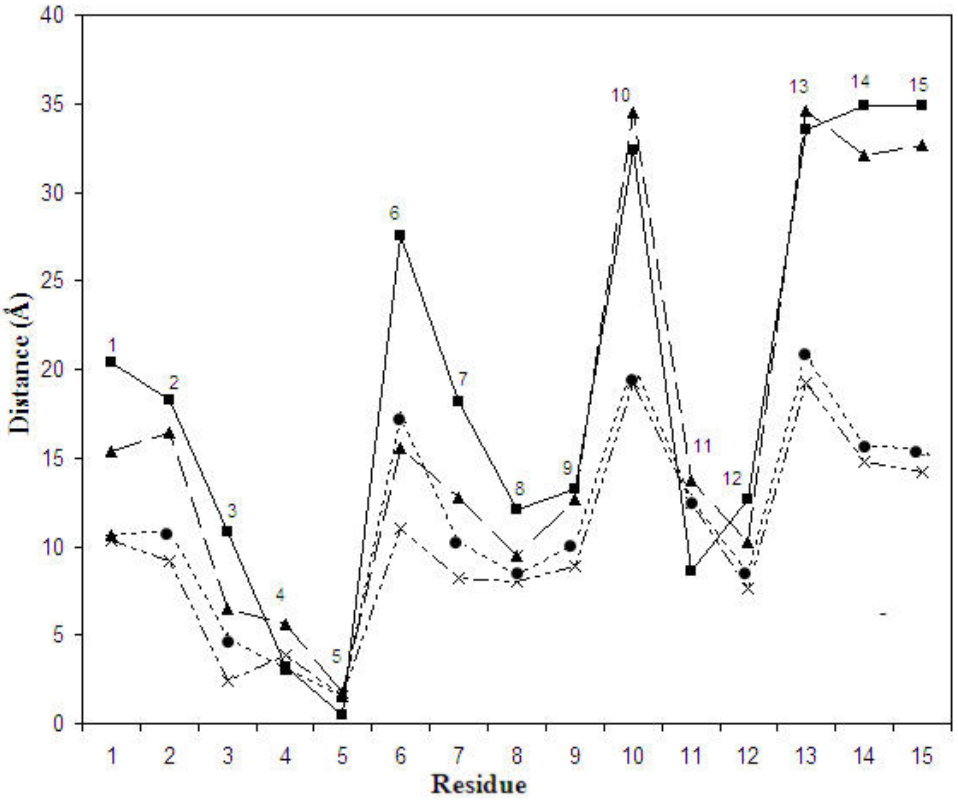


Figure 3.21: Conformational changes of the fur EC induced by DNA and Fe^{2+} binding. Distance between residues and helices on one fur subunit and the other. Apofur dimer (■). Apofur/DNA (▲). Fur/DNA in the presence of 4 Fe^{+2} ions (●) and Fur/DNA in presence of 8 Fe^{2+} ions (✕). Labels on the plots are as follows: N-terminal-N-terminal (1), α_1- α_1 (2), α_2- α_2 (3), Val25-Val25(4), Pro29-Pro29(5), α_3- α_3 (6), Glu49-Glu49 (7), Thr69-Thr69 (8), α_4- α_4(9), Gln85-Gln85(10), Ala53-Ile107 (11),; Arg112-Arg112 (12),; α_5- α_5(13), α_6- α_6(14), C-terminal-C-terminal(15).

Table 3. 7: Distances (in angstrom) between fur residues and AT of DNA. A) apofur dimer/ DNA (no iron present). B) Fur dimer /DNA + $4Fe^{+2}$, C) fur dimer /DNA + 8 Fe^{+2}.

Residue	(A) fur dimer/DNA	(B) fur dimer/DNA+ $4Fe^{+2}$	(C) fur dimer/DNA+ $8Fe^{+2}$
N-terminal	7.8	5.8	4.3
Ala 11	0.9	1.0	0.8
Gly 12	0.8	0.7	0.5
Leu 13	0.7	0.7	0.4
Pro 18	1.3	1.3	1.6
Arg 19	7.4	6.9	6.4
His 32	8.6	7.8	7.7
His 33	8.5	7.5	6.8
Arg 57	7.5	5.4	4.3
Gln 61	11.2	9.9	9.6
Phe 62	10.2	8.3	7.3
Ile 67	11.2	8.2	7.4
Arg 70	19.3	17.5	16.6
Phe 73	9.6	8.4	7.4
His 86	3.4	2.6	1.9
His 87	4.1	2.8	2.3
His 88	3.7	2.1	1.8
Arg 112	34.5	30.3	28.9
Ile 114	27.3	25.9	25.2
Ile 120	23.00	20.3	19.4
His 125	32.8	30.6	21.3
Gly 131	29.5	27.2	27.5
His 132	8.9	4.5	3.2
Asp 137	4.2	2.3	1.9
Arg 139	4.9	2.3	2.1
Glu 140	4.2	3.2	2.2
Asp 141	5.1	2.5	1.6
His 143	4.5	2.7	1.8
His 145	5.3	3.1	1.7
C-terminal	24.5	20.4	17.5

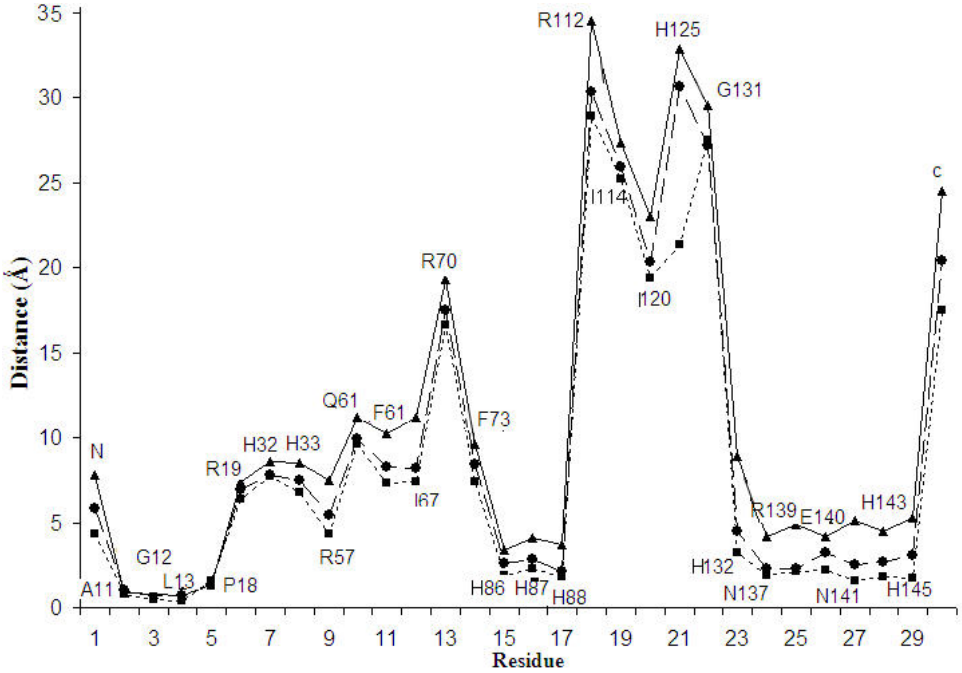


Figure 3.22: Conformational changes of the FurEC dimer And DNA binding. The Calculated distances between the amino acid residues of fur and the AT-unit in the B-canonical DNA (Table 3.7). Fur dimer and DNA fragment (▲) (continuous line). Fur dimer and DNA in the presence of 4 Fe^{+2} ions (●) (broken line). Fur dimer and DNA in presence of 8 Fe^{+2} ions (■) (dotted line). This plot show that residues A11, G12, L13 P18 and R19 near the N-terminal, His88 to R112, and the residues139—145 near the C-terminal are the closest to DNA.

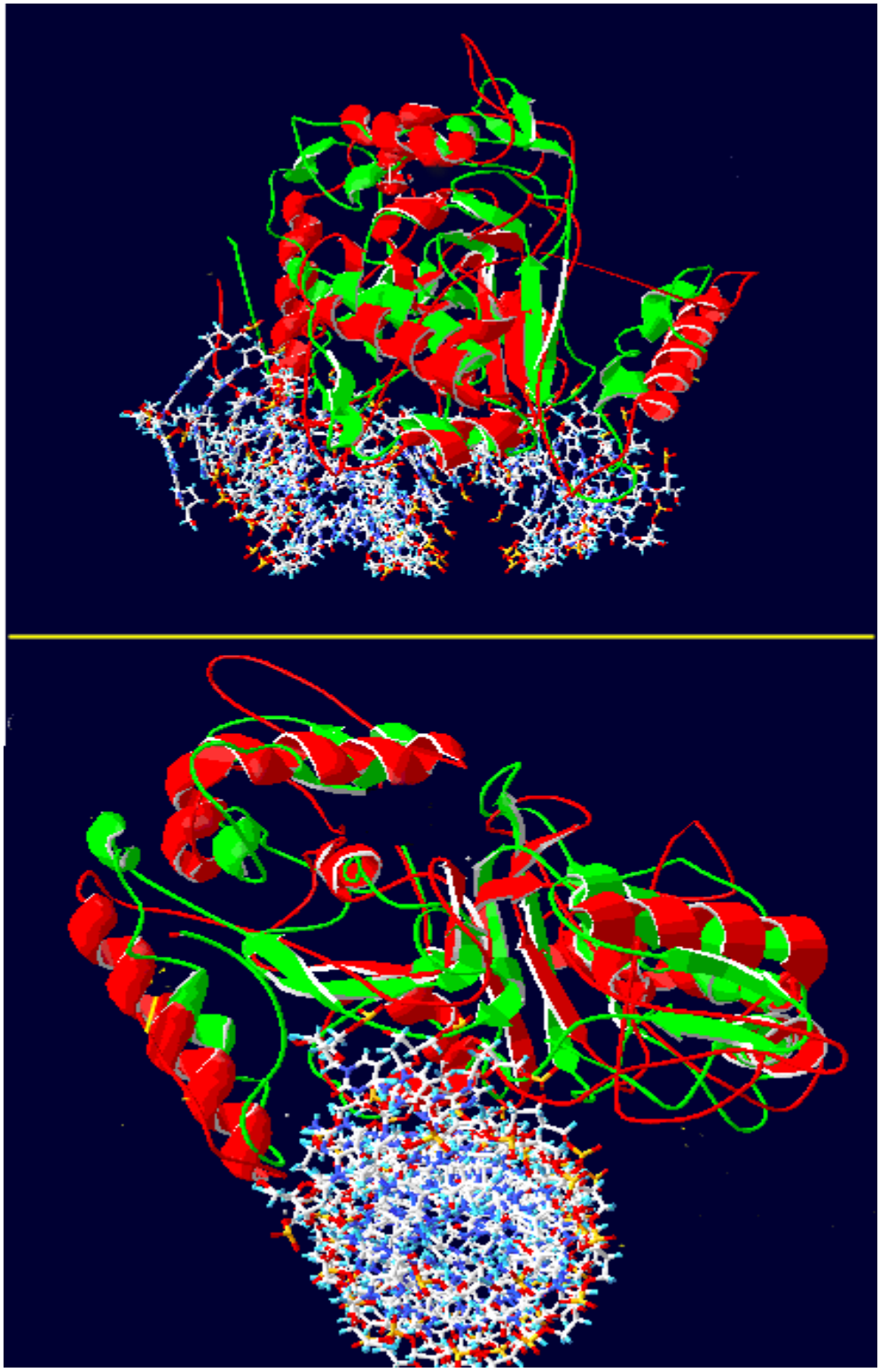


Figure 3. 21: The effect of iron concentration on the conformational changes of Fur. The green model shows Fur conformational at high concentration while red model at low concentration. The affinity binding of green model is much larger than the red model (red: low concentration, green: high concentration).

It was proposed that Lys76 and Lys 45 play a role in the iron box binding. According to our investigation, Lys76 and Lys45 do not appear to be close to the iron box or to the dimer domain. It was found that these two residues have high surface accessibility to solvent. VADAR and Phd server result agrees with our findings. Figure 3.24 show the calculated accessibility surface of the Fur amino acids to the solvent. It is clear that Lys76 and Lys45 have high values, i.e. they are easily exposed.

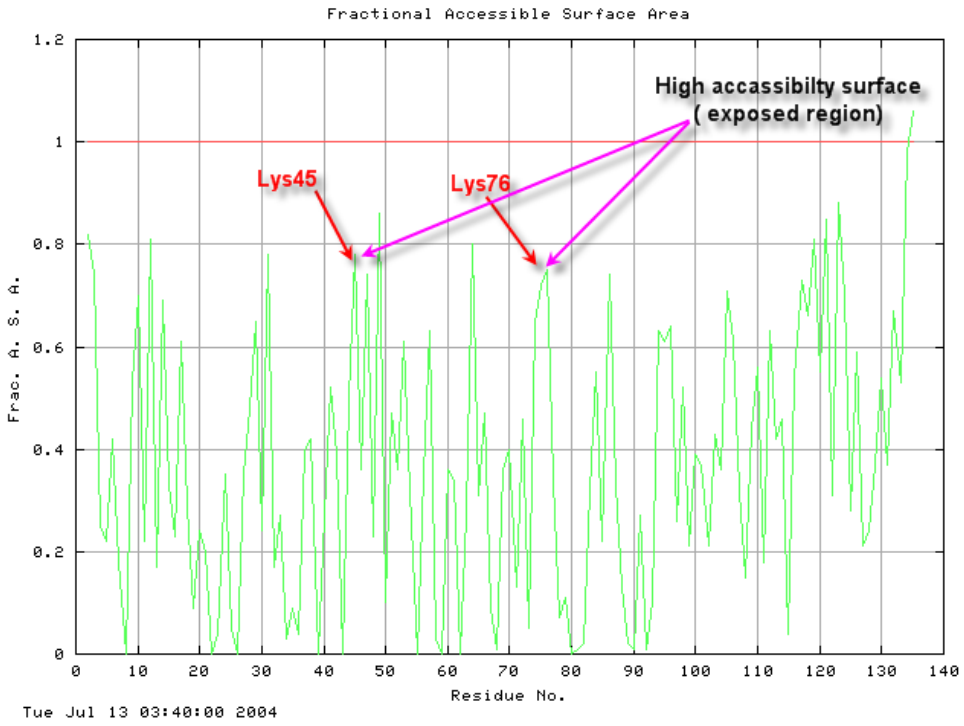


Figure 3. 22: The Fractional accessible surface area of the Fur amino acid sequence.

The His86-88 near the C-terminal was found to be close to the iron box as a result of increasing iron metals concentration. Table 3.7 shows the effect of the iron concentration on the binding of the His86-88. C-terminal and His86-88 which included in the C-terminal play a role in the

regulation of further iron metal uptake. This happens as a result of increasing binding affinity denotes an en increase in the iron concentration. The model shows that other Histidine residues play a role in regulation of further uptake. The His 143 and His 145 were found to be close enough to the iron box when the iron concentration became high (see Table 3.7). The distances between the iron box and Asp137, Asp 141, Arg 139 and Glu 140 decrease as a function of increasing the iron metal concentration. All these results agree with what was proposed before that the C-terminal residue plays a role in iron binding and also increase the binding affinity to the iron box as a result of increasing the iron metal concentration. This function of C-terminal prevents further uptaking of iron metal.

Chapter 4

Conclusion

The method of comparative protein modeling has been shown to effectively predict the structure of ferric uptake regulation protein. The comparative protein modeling was used to model the structure of Fur. Fur was found to consist of three domains: N-terminal, central and C-terminal domains. The prediction of conformational internal residues by molecular mechanics energy minimization helps to predict the tertiary structure. Internal residues form a core which defines the overall structure of the ferric uptake regulation protein (Fur) and thus allows a homologous protein to be aligned. This core gave a higher sequence identity and a lower RMSD for residues. This core of the ferric uptake regulation protein was used to predict its tertiary templates, which was used to determine which sequence can fit a given structural class.

The comparative protein modeling has been sufficient to predict the 3D structure of the ferric uptake regulation protein (Fur) with good accuracy. The main reasons for this accuracy are the increases in the numbers of known folds and the structure per fold family as well as the improvement in the fold recognition and comparative modeling techniques. The three structures of ferric uptake regulation protein (Fur) dimer reveal that each subunit consists of three domains, with a core

formed by the first two domains, which have distinctly different functions. The amino-terminal (aa1-aa76) domain is responsible for DNA-binding. The resultant model shows that DNA-Fur recognition which may be attributed to specific interaction between the residues of the recognition helix of the Fur and the exposed major groove of the DNA (iron box). The present study shows that Fur like others regulation proteins has an HTH domain interact in adjacent groove of the iron box. Fur shows interaction symmetric AT-AT unit in the iron box, which was proposed to be the best binding region. The extensive hydrophobic properties of Valine (Val 18), Leucine (Leu13) and Alanine(Ala11) residues shows hydrophobic interaction between the surface of the Fur and the edges of the bases of the suger-phosphate backbone in the groove of the iron box. While the hydrogen and a salt bridge with sugar-phosphate backbone inetraction occurs through an aromatic cycle of Proline (Pro 18) residue. During the docking process it's found that dimerization sites occur in central domain of the N-terminal. The residues from aa25-aa29 in the α-helix region near the N-terminal domain are responsible for the dimerization. The iron metal binds due His 71, Cys92 and Cys95 which presents in the C-terminal domain.

Our result shows that Fur undergoes conformational changes during the metal interaction. These conformational changes increase the affinity of interaction between Fur and DNA. Also the results show that when the metal concentration increase the affinity binding increase as shown in the results. The conformational changes in the C-terminal

reported to be higher than these in N-terminal. This explain how the Fur regulate the iron concentration in the bacteria's bodes.

References

[1] Schluz, G.E., Schirmer, R.H. **1978**. *Principles of protein structure.* New York: Springer Verlag.

[2] Stryer, L. **1988**. *Biochemistry*. 3Ed. New York: Freeman. 211.

[3] Zubay, G. **1993**. *Biochemistry*. 3Ed. Wm. C. Brown Publishers.

[4] Van Nuland, N., Hangi, I.W., van Schaik, R.C., Berendsen, H.J.C., van Gun Stteren, W.F., Scheek, R.M. and Robillard, G.T. **1994**. The high-resolution structure of the histidine-containing phosphor carrier protein HPr from Escherichia coli determined by restrained molecule dynamics from NMR nuclear over hauser effect data. *J. Mol. Biol*. 237:544-559.

[5] Kendrew, J.C. et al. **1958**. A three-dimensional Model of the Myoglobin Molecule Obtained by X-ray Analysis. *Nature* 181, 662-666 (1958).

[6] Ramachandron G.N., Sasisekharan, V. **1968**. Conformation of Polypeptides and Proteins. *Adv. Prot Chem.* vol 23, 283-437.

[7] Anfinsen, C.B. **1973**. Principles that govern the folding of protein chains. *Science* 181:223-230.

[8] Levinthal, C. **1968**. Are there pathways for proteins folding? *J. Chim. Phys*. 65:44- 45.

[9] Weiner, S.J., Kollman, P.A., Nguyen, D.T., Case, D.A. **1986**. An all atom force field for simulation of proteins and nucleic acids. *J. Comp. Chem.* 7:230-252.

[10] Pauling, L., Corey, R.B., & Branson, H.R. **1951**. Two Hydrogen-Bonded Helical Configurations of the Polypeptide Chain. *Proc. Natl. Acad. Sci. U.S.* 37, 205-211.

[11] Linderstrom-Lang, K.U., Shellman, J.A. **1959**. *Protein structure and enzyme activity*. The Enzymes, (P.D. Boyer, Ed.), Vol 1, 2nd ed., pp. 443-510. Academic Press, New York.

[12] David van der S. **1996**. *Structure and Dynamics of peptides: Theoretical Aspects of Protein Folding*. PhD Thesis, Univ. of Groningen.

[13] Burley, S.K., Petsko, G.A. **1989**. Electrostatics interaction aromatic oligopeptides contribute to protein stability. *Trends Biotech*, 7:354-359.

[14] Champan, M. **2002**. Protein Tertiary Structure, Molecular Mechanics & Dynamics. http://bio.fsu.edu/~stevet/BSC5936/Chapman.2002.pdf (accessed October 12, 2003).

[15] Atkins, P.W. **1990**. *Physical chemistry*. Fourth edition Ed. Oxford, UK: Oxford University Press.

[16] Besler, B.H., Merz Jr., K.M. and Kollman, P.A. **1990**. Atomic charges derived from semi empirical methods. *J. Comp. Chem*. 11:431-439.

[17] Weiner, S.J., Kollman, P.A., Nguyen, D. T. and Case, D.A.A. **1986**. All atoms force field for simulation of protein and nucleic acids. *J. Comp. Chem.* 7:230-252.

[18] Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. and Karplus, M. **1983**. CHARMM: a protein for macro-molecular energy, minimiziation, and dynamics calculation. *J. Comp. Chem*. 4:187-217, 1983.

[19] van der Spoal, D., van Buuren, A.R., Apol, E., Meulenho R.P.J., Tieleman, D. P., Sijbers, A.L., van Drunen, R. and Berendsen, H.J.C. **1996**. Gromacs User Manual version 1.3. Nijenborgh 4, 9747 AG. Groningen, Netherlands. http://rugmd0.chem.rug.nl/gmx

[20] Jorgensen, W.L., Tirado-Rives, J. **1988**. The OPLS potential functions for proteins. Energy minimizations for crystal of cyclic peptides and crambin. *J. Am. Chem. Soc*. 110-1657-1666.

[21] Laxmikant, K., Robert, S., Milind, B., Robert, B., Attila, G., Neal, K., James, P., Aritomo, S., Krishnan, V., and Klaus, S. **1999**. NAMD2: Greater scalability for parallel molecular dynamics. *Journal of Computational Physics*, 151:283-312.

[22] Levitt, M. **1976**. A simplified representations of protein conformations for rapid simulation of protein folding. *J. Mol. Biol.* 104, 59-107.

[23] Karplus, M., Shakhnovich, E. **1992**. Protein folding: Theoretical studies of thermodynamics and dynamicas. In: Protein Folding.Creighton, *T.E. ed.. Freeman* 127-196.

[24] Lathrop, R.H., Smith, T.F. **1996**. Global optimum protein threading with gapped alignment and empirical pair score functions. *J. Mol. Biol*. 255:641-666.

[25] Watson, J.D., Crick, F.H.C. 1953. Molecular structure of Nucleic Acids. Nature. 171: 737-738.

[26] Watson, J. D., Crick, F. H. C. **1953**. A Structure for Deoxyribose Nucleic Acids.

[27] Lathrop, R.H., Smith, T.F. **1996**. Global optimum protein threading with gapped alignment and empirical pair score functions. *J. Mol. Biol*. 255:641-666.

[28] Schmid, F.X. **1992**. Kinetics of unfolding and refolding of single-domain proteins. In: Protein Folding. Creighton, *T. E. ed.. Freeman* 197-241.

[29] Brooks III, C.L. **1993**. Molecular simulation of peptide and protein unfolding: in quest of a molten globule. *Curr. Opin. Struct. Biol.* 3:92-98.

[30] Dill, K.A., Fiebig, K.M., Chan, H.S. **1993**. Cooparativity in protein-folding kinetics. *Proc. Natl. Acad. Sci. USA* 90:1992-1946.

[31] Ohgushi, M., Wada, A. **1983**. Molten-globule state: a compact form of globuler proteins with mobile side-chains. *FEBS lett.* 164: 21-24.

[32] Levitt, M., Warshel, A. **1975**. Computer simulation of proteins folding. *Nature 253:694-698.*

[33] Privalov, P.L. **1992**. Physical basis of the stability of folded conformation of proteins. In: Protein foldin Creighton. *T. E. edn. Freeman* 83-126.

[34] Baldwin, R.L. **1996**. Why is protein folding so fast? *Proc. Natl. Acad. Sci. USA* 93: 2627- 2628.

[35] Fedorov, A.N., Baldwin, T.O. **1995**. Contribution of constranlational folding to the rate of formation of native protein structure. *Proc. Natl. Acad. Sci. USA* 92:1227-1231.

[36] Frydman, J., Nimmergern, E., Ohstuka, K. and Hartl, F.U. **1994**. Folding of nascent polypeptide chains in a high molecular mass assembly with molecular chaperones. *Nature* 370:11-117.

[37] DeBrunner, J.T.P., Munck, E. **1969**. Mossbauer Spectroscopy in Biological Systems: Proceedings of a meeting held at Allerton House, Monticello, Illinois. University of Illinois Press, 22-24.

[38] Dill, K.A. **1993.** Folding proteins: finding a needle in a haystack. *Curr. Opin. Struct. Biol*. 3:99-103.

[39] Kolinski, A., Skolnick, J. **1994**. Monte Carlo simulations of protein folding. II. Application to protein A, ROP, and crambin. *Proteins: Struct. Funct. Genet.* 18, 353-366.

[40] de Araújo A.F.P. **1999**. Folding protein models with a simple hydrophobic energy function: The fundamental importance of monomer inside/outside segregation. *The National Academy of Sciences: Proc. Natl. Acad. Sci. U S A.* 96(22): 12482–12487.

[41] Vieth, M., Kolinski, A., Brooks III and Skolnick, J. **1994**. Prediction of the folding path-ways and structure of the GCN4 leucine zipper. *J. MOL. Biol.* 237: 361-367.

[42] Srinivasan, R., Rose, G. D. **1995**. LINUS: A hierarchic procedure to predict the fold of a protein. *PROTEINS: struct. Funct. Gen.* 22:81-99.

[43] Pearlman, D.A., Case, D.A., Caldwell, J.W., Ross, W.S., Cheatham, T.E., Simmerling, C.L., Darden, T.A., Merz, K.M., Stanton, R.V., Cheng, A.L., Vincent, J.J., Crowley, M. Tsui, V., Gohlke, H., Redmer, R.J., Duan, Y., Pitera, J., Massova, I., Seibel, G.L. Singh, U.C. Weinner P.K. and Kollman, P.A. **2002**. AMBER7, University of caliornia, San Francisco. http://amber.scripps.edu

[44] Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G. Profeta, S. and Weiner, Jr.P. **1984**. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* 106, 765-784.

[45] Pearlman, D.A., Case, D.A., Caldwell, J.W., Ross, W.S., Cheatham, T.E., DeBolt, S., Ferguson, D., Seibel, G.L. and Kollman, P.A. **1995**. AMBER, a package of computer programs for 90 applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comp. phys. Commun.* 91, 1-41.

[46] Frenkel, D., Smit, B. **1996**. Understanding Molecular Simulation: From Algorithms to Applications. San Diego: Academic Press.

[47] Garrett, M.M., David, S.G., Ruth, H., William, E.H., Scott, H., Rik, B. and Arthur J.O. 2002. Automated Docking of Flexible Ligands to Receptors (Autodock). Molecular Graphics Laboratory, Department of Molecular Biology, 10550 North Torrey Pines Road, La Jolla, CA 92037-1000, Version 3.0.5. **http://www.scripps.edu/pub/olson-web/doc/autodock/**

[48] Christopher, D.R., Richard, K.B., Halliday, R.S. and William, E.H. **1997**. A Comparison of Global and Local Search Methods in Drug Docking. Morgan Kaufmann *Publishers, Inc., San Francisco, CA.* 221, 8.

[49] Benoît, D.A., Danièle, T., Beate, B., Jean, M.L., and Isabelle, M.S. **2002**. Direct inhibition by nitric oxide of the transcriptional ferric uptake regulation protein via nitrosylation of the iron. *The National Academy of Sciences Proc. Natl. Acad. Sci. U S A.* 99(26): 16619–16624.

[50] Teresa, E.C., Leslie, W.T. and Hans, J.V. **2001**. Structural Biology of Bacterial Iron Uptake Systems. *Current Topics in Medicinal Chemistry*, 1(1).

[51] Litwin, C.M., Calderwood, S.B. **1993**. Role of iron in regulation of virulence genes. *Clin. Microbiol Rev*. 6(2): 137–149.

[52] Jennifer, L.L., Mark, A. M. **2003**. *Architecture of a Fur Binding Site: a Comparative Analysis. American Society for Microbiology, J Bacteriol.* 185(7): 2194–2202.

[53] Hunt, M.D., Pettis, G.S. and McIntosh, M.A. **1994**. Promoter and operator determinants for Fur-mediated iron regulation in the bidirectional *fepA-fes* control region of the *Escherichia coli* enterobactin gene system. *J. Bacteriol.* 176:3944-3955.

[54] Stojiljkovic, I., Bäumler, A.J., and Hantke, K. **1994**. Fur regulon in Gram-negative bacteria: identification and characterization of new iron-regulated Escherichia coli genes by a Fur titration assay. *J. Mol. Biol.* 236:531-545.

[55] Noel, B., Helmann, J.D. **2002**. Recognition of DNA by Fur: a Reinterpretation of the Fur Box Consensus Sequence. *American Society for Microbiology. J. Bacteriology*, 5826-5832.

[56] *Frishman, D., Argos, P. **1996.** Incorporation of non-local interactions in protein secondary structure prediction from the amino acid sequence.* Protein Engineering*, Vol 9, 133-142.*

[57] *Kolade, O.O., Bellini, P., Wexler, M., Johnston, A.W., Grossmann, J.G. and Hemmings, A.M. **2002.** Structural studies of the Fur protein from Rhizobium leguminosarum.* Biochem Soc Trans. 30(4):771-4.

[58] Saito, T., Robert, J.P. Williams. **1990**. The Binding of the ferric uptake regulation protein to a DNA fragment. Eur. J. Biochem. 197, 43-47.

[59] Smith, A., Hooper, N.I., Shipulina, N., Morgan, W.T. 1996. Heme binding by a bacterial repressor protein, the gene product of the ferric uptake regulation (fur) gene of Escherichia coli. J. Protein Chem. 15(6):575-83.

[60] Coy, M., Neilands, J.B. **1991**. Structural dynamics and functional domains of the fur protein. Biochemistry. 30 (33):8201-8210.

[61] Allison, C.L., Phillip, A.D., Lynda, F., Geoffrey, R.M. and Stephen, S. 2002. The ferric uptake regulator of Pseudomonas aeruginosa has no essential cysteine residues and does not contain a structural zinc ion. *Microbiology*, **148**, 2449-2456.

[62] Mark, C., Sofia, E., Dilek, Tezcan, M., Jay, C.D.H., and Mikael, R. **2001**.Virulence gene regulation in Salmonella enterica. *The Finnish Medical Society Duodecim, Ann Med,* 33: 178–185.

[63] Erin, R.M., Amy, D., Kevin, T.M., and Terry, D.C. **1999**. Genetic Characterization of Wild-Type and Mutant *fur* Genes of Bordetella avium. *Infection and Immunity*, 6(67) 3160-3165.

[64] Schaffer, S., Hantke, K. and Braun, V. **1985**. Nucleotide sequence of the iron regulatory gene fur. *Mol. Gen. Genet.* 200(1):110-3.

[65] Lilian, J., Daniel, A., Annie, A., Jean-Louis, H., Jean-Marc, L. and Isabelle, M.S. **1998**. X-ray absorption spectroscopy of a new zinc in the Fur protein from Escherichia coli. *Biochemistry*, 37, 2564-2571.

[66] Lilian, J., Francois, D., Claudine, J., Jean-Louis, O., Emmanuel, P., Laurent, L.P., Daniel, A. and Jean-Louis, H. **2000**. First spectroscopic characterization of Fe^{II}-Fur, the physiological active form of the Fur protein. *J. Am. Chem. Soc.* 122:394-395.

[67] Yali, E.F., Mark, R.O.B. **2004**. The Ferric Uptake Regulator (Fur) Protein from Bradyrhizobium japonicum. Is an Iron-responsive Transcriptional Repressor *in Vitro. J. Biol. Chem.* 279(31): 32100-32105.

[68] Goragot, W., Randall, K.H. and Wim, G.J.H. **2004**. Crystal structure of an IdeR-DNA complex reveals a conformational change in activated IdeR for base-specific interactions. *J. Molecular Biology*. 342(4): 1155-1169.

[69] Narayanan, E., Bino, J., Nebojsa, M., Andras, F., Valentin, A.I., Ursula, P., Ashley, C.S., Marc, A.M., Madhusudhan, M.S., Bozidar, Y. and Andrej, S. **2003**. Tools for comparative protein structure modeling and analysis. *Nucleic Acids Research,* Vol. 31, No. 13 3375-3380. < http://nar.oupjournals.org/cgi/content/full/31/13/3375>.

[70] Torsten, S., Manuel, C.P. and Nicolas, G. SWISS-MODEL: An Automated Comparative Protein Modeling Server. Biozentrum, University of Basel, Switzerland. [Online] Available. http://swissmodel.expasy.org//SWISS-MODEL.html (accessed April, 2002)

[71] Joan, G., Andrew, K. and Frances M. **1992**. BLAST: Sequencing Projects Blast Search Services. Sanger Institute. http://www.sanger.ac.uk/Info/

[72] Brunak, S., John, S., Anna, T., Martin, V. and Gunner, H. FASTA: Sequence Comparison. European Bioinformatics Institute (EBI). http://www.ebi.ac.uk/fasta/

[73] Andrej, S., John, P.O. **1994**. Derivation of rules for comparative protein modeling From a database of protein structure alignments. Cambridge University Press**.** Protein Science, 3:1582-1596.

[74] Torsten, S., Jürgen, K., Nicolas G. and Manuel, C.P. **2003**. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Research*, Vol. 31, No. 13 3381-3385. http://nar.oupjournals.org/cgi/content/full/31/13/3381

[75] Nicolas, G., Manual, C.P. Principles of Protein Structure, Comparative Protein Modelling and Visualization. Glaxo Wellcome Experimental Research S.A. 16, chemin des Alux. 1228 Plan-les-Ouates. Switzerland.

[76] Steve, P., Kim, G., Patrick, M. and Jesse, M. CPH models. The Biomolecular Computing Resource (BIMCORE), Emory University. http://sleepy.bimcore.emory.edu/home/

[77] Shindyalov, I.N., Bourne, P.E. **2000**. Improving alignments in HM protocol with intermediate sequences. In: Forth meeting on the critical assessment of techniques for protein structure prediction. p. A-92. http://www.bimcore.emory.edu/home/Links/CPHModel.htm

[78] Wishart, D.S., Willard, L., Richards, F.M. and *Sykes*, B.D. **1995**. VADAR: A Comprehensive Program Suite for Protein Structural Analysis, PENCE / CIHR-Group Joint Software Centre. Version: 1.4 - Dec, 2002. http://www.pence.ca/software/vadar/latest/vadar.html

[79] Eswar, N., Ursula, P., Roberto, S. and Andrej S. ModWeb: Server of Comparative Protein Structure Modeling. Departments of Biopharmaceutical Sciences and Pharmaceutical Chemistry, and California Institute for Quantitative Biomedical Research University of California San Francisco. *Mission Bay Genentech Hall, Suite N472D, San Francisco,* CA 94143-2240.

[80] Arthur, J.O., MGL: Molecular Graphic Laboratory. Molecular Graphics Laboratory Department of Molecular Biology MB-5. The Scripps Research Institute. 10550 North Torrey Pines Road La Jolla, CA 92037-1000 USA.

[81] Hamed, Y.M. **1993**. *Binding of the ferric uptake regulation protein (fur) Fe(II) complex to DNA.* Journal of Inorganic Biochemistry. 51(594):1-2.

[82] Zheleznova, E.E., Crosa, J.H. and Brennan, R.G. **2000**. *Characterization of the DNA- and Metal-binding Properties of Vibrio anguillarum Fur Reveals Concervation of a Structural Zn^{+2} Ion*. Journal of Bacteriology, 6264-6267.

[83] Bagg, A., Neilands, J.B. **1987**. *Ferric uptake regulation protein acts as a repressor, employing iron(II) as a cofactor to bind the operator of an iron transport in Escherichia coli.* Biochemistry, 26: 5471-5477.

[84] Hamed, M.Y., Neilands, J.B. and Huynh,V. **1993**. *Binding of the ferric uptake regulation repressor protein (Fur) to Mn(II), Fe(II), Co(II), and Cu(II) ions as co-repressors: Electronic absorption, equilibrium, and ^{57}Fe Mössbauer studies*. Journal of Inorganic Biochemistry, 50 (3): 193-210.

[85] Qiu, X., Verlinde, L.M.J, Zhange, S., Schmitt, P.M., Holmes, R.K and Hol, W.G.J. **1995**. *Three dimensional structure of the diphtheria toxin repressor in complex with divalent cation co-repressors*. Current Biology Ltd., 3:87-100.

[86] Althaus, W.E., Outten, E.C, Olson, K.E., Cao, H. and O'Halloran, T.V. **1999**. *The Ferric Uptake Regulation (Fur) Repressor Is a Zinc Metalloprotein.* Biochemistry, 38, 6559-6569.

[87] Escolar, L., Perez-Martin, J. and de Lorenzo, V. **1998**. *Binding of the Fur (ferric uptake regulator) repressor of Escherichia coli to arrays of the GATAAT sequence*. Journal od Molecular Biology, 283(3): 537-547.

[88] Hamed, M.Y., Neilands, J. B. **1994**. *An electron spin resonance study of the Mn(II) and Cu(II) complexes of the Fur repressor protein*. Journal of Inorganic Biochemistry, 53(4): 235-248.

[89] Barondeau, D.P., Getzoff, E.D. **2004**. *Structural insight into protein-metal ion partnerships*. Current Opinion in Structural Biology, (Article in Press, Corrected Proof-Note to users).

[90] Jacquament, L., Dole, F., Jeandey, C., Oddou, J.L., Perret, E., La Pape, L., Aberdam, D., Hazemann, J.L., Michaud-Soret, I. and Latour, J.M. **2000**. *First spectroscopic characterization of Fe(II)-Fur, the Physiological active form of the Fur protein*. J. Am. Chem. Soc, 122: 394-395.

[91] de Peredo, A.G., Saint-Pierre, C., Latour, J.M., Michaud-Soret, I. and Forest, E. **2001**. *Conformational changes of the ferric uptake regulation protein upon metal activation and DNA binding; First evidence of structural homologies with the Diphtheria toxin repressor.* J. Mol. Biol., 310: 83-91.

[92] Lavrrar, L.J., McIntosh, M.A.**2003**. *Architecture of a Fur Binding Site: a Comparative Analysis*. Journal of Bacteriology. 185(7): 2194–2202.

[93] Escolar, L., Pérez-Martín, J. and de Lorenzo, V. **2000**. *Evidence of an Unusually Long Operator for the Fur Repressor in the Aerobactin Promoter of Escherichia coli.* J. Biol. Chem., 275(32): 24709-24714.

[94] PROF server- Secondary Structure Prediction System. University of Wales, Aberystwyth Computational Biology Group. Department of Computer Science, Aberystwyth SY23 3DB, Wales, UK.

[95] DeLano WL: The PyMol Molecular Graphics System on the World Wide Web. 2002 [http://www.pymol.org]

[96] Rost,B., Sander,C. and Schneider,R. **1994**. PHD—an automatic mail server for protein secondary structure prediction. *Comput. Appl. Biosci.* **10:** 53–60.

[97] Villeret, V., Tricot, C., Stalon, V. and Dideberg, O. **1995**. Crystal Structure of Pseudomonas aeruginosa Catabolic Ornithine Transcarbamoylase at 3.0-Å Resolution: A Different Oligomeric Organization in the Transcarbamoylase Family. National Academy of Sciences. 92: 10762-10766.

Appendix I
PDB Format Description Version 2.2

1- Symmetry Operations

The data type ***SymOP*** is used to succinctly describe crystallographic symmetry operations that may be performed on ***ATOM/HETATM*** coordinates. Symmetry operators applicable to a given entry are presented in ***REMARK 290***. Each operator is assigned a serial number. The ***SymOP*** is a number of up to six (6) digits that indicates the serial number of the symmetry operator and the cell translations along the x, y, and z axes.

The ***SymOP*** data type is of the form ***nnnMMM*** where 'n' is the serial number of the symmetry operator, and ***'MMM'*** is the concatenated cell translations along x, y, z with respect to the base number 555. Symmetry operators listed in ***REMARK 290*** operate on orthogonal crystallographic coordinates that appear in the entry.

As an example, the ***SymOP*** 2456 indicates that the second symmetry operation as listed in ***REMARK 290*** is applied with translation of -1 on x, and +1 on z. A program will be made available shortly that converts ***SymOP*** data into transformations that operate in the coordinate frame used in the entry.

The ***SymOP*** data type is used in ***SSBOND***, ***LINK***, ***HYDBND***, ***SLTBRG*** and ***REMARKs***

Template

```
1     2       3         4         5         6         7
12345678901234567890123456789012345678901234567890123456789
01234567890
REMARK 290
REMARK 290 CRYSTALLOGRAPHIC SYMMETRY
REMARK 290 SYMMETRY OPERATORS FOR SPACE GROUP: P 21 21 21
REMARK 290
REMARK 290       SYMOP   SYMMETRY
REMARK 290       NNNMMM   OPERATOR
REMARK 290       1555   X,Y,Z
REMARK 290       2555   1/2-X,-Y,1/2+Z
REMARK 290       3555   -X,1/2+Y,1/2-Z
REMARK 290       4555   1/2+X,1/2-Y,-Z
REMARK 290     WHERE NNN -> OPERATOR NUMBER
REMARK 290           MMM -> TRANSLATION VECTOR
REMARK 290
REMARK 290 CRYSTALLOGRAPHIC SYMMETRY TRANSFORMATIONS
REMARK 290 THE FOLLOWING TRANSFORMATIONS OPERATE ON THE
ATOM/HETATM
REMARK 290 RECORDS IN THIS ENTRY TO PRODUCE
CRYSTALLOGRAPHICALLY
REMARK 290 RELATED MOLECULES.
REMARK 290   SMTRY1   1  1.000000  0.000000  0.000000
0.00000
REMARK 290   SMTRY2   1  0.000000  1.000000  0.000000
0.00000
REMARK 290   SMTRY3   1  0.000000  0.000000  1.000000
0.00000
REMARK 290   SMTRY1   2 -1.000000  0.000000  0.000000
36.30027
REMARK 290   SMTRY2   2  0.000000 -1.000000  0.000000
0.00000
REMARK 290   SMTRY3   2  0.000000  0.000000  1.000000
59.50256
REMARK 290   SMTRY1   3 -1.000000  0.000000  0.000000
0.00000
REMARK 290   SMTRY2   3  0.000000  1.000000  0.000000
46.45545
REMARK 290   SMTRY3   3  0.000000  0.000000 -1.000000
59.50256
REMARK 290   SMTRY1   4  1.000000  0.000000  0.000000
36.30027
REMARK 290   SMTRY2   4  0.000000 -1.000000  0.000000
46.45545
REMARK 290   SMTRY3   4  0.000000  0.000000 -1.000000
0.00000
REMARK 290
REMARK 290 REMARK: NULL
```

2- Coordinate Systems and Transformations

The coordinates distributed by the Protein Data Bank give the atomic positions measured in Angstroms along three orthogonal directions. Unless otherwise specified, the default axial system detailed below is assumed. If a, b, c describe the crystallographic cell edges and A, B, C are unit vectors in the default orthogonal Angstrom system, then the following apply.

A, B, C and a, b, c have the same origin.

A is parallel to a.

B is parallel to (a X b) X A (cross product between C and A).

C is parallel to a X b (i.e., c*) (cross product between a and b).

The matrix which pre-multiplies the column vector of the fractional crystallographic coordinates to yield the distributed coordinates in the A, B, C system is:

```
  a     b(cos(gamma))     c(cos(beta))

  0     b(sin(gamma))c(cos(alpha)-cos(beta)  cos(gamma))/
sin(gamma)
  0     0        V/(ab sin(gamma))
```

V = abc(1 - cos2(alpha) - cos**2(beta) - cos**2(gamma) + 2(cos(alpha) cos(beta) cos(gamma)))**1/2**

The distributed entry will contain the following records.

- ORIGX - transformation from the distributed to the submitted coordinates.
- SCALE - transformation from the distributed to the fractional coordinates.

3- Atom Names

The following rules are used in assigning atom names.

- Greek letter remoteness codes are transliterated as follows: alpha = A, beta = B, gamma = G, delta = D, epsilon = E, zeta = Z, eta = H, etc.

- Atoms for which some ambiguity exists in the crystallographic results are designated A. This usually applies only to the terminal atoms of asparagine and glutamine and to the ring atoms of histidine.

- The extra oxygen atom of the carboxy terminal amino acid is designated OXT.

- Six characters (columns) are reserved for atom names, assigned as follows.

COLUMN	VALUE
13-14	Chemical symbol-right justified, except for hydrogen atoms
15	Remoteness indicator (alphabetic)
16	Branch designator (numeric)
77 - 78	Element symbol, right-justified

- Columns 73 - 76 identify specific segments of the molecule. The segment may consist of a complete chain or a portion of a chain. The importance of this new field can be appreciated if one considers an antibody structure having two molecules in the asymmetric unit. Since each chain must have a unique chain identifier, the two heavy chains and two light chains cannot currently be labeled to indicate their nature.

Segment id's of **CH, VH1, VH2, VH3, CL**, and **VL** would clearly identify regions of the chains and the relationship between them. Users of **X-PLOR** will be familiar with **SEGID** as used in the refinement application of **X-PLOR.**

Nucleic Acids

Atom names employed for polynucleotides generally follow the precedent set for mononucleotides. The following points should be noted.

- The asterisk (*) is used in place of the prime character (') for naming atoms of the sugar group. The prime was avoided historically because of non-uniformity of its external representation.
- The ring oxygen of the ribose is denoted O4 rather than O1.
- The extra oxygen atom at the free 5' and 3' termini are designated O5T and O3T, respectively.

4-Standard Residue Names and Abbreviations

Note that there will be a change to what are considered standard groups due to the adoption of the new **PDB** Het Group Dictionary. Only the twenty common amino acids and five nucleic acids plus inosine will be treated as "standard" with all others being treated as modified residues to be described by **MODRES** records.

No distinction is made between ribo- and deoxyribonucleotides in the **SEQRES** records. These residues are identified with the same residue name (i.e., **A**, **C**, **G**, **T, U**, **I**).

Amino Acids

RESIDUE	ABBREVIATION	SYNONYM
Alanine	ALA	A
Arginine	ARG	R
Asparagine	ASN	N
Aspartic acid	ASP	D
ASP/ASN ambiguous	ASX	B
Cysteine	CYS	C
Glutamine	GLN	Q
Glutamic acid	GLU	E
GLU/GLN ambiguous	GLX	Z
Glycine	GLY	G
Histidine	HIS	H
Isoleucine	ILE	I
Leucine	LEU	L
Lysine	LYS	K
Methionine	MET	M
Phenylalanine	PHE	F
Proline	PRO	P
Serine	SER	S
Threonine	THR	T
Tryptophan	TRP	W
Tyrosine	TYR	Y
Unknown	UNK	
Valine	VAL	V

Nucleic Acids

RESIDUE	ABBREVIATION
Adenosine	A
Modified adenosine	+A
Cytidine	C
Modified cytidine	+C
Guanosine	G
Modified guanosine	+G
Inosine	I
Modified inosine	+I
Thymidine	T
Modified thymidine	+T
Uridine	U
Modified uridine	+U
Unknown	UNK

Remarks 103 and 104 are included when an entry contains inosine.

5- Formulas and Molecular Weights for Standard Residues

These weights and formulas correspond to the unpolymerized state of the component. The atoms of one water molecule are eliminated for each two components joined.

Amino Acids

NAME	CODE	FORMULA	MOL. WT.
Alanine	ALA	C3 H7 N1 O2	89.09
Arginine	ARG	C6 H14 N4 O2	174.20
Asparagine	ASN	C4 H8 N2 O3	132.12
Aspartic acid	ASP	C4 H7 N1 O4	133.10
ASP/ASN ambiguous	ASX	C4 H71/2 N11/2 O31/2	132.61
Cysteine	CYS	C3 H7 N1 O2 S1	121.15
Glutamine	GLN	C5 H10 N2 O3	146.15
Glutamic acid	GLU	C5 H9 N1 O4	147.13
GLU/GLN ambiguous	GLX	C5 H91/2 N11/2 O31/2	146.64
Glycine	GLY	C2 H5 N1 O2	75.07
Histidine	HIS	C6 H9 N3 O2	155.16
Isoleucine	ILE	C6 H13 N1 O2	131.17
Leucine	LEU	C6 H13 N1 O2	131.17
Lysine	LYS	C6 H14 N2 O2	146.19
Methionine	MET	C5 H11 N1 O2 S1	149.21
Phenylalanine	PHE	C9 H11 N1 O2	165.19
Proline	PRO	C5 H9 N1 O2	115.13
Serine	SER	C3 H7 N1 O3	105.09
Threonine	THR	C4 H9 N1 O3	119.12
Tryptophan	TRP	C11 H12 N2 O2	204.23
Tyrosine	TYR	C9 H11 N1 O3	181.19
Valine	VAL	C5 H11 N1 O2	117.15
Undetermined	UNK	C5 H6 N1 O3	128.16

Iron (II) Triggered Conformational Changes in E.Coli Fur upon DNA Binding: A study using Molecular Modeling

Mazen Y. Hamed* and Salih Al-jabour

Computational science program, chemistry department Birzeit University, Po Box 14, Birzeit, Palestine

Abstract

The three dimensional structure of the ferric uptake regulation protein from E.Coli (Fur EC) was determined using homology modeling and energy minimization. The fur monomer consists of turn- helix - turn motif on the N-terminal domain, followed by another helix- turn- helix-turn motif, and two β strands separated by a turn which forms the wing. The C-terminal domain, separated by a long coil from the N-terminal, and consisting of two anti parallel β strands, and a turn-helix-turn-helix-turn motif.

Residues in central domain were found to aid the dimer formation, residues 45 to 70 as evident in the calculated distances; this region is rich in hydrophobic residues. Most interactions occur between residues Val(55), Leu(53), Gln(52), Glu(49) and Tyr(56) with closest contacts occurring at residues 49 to 56. These residues are part of an α-helix (α_4) near the N-terminal. The Fur EC dimer was docked onto DNA "iron box" , it was found to bind the A.T rich region and addition of iron (II) enhanced the fur binding of the helices near the N-terminal to major grove of DNA. Addition of high Iron (II) concentration triggered further conformational changes in both fur dimer as measured by distances between the two subunits, and mediated the fur binding by attaching itself to the DNA. DNA changed conformation as evident in the distortion in the backbone, and the shrinking of major grove distance from 11.4 Å to 9.3Å. Two major Iron (II) sites were observed on the C-terminal domain: site 1, the traditional Zn site, the cavity contains the residues Cys92, Cys95, Asp137, Asp141, Arg139, Glu 140 His 145 and His 143 at distances range from1.3 to 2.2 Å. Site 2 enclave consists of His71, Ile50, Asn72, Gly97, Asp105 and Ala109 at very close proximity to Fe (II).

The closest contacts between fur dimer and DNA at the A.T rich region were at residues A11, G12, L13, P18 and R19 mostly hydrophobic residues near the N-terminal domain. Close contacts repeated at H87, H88 and R112, and a third region near the C-terminal at N137, R139, E140, N141, H143, N141 and H145. Fur dimer has three major contact regions with DNA, the first on the N-terminal domain, a second smaller region at H87H88 R112 mediated by Fe^{2+} ions, and a third region on the C-terminal domain consisting mainly of hydrophobic contacts and mediated by Fe^{2+} ions at high concentration.

* for correspondence e-mail: mhamed@birzeit.edu, Fax: 97222982084, phone: 97222982003

Introduction

Fur protein from E. Coli K12 (Fur EC) is a 17 KDa, 148 amino acid residues protein [1]. Fur EC has attracted much attention in recent years [1-7] and it has been extensively studied as a repressor protein which uses Fe^{2+} as co-repressor to bind specifically to DNA [2-6], it was especially studied with the 19 bp iron box (5'-GAT AAT GAT AATC ATT ATC-3') [2, 8-13]. Other divalent transition metal ions such as Mn^{2+}, Co^{2+} were found to activate fur both in vitro and in vivo with varying degrees, while Zn^{2+}, Cd^{2+} and Cu^{2+} were found to bind fur strongly and could activate fur in vitro only [1, 5, 18]. The X-ray structure of fur protein from E.Coli is still not resolved; the NMR studies gave insight about the structure of fur EC and its relation to the fur function [13-15]. An x-ray structure on a member of the fur family from rhizobium leguminosarum was reported [16]. The first crystal structure of Fur from P.aeruginosa in complex with Zn^{2+} was determined at a resolution of 1.8Å [17]. X-ray absorption spectroscopic measurements and micro PIXE analysis were also performed [17] in order to characterize the distinct iron binding sites in solution and it was found to bind four Zn^{2+} ions per Fur dimer with N/O ligands at an average metal-ligand distance of 2.1Å.

Experimental work revealed many aspects about the fur structure-function relationship. The HTH motif near the N-terminus was suggested to play the DNA binding role similar to other repressor proteins (λ repressor, DtxR, lac repressor [20-22] and IdeR [24]. Other reports provided insight on the metal ion binding sites provided by fur and the role of metal ion in the DNA binding process [25]. Indeed, previous work based on, thermodynamic equilibrium gave evidence that Fe^{2+}, Mn^{2+}, and Co^{2+} ions are weakly bound to fur and ^{57}Fe Mössbauer study showed that Fe^{2+} is present in an axially distorted octahedral environment with $\partial = 1.3 mms^{-1} and\ \Delta = 3.3 mms^{-1}$ [5, 7, 26] These values, when compared with reported values for Fe^{2+} sites, indicated a moderately bound Fe^{2+} to oxygen and/or nitrogen ligands [23]. This is consistent with the reversible metal ion binding (K_d value 55 μM [5]) which agrees well with the role of fur protein as metal ion sensor. Site multiplicity and flexibility was not ruled out as more than one ion was found to bind per fur [5]. Other metal ions could replace Fe^{2+} as co-repressors and was active in various degrees [1].The proposed role of metal ion was interpreted as to trigger conformational changes in the fur protein dimer and consequently facilitate DNA binding. Coy [12], basing his study on proelytic cleavage suggested that the metal ion role was to induce conformational changes, and also proposed that both DNA binding and N-terminal sensitivity of fur were dependent on the metal ion concentration. He also suggested that the C-terminal was responsible for metal ion binding [12]. Most workers [12, 17, 24, 25] tend to agree that fur has three major domains based on its function; an N-terminal which is responsible for the DNA binding process, a middle domain which plays a role in the dimerization of fur and the C-terminal which contains the metal ion binding sites. C-terminal plays the role of metal ion concentration sensing and binding. In this work, the three dimensional structure of fur was built using molecular dynamics. The dimerization of fur was performed in water to produce the fur dimer. The dimer was studied in the presence of DNA with and without the presence of Fe^{2+} ion. The effect of metal ion on the conformational changes of fur and how does this act to enhance the DNA binding process at elevated Fe^{2+} concentration and the unbinding of fur dimer to DNA at reduced Fe^{2+} concentration.

Three dimensional structures of protein molecules are, generally represented, as coordinates of all atoms in space. Structures can be, uniquely, determined by specificity of amino acid sequence. It has been reported that proteins with high sequence similarity have similar structures and may play similar functions. It is generally accepted that the final structure is the one which occurs at minimum energy. Comparative protein procedure depends on alignment of an unknown protein sequence with other known

proteins in Data Banks. The aim of alignment process is to find optimal superposition.

Computations and homology modeling

All the molecular dynamics (MD) simulations were performed using AMBER7 molecular simulation package [27, 28]. An AMBER force field was used for molecular minimization and molecular dynamics. The analyses of MD trajectories were also preformed by AMBER7. Pymol molecular viewer package was used for visualization [29]. All other calculations were performed on a single-CPU Pentium III machine with Linux platform.

Homology modeling of Fur protein:

The known Fur sequence (from *E.coli)* was submitted to different modeler servers in order to predict the three dimensional structure. SWISS-MODEL [30], PHD, 3DPSSM [ ] and VADAR servers were used to align the fur sequence with similar known proteins Data Bank. Several templates for fur protein were generated while the sequence with high similarity served as a reference sequence. The superposition of each atom was optimized by maximizing Cα in the common core while minimizing their relative mean square value deviation (RMSD) at the same time. Spare part algorithm was used to search for fragments that can be accommodated into the framework of the Brookhaven Protein Data Bank (PDB). The coordinates of central backbone atoms (N, O and C) were averaged, and then added to the target model. The side chains were added according to the sequence identity between the model and the template sequence. AMBER7 was used to idealize the geometry for bonds and also to remove any unfavorable non-bonded contacts. This was done by minimizing the energy. All hydrogen atoms were added and the apofur structure was subjected to a refinement protocol with constraints on the fur structure gradually removed. 100 steps of steepest descent, followed by 300 steps of conjugate gradient algorithm were applied during energy minimization. The energy minimization process on the apofur model was performed, first in vacuum and second in H_2O as solvent, nine Na^+ ions were added to the model to neutralize the system.

Building the Fur dimer:

AUTODOCK 2.4 [32] was used to generate the apofur dimer. Two molecules of the previously determined structure for the apofur monomer were docked on each other, and the best docking sites were predicted. Monte Carlo (MC) Simulated Annealing (SA) algorithm was used for exploring the fur configuration by a rapid energy evaluation technique using a grid based molecular affinity potential. The energy of interaction, affinity and the grid for electrostatic potential were evaluated using the Poisson-Boltzmann finite difference method and were assigned to each atom.

Docking of ApoFur dimer onto a 19 Bp fragment representing the DNA:

Nucgen suite program (part of the AMBER7 package [28]) was used to build the Cartesian coordinates for canonical B- model of the iron box (a19-bp inverted repeat sequence designated the iron box (5' GATAATGATAATCATTATC 3'); the proposed recognition site of fur on the DNA. The right handed B-DNA duplex conformation was applied for the model. The iron box was docked to the Fur-dimer using the AUTODOCK program. The energy minimization was applied to the resultant model in order to refine the Fur dimer –DNA complex. The parameters file for the iron metal was built manually and inserted into AMBER7 as a library file. The first scenario was using 4 Fe^{+2} ions per fur dimer-DNA complex in the water environment and adding Na^+. MD simulations were carried out at 300K. Explicit solvent model WATBOX216 water was used as solvent model. The models were solvated with a 10 Å water cap from the center of mass of the ligands. The dynamics simulation was applied for 25 ps time limit. In a second scenario, the same was repeated using 8 Fe^{2+} ions and simulation was applied for 25 ps.

Results and Discussion

Predicting Fur 3D structure:

The fur sequence was submitted to several servers in order to study the preserved amino acid residues in the fur family. The results of alignment Figure 1 showed highly preserved residues in both the N-terminal and C-terminal domains, these residues must play crucial role in the fur function as an iron regulator. Comparison of the amino acid sequence of homologous proteins indicates which of the proteins' residues are essential to its function, which are of less significance and which have little specific function, invariant residues uniquely suit essential function of the protein, other residues, conservatively substituted have less stringent side chain requirements [17, 22]. On the other hand, other amino acid residues have nonspecific function "hypervariable". The main feature is the preserved hydrophobic residues (AGLIV) on 17 positions on the N-terminal domain and to less extent (4 major positions) on the C-terminal domain. Hydrophilic basic residues K and R(residues K9,K10), K14, R19X K 21), (K41, K(R)42), R57 ,R70, K77 are repeated 7 times on the N-terminal domain, and once on the C-terminal R110XK112. We can say that proteins in the fur family are mostly hydrophobic and their N terminal domains are more hydrophobic than their C terminal domains. All proteins in the fur family appear to be Histidine rich [14, 15], H32 (replaced by Q or E in some proteins), H33 is preserved and important to fur function. Indeed, the H33L mutant reported to be inactive in vivo [6]. H71, H86, H87, motif H88DH90, and H135 are fully conserved, while H142 occurrence is less frequent. The unit C93LDC96G is present in a coil folding and is highly preserved in the fur family. Its worth noting that it was reported by Coy et al [6] that the C92S C95S mutations altered the fur activity drastically , which confirms that C92 C95 residues are essential to the fur activity. Glutamic acid 81and Cys132 are also preserved in the fur family. Some of the homology modeling results for folding coincided with those predicted by nmr [13-15] for coil T_2, α_3, T_3, α_4, and α_5.

The fur secondary structure was predicted Figure 2, especially the conserved region, and compared with those predicted by nmr [13-15]. The results of the homology modeling [35, 36, 37] (Figure 2)(Table 1) from different servers coincided with each other to a great extent and this allowed us to propose a three dimensional structure for the fur monomer Figure 3. The fitted structure was in good homology with winged helix proteins with an RMSD value of 1.3 Å which falls within the accepted value for protein alignment 1-2 Å. The final 3D structure of fur agrees with its proposed function; the N-terminal domain contains the HTH motif. Most servers gave an α helix for the residues 4 to 6 with good confidence level, a coil for residues 11 to 16 and another α-helix for residues 17 to 27 another coil 29 to 35. Another, α helix for the residues 49-59 followed by coil (60-64), these regions included in the central domain which was reported to be responsible for fur dimerization [8, 12]. In the C-terminal domain two α helices were found in the region 108-113 and 134-136 separated by a β strand in the region 121-132 and a coil between 118-120. The comparative protein calculations gave 67.57% of the fur residues are exposed to solvent, and this is especially clear for residues forming the loops and residues at both C- and N-terminal domains. 32.43% of fur protein residues were buried (Figure 2). The fur EC sequence was aligned with the Fur sequence from Pseudomonas aerugnosa with known crystal structure [17] which was found to bind $zinc^{2+}$ in two different binding sites and does not have sequence similarity with fur EC, the results of alignment Figure 1b gave 62.9 % sequence identity. The Similarity with high confidence level was for residues Lys10-Pro19, Gly48-Thr54, His71-Ser79, Thr84-Ala110, and Arg121-Gly136. Most important preserved residues are H89H90DH91 and C92 (Figure 1b). Calculated surface area for Fur EC using Spdv was 7016 $Å^2$ and the volume was 16863 $Å^3$, a cavity of volume 14 $Å^3$ and area 34 $Å^2$ was formed by residues CYS93XY CYS96*, His 71 -Glu74, and His 86 to His 90

The amino terminal domain of Fur shares considerable similarity with DtxR [22]; both proteins are iron-dependant repressor proteins but differ in their DNA specific binding [20, 22]. Although both of these proteins regulate iron uptake. The fur monomer Figure 3

* Usually referred to in the literature as C92 and C95

resembles a great deal the determined structure of DtxR [22] which contains two clearly defined domains; the amino terminal domain consists of 72 residues and contains three helices, two antiparallel **β** strands plus the first half of α_4. The second domain (70 residues) contains α_4, α_5 and α_6. The structure contains helix-helix interactions; α_1 with α_4 and α_5, α_2 with α_4, and $\alpha_1$with α_5 thought to be crucial for protein function, some interactions between helices were observed by nmr but not very pronounced [22, 13-15].

Molecular Modeling of the Fur protein using Molecular dynamics:

The three dimensional structure of the fur monomer which resulted from homology modeling was used as the starting structure in calculations using Amber software, the calculated three dimensional structure for fur monomer at minimum energy is shown in Figure 4. The energy minimization idealized the geometry of bonds and removed unfavorable connections. Energy minimization was applied in a water box. Indeed an X-ray structure of fur protein dimer from rhizobium leguminosarum [16] has shown two discrete domains with N-terminal formed from association of two HTH motifs, a flexible hinge linked a compact C-terminal consisting of **α/β** domain, and a solution X-ray scattering in reducing environment [18] showed that the two domains are flexibly arranged with respect to each other, and no structural homology with DtxR [22] or IdeR [24] apart from that expected HTH motif in the N-terminal. There is also an interface region consisting of polar residues with large void in the core lined by basic residues. In contrast to the N-terminal, the C-terminal formed from a large and stable domain subunit with the role of maintaining the dimerization of fur. The classic HTH motif consists of two helices (α_1, α_2) joined by loop. It is found that HTH is a conserved domain which binds the DNA [46]. The HTH motifs alone is apparently insufficient for independent folding, a third helix (α_3) stabilizes the motif as a compact, globular domain. The HTH motif followed by two β-hairpin wings reported in the Fur structure which shows a high similarity with winged-helix family.

The folding as resulted from *Amber* minimization

(1-8) Coil T1, (9-17) helix* α 1, (18-22) coil T2*, (23-26) helix* α 2 (27-29) coil T3, (30-33) helix α 3, (34-36) coil T4, (37-40) strand^{+} β 1, (41-42) coil^{+} T5 (43-46) strand^{+} β 2, (47-55) coil T6, (56-60) helix α 4, (61--90) coil T7(91-95) strand β 3, (96-107) coil T8, (108-111) strand β 4, (112-121) coil T9 (122-127) helix α 5, (128-131) coil T10, (132-141) helix α 6, (142-148) coil T11

* Helix turn helix motif
$^{+}$ Wing

Fur Dimer Structure:

Two fur monomers were docked on each other using AutoDock [32] and minimizing the energy. The features of the fur structure are in good agreement with its function as a repressor protein which uses Fe^{2+} or other divalent transition metal ion as co-repressors, i.e. binds the DNA at high Fe^{2+} concentration and falls off the DNA at lower iron concentrations. The Structure of fur dimer shows that each subunit is composed of an amino-terminal DNA-binding domain, an interface-domain in the middle and a carboxyl- terminal which contains the metal binding sites. Each DNA-binding domain contains the helix-turn-helix motif with a topology similar to other repressor proteins (DtxR, λ repressor,) [22, 24]. The resultant apofur dimer model shows helix-helix interactions at residues 45 to 60 between the two monomer subunits. This behavior is similar to other proteins; i.e. helix-helix interactions are found in the dimerization domain [17, 22, 24].

Residues in central domain were found to aid the dimer formation, specifically residues 45 to 70 as evident in the calculated distances (Table 2) (Figures 5 and 11), this region is rich in hydrophobic residues. Most interactions occur between residues Val(55), Leu(53), Gln(52), Glu(49) and Tyr(56) with closest contacts occurring at residues 49 to 56. These residues are part of an α-helix (α_4) near the N-terminal. Coy et al. [12] and Klode et al.

[16] suggested that the helix-helix interactions occur at the central domain closer to the N-terminal. The negatively charged and highly polar Glumatic acid residue seems to aid the establishment of hydrogen bonding a cross fur subunits. Extensive hydrophobic interactions occur between the two monomers aided by the hydrophobic properties of valine and leucine. The aromatic ring of Tyrosine also helps to establish hydrogen bonding between the two monomers [17, 38]. To the Contrary of what was predicted by nmr [13], the N-terminal from each subunit is close to the other and at large distance from the C-terminal.

DNA Binding:

The Fur EC dimer was docked onto a DNA Iron Box (5' GATAATGATAATCATTATC 3') in the presence of water and Na^+ ions and measuring the contacts between Fur residues and DNA, the results are shown in Table 2 and Table 3. When Fe^{2+} ions were added to the Fur/DNA complex, it resulted in an obvious tuning of the fur structure, this constituted a conformational change, obviously triggered by the addition of Fe^{2+} ions (see Table 2). The outcome was to bring the HTH motif near the N-terminal in close proximity to the Major grooves of the DNA. As a result of this process, the fur dimer engulfed the DNA, see Table 3 and Figure 6. Upon the addition of another four Fe^{2+} ions, the change in conformation was more evident and the helices moved closer to the major groove of DNA. This proved without doubt that the process, i.e. the Fur dimer specific binding to DNA depends on the concentration of Fe^{2+} [11, 12, 13, 19, 25]. A critical issue in terms of the structure-function relationship of Fur is how the regulator interacts with its operator site to block the access to the promoter region of an iron-responsive gene [25]

The Fur dimer/DNA model clearly suggested that the putative DNA-binding helices α_2 and α'_2 contact the major groove of DNA [11, 12, 25]. The model shows that $\alpha 2$ and α'_2 fit well into the major grove (Fur changes conformation to prevent their overlap). Recognition and binding is the result of direct interactions between the base pairs in the major groove of DNA and the amino acid side chains of α_2 and α'_2 helices (Figure 6). The calculated distances showed specific contacts taking place by the side chains of Val15, Leu13, Ala11 and Pro 18 and DNA, Table3. The aromatic ring of Proline 18 undergoes hydrogen bonding to the AT base pairs spaced by 4 base pairs [2, 24, 25, 39, 40, 41]. While the hydrophobic properties of Valine, Leucine and Alanine residues made the hydrophobic interactions between the fur and edges of the bases and sugar-phosphate backbone of DNA groove possible [25]. These interactions induce an affect on the DNA by over winding the four base pairs in the middle (Figure 6c). As a result the minor groove in the center of the operator was compressed in a way that the phosphate to phosphate distance was reduced from 11.4 Å for canonical B-DNA to 9.3 Å upon Fur dimer binding (Figure 6c) [25].

Types of Fur contacts with DNA Operator sequence were analyzed experimentally by several workers [9,10, 11,12, 42] using ethylation and hydroxyl radical foot printing and was found to be similar to the unique HTH motif and these contacts were found to be on one face of DNA [42] and span three major grooves[11,12], indeed this is clearly observed in our calculated structure shown in Figures 6 and 7, the fur dimer clamps around the major grooves of DNA using an α_2 helix from each monomer. When the nature of the residues which contacts with DNA were analyzed the following can be said about the fur DNA complex: A striking structural feature (a pair of two-fold α helices were tilted and has center to center separation of 2.4 Å. α_2 helices were also located at very close proximity to DNA so that the N-terminal chain and side chains were able to make nonspecific contacts with phosphate diester backbone see Figures 6 and Tables 2 and 3 , the common DNA binding structure is still the HTH motif in which the contacts can result from hydrogen bonds, salt bridges and van der Waal forces. All these forces account for site recognition and specific binding. Additional nonspecific contacts attributed to the loose loops on both ends of fur dimer: residues near the C-terminal (see Table 3) loops work as an arm to engulf the DNA.

The Change in DNA conformation is worth noting as the tilting which took place upon fur binding in the presence of Fe^{2+}, H_2O and Na^+ is evident and the major groove distance shrunk from 11.4 Å to 9.3 Å, a notable conformational change is evident as can be

seen in Figure 6c this was interpreted in some reports as a hand shake between fur and DNA [25].

Iron (II) binding sites on the fur dimer:

The addition of Fe^{2+} ions to the Fur dimer/ DNA complex induced a change in conformation of the fur structure as evident in the distances between residues and helices of the fur subunits in the dimer (Table 2) (Figures 7a, 7b and 11b). The N-terminal domains were at 20.6 Å apart in the apofur dimer, they moved closer to each other by 5 Å upon addition of DNA. Upon adding the first $4Fe^{2+}$ ions a significant move took place; the N-terminals became at 10.6 Å apart. At the same time residues moved closer to the DNA. The addition of the first 4 Fe^{2+} ions per fur dimer could produce a significant change in fur conformation. The Fur dimer/DNA complex in the presence of water and Na^{+} ions, could take up to 8 Fe^{2+} ions per complex, the more Fe^{2+} ions added, the closer the fur subunits became to the DNA. This was accompanied by conformational changes in both fur dimer and DNA.

The nature of ligands provided by the fur dimer to metal ion, and the number of metal ion sites were always a matter of debate [5] and it is worth the attention as it plays a key role in the whole process. There are two major sites provided by the fur dimer to Fe^{2+}, site 1 which involves Cys92 and Cys95 and other residues with N or O ligands (Table 4)(Figure 9). Cys92 and Cys95 were always reported to play a crucial role in metal ion binding and fur function [5, 6, 12, 13, 18, 19, 23, 43]. Indeed a fur mutant with either or both Cys92 and Cys95 replaced by Ser lost its repressor activity and failed to bind the DNA [6]. Both Cys92 and Cys95 are present in a β strand and a loop, respectively near the C-terminal domain and they are relatively buried inside the protein as can be seen in Figure 2. EXAFS results [18] suggested a metal environment consisting of a total of 5 oxygen and nitrogen atoms at an average distance of 2.13 Å (either 2O at 2.05Å/3N at 2.17 Å or 3O at 2.08Å/2N at 2.19Å). In our study the calculated distances were Fe^{2+} -Cys92 = 2.2Åand Fe^{2+}-Cys95 = 1.6Å (Figure 8c). Cysteins are probably bound through H-bonded H_2O intermediate or a protonated SH as indicated by the weak binding evident in the Mössbauer parameters for Fe^{2+} and the reported dissociation constant which ruled out the presence of strong sulfur –Fe^{2+} bonds [5]. His143 and His145 were close to the DNA and it seems that they form part of the iron binding environment [43] (Figure 8d). The calculated distances show that these residues moved closer to the DNA upon metal binding. Aspartic (Asp137-Asp141), Arg 139and Glutamic acid (Glu140) complete the distorted octahedral environment around Fe^{2+}. Another Fe^{2+} is coordinated by the side-chains of residues His 71 (end of β strand), Asp105 (coil), Ala109 (α helix), Asn72 (β strand) and Ile50 (coil) [43] (Figure 8b). This site is probably site 2 with O and N bound to Fe^{2+} in a distorted octahedral environment. Table 3 shows the calculated distances between the donor atoms of these residues and Fe^{2+} ions, His71 plays an important binding role to Fe^{2+}. Recent experimental reports suggested that apofur contains at least one Zn^{2+} ion per monomer coordinated to Cys92 and Cys95 and another metal ion binding site which contains iron [17-19, 23, 34 43]. Site 1 is the Zn^{2+} binding site while (Fur was reported to contain structural Zn^{2+} ion per monomer [34]) the other site is an Fe^{2+} site. Another reported Zn^{2+} binding site which involves Cys132 and Cys137 in the C-terminal domain [17] could not be found in our study. The excess Fe^{2+} bind the Phosphate backbone in AT- rich region of the minor grove, see Figure 8c, 8d and Figure 9. It is evident that the Fe^{2+}, in this case, acts as mediator for the binding of fur residues to the DNA, and at the same time participate in conformational changes of DNA.

The metal ion and HTH binding to major grooves play an important role in inducing conformational changes of the canonical B-DNA [25]. Recent studies proved the presence of strongly bound Zn^{2+} ion to the Fur [18] the suggested site is 1 and its tetrahedrally bound to both C92 and C95 and other residues .This made what used to be apofur dimer to be active in vitro without adding Fe^{2+} [43] .

Evidence for conformational changes triggered by DNA binding and metal ion binding:

In the presence of DNA the fur dimer changes conformation before adding the Fe^{2+} as can be seen in Figure 11. Residues on the

sub units move closer together except for Val 25-Val25, Pro 29-Pro 29, Gln85-Gln85, Ala 53-Ile 107 and α_5-α_5 moved apart. Upon adding low Fe^{2+} concentrations all residues and helices on the fur subunits move closer together causing a drastic change in conformation. The addition of larger concentration of Fe shifted the subunits closer but the move was less drastic than when the first Fe was added.

The N and C terminals behave in different manner, the N-N moved drastically towards each other upon adding the DNA and the first Fe^{2+} addition but the second Fe^{2+} addition did not cause much change in the N-N distance. The C-C distance shifted slightly upon DNA binding, while the drastic shift in distance was when the low Fe^{2+} concentration was added and a similar shift occurred when more Fe was added. The inter phase region showed considerable rigidity as can be seen in Figure 11 (Val 25, Pro 29 and α_4-α_4), no considerable change in distances was observed.

The closest contacts between fur dimer and DNA at the A.T rich region were at residues A11, G12, L13, P18 and R19 mostly hydrophobic residues near the N-terminal domain. Anther close contacts repeated at H87 H88 and R112 and a third region engulfs the DNA near the C-terminal at N137, R139, E140, N141, H143, N141 and H145. As can be observed in Figure 11b fur dimmer has three major contact regions with DNA, the first on the N-terminal domain, a second smaller region at H87H88 R112 mediated by Fe^{2+} ions as shown in figure8c and a third region on the C-terminal domain consisting mainly of hydrophobic contacts and mediated by Fe^{2+} ions at high concentration.

Aknowledgement: We wish to thank the Amber group in UCSF for providing the software, and Pymol Prof. Arthur J. Olson, DeLano Scientific California for the Pymol viewer.

References

[1] Bagg A, Neilands, JB (1987). Biochemistry. 26: 5471-5477.

[2] Escolar, L, Perez-Martin, J and de Lorenzo, V (1998). Journal of Molecular Biology. 283(3): 537-547.

[3] Hamed, YM (1993). Journal of Inorganic Biochemistry. 51(594):1-2.

[4] Hamed, MY, Neilands, JB (1994). Journal of Inorganic Biochemistry. 53(4): 235-248.

[5] Hamed, M.Y., Neilands, J.B. and Huynh,V. 1993. Journal of Inorganic Biochemistry. 50 (3): 193-210.

[6] Coy, M., Doyle, C., Besser, J. and Neilands, B.J. (1993). BioMetals. 7: 292-298.

[7] Jacquament, L., Dole, F., Jeandey, C., Oddou, J.L., Perret, E., La Pape, L., Aberdam, D., Hazemann, J.L., Michaud-Soret, I. and Latour, J.M. (2000). J. Am. Chem. Soc. 122: 394-395.

[8] Neilands, J.B. (1992). Can J. Microbiol. 38(7):728-33.

[9] Le Cam, E., Frechon, D., Barray, M., Fourcade, A. and Delain, E. (1994). Proc-Natl Acad Sci, USA. 91(25):11816-20

[10] Frechon, D., Le Cam, E. (1994). Biochem Biophys Res. Common. 201(1): 346-55.

[11] Coy, M. (1995). Biochem Biophys Res. Commun. 212(3): 784-92.

[12] Coy, M., Neilands, J.B. (1991). Biochemistry. 30 (33):8201-8210.

[13] Saito, T., Williams, R. J.P. (1991). Eur. J. Biochem. 197: 43-47.

[14] Saito, T., Womlad, M.R. and Williams, R.J.P. (1991). Eur. J. Biochem. 197: 29-39.

[15] Saito, T., Williams, R.J.P and Duly, D. (1991) . Eur. J. Biochem. 197: 39-42.

[16] Kolade, O.O., Bellini, P., Wexler, M., Johnston, A.W.B., Grossmann, J.G. and Hemmings, A.M. (2002). Biometals. 30(4):771-774.

[17] Pohl, E., Haller, J.C., Mijovilovich, A., Meyer-Klaucke, W., German, E. and Vesil, M.L. (2003). Molecular Microbiology. 47(4):903-915.

[18] Jacquamet, L, Aberdam, D, Adrait A, Hazemann, JL, Latour, JM and Michaud-Soret, I. (1998). Biochemistry. 37: 2564-2571.

[19] Zheleznova, E.E., Crosa, J.H. and Brennan, R.G. (2000). Journal of Bacteriology. 182(21): 6264-6267.

[20] Schmitt, M.P., Twiddy, E.M. and Holmes, R.K. (1992). Proc. Natl. Acad. Sci USA. 89(10): 7576-80

[21] Reidhaar-Olsin, JF, Sauer, RT, (1990). PROTEINS: Structure, Function and Genetics. 7: 306-316.
[22] Qiu, X., Verlinde, C.L.M.J, Zhang, S., Schmitt, M.P., Holmes, R.K. and Hol, W.G.J. (1995). Structure. 3(1):87-100.
[23] Jacquamet, L., Dole, F., Jeandey, C., Oddou, J.L., Perret, E., Le Pape, L., Aberdam, D., Hazemann, J.L., Michaud-Soret, I. and Latour, J.M. (2000). J. Am. Chem. Soc. 122: 394-395.
[24] Wisedchaisri, G., Holmes, R.K. and Hol W.G.J. (2004). Journal of Molecular Biology. 342(4): 1155-1169.
[25] Brennan, R.G, Matthews, B.W. (1988). TIBS. 80: 6513-6517
[26] Barondeau, D., Getzoff, E.D. (2004). Curr Opin. Struct. Biol. 14(6):765-74.
[27] D.A. Case, D.A. Pearlman, J.W. Caldwell, T.E.Cheatham III, J. Wang, W.S. Ross, C. L. Simmerling, T.A. Darden, K.M Merz, R.V. Stanton, A.L Cheng, J.J Vencent, M. Crowley, V. Tsui, H. Gohlke, R.J. Radmer, Y.Duan, J.Pitera, I.Massova, G.L. Seibel, U.C. Singh, P.K. Weiner and P.A. Kollman (2002), AMBER 7, University of California, San Francisco.
[28] Pearlman, D.A., Case, D.A., Caldwell, J.W., Ross, W.R., Cheatham, T.E., DeBolt, III S., Ferguson, D., Seibel, G. and Kollman, P. (1995). Comp. Phys. Commun. 91, 1-41.
[29] DeLano WL. (2002): The PyMol Molecular Graphics System on the World Wide Web.
[30] Torsten, S., Jürgen, K., Nicolas G. and Manuel, C.P. (2003). SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Research*, Vol. 31, No. 13 3381-3385.
[31] Kelley LA, MacCallum RM & Sternberg MJE (2000). Enhanced Genome Annotation using Structural Profiles in the Program 3D-PSSM. J. Mol. Biol. 299(2), *501-522*.
[32] Garrett, M.M., David, S.G., Ruth, H., William, E.H., Scott, H., Rik, B. and Arthur J.O. 2002. Automated Docking of Flexible Ligands to Receptors (Autodock). Molecular Graphics Laboratory, Department of Molecular Biology, 10550 North Torrey Pines Road, La Jolla, CA 92037-1000, Version 3.0.5.
[33] Zheleznova, E.E., Crosa, J.H., Brennan, R.G., (2000), J. Bacteriol. 182(21): 6264-6267
[34] Althaus, E.W., Outten C.E, Olson, K.E, Cao,H., O'Halloran,T.V (1999),Biochemistry 38: 6559-6569
[35] Lesser, G.J., Rose, G. D. (1990) Proteins: structure, Function, and Genetics. 8:6-13
[36] Schiffer, C.a., Calwell, J.W., Kollman, P.A. and Stroud, R.M. (1990). PROTEINS: Structure, Function, and Genetics. 8:30-43.
[37] Akke, M., Forsen, S. (1990). PROTEINS: Structure, Function, and Genetics. 8: 23-29.
[38] De Peredo, A.G., Saint-Pierre, C., Latour, J.M., Michaud-Soret, I. and Forest, E. (2001). J. Mol. Biol. 310: 83-91.
[39] Fuangthong, M., Helmann, J.D. (2003).American Society for Microbiology. 185(21):6348-6357.
[40] Huffman, J.L., Brennan, R.G. (2002). Current Opinion in Structural Biology. 12(1):98-106.
[41] Lavrrar, J.L, Christoffersen, C.A and McIontosh, M.A. (2002). J. Mol. Biol. 322: 983-995.
[42] de-Lorenzo,V., Giovannini, F. Herrero,M.and Neilands, J.B(1987) J.Mol.Biol 203,875-884
[43] Bsat, N., Helmann, J.D. (1999). Journal of Bactriology. 181(14):4299-4307.
[44] Escolar, L., Perez-Martin, J. and de Lorenzo, V. (1999). Journal of Bacteriology. 181(20): 6223-6229.
[45] Adrait, A., Jacquamet, L., Le Pape, L., de Peredo, A.G., Aberdam, D., Hazemann, J.L., Latour, J.M. and Michaud-Soret, I. (1999). 38: 6248-6260.
[46] Baichoo, N, Helmann, JD (2002). JOURNAL OF BACTERIOLOGY. 184(21): 5826-5832.

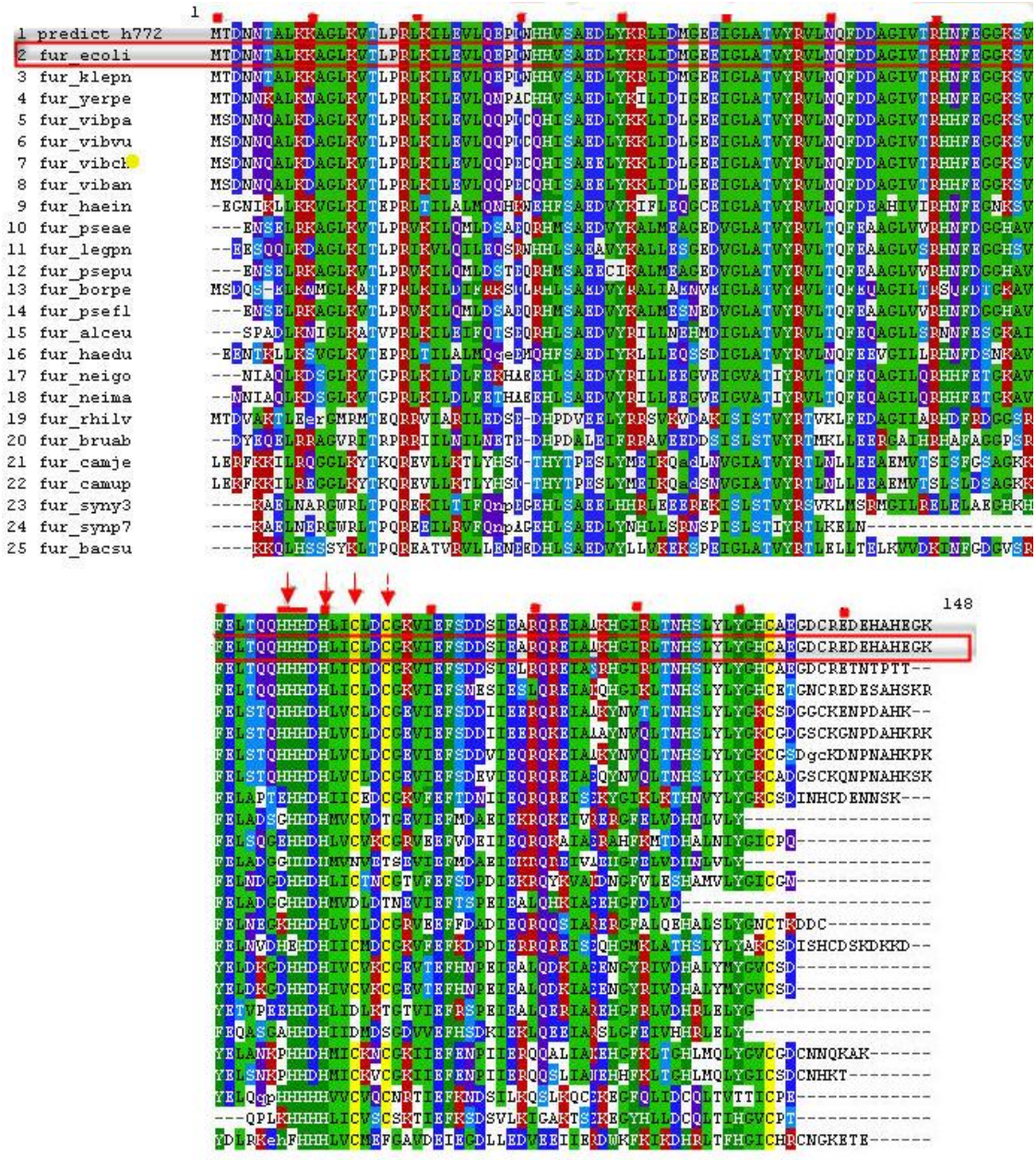


Figure 1(a) : An alignment of iron acquisition subfamily and that of Fur_E.Coli protein.Domain predicted using SWISS-MODEL server., hydrophobic residues, Green. Cystein yellow, hydrophobic acidic(D and E), dark blue. hydrophilic basic K and R, red. Polar uncharged,purpol and light blue. His, Green back ground with white text.

```
TARGET    2        TDNNTAL KKAGLKVTLP RLKILEVLQE PDNHHVSAED LYKRLIDMGE
1mzbA     1        MVENSEL RKAGLKVTLP RVKILQML-D SAQRHMSAED VYKALMEAGE
                     .*. * .********* *.***..* . .   .*.**** .** *.. **

TARGET    49     EIGLATVYRV LNQFDDAGIV TRHNFEGGKS VFELTQQHHH DHLICLDCGK
1mzbA     48     DVGLATVYRV LTQFEAAGLV VRHNFDGGHA VFELADSGHH DHMVCVDTGE
                 ..******** * **. **.*  ****.** . ****..  ** **..*.* *

TARGET    99     VIEFSDDSIE ARQREIAAKH GIRLTNHSLY LYGHCAE
1mzbA     98     VIEFMDAEIE KRQKEIVRER GFELVDHNLV LYVRKKK
                 **** *  **  **.**   . *. * .*.*  ** .
```

Figure 1(b): Alignment of Fur E.C with Fur P.A [17] PDB code :1MZB

10	20	30	40	50	60
MTDNNTALKK	AGLKVTLPRL	KILEVLQEPD	NHHVSAEDLY	KRLIDMGEEI	GLATVYRVLN
eee..e.bee	.e.e...e..	.bbe.bee.e	eee..be.bb	e.b.e.eeeb	.bbbb..bbe
eee..e.b.e	.eie.....	.b.e.b.e.e	....i.eib.	eib.e...e.	....b.ibbe
...HHHHHH.	CCC.CCHHHH	HHHHHHH.CC	CCCCCHHHHH	HHHH..CCCC	C..HHHHHHH

70	80	90	100	110	120
QFDDAGIVTR	HNFEGGKSVF	ELTQQHHHDH	LICLDCGKVI	EFSDDSIEAR	QREIAAKHGI
.bee.bbb..	.ebeeee..b	eb.ee.eb.b	bbb.e.eeb.	e.eeeebee.	.eebeee.eb
ib.e.....	....e.....	....	..b.....	....	.e...e...b
HHH.CCC.EE	EE.CCC.EEE	E..CCCCCC.	EEECCCCCEE	ECCCCCCHHH	HHH....CCC

130	140	148
RLTNHSLYLY	GHCAEGDCRE	DEHAHEGK
eb.e.eb.b.	bbb.e.eeee	eeeeeeee
........b.	b........	.e...eee
EEEEEEEEEE	EE.HHH...C	CCCCCCCC

Figure 2: Fur secondary structure : Row 1 the amino acid residues: DNA binding residues Dimerization region Iron (II) binding region. Row2 contains the predicted solvent accessibility composition (core/surface ratio) for fur protein : e: residues exposed with more than 16% of their surface, b : all other residues. Row 4 contains the observed relative solvent accessibility, where b = 0-9%, i = 9-36%, e = 36-100% . predicted solvent accessibility composition and observed relative solvent accessibilty calculated by PROF server [?].Row5 contains the predicted secondery structure from different servers (high confidance predictions only). Helix, H, Coil, C Beta strand, E.

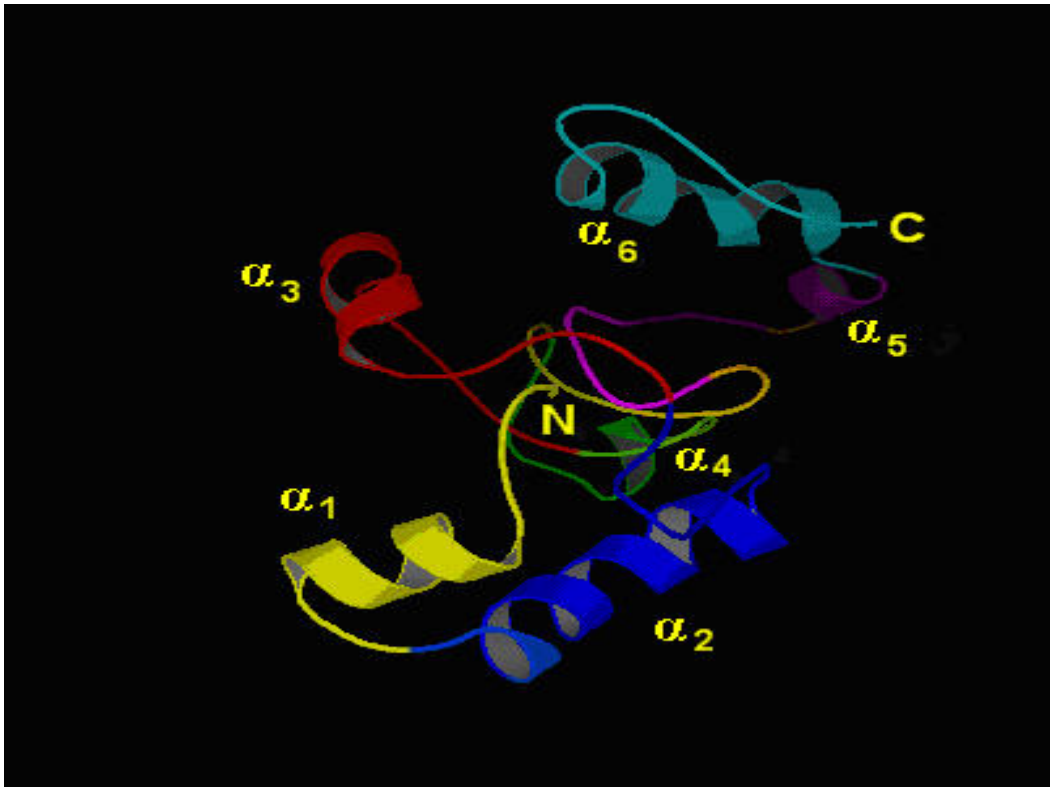


Figure 3: The three dimensional structure of the Fur monomer from E.Coli. A structuutre generated using homology modeling procedure. SWISS MODEL server Starting from N-terminal coil , α_1 yellow, α_2 blue, α_3 red, α_4 green, α_5 magenta, and α_6 aquamarine. This labeling was in comparison with the DNA binding- domains of DtxR, CAP, λ-repressor and GH5 in reference [22]**.**

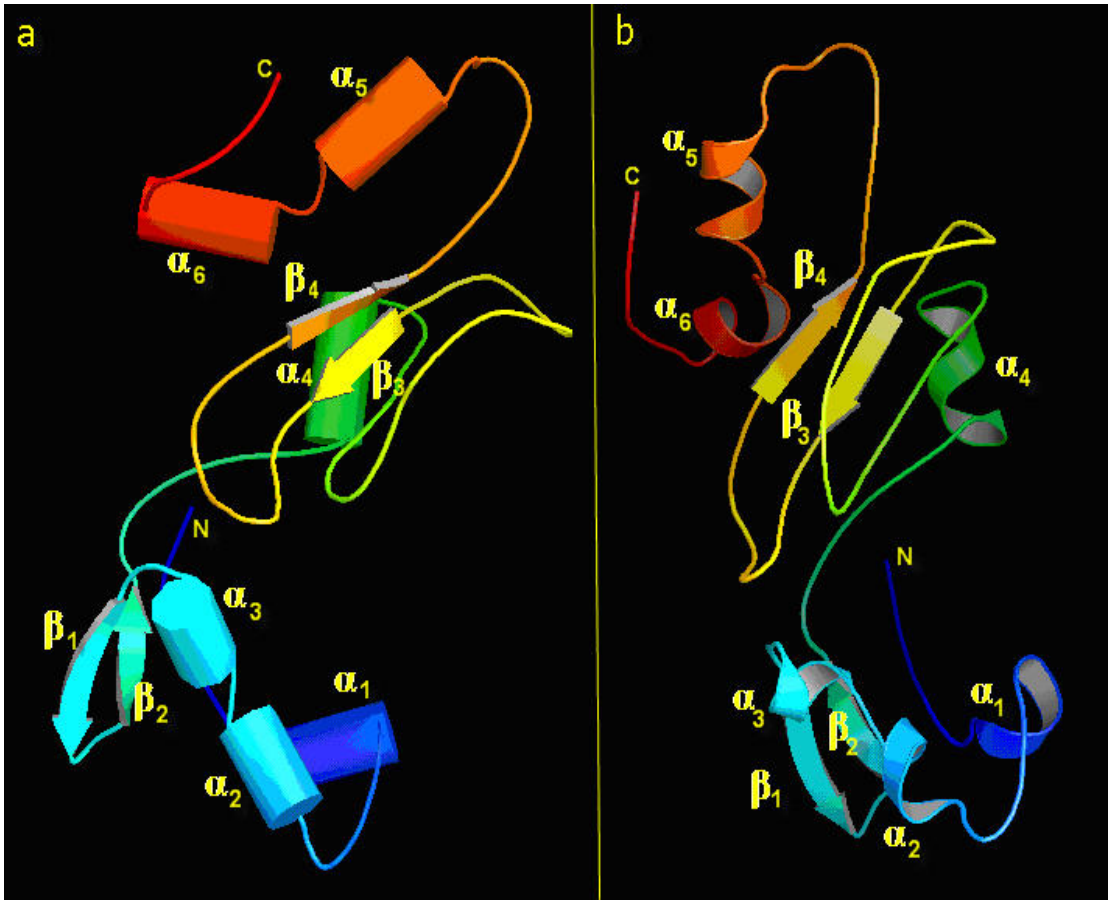


Figure 4: The three dimensional structure of Fur protein monomer from E.Coli at minimum energy calculated using AMBER7 in a water box.a) Three dimensional structure of Fur using cartoon representation. b) Using Ribbon display. Publication Colors: starting from the N-terminal domain: Blue α_1, cyan α_2, α_3, β_1, β_2 Cyan-gradual to green , α_4 green, β_3 yellow β_4 orange $\alpha 5$ light red , α_6 red**.**

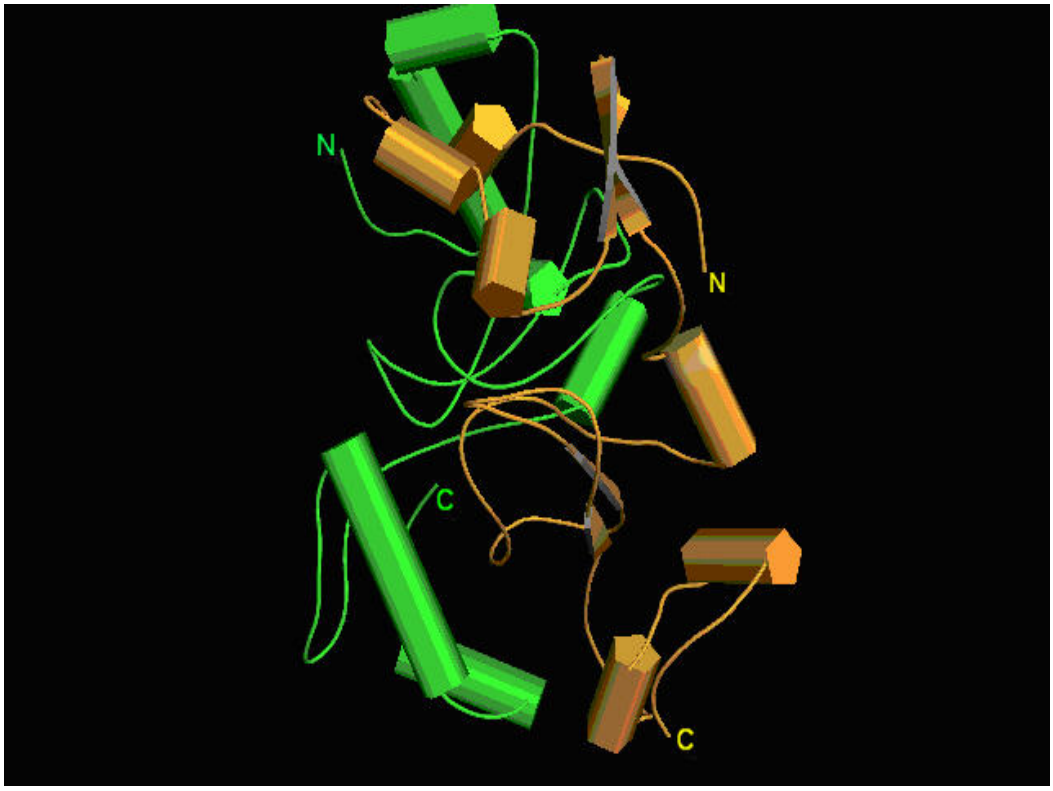


Figure 5: Fur dimer structure in a water box generated using Autodock. It shows that the closest distances occur at the central domain of each monomer while N- and C-terminal domains in each monomer are pointing away from each other. Colors: one subunit is gold color, the other is green.

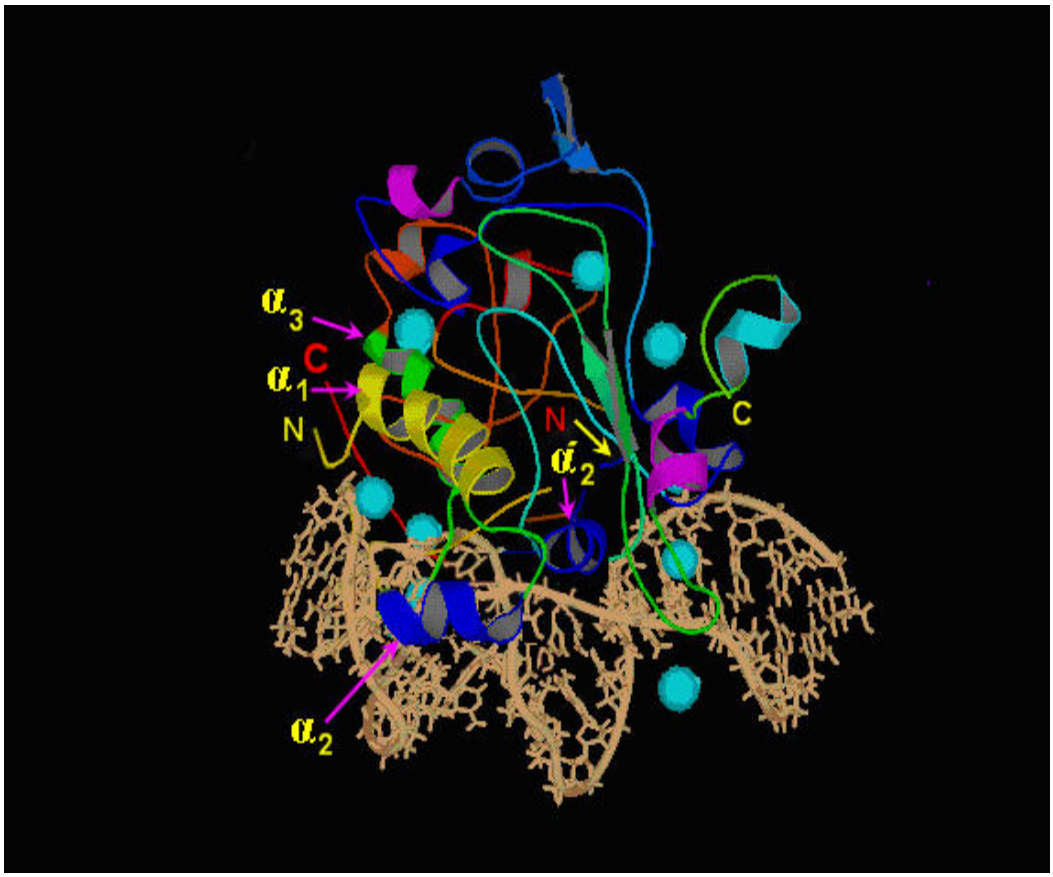


Figure 6(a): The interaction of the Fur dimer with DNA in the presence of Na^{+}, $8Fe^{2+}$ ions and using H_2O as solvent. The α_2 and α'_2 helices (Blue) interact with the AT-rich region of the major groove of the conical B-DNA (iron box).

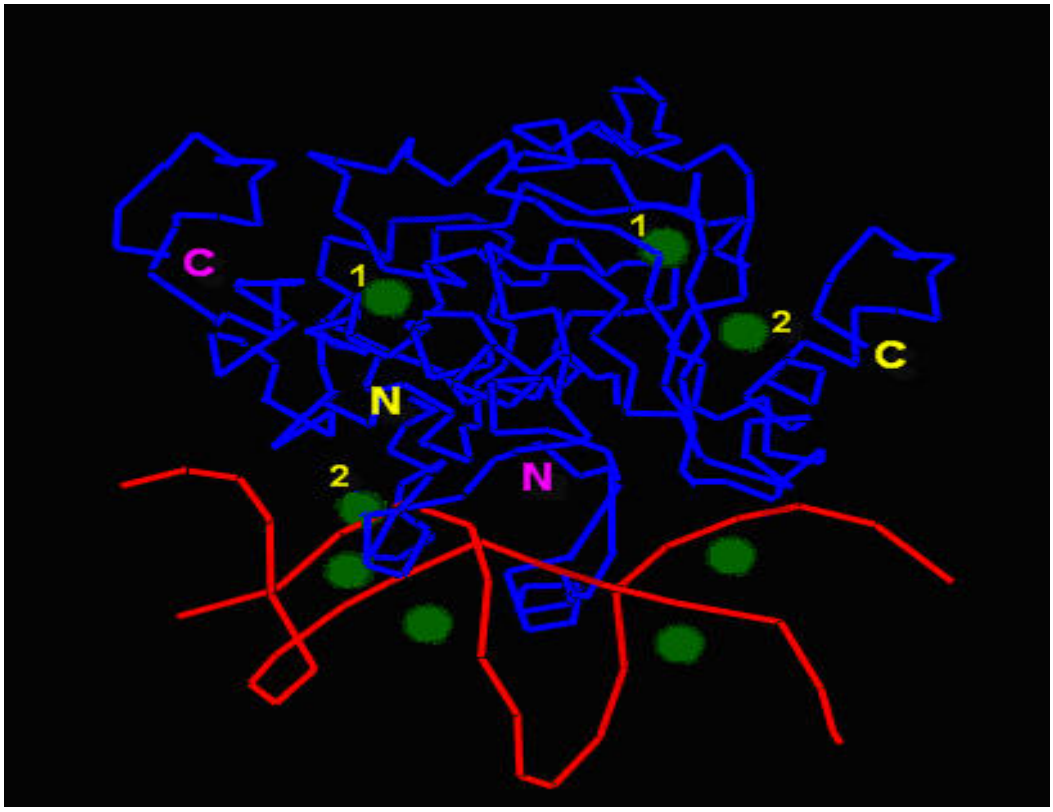


Figure 6(b): line structure of the Fur dimer interacting with DNA, conditions as in 6(a). The figure shows the major Fe^{2+} sites 1 and 2, and the other four Fe^{2+} ions are close to the DNA.

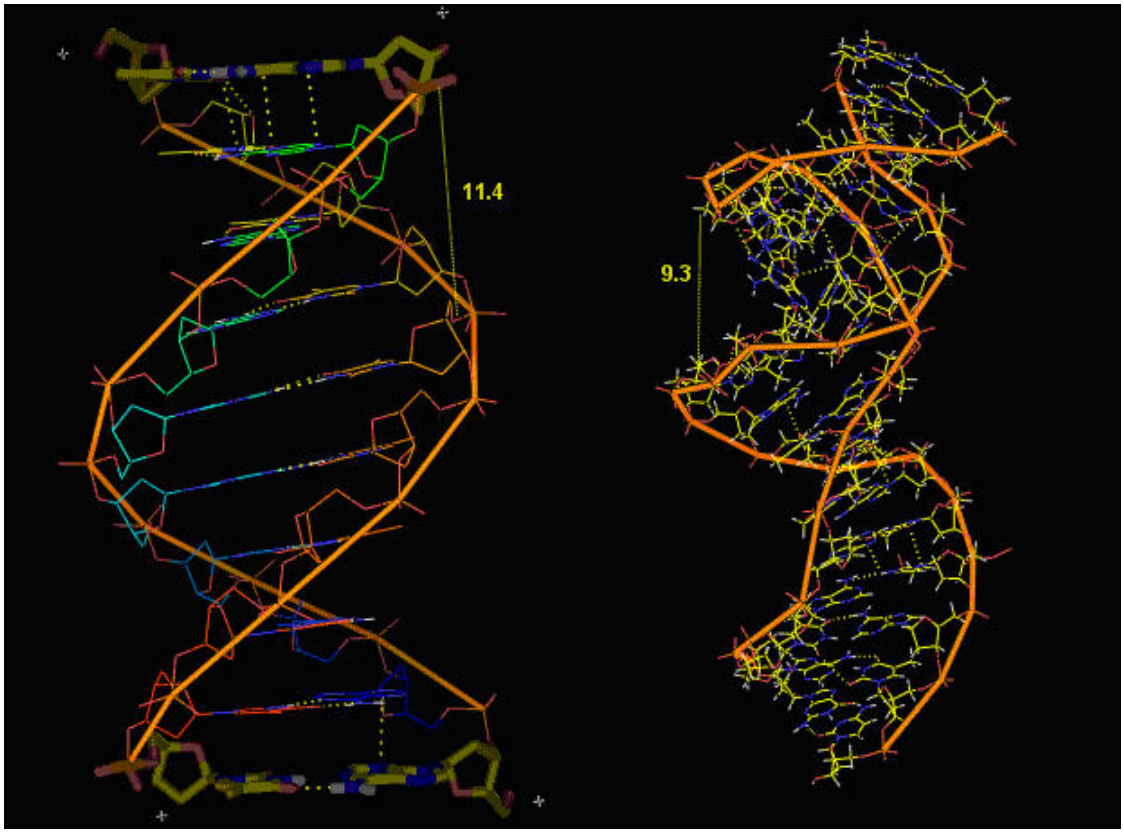


Figure 6(c): Tilting of DNA in the presence of Fur dimer,8 Fe^{2+}, Na^{+} in water: The three diemesional structure of the conecial B-DNA, before binding to the fur dimer (left) and after binding the fur dimer (right). The calculated distances between phosphates in the backbone in the first major grove of the two models are shown.

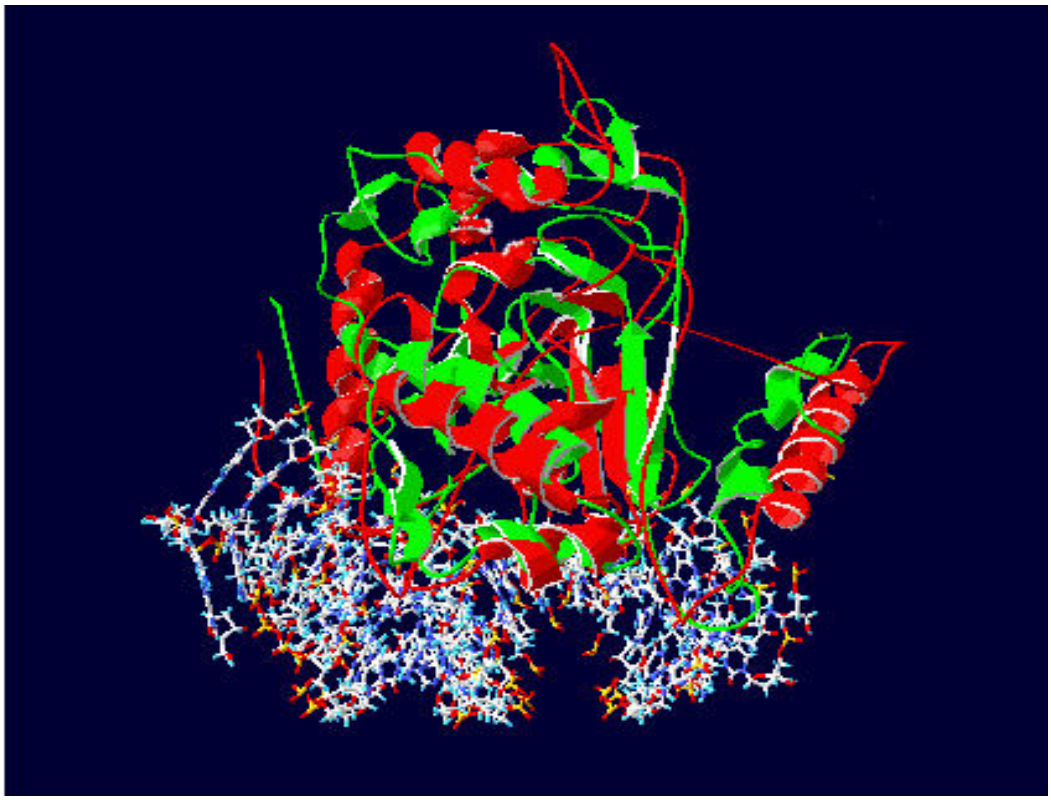


Figure 7(a): Conformational changes in Fur dimer upon binding to Fe^{2+} and DNA:
The fur dimer DNA complex no Fe^{2+} present (red) and after adding $8Fe2^{+}$ions (green).

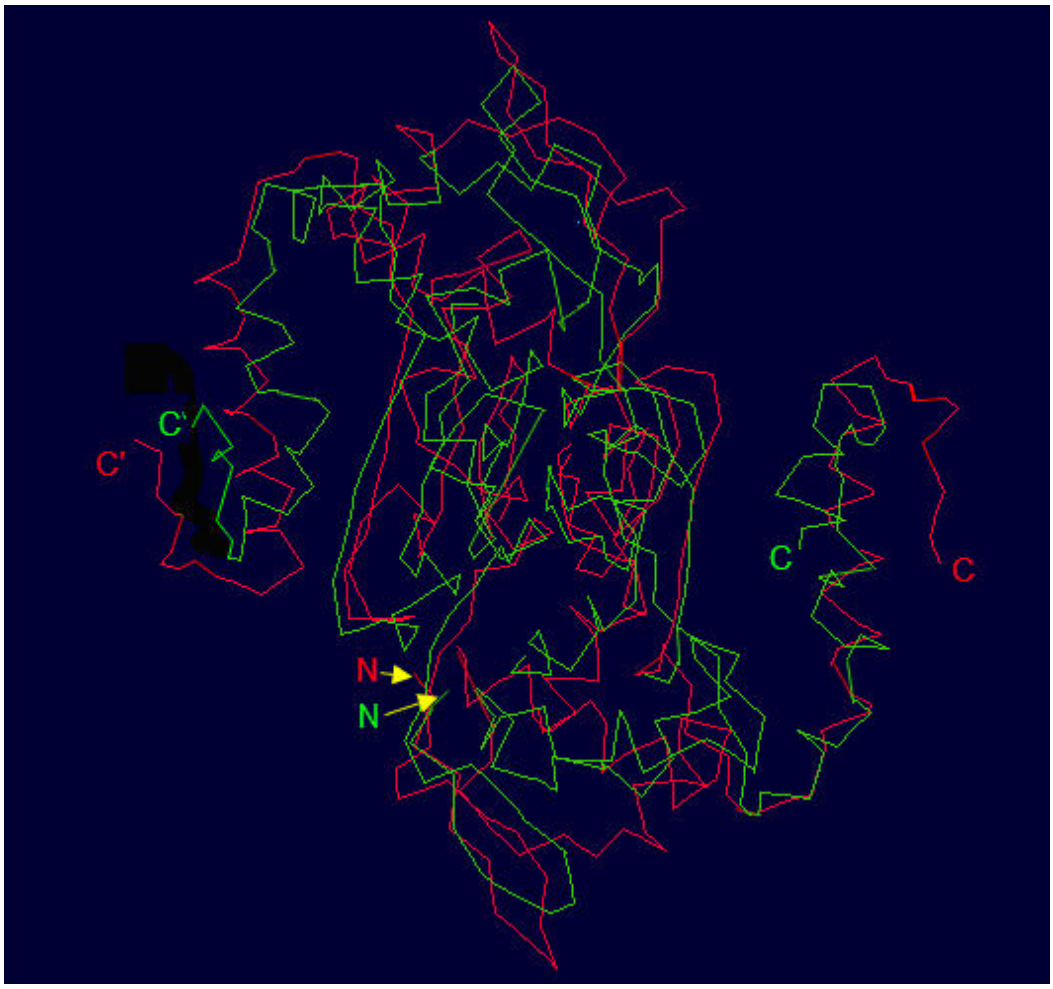


Figure 7(b): Two models of the three dimensional structure of the Fur dimer displayed in line mode. Before adding Fe^{2+} (red) the conformational changes appear between the two models. After adding $8Fe^{2+}$ ions and DNA binding (green).

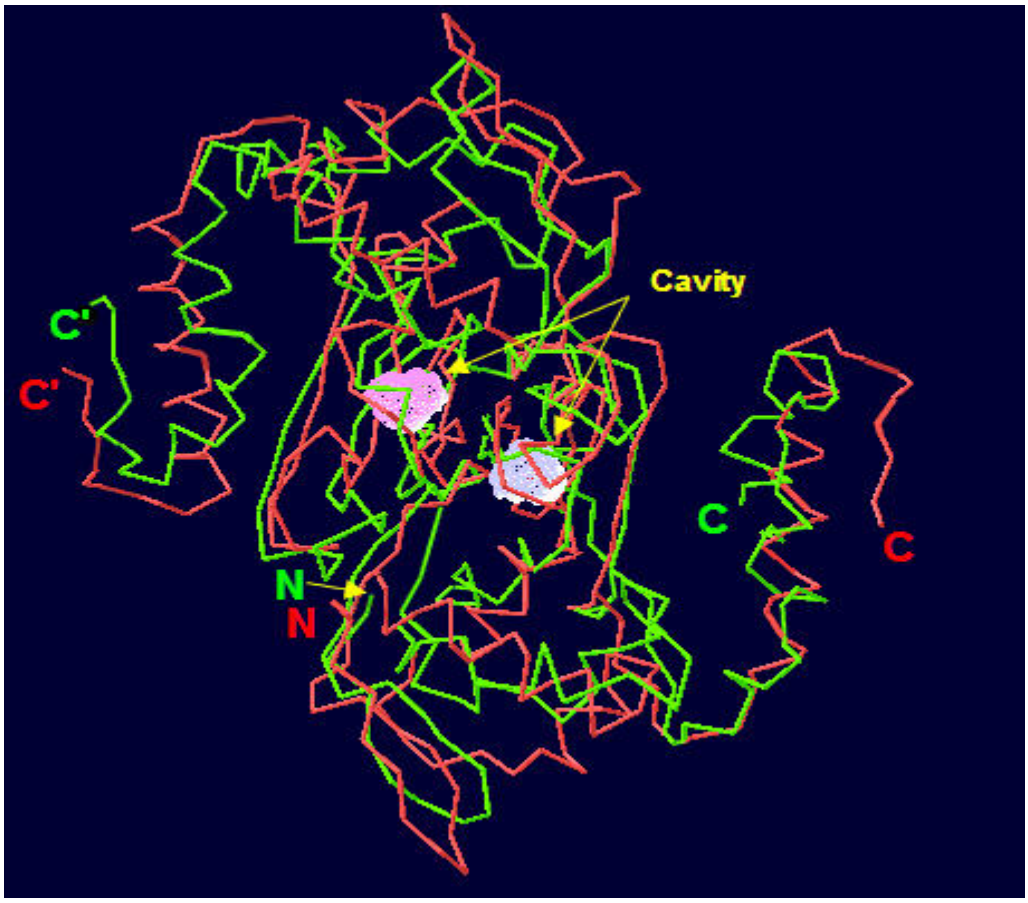


Figure 7(c): The same model as in Figures 6(b) showing the cavities Lilac colored cavity is for red model before adding Fe^{2+}, and the off-gray cavity for the green model after adding the $8Fe^{2+}$, the shift in cavity position upon adding iron is apparent.

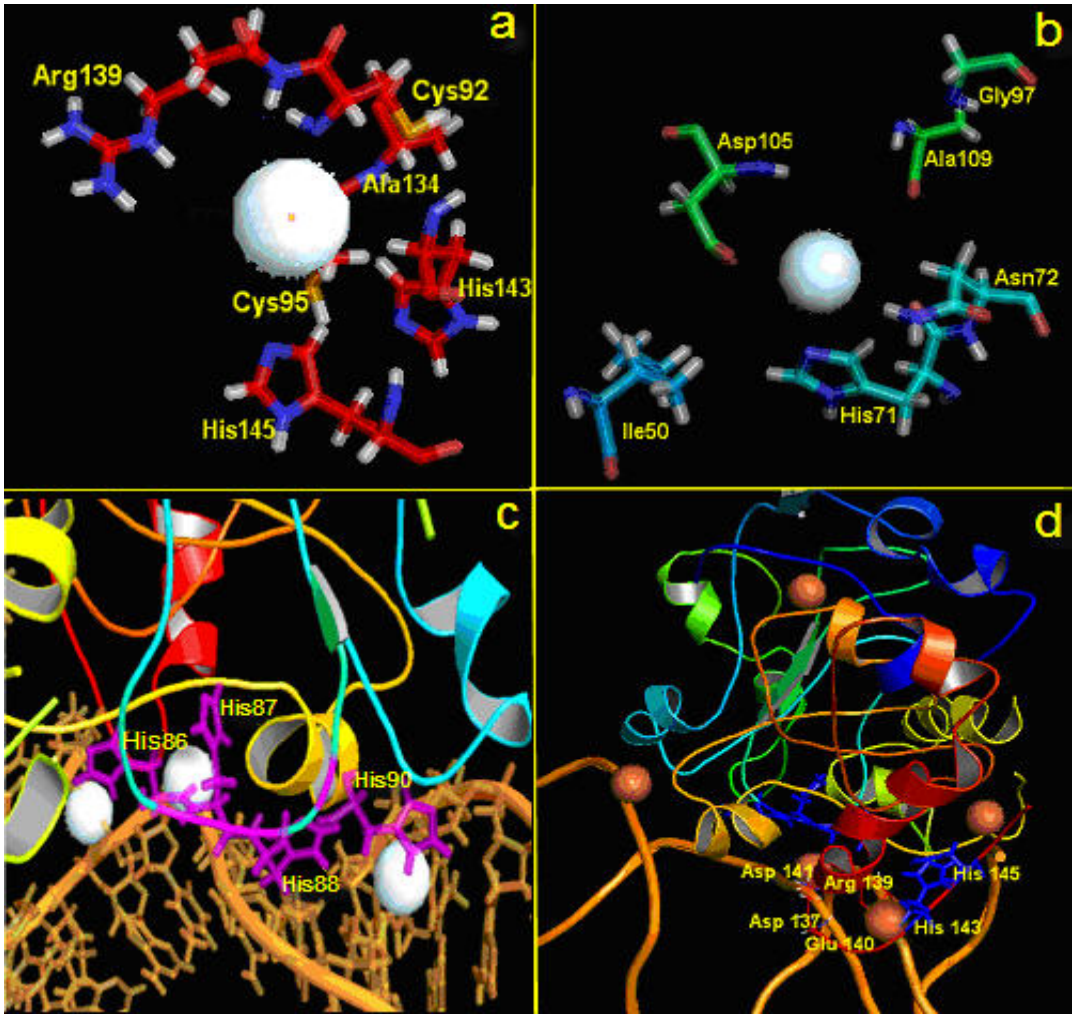


Figure 8: (a) Close-up View of the coordination at metal binding site 1. (b) Close-up view of metal site 2. **(b):** Close-up View of the residues and Fe^{2+} near the DNA. A metal ion is present between His 86, His 87, His 89 and His 90 and AT of DNA (for distances see Tables 3 &4). The recognition site for Fe^{2+} the motif H86H87H88D89H90 binds DNA mediated by Fe^{2+} [43]. **(c):** Close-up view of Fe^{2+} site 1 close to the DNA. Ligands provided by C-terminal are Asp 137, Arg 139, Glu 140, , Asp 141, and His 145, Fe^{2+} ion in DNA groove shown in bright green (calculated distances are shown in Table 3 and 4)

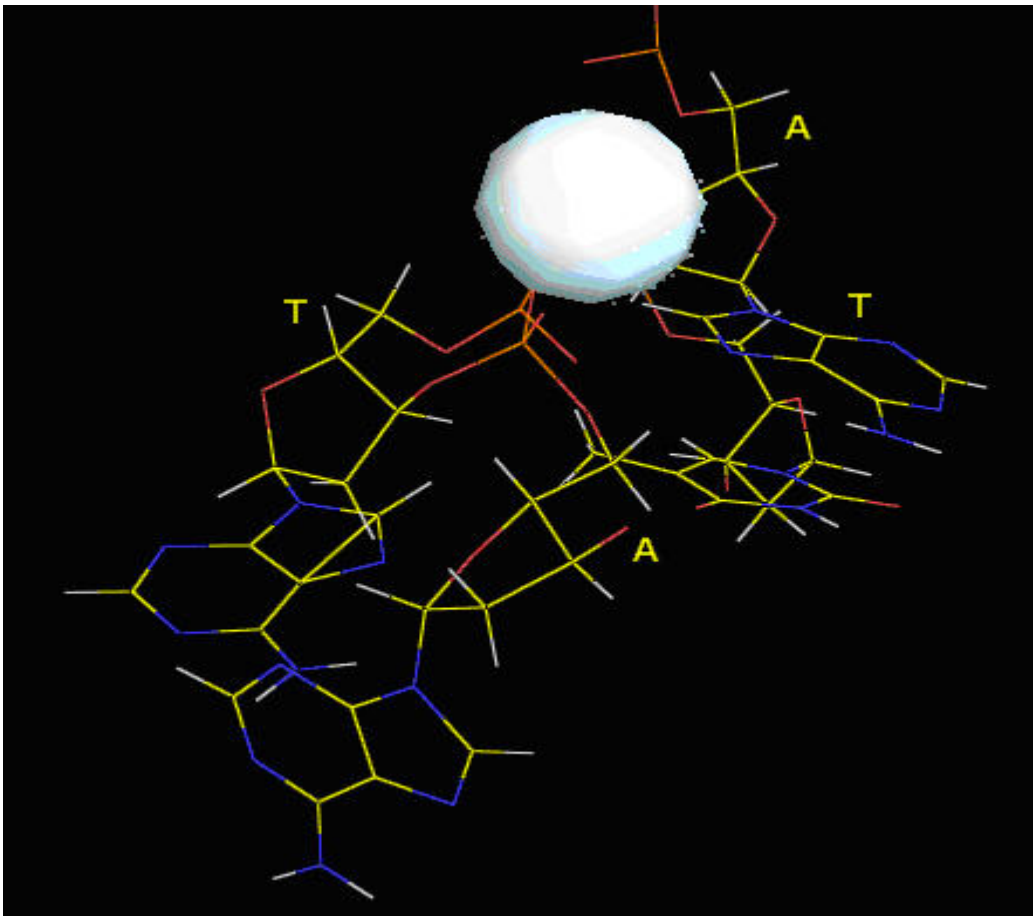


Figure 9: Close-up view of Fe2+ binding to DNA (A.T region) at elevated concentration, conditions as in Figure 6

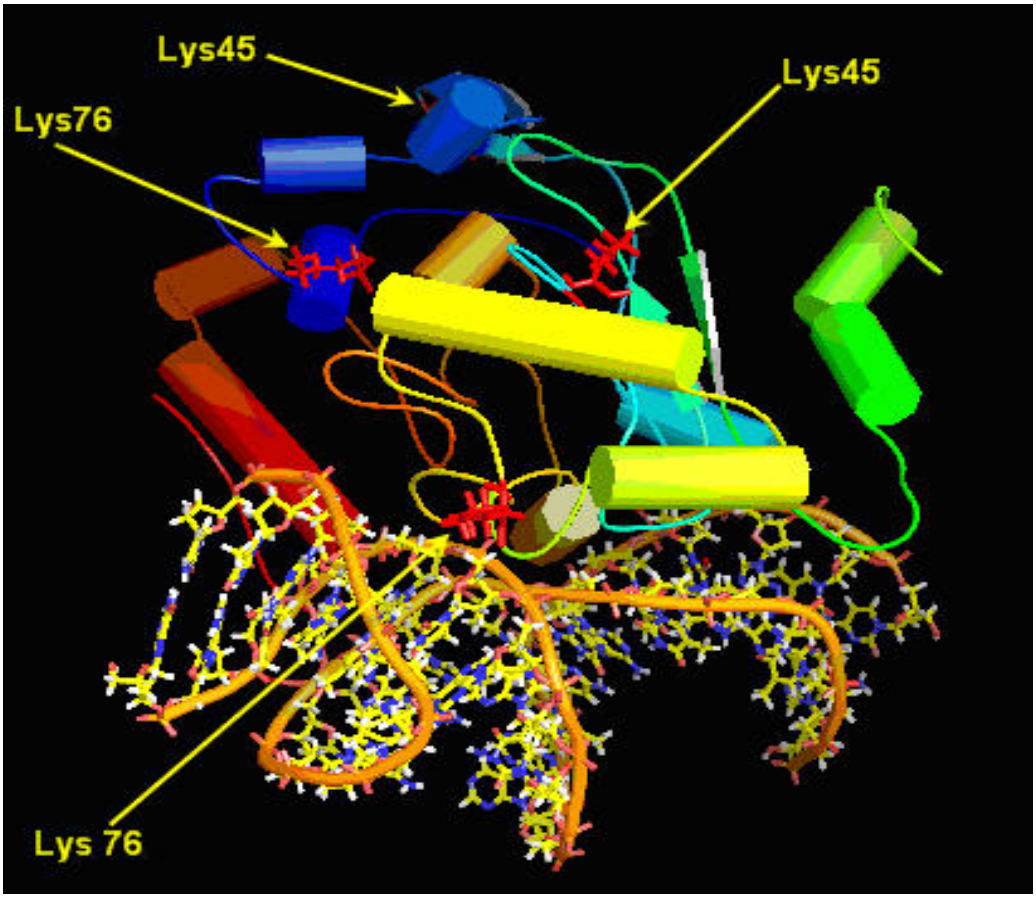


Figure 10: The fur dimer binding to DNA in the presence of Na+, H2o and 8Fe2+. Testing the effect on Lysine 45 and Lys 76 hydrophobic residue which was reported by de Peredo et al [38]. Lys 76 proved to be highly protected from modification upon fur DNA binding (Lys76 present in the wing and may interact with DNA). The result was interpreted as change in Fur conformation upon activation.

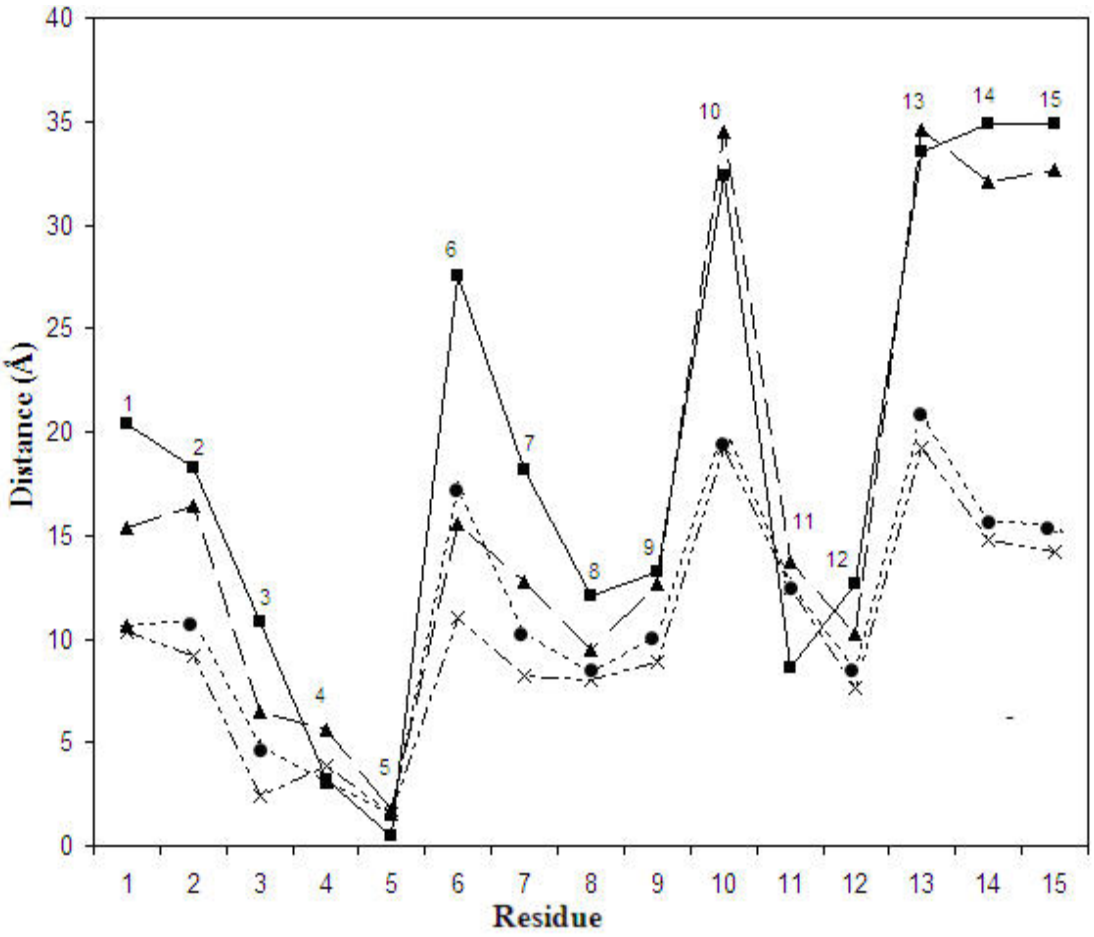


Figure 11(a): Conformational changes of the fur EC induced by DNA and Fe^{2+} binding. Distance between residues and helices on one fur subunit and the other. Apofur dimer (■). Apofur/DNA (▲). Fur/DNA in the presence of 4 Fe^{+2} ions (●) and Fur/DNA in presence of 8 Fe^{2+} ions (⨯). Labels on the plots are as follows: N-terminal-N-terminal (1), α_1- α_1 (2), α_2- α_2 (3), Val25-Val25(4), Pro29-Pro29(5), α_3- α_3 (6), Glu49-Glu49 (7), Thr69-Thr69 (8), α_4- α_4(9), Gln85-Gln85(10), Ala53-Ile107 (11),; Arg112-Arg112 (12),; α_5- α_5(13), α_6- α_6(14), C-terminal-C-terminal(15).

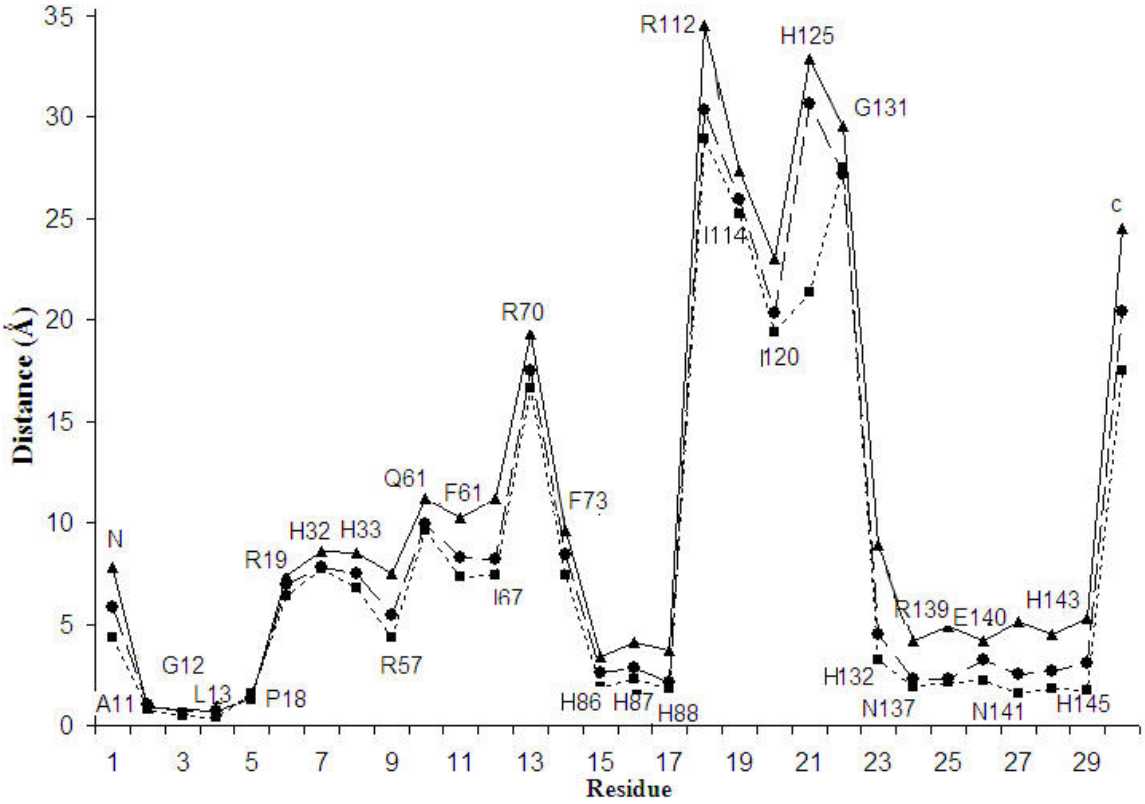


Figure 11(b): Conformational changes of the FurEC dimer And DNA binding. The Calculated distances between the amino acid residues of fur and the AT-unit in the B-canonical DNA (Table 3). Fur dimer and DNA fragment (▲) (continuous line). Fur dimer and DNA in the presence of 4 Fe^{+2} ions (●) (broken line). Fur dimer and DNA in presence of 8 Fe^{+2} ions (■)(dotted line). This plot show that residues A11, G12, L13 P18 and R19 near the N-terminal, His88 to R112, and the residues139—145 near the C-terminal are the closest to DNA.

Table1:
Results of homology modeling of fur from different sources compared to that predicted by NMR study [13-15] ; column 2 shows the proposed role reported for each domain in literature .

Residues near N-terminal♣		Folding predicted by NMR [15]	folding	Confidence level
4 to 6	**DNA binding HTH motif**		**α 1 helix H1**	**9**
11 to 16			**Coil T1**	**8 to 9**
17 to 27			**α 2 helix H2**	**9**
29 to 35		**coil**	**Coil T2**	**9**
36- 44	**wing**	**helix α**	**α 3 helix H3**	**9**
47-51 (G)	**Dimerization Region**	**44-48 coil**	**Coil T3**	**9**
52-63	**Y55-F61 suggested DNA binding domain [15]**	**49-59 helix 60-64 coil**	**α 4 helix H4**	**9**
65,66,67		**65-74 helix**	**Coil T4**	**8 to 9**

Residues near C-terminal		**NMR predicted**	**folding**	**Confidence level**
69-72			**β 1 sheet**	**6 to8**
74-76 LYS	**Metal ion binding sites**		**Coil T5**	**8 -9**
78-81			**β 2sheet**	**9**
83-89 contain His			**Coil T6**	**8**
90-93 contain His and Cys 92			**β 3sheet**	**8**
94-98, Cys 95			**Coil T7**	**8-9**
99-101			**β 4 sheet**	**8 -9**
102-107			**Coil T8**	**8-9**
108-113		**107-117 helix α**	**α 5 helix H5**	**7, 8 to9**
118-120			**Coil T9**	**7,9,6**
121-132			**β 5 sheet**	**9**
134-136			**α 6 helix H6**	**8,8,7**
140-148			**Coil T10**	**7, most 9**

♣ Residue numbering is shifted by one in our case , in literature reports ,usually the first residue M is ignored For example C92 is labeled C93.

Table2:
Conformational changes of fur EC dimer induced by DNA and Fe^{2+} binding as indicated by distances between the residues on one subunit relative to the accompanying residue on the other: The calculated distances between residues on each monomer of the Fur dimer. The first column for apofur dimmer, 2nd column for apofur dimer with DNA; the last two columns show the distances after adding Fe^{2+}.

Residue	ApoFur dimer	ApoFur dimer/DNA	Fur dimer/DNA+ $4Fe^{2+}$	Fur dimer/DNA + 8 Fe^{2+}
N-terminal-N-terminal	20.4	15.4	10.6	10.3
$\alpha_1 \rightarrow \alpha_1$[1]	18.3	16.4	10.9	9.2
$\alpha_2 \rightarrow \alpha_2$	10.8	6.5	4.8	2.4
Val 25- Val 25	3.2	5.60	3.10	3.9
Pro29-Pro29	0.5	1.7	1.5	1.5
$\alpha_3 \rightarrow \alpha_3$	27.5	15.6	17.3	11.0
$\alpha_4 \rightarrow \alpha_4$	13.2	12.7	10.1	8.9
Leu52-Leu82	0.7	1.2	1.6	1.2
Gly51-Gln85	0.02	0.02	0.5	0.5
Glu49-Glu81	0.02	0.4	0.09	0.09
Thr54-Thr83	0.5	1.2	0.9	0.7
Glu49-Glu49	18.2	12.8	10.2	8.2
Thr69-Thr69	12.1	9.5	8.4	8.0
Gln85-Gln85	32.4	34.5	20.4	19.2
Ala53-Ile107	8.60	13.7	12.7	12.5
Thr54-Glu108	9.50	11.8	9.3	8.9
$\alpha_5 \rightarrow \alpha_5$	33.5	34.6	20.8	19.2
Arg112-Arg112	12.7	10.2	8.5	7.6
$\alpha_6 \rightarrow \alpha_6$	34.9	32.1	15.7	14.8
C-terminal-C-terminal	34.9	32.7	15.6	14.2

[1] Helix-Helix distance was measured centre to centre

Table 3: Distances between fur residues and AT of DNA: Column (A) apofur dimer/ DNA (no iron present). Column (B) fur dimer /DNA + $4Fe^{2+}$. Column (C) fur dimer /DNA + 8 Fe^{2+}.

Residue	(A) fur dimer/DNA distance (Å)	(B) fur dimer/DNA+ $4Fe^{2+}$ distance (Å)	(C) fur dimer/DNA+ $8Fe^{2+}$ distance (Å)
N-terminal	7.8	5.8	4.3
Ala 11	0.9	1.0	0.8
Gly 12	0.8	0.7	0.5
Leu 13	0.7	0.7	0.4
Pro 18	1.3	1.3	1.6
Arg 19	7.4	6.9	6.4
His 32	8.6	7.8	7.7
His 33[a]	8.5	7.5	6.8
Arg 57[b]	7.5	5.4	4.3
Gln 61[c]	11.2	9.9	9.6
Phe 62[b]	10.2	8.3	7.3
Ile 67[c]	11.2	8.2	7.4
Arg 70	19.3	17.5	16.6
Phe 73	9.6	8.4	7.4
His 86	3.4	2.6	1.9
His 87	4.1	2.8	2.3
His 88	3.7	2.1	1.8
D89[d]	3.9	3.2	2.5
H90[d]	4.1	3.4	2.9
Arg 112	34.5	30.3	28.9
Ile 114	27.3	25.9	25.2
Ile 120	23.0	20.3	19.4
His 125	32.8	30.6	21.3
Gly 131	29.5	27.2	27.5
His 132[a]	8.9	4.5	3.2
Asp 137	4.2	2.3	1.9
Arg 139	4.9	2.3	2.1
Glu 140	4.2	3.2	2.2
Asp 141	5.1	2.5	1.6
His 143[e]	4.5	2.7	1.8
His 145	5.3	3.1	1.7
C-terminal	24.5	20.4	17.5

Table 4: Calculated distances between Fe(II) and closest residues on the fur for the first two iron ions added.

Residue	Position of residue in structure	Donor atom (type of interaction)	Residue-Fe(II) Distance (Å)
Site 1 (Zn site)			
Fe-Cys 92	**coil**	**H-bonded H_2O**	**2.2**
Fe-Cys 95	**coil**	**H-bonded H_2O**	**1.6**
Fe-Asp 137	**coil**	**O**	**1.3**
Fe-Asp 141	**coil**	**O**	**1.5**
Fe-Arg 139	**coil**	**N**	**1.7**
Fe-Glu 140	**coil**	**O**	**1.3**
Fe –His 145	**coil**	**N**	**1.2**
Site 2			
Fe-His 71	**End of β strand**	**N**	**1.3**
Fe-Ile 50	**Coil**	**hydrophobic**	**2.3**
Fe-Asn 72	**β strand**	**N**	**1.5**
Fe-Gly 97	**Coil**	**polar**	**2.3**
Fe-Asp 105	**Coil**	**O**	**1.4**
Fe-Ala 109	**α helix**	**hydrophobic**	**2.1**
Other Residues at close proximity to Iron			
Fe-His 32	**α helix**	**N**	**3.6**
Fe-His 33[a]	**α helix**	**N**	**4.2**
Fe –Arg 57	**α helix**	**N**	**5.1**
Fe –Gln 61	**coil**	**N, O**	**4.9**
Fe –Phe 62	**coil**	**hydrophobic**	**7.9**
Fe –Ile 67	**coil**	**hydrophobic**	**8.3**
Fe –Arg 70	**coil**	**N**	**3.4**
Fe –Phe 73	**coil**	**hydrophobic**	**3.1**
Fe –Ile 114[b]	**coil**	**hydrophobic**	**4.9**
Fe –Ile 120[b]	**coil**	**hydrophobic**	**6.2**
Fe –His 132[a]	**α helix**	**N**	**5.4**
Fe-His 86	**coil**	**N**	**4.1**
Fe-His 87	**coil**	**N**	**3.7**
Fe-His 88	**coil**	**N**	**4.2**
Fe-H90[c]	**coil**	**N**	**3.9**
Fe-D89[c]	**coil**	**O**	**4.2**

[a] The largest effect on nmr shift was observed for H33 upon addition of Mn^{2+} [13]
[b] Considerable change in nmr shift was observed upon titrating Fur-Mn^{2+} with DNA [13]
[c] Possible ligands for iron(II) in regulatory site in vivo as reported by Bsat et al [43]