Scuola Normale Superiore

Classe di Scienze

**Tesi di perfezionamento in Chimica**

# NMR study of CyaY and YFHJ, two proteins involved in iron metabolism

*Candidata*: *Chiara Pastore*

*Relatori: Prof. Piero Salvadori*

*Prof. Lorenzo Di Bari*

Anno Accademico 2005-2006

# ACKNOWLEDGEMENTS

**TABLE OF CONTENTS**

# Summary

In this thesis we apply Nuclear Magnetic Resonance to gain insights into the structures and functions of the two proteins CyaY and YFHJ. The overall material is organised in eight chapters, an introduction and a conclusion.

**Chapter 1**: In this chapter the basis of Nuclear Magnetic Resonance are provided, with reference to its application to the study of biological macromolecules

**Chapter 2**: This chapter describes the 2D and 3D NMR experiments that allow the assignment of a protein's spectra.

**Chapter 3**: Here a short description of the bioinformatic tools and software packages available to extract information on macromolecules is provided. The material is organised following the proteins' level of structure it refers to: first there is a short description of the web databases and the programs that can be used to extract information from the primary sequence of a protein, then an elucidation of the programs useful to analyse the secondary structure and some details on the one applied in this study to calculate the 3D structure; finally, the available tools for structure comparison and refinement are reported.

**Chapter 4**: This chapter summarises what is known about Friedreich's ataxia and frataxin.

**Chapter 5**: Here the study of CyaY iron binding site by means of NMR is presented. In the first part of the chapter the problems and advantages of using paramagnetic probes in NMR are briefly discussed.

**Chapter 6**: This chapter reviews the systems involved in iron-sulphur clusters' biosynthesis with particular attention to the proteins encoded by the bacterial Iron Sulphur Cluster operon.

**Chapter 7**: An elucidation of the 3D structure determination process of the protein YFHJ is reported. In the first part circular dichroism and its application in biomolecular studies are discussed.

**Chapter 8**: This chapter illustrates the functional characterisation of the protein YFHJ by means of NMR, mass spectrometry and fluorescence polarisation. The techniques used are elucidated in the initial paragraphs.

# Introduction

Nuclear Magnetic Resonance is a well established tool to investigate biomolecular problems [1-3]. In the following work we apply this technique to the study of two proteins involved in iron metabolism. In the first part of the thesis NMR is used to characterise the iron binding site of the bacterial orthologue of the protein frataxin, while in the second it is employed to carry out a structural determination and functional investigation of the Iron Sulphur Cluster (ISC) protein YFHJ. The functional studies in this case are integrated with other biophysical techniques such as CD spectroscopy, fluorescence and mass spectrometry.

Iron is a ubiquitous element in biology and defining its homeostasis, i.e. how cells and organisms regulate their iron content and how diverse tissues orchestrate its allocation, would be of capital importance for understanding why the mechanism sometimes can be so much compromised to cause pathological conditions, from haematological, to metabolic to neurodegenerative diseases [4].

Eukaryotic and most of the prokaryotic organisms depend on iron for survival and proliferation; the metal ion in fact is the constituent of hemoproteins, iron-sulphur proteins, and proteins that use it in other functional groups to carry out their housekeeping functions.

The concentration of iron in the cells must be tightly regulated: if on one hand iron deficiency is lethal, on the other iron accumulation is highly toxic. In fact, free iron in the cytosol, normally present in its reduced form, can participate to the Fenton chemistry, i.e. can be oxidised by $H_2O_2$ with consequent production of free radicals and damage of lipid membranes, proteins and nucleic acids.

Most of the iron within the eukaryotic cells is routed to the mitochondria, where a substantial amount is employed for the biosynthesis of heme and Fe-S clusters. The latter process in particular, is fundamental for maintaining mitochondrial iron homeostasis. Work done on yeast showed that strains depleted of proteins crucial for the clusters' biosynthesis suffered from a marked iron accumulation in their mitochondria, as well as mitochondrial oxidative damage.

Because Fe-S clusters are key factors for the sensing and regulation of many bacterial transcription proteins [5], it was proposed that they are also involved in a feedback regulation mechanism of mitochondrial iron import/export. A Fe-S protein

would behave as a sensor for the mitochondrial iron status, and, depending on the oxidation state of its cluster, would signal the need of importing or exporting iron. In this scenario, a defect in the synthesis of Fe-S clusters would compromise the regulation process, and the decreased concentration of the signalling protein could be interpreted as a signal for iron import. In this way the cytosolic iron pool would be depleted, and the cell would enter a vicious cycle which would exacerbate mitochondrial iron overload.

Abnormal Fe-S protein biogenesis and mitochondrial iron accumulation in heart and neurones are typical phenotypes of a very serious form of genetic neurodegeneracy, Friedreich's ataxia [6,7]. This pathology is caused by deficiency of the protein frataxin, a mitochondrial matrix protein, ubiquitously expressed and extremely conserved throughout the species [8]. The function of the protein has not been clarified yet but its interaction with iron was demonstrated *in vitro* with different biophysical techniques. Based on evidences from the *in vitro* and *in vivo* studies, it was suggested that frataxin has a role in iron storage or that it constitutes the iron chaperone in the biosynthesis of heme or Fe-S clusters [9-17].

Some literature reports state that human and yeast frataxin form high molecular weight aggregates in the presence of iron *in vitro* and that therefore the protein has a ferritin-like behaviour [13,18-20]. Other studies made on yeast though, show that aggregation is not essential *in vivo*, supporting therefore the chaperone role of the molecule [20]. Despite its involvement in iron metabolism, frataxin does not possess any of the typical motifs designed by macromolecules to chelate the metal ion [21]. Although depletion of CyaY alone does not cause iron accumulation in bacterial cells [22], the high conservation of the sequence and of the 3D fold amongst the frataxin family members [23], supported our choice to use CyaY, a protein extremely suitable for NMR, to carry out our study. We used NMR chemical shift perturbation to identify and characterise the iron binding site. The specificity issue is also addressed, with NMR titrations performed using a wide range of metal ions. The final aim is to provide structural evidence for discriminating between the current working hypotheses on frataxin's role.

As briefly mentioned earlier, CyaY is thought to have an active role in the Fe-S clusters' biosynthesis, although its actual function within the mechanism is still a matter of debate. Other aspects of the clusters' synthetic pathway have not been elucidated yet, in particular the specific function of many of the involved proteins. In order to further investigate the mechanism in simple organisms, we decided to carry out a structural and functional characterisation of the protein YFHJ from *E. Coli*. YFHJ is a hypothetical protein, whose open reading frame (ORF) has been found in

the *isc* operon of *A. Vinelandii* upon genome sequencing [24]. The coding sequence is conserved in some classes of prokaryotes and of very elemental eukaryotes, but it is not present in superior organisms. It is possible that YFHJ has been substituted by another protein in its function or, alternatively, that it represents the remains of a primitive requirement of certain species, no longer necessary in Eukaryotes. In the second part of the thesis the determination of the 3D solution structure of the protein by biomolecular NMR is described. Knowing the structure of a protein is valuable because it contains the details of its function and because it provides the starting point for analysing at a molecular level its interactions with the surrounding molecules [2]. In the present work the interactions of YFHJ with the desulphurase IscS, CyaY and iron will be studied by means of NMR chemical shift perturbation, fluorescence polarisation and mass spectrometry.

These studies will hopefully provide important pieces of information useful to understand how this protein is involved in the clusters' synthesis and why it has been lost during evolution.

# Biomolecular NMR 1

In the past years, Nuclear Magnetic Resonance (NMR) has proved to be a valid methodology for the determination of the three-dimensional structure of macromolecules, together with X-ray diffraction (XRD). The two techniques are in any respects complementary. On one hand XRD allows a very detailed description of the atomic positions of macromolecules and their complexes, with no size limitations. The bottleneck step of XRD is crystallisation: not all the proteins in fact form the large, individual and well ordered crystals necessary for obtaining a good diffraction pattern, and very often a long screening process is necessary to find the best crystallisation conditions. However, once the crystals have been produced, the structure determination is straightforward, and leads to a very resolute, albeit static, structure of the molecule. On the other hand, NMR can be performed in solution, in almost native conditions and does not require a cristalline sample. It can be extremely effective to study small molecules although data acquisition and manipulation can be a lengthy process. This implies that the macromolecule of interest must be stable in solution over the time. Moreover it presents some drawbacks with high molecular weight species: molecules larger than 40 kDa are difficult to study with this technique essentially because of the increased complexity of the spectra and the large line broadening caused by fast transverse relaxation. Application of deuterium labelling and of the TROSY technique [25,26] can in part overcome these problems, allowing structure determination of up to 100 kDa molecules.

Nevertheless, NMR has the unique capability to investigate dynamic processes. The effect of external factors, such as pH, temperature, denaturing agents can be easily monitored and the dynamic properties of the molecules can be investigated over a range of different time scales. It can thus be exploited to analyse a protein function, folding and intra- and inter-molecular interactions [27].

## 1.1 *Physical principles of NMR*

Nuclear Magnetic Resonance phenomenon arises from the interaction of a magnetic nuclear dipole with electric or magnetic fields originated by an external apparatus or by the sample itself (*external* and *internal spin interactions*) [28].

The origin of the magnetic nuclear dipole $\underline{\boldsymbol{\mu}}$ is the non-zero intrinsic angular momentum of nuclei with spin number I $\geq$ ½. Equation 1.1 describes the relationship between the two properties:

$$\underline{\boldsymbol{\mu}} = -\hbar\gamma_I\underline{\mathbf{I}}$$

(1.1)

where $\gamma_I$ is the magnetogyric ratio specified in units of rad s$^{-1}$ T$^{-1}$, and $\underline{\mathbf{I}}$ is the observable associated with the quantum mechanical spin operator $\underline{\hat{\mathbf{I}}}$. The projection of $\underline{\hat{\mathbf{I}}}$ on the z axis $\hat{\mathbf{I}}_z$ possesses quantized eigenvalues: for a spin ½ particle these values are + ½ and - ½.

In the absence of external magnetic fields, the *spin polarisation axes* of each spin, i.e. the directions of every spin angular momentum, point in all possible directions, and the states described by the $\hat{\mathbf{I}}_z$ eigenvalues are completely degenerate in energy.

The NMR method consists in introducing a strong external magnetic field $\underline{\boldsymbol{B_0}}$ to perturb the equilibrium of the system and remove the denegeneracy of the spin levels. In this situation, each spin polarisation axis will tend to align in the same direction of the external field, the *z* direction, thus originating a *net magnetisation*: this corresponds to a difference in spin state populations. The net magnetisation may precess around the z axis at the Larmor frequency, giving rise to the NMR signal. The Larmor frequency depends on the external field and on the unique properties of the considered spins. Since each spin in a molecule has a particular chemical environment and thus senses an effective magnetic field slightly different form the one applied, the assignment of each frequency to the corresponding nucleus allows gaining access to each spin of the molecule individually.

The actual NMR experiments consist of a series of radiofrequency pulses given by means of a field $\underline{\boldsymbol{B_{RF}}}$. Under action of these pulses the net magnetisation follows a particular pathway in the spin space, being transferred from one spin to another. The number of pulses, their parameters and the delay times between them, depend on the particular experiment and on the coherence pathway one wants to select. At the end of the pulse sequence the resulting magnetisation is measured in a direction perpendicular to the z axis, during the acquisition time. In biomolecular NMR spectroscopy the most interesting nuclei are $^1$H, $^{13}$C and $^{15}$N. Because in natural abundance $^{13}$C and $^{15}$N represent only the 1.1% and 0.35% of the total content of

carbon and nitrogen respectively, heteronuclear NMR experiments are performed on enriched samples (*labelled* proteins) [29].

## 1.2 *External and internal spin interactions*

Each external and internal spin interactions can be represented by a Hamiltonian which contributes to the total NMR Hamiltonian [29]. In the following paragraphs a short description of the most important of these contributions is reported, with reference to spin-½ nuclei in an isotropic solution.

### 1.2.1 Zeeman interaction

The Zeeman interaction is the magnetic interaction of nuclear spins with the static external magnetic field of the spectrometer, $\underline{\boldsymbol{B}}_0$.
The Zeeman Hamiltionian for a single spin is given by:

$$\hat{\mathbf{H}}_z = -\hbar \gamma_I B_0 \hat{\mathbf{I}}_z \tag{1.2}$$

The subscript $z$ is due to the fact that conventionally the external magnetic field is along the z axis. The product $\hbar \gamma_I B_0$ corresponds to the Larmor frequency typical of spin $\mathbf{I}$ [29].

### 1.2.2 The radiofrequency field

As already mentioned, the radiofrequency (RF) field is fundamental in an NMR experiment to bring the system out of the equilibrium and create coherences between the spin levels. The oscillating radiofrequency field is a planar wave, linearly polarized perpendicular to the external field and has the form of a pulse, characterised by a particular frequency $\Omega$, strength $2B_1$ and length $t$:

$$\underline{\mathbf{B}}_1(\mathbf{t}) = 2B_1 \cos(\Omega t)\, \underline{\mathbf{i}} \tag{1.3}$$

where $\underline{\mathbf{i}}$ is the unit vector along x or y.
The RF field can also be seen as the superimposition of two circularly polarised components, counter-rotating in the *xy* plane. Only the component that rotates in the

same direction of the Larmor frequency creates coherences. Supposing that the RF field is given along the y-axis, the Hamiltonian is:

$$\hat{\mathbf{H}}_1(t) = \hbar B_1 \left\{ \cos(\Omega t)\gamma_I \hat{\mathbf{I}}_y - \sin(\Omega t)\gamma_I \hat{\mathbf{I}}_x \right\} = -\hbar\gamma_I B_1 e^{i\Omega t\hat{\mathbf{I}}_z} \hat{\mathbf{I}}_y e^{-i\Omega t\hat{\mathbf{I}}_z} \tag{1.4}$$

The effect of a RF pulse with frequency $\Omega$ and length $t$ given along the y axis is therefore the rotation of the net magnetisation by an angle $\Omega t$ around the y axis.

### 1.2.3  Chemical shift

The origin of the chemical shift is due to the electron distribution around the spins and to the induced local magnetic fields generated under the perturbing effect of $B_0$. The effective magnetic field experienced by a given nucleus can be written as:

$$\underline{\mathbf{B}}_{eff} = \underline{\mathbf{B}}_0 (1 - \underline{\underline{\sigma}}_{\mathbf{I}}(\mathbf{t})) \tag{1.5}$$

$\underline{\underline{\sigma}}_{\mathbf{I}}(t)$ is the chemical shift tensor acting on spin $\hat{\mathbf{I}}$. Its isotropic, time-independent part is different for each nucleus in a spin system and is responsible for the frequency shift observed in high-resolution spectra.

$$\sigma^{iso} = \frac{1}{3} Tr \left\langle \underline{\underline{\sigma}}(t) \right\rangle \tag{1.6}$$

Upon fast isotropic reorientation, the anisotropic part constitutes a relaxation mechanism [29].
The chemical shift Hamiltonian can be expressed as:

$$\hat{\mathbf{H}}_{cs}^{iso} = \hbar B_0 \gamma_I \sigma_I^{iso} \hat{\mathbf{I}}_z \tag{1.7}$$

For protein structure determination, the comparison between the chemical shift of the single residues when in a disordered (random coil) conformation with the one adopted in the folded protein, is a very important qualitative and quantitative parameter for understanding the general features of a protein spectrum and defining the secondary structure elements, as it will be further discussed in chapter 3. A widespread distribution of the peaks in the 1D proton spectrum and in the 2D [$^1$H,$^{15}$N]

HSQC spectrum of a protein indicates the presence of a well folded biopolymer. In fact, the chemical shifts of the amino acids' protons fall in the same region of the spectra, when residues are considered as separated entities; however when they are folded into a structured conformation, an appreciable chemical shift dispersion can be observed. Interior peptide segments are in fact shielded from the solvent and are nearest neighbours to other peptide segments, so different residues experience different chemical micro-environments and, therefore, a different effective magnetic field.
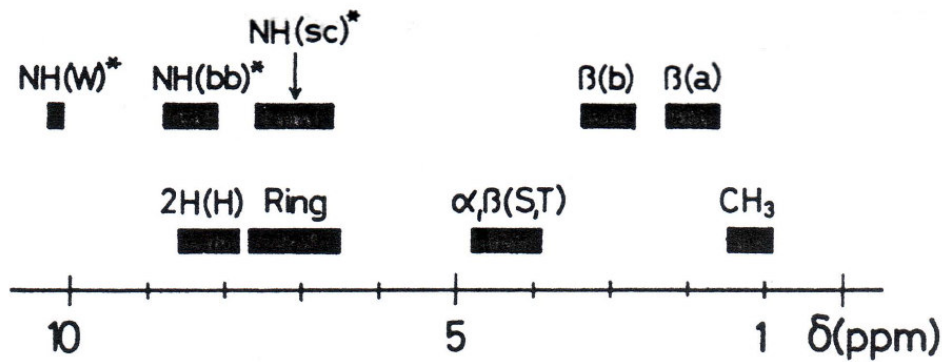


**Figure 1-1: Groups of hydrogen atoms in the common amino acid residues with similar random coil ^{1}H chemical shift [3].**

### 1.2.4  J-coupling

The *J*-coupling is the indirect nucleus-nucleus coupling, mediated by the presence of the electrons. The time dependent Hamiltonian for a couple of spins $\mathbf{I}$ - $\mathbf{S}$ is given by:

$$\hat{\mathbf{H}}_j(t) = h\hat{\underline{\mathbf{I}}} \cdot \underline{\underline{\mathbf{J}}}_{\mathrm{IS}}(t) \cdot \hat{\underline{\mathbf{S}}} \tag{1.8}$$

with $\underline{\underline{\mathbf{J}}}_{\mathrm{IS}}(t)$ the coupling tensor expressed in Hz.

The isotropic part of the coupling tensor is responsible for the spin-spin splitting observable in NMR spectra:

$$J_{IS}^{iso} = \frac{1}{3}Tr\left\langle \underline{\underline{\mathbf{J}}}_{\mathrm{IS}}(t)\right\rangle \tag{1.9}$$

The anisotropic term on the *J*-coupling instead arises from the screening by electrons of one spin as seen by the other, and is a time dependent matrix. It provides a relaxation pathway [29].

The *J*-coupling is field independent and it is extremely useful in protein structure determination by NMR.

Many NMR heteronuclear experiments, in fact, rely on J couplings for transferring the magnetisation from one atom to the other, through the chemical bonds [30]. In Figure 1-2 values for the typical $^1J$- and $^2J$-couplings useful for biomolecular NMR experiments are reported.
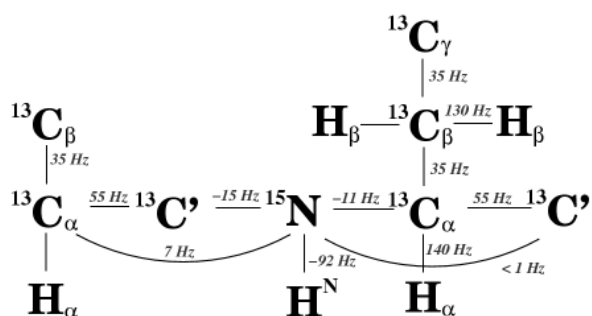


**Figure 1-2: Typical values of $^1$J and $^2$J-coupling in proteins [31].**

Although fundamental for designing specific NMR pulse sequences, one- and two-bonds J couplings do not provide any structural information.

On the contrary, the three bonds J couplings are strictly related to molecular structural parameters such as bond angles and torsion angles. In particular, $^3J_{HNH\alpha}$ between $H^N$ protons and $H^\alpha$ protons in proteins have a well defined dependence on the molecular torsion angle $\phi$ -the Karplus relationship-, and are therefore widely exploited as structural restraints [30,32].

$$^3J_{HNH\alpha} = A\cos^2\phi + B\cos\phi + C \tag{1.10}$$

where A, B and C are semi-empirical constants that can have different values according to the calculation method [30].

$^3$J couplings are also exploited for amino acid type recognition in the TOCSY-type experiments, where the proton spin system of each residue, i.e. a group of protons connected by scalar spin-spin couplings, is recorded. All but five amino acids form a single unique spin system; the exceptions are methionine, whose $\varepsilon$-$CH_3$ is not connected with the rest of the side chains, and the four aromatic residues where the aromatic ring is not connected with the $\alpha CH$-$\beta CH_2$ fragments.

18

### 1.2.5  Dipolar interaction

The dipolar interaction is the direct, through space magnetic interaction of nuclear spins, each acting as a magnetic dipole, i.e. without involving the electron cloud.
The time-dependent dipolar Hamiltonian is given by:

$$\hat{\mathbf{H}}_{DIP}(t) = \frac{\gamma_I \gamma_S \hbar^2}{r^3(t)} \left[ \hat{\underline{\mathbf{I}}} \cdot \hat{\underline{\mathbf{S}}} - \frac{3}{r^2} \left[ \hat{\underline{\mathbf{I}}} \cdot \underline{\mathbf{r}}(t) \right] \left[ \hat{\underline{\mathbf{S}}} \cdot \underline{\mathbf{r}}(t) \right] \right] \tag{1.11}$$

The dipolar interaction averages to zero in isotropic liquids, but is responsible for the dipolar relaxation pathway and the nuclear Overhauser effect, NOE [29].

## 1.3  *Relaxation*

After a radio frequency pulse, the spin system returns to equilibrium mainly through two types of processes: the spin-lattice, or transverse relaxation, and the spin-spin, or longitudinal, relaxation.

For spins ½, relaxation is caused by the fluctuating fields at the site of the nuclear spins due to the thermal motion of the molecules, to the dipole-dipole interaction and to the chemical shift anisotropy. In spins > ½, the quadrupolar couplings are also involved [28].

Longitudinal relaxation, characterised by the $T_1$ relaxation time, is responsible for the recovery to thermodynamic equilibrium of the magnetisation, initially perturbed by a radio frequency pulse. Transverse relaxation, characterised by $T_2$ relaxation time, is concerned with the decay of the coherences, i.e. with the decay of the effective magnetisation observed in the *xy* (transverse) plane after a 90° pulse.

$T_1$ is typically in the range 100 ms-100 s and is correlated with the overall rotational tumbling of the molecule in solution. It may be affected by intramolecular mobility in flexible structures. $T_2$ is always shorter than or equal to $T_1$. Both relaxation times are correlated with the dynamical features of the molecule and with its $\tau_c$. The $T_1$ and $T_2$ relaxation times of amide protons are extremely useful to identify flexible regions within a protein's backbone.

### 1.3.1  Dipolar relaxation

The principal contribution to the relaxation for spins-½ is generally the through space dipolar coupling between the spins.

If we consider the simple case of a two homonuclear spin system, with a Larmor frequency $\omega_0$, the energy diagram shown below can be drawn.
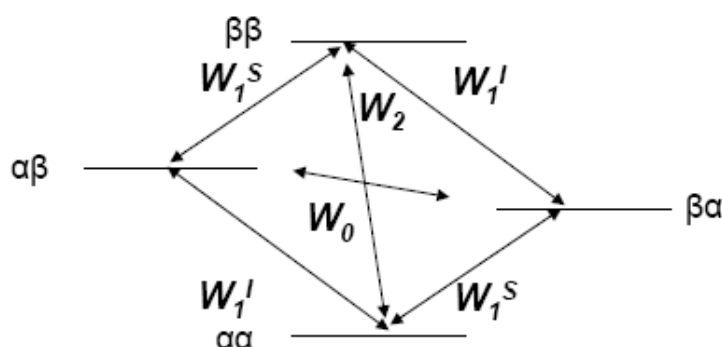


**Figure 1-3: Relaxation pathway of a homonuclear system of two spins not scalar coupled to each other [33].**

Relaxation can take place in several ways, through single quantum transition, whose relaxation rate is indicated by $W_1$, or through double quantum and single quantum transitions, $W_2$ and $W_0$. While single quantum transitions are responsible for observable peaks in an NMR spectrum, zero and double quantum transitions contribute exclusively to the relaxation pathway. When the only relaxation mechanism is the intramolecular dipole-dipole interactions of the spins with each other, we have:

$$W_1 = \frac{3}{20} b^2 J(\omega_0)$$

$$W_2 = \frac{3}{5} b^2 J(2\omega_0) \qquad (1.12)$$

$$W_0 = \frac{1}{10} b^2 J(0)$$

b is the dipole-dipole coupling constant, given by:

$$b = -\frac{\mu_0}{4\pi} \frac{\hbar \gamma^2}{r^3}$$

where $r$ is the distance between the spins. The dependence of the rate constants on $1/r^6$ justifies the importance of the dipolar relaxation for deriving structural restraints.

20

$J(\omega_0), J(2\omega_0)$ and $J(0)$ are the normalised spectral densities at the Larmor frequency, at twice the Larmor frequency and at zero frequency respectively. $J(\omega)$ is given by the following equation::

$$J(\omega) = \frac{\tau_c}{1 + \omega^2 \tau_c^2} \tag{1.13}$$

The relaxation of a two-spins system can be described by means of the Solomon equations:

$$\frac{d\mathbf{I}}{dt} = -\rho_I (\mathbf{I} - \mathbf{I}_0) - \sigma_{IS}(\mathbf{S} - \mathbf{S}_0)$$
$$\frac{d\mathbf{S}}{dt} = -\rho_I (\mathbf{S} - \mathbf{S}_0) - \sigma_{IS}(\mathbf{I} - \mathbf{I}_0) \tag{1.14}$$

where $\rho = W_0 + 2W_1 + W_2$ is the *auto-relaxation* rate constant and $\sigma_{IS} = W_2 - W_0$ is the *cross-relaxation* rate constant.

In the case of intra-molecular dipole-dipole relaxation, the longitudinal and transversal rate constants can be described as:

$$T_1^{-1} = \frac{1}{R_{auto} - R_{cross}} = \frac{1}{2W_1 + 2W_2} = \frac{3}{10}b^2 \{J(\omega_0) + 4J(2\omega_0)\}$$
$$T_2^{-1} = \frac{3}{20}b^2 \{3J(0) + 5J(\omega_0) + 2J(2\omega_0)\} \tag{1.15}$$

The values of the relaxation rates are therefore dependent on the value of the correlation time $\tau_c$. In Figure 1-4 the experimental plot of $T_1$ and $T_2$ with respect to $\tau_c$ is shown. In the region of extreme narrowing, when $\omega_0 \tau_c \ll 1$, $T_1 = T_2$ and both relaxation times increase proportionally to $\tau_c^{-1}$. For macromolecules that rotate slowly relative to $\omega_0$, so that $\omega_0 \tau_c \gg 1$, $T_1$ increases proportionally to $\tau_c$, while $T_2$ is proportional to $\tau_c^{-1}$. In the intermediate range the relaxation times have a more complicated dependence on the correlation time. The monotonical dependence of $T_2$ on increasing molecular size (high $\tau_c$) represents one of the challenges in protein

NMR because it is the cause of the severe line broadening of the peaks in macromolecules with MW >20 kDa.

A thorough treatment of dipolar relaxation and Solomon equations can be found in references [28] and [34].
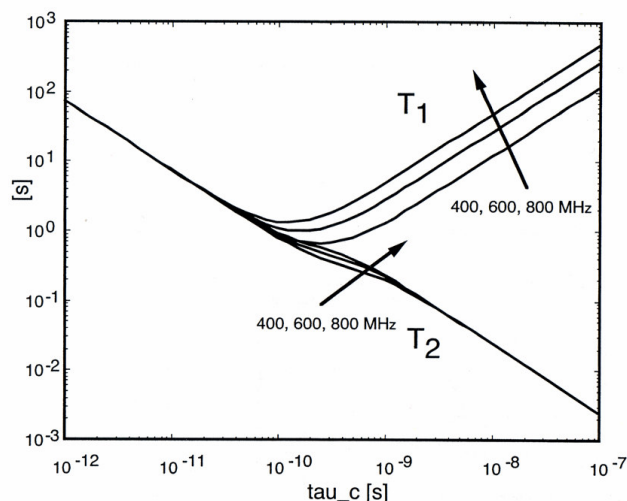


**Figure 1-4: Plot of the relaxation times $T_1$ and $T_2$ versus the correlation time $\tau_c$. The curves have been computed for dipolar relaxation of two protons with identical Larmor frequencies (400, 600 or 800 MHz) at a distance of 2 Å [35].**

### 1.3.2 The Nuclear Overhauser Effect

The Nuclear Overhauser Effect (NOE) is the change in intensity of one nucleus resonance when another one is perturbed, e.g. by selective saturation (steady state NOE) or inversion (transient NOE). The phenomenon is due to the presence of the cross-relaxation mechanism, generated by the dipolar interaction, and it is diagnostic of the proximity of pairs of spins.

In the same spin system illustrated in Figure 1-3, $\eta$ ,the NOE enhancement of spin $I$ upon saturation of spin $S$ is derived from the Solomon equation, considering that at steady-state $\mathbf{S} = 0$ and $\dfrac{d\mathbf{I}}{dt} = 0$ [33]:

$$\eta = \frac{\mathbf{I}_z^{ss} - \mathbf{I}_z^0}{\mathbf{I}_z^0} = \frac{\gamma_S}{\gamma_I} \frac{W_2 - W_0}{2W_1^I + W_2 + W_0} = \frac{\gamma_S}{\gamma_I} \frac{\sigma}{\rho_I} \qquad (1.16)$$

22

$\mathbf{I}_z^0$ is the intensity of the original signal, measured with an experiment recorded without saturation.

The expression for the relaxation rates in the heteronuclear case are similar to those in (1.12), modulated with the Larmor frequencies of both spins:

$$W_1^I = \frac{3}{20}b_{IS}^2 \frac{\tau_c}{\left(1+\omega_I^2\tau_c^2\right)}$$

$$W_0 = \frac{3}{5}b_{IS}^2 \frac{\tau_c}{\left(1+\left(\omega_I-\omega_S\right)^2\tau_c^2\right)} \qquad (1.17)$$

$$W_2 = \frac{1}{10}b_{IS}^2 \frac{\tau_c}{\left(1+\left(\omega_I+\omega_S\right)^2\tau_c^2\right)}$$

where the equation was used for the spectral densities and $b_{IS} = -\frac{\mu_0}{4\pi}\frac{\gamma_I\gamma_S\hbar}{r^3}$.

$\eta$ depends on the correlation time and hence on the molecular motion of the molecules. This is an important feature in biomolecular NMR, where the steady state $^1$H-$^{15}$N NOE is used to gain information on the dynamics of the protein backbone. When $I$ is $^{15}$N and $S$ $^1$H, in fact, the NOE enhancement changes sign, from negative to positive, as the molecular motion is slowed down (long $\tau_c$). Negative values of NOEs are therefore indicative of particularly flexible regions, while residues embedded in a rigid structure are characterised by positive values of NOE [28].

The transient NOE instead consists on the enhancement of spin $I$ upon inversion of the spin $S$. After a mixing time that allows magnetisation transfer to spin I upon cross relaxation, the spectrum is recorded.
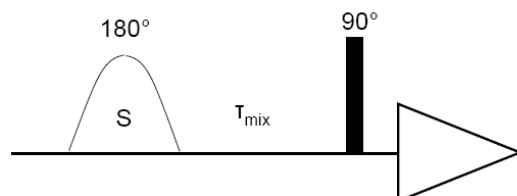


**Figure 1-5: Pulse sequence for recording transient NOE enhancement. The sequence involves selective inversion of spin S, by means of a shaped pulse.**

In the steady-state NOE the relaxation rate of spin **S** does not influence the mechanism, since this is constantly irradiated. In the transient NOE instead, two processes are active during time: the cross-relaxation, which is the cause of the NOE enhancement, and the $T_1$ relaxation of the inverted spin, which will eventually re-establish the initial equilibrium. The NOE enhancement therefore, will not reach a steady state, but, instead, it will grow to a maximum value and then decay with increasing time.

Although the temporal dependence of the NOE curve is fairly complicated, its initial build-up rate is proportional to the cross-relaxation term $\sigma_{IS} = W_2 - W_0$. In the *initial rate approximation* the Solomon equations (1.14) can in fact be solved with the initial conditions:

$$I_z(0) = I_z^0$$
$$S_z(0) = -S_z^0$$

(1.18)

In this way:

$$\left. \frac{d\mathrm{I}_z(t)}{dt} \right|_{init} = 2\sigma_{IS} S_z^0$$

(1.19)

[34]

In this condition therefore, the intensities of the NOEs are directly proportional to $1/r^6$, with *r* the inter-atomic distance between the two spins, and can be directly exploited to obtain structural information. Since the cross correlation term is difficult to determine, the distance scale is usually calibrated by using proton pairs separated by known distances in rigid parts of the molecule [28,33].

Transient NOE represents the basis for the NOESY pulse sequence.
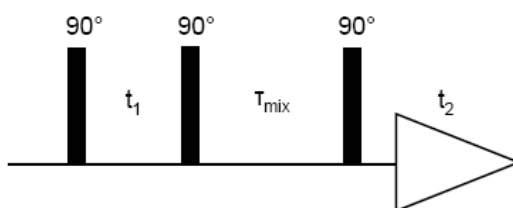


**Figure 1-6: Pulse sequence for the NOESY experiment.**

In this experiment, the magnetisation of spin S, instead of being inverted at the start of the mixing time, has an amplitude label which depends on the evolution during a variable time delay $t_1$. The development of NOESY cross-peak intensity with time for a three spins system is shown in Figure 1-7.

The application of NOESY-type experiments will be further described in the next chapter.
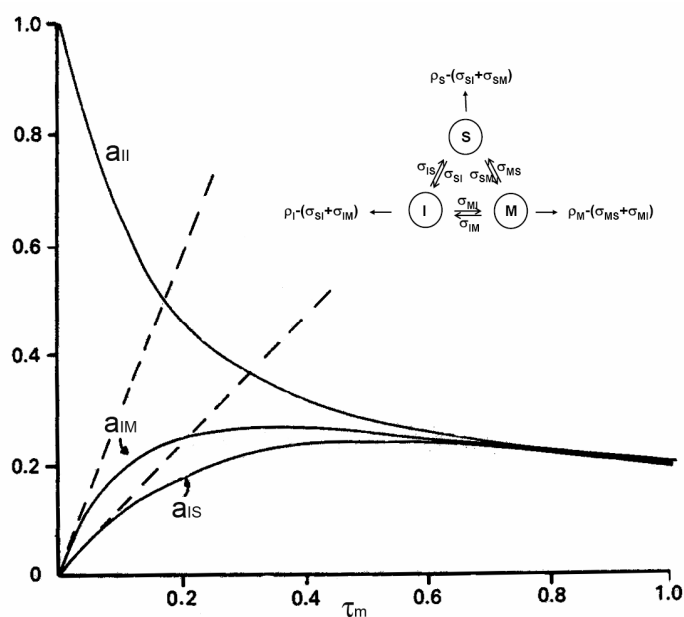


Figure 1-7: Time development of the intensities of the spin *I* diagonal peak ($a_{II}$) and the S and M spin cross-peaks ($a_{IS}$ and $a_{IM}$) as a function of mixing time ($\tau_m$) of the NOESY sequence. The picture refers to a three spin system, with the following parameters: $\sigma_{SM} = \sigma_{MS} = 3.28$ s$^{-1}$, $\sigma_{IS} = \sigma_{SI} = 1.47$ s$^{-1}$, $\sigma_{IM} = \sigma_{MI} = 3.28$ s$^{-1}$, $R_I = R_S = R_M = 0.5$ s$^{-1}$, $\rho_n = \sum_m \sigma_{nm} + R_n$, where Rs are the total direct relaxation rates of the spins [33].

## 1.4 *Resonance linewidth*

The resonance linewidth in NMR spectra is a crucial parameter since it determines the spectral resolution that can be attained at a certain field $\mathbf{B}_0$. The linewidth corresponds to the peakwidth at half height, and is equivalent to $1/\pi T_2$ (Hz). A rapid decay of the transverse magnetisation (large $T_2$) corresponds therefore to a broad

signal [3,28]. In many cases, however, the linewidth is influenced by additional factors. For example, the presence of an exchange process can severely broaden the NMR peaks. Moreover, experimental factors, such as those arising from the inhomogeneity of the static magnetic field $\mathbf{B}_0$, can heavily contribute to the peaks' linewidth. If $\mathbf{B}_0$ is inhomogeneous over the sample volume in fact, different spins in the macroscopic sample will precess at slightly different frequencies, resulting in an effective coherence loss, which contributes to the observed $T_2$ [3].

## 1.5  *NMR time scales*

One of the most important and promising features of NMR spectroscopy is its ability to probe different types of molecular motion, over a wide range of timescales, from picoseconds to seconds. Below a schematic picture of the dynamic processes that can be analysed with NMR and their timescales are reported.
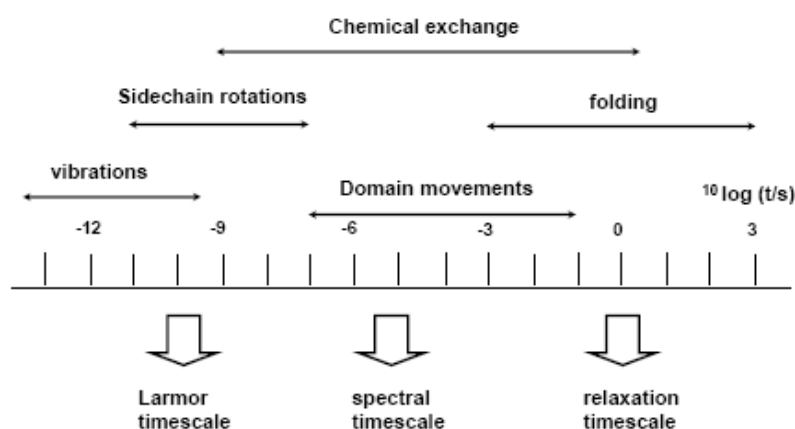


**Figure 1-8: Typical motional timescales for physical processes (adapted from [28])**

Molecular vibrations and librations occur at fraction of picoseconds, while folding, macroscopic diffusion and some chemical exchange processes are very slow, in the order of seconds or even minutes. The effects of the motional processes depend on their relationship with the characteristic timescales of the nuclear spin system, namely the Larmor, the spectral and the relaxation timescales.

The characteristic Larmor timescale is given by $\tau_0$, where:

$$\left| \omega_0 \tau_0 \right| \sim 1$$

and indicates the time required for the spins to precess through 1 radian in the magnetic field.

The characteristic time of the spectral timescale corresponds to the inverse width of the NMR spectrum, in frequency units. If a system contains two spins with chemical shift $\Omega_1^0$ and $\Omega_2^0$, the spectral timescale is given by:

$$\left| \left( \Omega_1^0 - \Omega_2^0 \right) \tau_{spec} \right| \sim 1$$

Finally, the relaxation timescale is characterised by the spin-lattice relaxation time $T_1$. The processes that are on the Larmor timescale, with characteristic times that range from $10^{-12}$-$10^{-7}$ seconds, are directly manifested in the relaxation parameters $T_1$ and $T_2$. Much slower NMR time scales result from reversible exchange of nuclei between two or several different environment, by chemical or physical exchange or other dynamic processes. These processes generally are on the spectral timescale, and strongly affect the NMR lineshapes [28].

### 1.5.1 Exchange processes

Considering the exchange of an isolated spin between two chemical environments A and B, with different chemical shifts $\Omega_a^0$ and $\Omega_b^0$, we have:

$$A \underset{k_-}{\overset{k_+}{\rightleftharpoons}} B$$
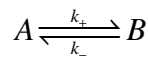
Supposing that the transition between A and B does not involve any intermediate state and that $k_+ = k_- = k$, three different exchange regimes can be defined, based on the process rate constant and on the chemical shift difference $\Omega_a^0 - \Omega_b^0 = \Omega_\Delta$:

- *slow exchange process* : $k \ll 2\pi\Omega_\Delta$
- *intermediate exchange process*: $k \approx 2\pi\Omega_\Delta$
- *fast exchange process:* $k \gg 2\pi\Omega_\Delta$

In the first situation, separate narrow resonances can be observed for the nuclear spin in the two chemical environments. In the fast exchange, a single narrow

resonance can be detected, whose position is given by the average of the two chemical shifts weighted by the equilibrium concentration of the two species.

The intermediate exchange is characterised by a continuous transition between the two limiting conditions, with pronounced exchange broadening of the lines. The jump of the nuclear spins between the two states causes in fact dephasing of the transverse magnetisation with consequent *motional broadening*. [3,28].
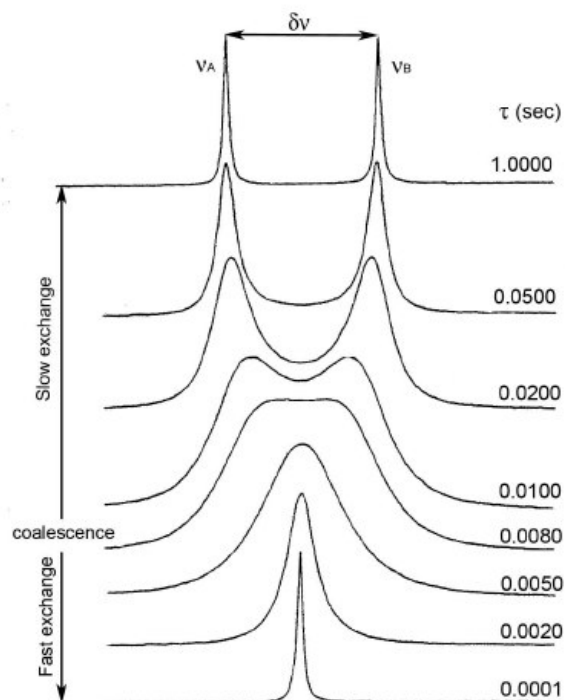


**Figure 1-9: Typical spectra reflecting exchange at different rates relative to the chemical shifts. [33].**

# Protein Assignment and Structure Calculation  2

In the initial stage of every NMR investigation, each nuclear magnetic resonance must be associated with a specific nucleus in the protein under investigation. Resonance assignment must be *sequence specific*, in that each resonance must be assigned to a spin in a particular amino acid residue in the protein sequence. The complete assignment is reached in two steps: one is the so called *backbone sequential assignment*, which consists in aligning the amide signals from the NMR spectra within the protein sequence, and the other is the *side chain assignment*, i.e. the assignment of every spin system connected to the amide resonances.

NMR spectroscopy provides three types of information useful for spectral assignment: through-bond interactions, via scalar coupling, through-space interactions, via dipolar coupling, and chemical environment, via chemical shift [30].

This chapter will specifically describe the assignment procedure of proteins, but in the first part a short introduction of the modern coherence selection methodologies and of the water suppression technique used in this study will be provided.

## 2.1 *Pulsed-field gradients*

The optimal result of a multi-pulsed NMR experiment is a spectrum whose signals correspond to a well defined coherence transfer pathway. However, the RF pulses may be rather indiscriminate in their effect on the spin systems and lead to the appearance of several alternative coherence pathways other than the desired one. A selection method is therefore necessary to obtain a spectrum clean from the unwanted signals.

The traditional selection method is the *phase cycling* Here the pulse sequence is repeated with systematic variation in the phase of the radio frequency pulse and the signals observed for each repetition are combined in such a way that only the ones arising from the desired coherence transfer pathway add, all the others being cancelled. Phase cycling relies on a difference method to achieve the selection of the desired pathway: any changes in amplitude or phase between the signals recorded in different repetitions will result therefore in incomplete subtraction. This can be troublesome if the signals from the wanted pathway are much weaker than those from the unwanted pathways [36].

An alternative methodology, which does not require the repetition of the pulse sequence, is based on the use of *pulsed-field gradients*During a pulsed-field gradient the applied magnetic field is deliberately made inhomogeneous for a short time. As a result, all the coherences dephase across the sample. They can, however, be refocused by application of a second gradient that reverts the dephasing process, thus restoring the coherence. The crucial property of the dephasing process is that it occurs at different rates for different coherences. By means of the application of gradient pulses of different length and strength it is therefore possible to refocus exclusively a certain type of coherences, losing the unwanted ones through the dephasing process [36].

The pulse field gradients are incorporated in all the modern pulse sequences, and are fundamental for the WATERGATE sequence, described in the next paragraph, which allows the suppression of the strong water signal.

A more detailed description of the theory at the base of the pulse field gradients and of their application in the NMR technique can be found in [36].


## 2.2 *WATERGATE*

The need to record NMR spectra in aqueous solutions has prompted the development of a variety of methods for suppressing the large water signal, which otherwise would jeopardise the quality of the protein spectrum.

One of the most commonly used water suppression sequence is the so called WATERGATE sequence (WATER suppression by GrAdient-Tailored Excitation) [37]. The sequence can be incorporated into most nD NMR experiments, and is a gradient echo sequence composed of two parts. In the first part a non-selective $90°(x)$ RF pulse excites all resonances uniformly, regardless of their chemical shift. The pulse is followed by a symmetrical echo segment, formed by two field-gradient pulses of the same amplitude and sign with a centrally placed $180°(x)$ selective RF pulse. All coherences dephased by the first field gradient are refocused by the second one, provided that they are rotated of $180°$ by the RF pulse. The two flanking rectangular $90°(-x)$ pulses selectively rotate the water magnetisation in the opposite direction, thus preventing its rephasing during the subsequent gradient. The scheme of the pulse sequence is presented in Figure 2-1.

30
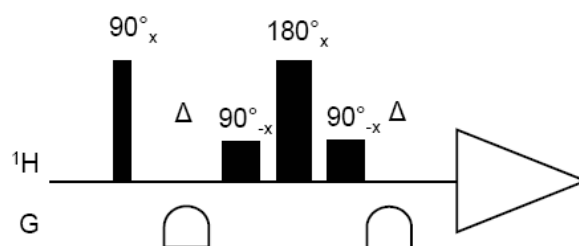
**Figure 2-1: Scheme of the Watergate pulse sequence [37].**

## 2.3 *Assignment procedure*

The protein assignment procedure is different according to the labelling extent of the analysed biomolecule. Unlabelled proteins are assigned using two-dimensional homonuclear spectra, which however are informative only for small proteins. The assignment of larger proteins requires extensive labelling and the use of triple resonance experiments.

### 2.3.1 Homonuclear 2D experiments

Unlabelled proteins are assigned using homonuclear 2D experiments, such as COSY, TOCSY and NOESY.

COSY peaks arise from coherence transfer between scalar coupled spins. In its simplest version, the pulse sequence consists of two 90° pulses separated by an incrementable delay $t_1$, during which the resonances are correlated to each other through the scalar coupling. In this way, coherence transfer is limited only to protons separated by two or three bonds. On the contrary, in a TOCSY experiment magnetisation can be transferred through several bonds during the so-called *isotropic mixing time*, owing to the application of a suitable modulated RF field. In the absence of relaxation, cross peaks are potentially generated between all resonances within a spin system. Finally, in the NOESY sequence, magnetisation transfer occurs through dipolar interaction during the mixing period $\tau_m$ [30].

The spin systems attached to every $H^N$ proton, thus within the same amino acid, are identified in the COSY and TOCSY spectra, while the NOESY experiment is used for the sequential assignment, i.e. for jumping from one amino acid unit to the next [31]. In a first step, amino acid spin systems are identified using scalar couplings between $H^N$, $H^\alpha$ and aliphatic side chain resonances. The amino acid type is suggested by the

chemical shifts' pattern. Every spin system is then associated to a residue within the polypeptide chain using through space dipolar interaction, thus achieving the sequential assignment. Intense NOEs from $^1H^N$, $^1H^\alpha$ and/or $^1H^\beta$ of one spin system on the $^1H^N$ of another indicate that the two amino acids are adjacent to each other [3]. In alternative, the scalar coupling connectivities are initially used to identify $^1H^N$- $^1H^\alpha$ -$^1H^\beta$ units only, that are then aligned using NOESY spectrum. The side chain assignment is then carried out using TOCSY connectivities [30].

### 2.3.2  Heteronuclear experiments

2D homonuclear NMR experiments are ineffective for proteins with molecular weights larger than approximately 10-12 kDa. In fact the number of protons scales with the molecular mass, and so does the rotational correlation time of globular proteins. The increased number and linewidth of proton resonances result in extensive chemical shift overlap, which prevents the analysis of the signals.

Heteronuclear NMR experiments allow the complete assignment of proteins up to 30 kDa, provided the proteins can be uniformly labelled with $^{13}C$ and $^{15}N$. Spectral resolution is improved by increasing the dimensionality of the spectrum, so that the highly overlapped $^1H$ resonances present in the $^1H$ 2D spectra are well resolved in 3D or 4D spectra, where each cross peak is labelled with three or four frequencies. In addition, the efficiency of the magnetisation transfer in these spectra is increased because they rely on large $^1J$ and $^2J$ scalar coupling interactions between pairs of heteronuclei and between heteronuclei and their directly attached protons.

### 2.3.3  Sequential assignment

The basic building block for the 3D experiments is the HSQC (Heteronuclear Single Quantum Correlation) experiment, that allows the detection of protons directly attached to a heteronucleus ($^{13}C$ or $^{15}N$). The use of [$^1H$,$^{15}N$] HSQC is widespread in biomolecular NMR because it provides a sort of fingerprint map of each protein, through its amide protons. A detailed discussion of the HSQC experiment can be found in [30] and [38].
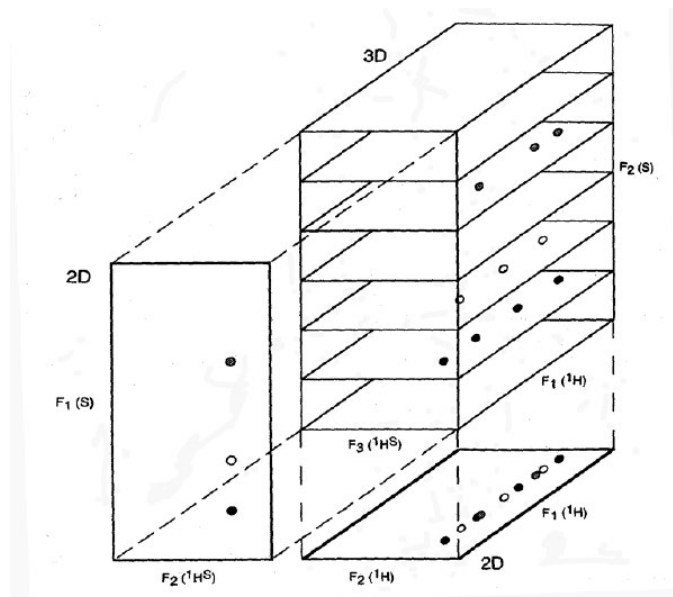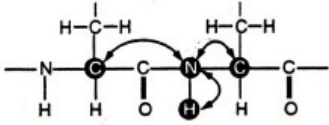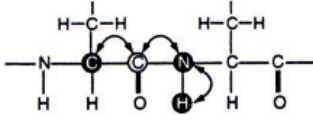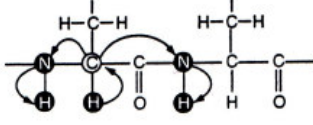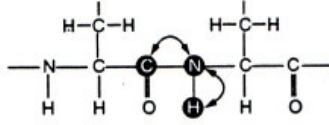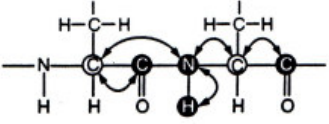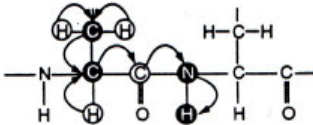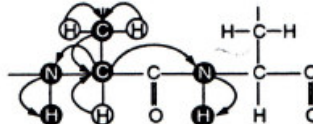
**Figure 2-2: schematic illustration showing the relationship between a 3D heteronuclear-edited spectrum and 2D homonuclear and heteronuclear correlation spectra. Cross peaks representing three different spin systems, with degenerate $^1H_S$ chemical shifts, but differing S spin chemical shifts , are indicated. The 3D spectrum is represented as a series of $F_1(^1H)$-$F_3(^1H^S)$ slices edited by the chemical shift of the directly attached S spins along the $F_2$ axis between a 3D heteronuclear-edited spectrum and 2D homonuclear and heteronuclear correlation spectra [30].**

The typical triple-resonance experiments used for backbone sequential resonance assignment are listed in Table 2-1. The nomenclature reflects the magnetisation transfer pathway of the experiments. Nuclei that are involved in magnetisation transfers form the name of the experiment. Spins whose chemical shifts do not evolve are put in parentheses. For the 'out-and-back' type of experiments, where the magnetisation of a spin is transferred to a remote spin and then brought back the same way, only the first half of the magnetisation transfer is used for its name [31].

The paragraphs that follow shortly describe the experiments that were used in this study for the protein assignment. A detailed discussion of all the 3D experiments listed in Table 2-1 can be found in [30] and [31].

**Table 2-1: Triple resonance experiments used for sequential resonance assignment [30].**

| Experiment | Correlations observed | Magnetisation transfer | J Couplings |
|---|---|---|---|
| HNCA | $^1H^N_i$–$^{15}N_i$–$^{13}C^\alpha_i$ <br> $^1H^N_i$–$^{15}N_i$–$^{13}C^\alpha_{i-1}$ | | $^1J_{NH}$ <br> $^1J_{NC^\alpha}$ <br> $^2J_{NC^\alpha}$ |
| HN(CO)CA | $^1H^N_i$–$^{15}N_i$–$^{13}C^\alpha_{i-1}$ | | $^1J_{NH}$ <br> $^1J_{NCO}$ <br> $^1J_{C^\alpha CO}$ |
| H(CA)NH | $^1H^\alpha_i$–$^{15}N_i$–$^1H^N_i$ <br> $^1H^\alpha_i$–$^{15}N_{i+1}$–$^1H^N_{i+1}$ | | $^1J_{C^\alpha H^\alpha}$ <br> $^1J_{NC^\alpha}$ <br> $^2J_{NC^\alpha}$ <br> $^1J_{NH}$ |
| HNCO | $^1H^N_i$–$^{15}N_i$–$^{13}CO_{i-1}$ | | $^1J_{NH}$ <br> $^1J_{NCO}$ |
| HN(CA)CO | $^1H^N_i$–$^{15}N_i$–$^{13}CO_i$ <br> $^1H^N_i$–$^{15}N_i$–$^{13}CO_{i-1}$ | | $^1J_{NH}$ <br> $^1J_{NC^\alpha}$ <br> $^2J_{NC^\alpha}$ <br> $^1J_{C^\alpha CO}$ |
| CBCA(CO)NH | $^{13}C^\beta$–$^{13}C^\alpha_i$–$^{15}N_{i+1}$–$^1H^N_{i+1}$ | | $^1J_{CH}$ <br> $^1J_{C^\alpha C^\beta}$ <br> $^1J_{C^\alpha CO}$ <br> $^1J_{NCO}$ <br> $^1J_{NH}$ |
| CBCANH | $^{13}C^\beta/^{13}C^\alpha_i$–$^{15}N_{i+1}$–$^1H^N_i$ <br> $^{13}C^\beta/^{13}C^\alpha_i$–$^{15}N_{i+1}$–$^1H^N_{i+1}$ | | $^1J_{CH}$ <br> $^1J_{C^\alpha C^\beta}$ <br> $^1J_{NC^\alpha}$ <br> $^2J_{NC^\alpha}$ <br> $^1J_{NH}$ |

### 2.3.3.1   HNCA and HN(CO)CA

The HNCA experiment correlates the amide $^1H$ and $^{15}N$ chemical shifts with the intraresidue $^{13}C_\alpha$ shift by using the small one-bond $^{15}N$-$^{13}C_\alpha$ J coupling (7-11 Hz). In addition this experiment provides sequential connectivities by transferring the coherence from the $^{15}N$ spins to the $^{13}C_\alpha$ of the preceding residue by means of the

interresidual two-bonds $^{15}$N-$^{13}$C$_\alpha$ J coupling (up to 9 Hz). The HN(CO)CA represents the complementary approach to the HNCA: it provides in fact exclusively sequential correlations between the amide $^1$H and $^{15}$N chemical shifts of residue i and the $^{13}$C$_\alpha$ chemical shift of the preceding residue by transferring coherence via the intervening $^{13}$CO spin.

### 2.3.3.2    CBCANH and CBCA(CO)NH

In the CBCANH experiment the magnetisation is transferred from the amide proton of a residue to its own $^{13}$C$_\alpha$ and $^{13}$C$_\beta$ and to the $^{13}$C$_\alpha$ and $^{13}$C$_\beta$ of the preceding amino acid; the peaks of the $^{13}$C$_\alpha$ and of the $^{13}$C$_\beta$ have opposite phase.

Information regarding the amino acid type can be obtained from the $^{13}$C$_\alpha$ and $^{13}$C$_\beta$ chemical shifts, providing useful links between the sequence of the protein and the sequence of the spectral signals. For example while the $^{13}$C$_\beta$ of almost every residue is upfield shifted with respect to the $^{13}$C$_\alpha$, Thr and Ser have the opposite patterns. The $^{13}$C$_\beta$ of alanines is also very peculiar since it resonates at particularly low frequencies (Figure 2-3). Glycines do not have a $^{13}$C$_\beta$ atom and their $^{13}$C$_\alpha$ has opposite phase with respect to all the other $^{13}$C$_\alpha$s.

The CBCA(CO)NH is the complementary experiment of the CBCA(CO)NH and provides inter-residual correlations between the amide proton and the $^{13}$C$_\alpha$ and $^{13}$C$_\beta$ of the preceding residue.

### 2.3.3.3   HNCO

The HNCO experiment correlates the amide $^1$H and $^{15}$N chemical shifts of one residue with the $^{13}$CO chemical shift of the preceding one, by using the one-bond $^{15}$N-$^{13}$C J coupling (~ 15 Hz) to establish the sequential correlations.

## 2.3.4   Side chain assignment

Once the sequential assignment has been completed, side chain assignment can be achieved using the TOCSY-HSQC and the HCCH-TOCSY spectra.
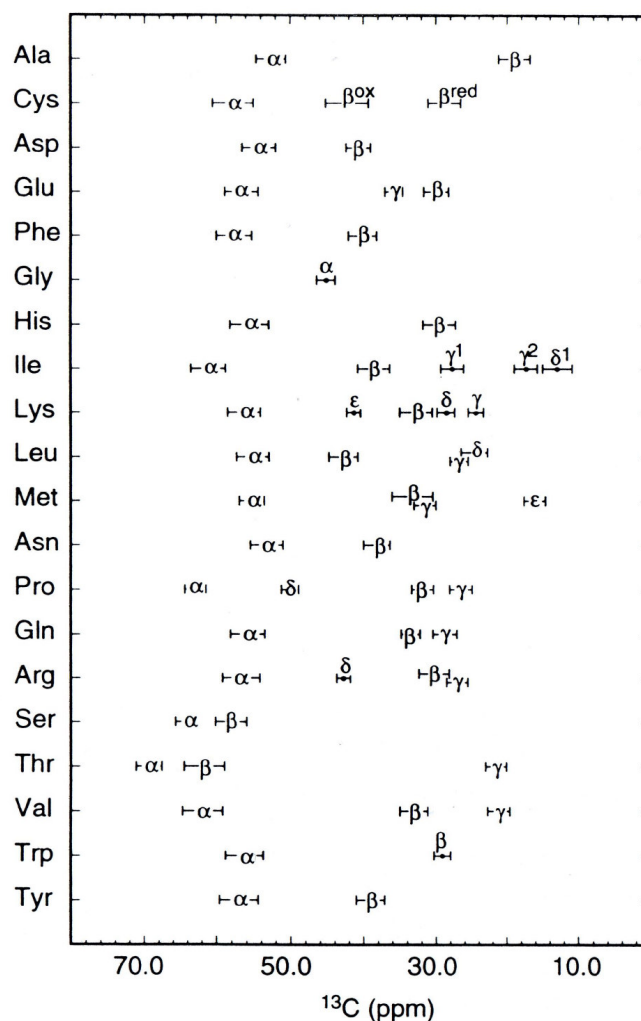
**Figure 2-3:** [13]C chemical shift of the side chains carbon of the 20 amino acids [30].

### 2.3.4.1   TOCSY-HSQC and HCCH-TOCSY

The information obtained with a TOCSY-HSQC spectrum is the same as the one obtained from a 2D TOCSY [1]H-^{1}H TOCSY, but it is edited in the third dimension according to the ^{15}N chemical shift associated with the amide [1]H$^N$. It provides therefore a direct correlation of each spin system with the amidic protons observed in the [^{15}N, ^{1}H] HSQC spectrum, previously sequenced.

The HCCH-TOCSY experiment is used to assign aliphatic ^{1}H, ^{13}C spin systems and to link them to the sequentially assigned backbone. In an $H_nC(i)$-...-$H_mC(j)$-spin system a 3D HCCH-TOCSY experiment connects chemical shifts H(i), H(j) and C(j). This connectivity information yields to complete assignment of the ^{1}H and ^{13}C resonances because the aliphatic proton chemical shifts of a given spin system are found at the ^{13}C chemical shift of all the carbon frequencies involved in that spin system [31].

36

### 2.3.4.2 Assignment of aromatic residues

Aromatic spin systems are very important for structure determination by NMR since these residues are usually located in the hydrophobic core of a protein. They show many long range NOEs, that provide important distance restraints to define the tertiary fold. The correspondent side chain resonances can be identified using experiments based on scalar connectivities between the $^{13}C\beta$ and the $^{1}H\delta/\varepsilon$. HCCH-TOCSY optimised for aromatic spin system can then be used to achieve the complete assignment of the ring resonances [31,39].

## 2.3.5  Structural restraints from NOEs based spectra

The achievement of the complete resonance assignment of a protein represents the first step towards the complete determination of the protein structure in solution.

Atomic distance restraints are obtained through the analysis of spectra based on dipolar coupling interaction between resonances, i.e. NOESY-HSQC, both $^{15}N$ and $^{13}C$ edited. The $^{15}N$ edited NOESY-HSQC can also be useful for the sequential assignment, in a similar way to the correspondent homonuclear 2D spectrum.

The presence of NOEs between groups that are close in the covalent structure can help defining the secondary structure, together with information from the $^{3}J_{HNH\alpha}$ coupling constants. Intense $d_{NN}$ NOEs and small $^{3}J_{HNH\alpha}$ couplings (<6 Hz) are indicative of a helical conformation, whereas intense $d_{\alpha N}$, weak $d_{NN}$ and $d_{\beta N}$ NOEs and large $^{3}J_{HNH\alpha}$ coupling constants are characteristic of extended β-strand polypeptides [30].

Long range NOEs are informative of the tertiary structure of the proteins. In fact, in the initial rate approximation, NOEs are proportional to the inverse sixth power of the distance between two interacting protons, as already mentioned in paragraph 1.3.2.

Thus, if one interproton distance is known, $r_{ref}$, then another unknown interproton $r_i$ can be determined from the relationship:

$$r_i = r_{ref} \left( \frac{S_{ref}}{S_i} \right)^{1/6}$$

(2.1)

where $S_{ref}$ and $S_i$ are the integrated cross peaks' intensities. The distance scale is calibrated by using proton pairs separated by known distances in rigid part of the

molecule. Equation 2.1 neglects differential internal mobility, i.e. different values of $\tau_C$ for the interacting protons. Spin diffusion mechanism also is ignored and it is assumed that only the dipolar relaxation between the two considered protons contributes to the NOE.

In practice, NOESY mixing time short enough to fully satisfy the initial-rate approximation are not possible because the cross peaks would have too low a signal-to-noise ratio. A precise calculation of $^1$H-$^1$H separations should therefore be obtained from a much more involved numerical analysis of the NOE peaks intensities. In general, it appears sufficient to grossly group NOE cross peaks on the basis of their intensities into three main categories, strong, medium and weak, with each category associated with an upper bound separation between the interacting spins. Usually strong NOEs imply an inter-atomic distance < 2.5 Å, medium NOEs < 3.5 Å and weak NOEs < 5 Å .The cross peak volume limits and the upper bound distances for the categories are estimated from the NOE intensities observed for protons of known covalent geometry or between protons in regions of regular secondary structure. Conservative upper bound distances compensate for cross peaks affected by spin diffusion or partial overlap.

3D structure can be generated using the distance geometry approach (DG) or the restrained molecular dynamics approach (rMD). Usually, the DG approach is used to generate the first structures that are subsequently refined (annealed) with molecular dynamics. In this approach force fields are supplemented by square-well pseudoenergy obtained from the NMR data. Those potentials drive the structure towards a minimum energy conformation, with the smallest number of restraints violations, by means of a heatup-cooldown annealing cycle [30].

# Bioinformatic Tools 3

In this chapter I shall go through all the bioinformatic tools which have assisted all stages of my work.

## 3.1 *Primary sequence analysis*

### 3.1.1 Protein databases

Proteins' and nucleic acids' sequences are stored and archived in virtual databanks, easy accessible through the Web.

The easiest way to access such archives is to enter the ENTREZ service. ENTREZ is maintained by the National Center for Biotechnology Information (NCBI) and offers access to various databases, from proteins and peptides, to nucleotides, genomes and taxonomy. Here one can retrieve the accession numbers for the source databanks of every known macromolecule. UniProt (Universal Protein Resource) instead, offers access to protein databases exclusively.

The Swiss Institute of Bionformatics maintains another very useful service for information retrieval and analysis, called ExPASy (Expert Protein Analysis System). From the ExPASy main page many databases and services can be accessed. Amongst them SWISS-PROT is an annotated database of amino acid sequences, created in collaboration with the EMBL data library, which contains a nucleic acids sequence archive. Translation of nucleotide sequences from the EMBL Data Library not yet fully integrated into SWISS-PROT can instead be found in the database TrEMBL. A useful database is PROSITE, a collection of signature patterns that can be exploited to identify patterns of residues shared by proteins. The information stored in any of these libraries (swiss-prot, embl, prosite, tremebl, etc.) can be retrieved by the Sequence Retrieval System (SRS).

Amongst the software programs available at the ExPASy website, a very important one is ProtParam, which uses the sequence of a protein to compute its physicochemical properties, such as the isoelectric point and the molar exctinction coefficient.

Similarly to SwissProt, the PIR database, developed at Georgetown University USA, maintains several databases about proteins [40].

A very useful tool is PFAM, which is a large collection of multiple sequence alignments that covers many proteins and families. Other than providing a multiple alignment, it also allows one to view the protein domain architecture and to examine its distribution amongst the species. All the domains found in a non redundant protein database and taxonomic information are stored in a relational database system called SMART (Simple Modular Architecture Research Tool). It contains domain architecture, annotated with respect to phylogenetic distributions, functional class, tertiary structure and functionally important residues.

A specific database classifies conserved protein domains, the COG (Cluster of Orthologous Groups of Proteins), from NCBI. It compares sequences encoded in complete genomes representing major phylogenetic lineages. Each COG consists of individual proteins or groups of paralogues from at least three lineages, corresponding therefore to an ancient conserved domain.

Moreover, known and predicted protein-protein interactions are stored in the STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) database. The interactions include direct (physical) and indirect (functional) associations. STRING quantitatively integrates data from genomic context, high throughput screens, protein coexpression for a large number of organisms and transfers information between these organisms when applicable (font: http://string.embl.de/)

Protein 3D structures also are classified. The principal structure database is the Protein Data Bank (PDB) at Brookhaven National Laboratories, that keeps both crystallographic and NMR structures. The BioMagResBank (BMRB) at the Department of Biochemistry, Madison, Winsconsin, is a specific archive for NMR data, such as protein assignments.

### 3.1.2   Sequence alignment

Analysis of the differences and the similarities amongst amino acid sequences aims at inferring structural, functional and evolutionary relationships amongst proteins. The similarity between two sequences in fact, allows drawing conclusions on their homology, i.e. a common evolutionary history. The most common comparative method is *sequence alignment* protein sequences may be aligned and compared pairwise or with multiple copies (*pairwise* or *multiple alignment*) [41].

At this point it seems important to give some definitions on proteins' nomenclature: two proteins are called *homologues* when their sequences and the organisms in which they occur are descended from a common ancestor, with the consequence that the similarities between them are shared ancestral characteristics. *Orthologues*

are homologues in different species that often perform the same function. *Paralogues* instead, are related genes that have duplicated within the same genome and have in many cases diverged to provide separate functions in descendant species [40].

Comparison of proteins' sequences can be important for understanding the role of each residue. Different degree of conservation amongst the amino acids of the sequences is in fact suggestive of which residues are crucial for maintaining a protein's function or structure, although conserved positions from closely related species may simply correspond to residues that have not had time to diverge.

An alignment that spans the full extent of the input sequences is called *global alignment*; in the case of modular proteins a *local alignment*, consisting in aligning sub-sequences flanked by unrelated residues, is preferred to the global one.

Because for the majority of the problems the total number of distinct alignments is extremely large, it is necessary to assign to each of them a quality score, so that the optimal alignment can be identified. Changes that occur during diversion from a common ancestor can be categorised as substitutions, insertions and deletions. During the alignment process, a score is calculated for each of these changes and the single contributions are summed up to a global score, which is indicative of the alignment quality.

To calculate the scores for the amino acids' substitution it is important to consider that residues with similar physico-chemical characteristics can substitute easily for one another in related proteins. Examples of the so called *'conservative mutations'* include isoleucine for valine or threonine for serine, thus preserving the hydrophobic and hydrophilic character within each pair respectively. Moreover, different sets of values may be necessary for comparing very similar sequences as opposed to highly divergent ones. These considerations are dealt with in a flexible manner by the so called *substitution matrices*, that provide the score for any pair of amino acids mutation [41].

The matrices based on point-accepted mutation (PAM) model of evolution were developed by Dayhoff and co-workers [42]. One PAM corresponds to one percent accepted mutation; two sequences one PAM apart have 99% identical residues. When two sequences are within the PAM 1 level of divergence it is very likely that there has been no more than one change at any position. By collecting statistics from pairs of sequences as closely related as these and correcting for the different frequency of each amino acid, the PAM 1 substitution matrix can be built. For more divergent sequences, powers of the PAM 1 matrix are adopted. The PAM 250 level corresponds to 20 % overall sequence identity and represents the lowest level at which a sequence alignment can be considered correct. It is therefore the most used

substitution matrix. The score for the mutation of the amino acid **i** into the amino acid **j** is expressed as a *log-odds* value, i.e. natural log of the ratio of the observed substitution over the mutational rate expected from the amino acid frequencies. Since the probability of two independent mutational events is the product of their probabilities, the score for multiple substitutions is calculated adding up the single mutation scores.

BLOSUM matrices instead are based on blocks of local multiple alignment involving distantly related sequences. Similarly to the Dayhoff matrices, the score provided is the natural log of the ratio between the number of observed pairs of amino acids at any position to the number of pairs expected from the overall amino acid frequencies. Sequences having a degree of similarity higher than a certain threshold are considered as a unique sequence, so that the substitution frequencies are more heavily influenced by sequences that are more divergent than this cutoff. A threshold of 62% defines the matrix BLOSUM62, which is the most used BLOSUM matrix [40,41].

As for insertion and deletion of residues, a gap penalty is introduced. Deletion of a succession of contiguous amino acids is a more probable event that deletion of the same number of amino acids at non-contiguous positions. Gap initiation have therefore a higher penalty score with respect to gap extension [40,41].

Once the optimal alignment is obtained, its significance can be assessed from the comparison of the observed alignment score to those of many alignments made from random sequences of the same length and composition as those under study [40,41]. Sequences potentially related to the one of interest can be retrieved from the web by means of a database similarity search. The bioinformatics methods available today for carrying out such a search sequentially align a query sequence to each subject sequence in the database, and report the results as a ranked list, providing the score and statistical parameters. One of the most commonly used methods, BLAST, is described in the next paragraph.


### 3.1.3 BLAST

BLAST, that stands for 'Basic Linear Alignment Sequence Tool', is a powerful tool for searching the sequence databases with a probe sequence. BLAST seeks for regions of local similarity between two sequences, using a particular variation of the Dayhoff matrix to calculate the similarity scores [43]. A maximal segment pair (MSP), defined as the highest scoring pair of identical length segments chosen between two sequences, provides a measure of local similarity for any pair of sequences. The

boundaries of the MPS are not fixed, but they are chosen to maximise the score. BLAST looks for MPS with score above a certain cut-off S; because the local alignment produced by the program does not contain gaps, it will be broken into several MPSs. To increase the speed of the search, instead of requiring words to match exactly, BLAST seeks only segment pairs that contain a word with a score of at least T when a substitution matrix is used to compare the word from the query. In this way the length of the word *w* can be kept long and the search speed elevated, without losing sensitivity. If the value of T is high, the program will run faster but it will find a big number of hits. On the other hand, a small value of T increases the execution time but permits to identify distant relationships. When a word hit is identified, the program attempts at determining a locally optimal and ungapped alignment, whose score is greater or equal to S, by extending to the left and to the right the alignment and accumulating the scores. The score dropoff threshold hit *X* allows unpromising aligment to be discarded: when the reduction of the score (relative to the maximum value encountered) exceeds *X*, the extension of a hit is terminated.

In a more recent version of the program, two words hit must be found within a certain window of residues in order to be considered. With this strategy, many random hit can be avoided and the ones that are processed are more likely to lead to a significant alignment. Moreover gaps are treated explicitly, resulting in a higher sensitivity towards distant sequence relationships. After initial standard comparison of the MPSs, the column that is central to the highest score alignment is identified, and the local gapped alignment is extended in both directions, and abandoned when many negative substitution scores are encountered [41].

Position-specific iterative BLAST (PSI-BLAST) is extremely powerful in picking up distant relationships. The program probes each sequence in the chosen database independently for local similarity to the query sequence, using a BLAST-type search, allowing gaps. The significant local matches are used to create a multiple sequence alignment table with the query sequence. Then PSI-BLAST uses the table to create a *profile*. A profile is a *position-specific scoring matrix* that lists the frequencies of finding each of the 20 amino acids at each position in a conserved sequence, weighted according to the substitution probability (defined by the amino acid substitution matrices). A particular amino acid could score high if it appears frequently in the inventory at this position or if it has a high probability of arising by mutation from residue types that are common at this position. The program then re-probes the database considering the profile, and the process is iterated until a cycle produces no changes [40].

The databases searched by BLAST and PSI-BLAST are listed in the table below.

**Table 3-1 Databases used by PSI-BLAST. In the non-redundant database, redundancy is reduced by merging sequences that are completely identical [40].**

| Database | Description |
| --- | --- |
| nr | Non–redundant merge of Swiss-Prot, PIR, PRF, and proteins derived from GenBank coding sequences and PDB atomic coordinates |
| month | Subset of the *nr* that is new or modified within the last 30 days |
| swissprot | The Swiss-Prot database |
| pdb | Amino acid sequences parsed form atomic coodinates |
| yeast | Complete set of proteins encoded by the *S. Cerevisiae* genome |
| ecoli | Complete set of proteins encoded by the *E. Coli* genome |

### 3.1.4 ClustalW and ClustalX

ClustalW is a multiple sequence alignment program for protein and nucleotide sequences [44]. As in many other approaches, the alignment is built progressively by a series of pairwise sequence alignments. First the most closely related sequences are aligned, and gradually the distant ones are added, following the branching order of a phylogenetic tree. In this way a multiple sequence alignment is constructed.

This approach however may encounter two main problems. One is defined as the 'local minimum problem', which stems from the fact that any mistake made in the initial steps of the alignment, normally because of an incorrect branching order in the guide tree, cannot be repaired later. Therefore it is not guaranteed that the multiple alignment obtained is the optimal one.

Another major problem is the choice of the alignment parameters. In general, a weight matrix and two gap penalties (one for opening a gap and one for extending an existing one) are adopted throughout the whole alignment procedure. This can be reliable as far as the sequences are closely related. For divergent sequences however, the scores given to non identical residues is critical, as the number of mismatches overcomes the one of identities. Different weight matrices would be optimal at different evolutionary distances or for different classes of proteins. Moreover the values of the gap penalties acquire a big importance for distant proteins, while in evolutionary close sequences they don't count as much.

ClustalW attenuates the problem of the parameters' choice by allowing variation of the gap penalties in a residue and position specific manner, and introducing the possibility of choosing different weight matrices as the alignment proceeds.

As for the gap opening penalties, they are adjusted considering the observed relative frequency of gaps adjacent to each of the 20 amino acid. In addition, they are reduced in stretches of five or more hydrophilic amino acids, which indicate the presence of a loop or a random coil region. Instead a high penalty is given when a gap is opened within a sequence of eight amino acids, because it was observed that two non-contiguous gaps tend not to be closer than eight residues on average.

Based on the estimated distances between the sequences that are aligned, ClustalW uses different weight matrices. The default is the BLOSUM matrix but the Dayhoff PAM matrix is available as well. In each case, there is a choice of matrix ranging from strict ones, useful for comparing very closely related sequences to very "soft" ones that are useful for comparing very distantly related sequences. In addition, sequences that are very similar to other sequences already present in the database are down-weighted while very distant ones are up-weighted, to correct for the uneven sampling across the evolutionary distances in the dataset. Weights are calculated directly from the branch lengths in the initial guide tree [44].

ClustalX is a graphical interface widely used to display alignment results. It also provides tools for alignment quality analysis, which marks low scoring regions in the alignment. This is very important because it permits refinement of the local alignment, either automatically or manually. A sequence colouring scheme based on physico-chemical properties of the amino acids then, highlights conserved features in the alignment [45].

## 3.2 *Secondary structure analysis*

### 3.2.1 CSI: Chemical Shift Index

CSIs (secondary chemical shifts, i.e. the deviation of the measured chemical shifts of the protein residues from the correspondent random coil tabulated values) are good parameters to identify the secondary structure motifs.

^{1}H NMR chemical shifts are strongly dependent on the character and nature of the secondary structure of a protein. In particular, the chemical shifts of the α-CH protons of all the natural occuring amino acids experience an upfield shift, with respect to the random-coil values, when in a helical conformation and a comparable downfield shift when in an extended conformation.

Wishart et al. introduced a reliable and empirical method that allows a rapid and quantitative determination of the identity, extent and location of secondary structure

elements in proteins based on the simple inspection of the $H^{\alpha}$, $^{13}C^{\alpha}$, $^{13}C^{\beta}$ and carbonyl $^{13}C$ chemical shifts [46]. The technique develops in two stages. In the first stage the program calculates the differences between the input and the random coil values, assigns a chemical shift index (CSI) for each type of atom and plots the results against the residue number. When the chemical shift of a particular atom is greater or smaller than a tabulated range, a CSI of 1 or -1, respectively, is assigned to that atom; when the chemical shift is within the range, the assigned CSI is 0. In a second time, secondary structure elements are identified according to the densities of the CSIs. A dense grouping of four or more -1's not interrupted by 1, is a helix, while a dense grouping of three or more 1's corresponds to an extended conformation. All the other regions are designated as coil. A local density of more than 70% of non zero CSI is needed to define a secondary structure element. The termination points of helices or strands are determined by the first appearance of a CSI of opposite magnitude to the one of the corresponding secondary structure conformation or by the presence of two consecutive zero-valued CSI. When the CSIs calculated at least for three atom types are consistent, the program calculates a consensus secondary structure [46-48].

### 3.2.2 TALOS: Torsion Angle Likelihood Obtained from Shift and sequence similarity

TALOS is a database system for empirical prediction of $\psi$ and $\varphi$ torsion angles. It reads the experimental $H_{\alpha}$, $C_{\alpha}$, $C_{\beta}$, CO and N chemical shifts and converts them in secondary chemical shifts, before entering them in the database. Then it evaluates the similarity in amino acids sequence and secondary chemical shifts for a string of three sequential amino acids relative to all strings of sequential residues taken from proteins with defined structures contained in its database. For Gly residues $^{1}H_{\alpha}$ chemical shift is considered the average between $^{1}H_{\alpha 2}$ and $^{1}H_{\alpha 3}$.

If amino acids' triplets in the target protein have similar secondary chemical shifts and sequences to those present in a known protein, then the $\psi$ and $\varphi$ angles of the known protein are useful predictors of the angles in the target.

The similarity is measured with a score based on the weighted sum of squares differences between the shifts in the target protein and in the database entries, so that the lower score indicates high similarity. The score also includes a qualitative residue type term that biases the errors due to roughly similar sequences.

TALOS searches a database of 78 known proteins, providing more than 9000 residue triplets; the output for φ and ψ backbone angles for the centre residue in each string consists of the average of the corresponding angles in the 10 strings in the database with the highest degree of similarity. The program considers only those values for which at least 9 out of 10 predictions fall in the same populated region of the Ramachandran plot and none of the centre residues in the 10 strings has a positive φ angle. If a single residue falls outside the Ramachandran region in which the remaining 9 residues are located, its φ/ ψ values are excluded from calculating the average and the rmsd.

**UNKNOWN STRUCTURE**

$^{13}CO$, $^{13}C_\alpha$, $^{13}C_\beta$, $^{1}H_\alpha$, $^{15}N$
Chemical shifts

Residues (i-1, i, i+1)

**TALOS PROGRAM**

**Uses:**

– residue similarity matrix

–Compensation factors for missing shifts

–Secondary shift weighting factors

**Calculates:**

– secondary chemical shifts

– differences in secondary chemical shifts for all 5 atoms between residues (i-1,j-1),(i,j),(i+1,j+1)

–Overall similarity score, $S$(i,j)

–Dihedral angles of the 10 lowest scores

–Average (Φ,Ψ) and the st. dev. of the 10 lowest scores

**detects:**
missing assignments

**DATABASE**

Currently: 78 proteins,
        more than 9000 residues
**NMR** Chemical Shifts:
$^{13}CO$, $^{13}C_\alpha$, $^{13}C_\beta$, $^{1}H_\alpha$, $^{15}N$
**X-Ray** structures (PDB):
Resolution better than 2.2 ?
Residues: (j-i,j,j+1)

Predicted (Φ,Ψ)
for residue i

Interactive inspection of predictions, elimination of outlieres, selection of useful predictions

**Figure 3-1: Flow-chart of the TALOS program (adapted from [49]).**

However, if several predictions fall outside the most populated area of the Ramachandran map but generally cluster well with the other φ/ψ predictions, the prediction is accepted.

The standard deviations and the range of (φ, ψ) values in the 10 most similar database strings provide a measure of the uncertainty in these averages. When the

standard deviation exceeds 45° the prediction must be labelled as ambiguous. In this case one should discard the result without further inspection of other data, such as the $d_{\alpha N(i-1,i)}/d_{\alpha N(i,i)}$ NOE intensities or $^3J_{HNH\alpha}$.

TALOS provides three different graphical interface windows: the sequence display, the prediction display, the Ramachandran display. The first window lists the residues in the protein whose backbone angles have been predicted, differently coloured according to whether the prediction for the given residue was designated as good, bad or ambiguous. The selection of a residue from this window will recall the prediction window, where the parameters relative to the prediction, the angles values, the standard deviation and the similarity factor, are listed, together with the residue numbers and the names of the proteins from which the triplets were taken. The 10 $\varphi/\psi$ values can then be pictured in the Ramachandran plot. By clicking on an individual match in the Ramachandran display it is possible to include or remove the entry form the overall prediction [49].

## 3.3 *3D Structure analysis*

### 3.3.1 Structure calculation: Ambiguous Restraints for Iterative Assignment (ARIA)

The program derives distance restraints from NOESY spectra and calculates molecular conformations that are compatible with the observed cross peaks. It uses CNS employing simulated annealing strategy to calculate the structure [50]. What is essential for the program to run is the definition of the molecular system, which usually is the sequence of the interested protein, a NOESY spectrum and a list of chemical shift assignments. Other kinds of restraints can be added to refine the structure calculation.

The first thing ARIA does is creating the molecular topology file. This file defines the name, chemical type, charge and mass of each atom, as well as the covalent connectivities. An extended conformation molecule, whose coordinates are written in a PDB file, is constructed. The extended conformation is then randomized by default and used as starting structure for the calculation.

ARIA then checks for inconsistencies between the list of chemical shift assignments provided and the peaks on the spectrum. Whenever it finds a mismatch, it ignores the peak and deletes it from the current list used for the structure calculation. It is therefore very important to check carefully the assignment before proceeding with the

48

calculation. ARIA will discard diagonal peaks since it cannot derive any distance restraints from them.

The step that follows is the assignment of the NOEs. When a protein's spectrum is manually assigned, many peaks can be misassigned, because of the severe overlap typical of a NOESY spectrum. A very nice feature of the software is the capability of assigning the peaks automatically, starting from the chemical shift list. The beauty of this feature is that the program provides, for each peak, all the possible assignments consistent with its chemical shifts. A proton is assigned to all cross-peaks whose frequency is included in a proton chemical shift window, defined by the user; in some cases the assignment is unambiguous, being the only possible one for that cross peak, but in many other cases the cross-peak can be attributed to different protons pairs, thus being ambiguous. The user should choose very carefully the frequency windows, because selecting a too narrow window would result in discarding many possible assignments, while having too broad a window could cause severe convergence problems. Moreover it is important at this point to have a list of chemical shifts as much complete as possible, because atoms with missing resonances cannot be assigned to any cross-peaks.

The value of an ambiguous NOE cross-peak at the chemical shift coordinates F1, F2, contains contributions from all proton pairs with the same chemical shift assignment

$$NOE_{F1,F2} = \sum_{i\in(F1,\Delta 1), j\in(F2,\Delta 2)} NOE_{i,j}$$

(3.1)

where $\{F,\Delta\}$ is defined as the set of protons with chemical shift between F-$\Delta$ and F+ $\Delta$. In the isolated spin pair approximation (ISPA), an ambiguous NOE is related to a summed distance $\overline{D}$:

$$NOE \propto \overline{D}^{-6}$$

(3.2)

$$\overline{D}_{F1,F2} = \left( \sum_{i=1}^{N(F1,F2)} D^{-6}_k \right)^{-1/6}$$

(3.3)

where $k$ runs through all N(F1,F2) contributions to a cross-peak at frequencies F1, F2 and $D_k$ is the distance between two protons corresponding to the contribution $k$. The distance can be calculated from the coordinates of a model structure. Structure

calculation and refinement then proceed by restraining $\overline{D}$ by means of an appropriate target function to upper and lower bounds U and L derived from the size of the NOE:

$$L \leq \overline{D} \leq U \qquad (3.4)$$

Once the structure is obtained, the small contributions that correspond to large distances in the structure can be discarded.

The assignment of a peak is made in such a way to take into consideration all the contributions from the different protons pairs, appropriately weighted. For each contribution $k$ to the ambiguous NOE the minimum distance in the ensemble of converged structures $D_{\min}^k$ is defined. The contribution $C^k$ of the assignment $k$ to the cross peak is estimated as:

$$C^k = \frac{D_{\min}^{k\ -6}}{\sum_i^{N(F1,F2)} D_{\min}^{i\ -6}} \qquad (3.5)$$

The $C^k$ are then sorted according to their size. The $N_p$ largest contributions are chosen such that:

$$\sum_i^{N_p} C^i > p \qquad (3.6)$$

where $p$ is a parameter set by the user that can be narrowed during the iterations [51].

In order to calculate distances from the NOEs' values, a calibration factor must be defined. In each iteration after iteration 0, the calibration factors are based on distances calculated from the ensemble of the S lowest energy structures derived from the previous iteration. For each proton pair, a distance characteristic for the ensemble, $\hat{d}_{i,j}$, is calculated as an arithmetic average:

$$\hat{d}_{i,j} = \frac{1}{S}\sum_{S=1}^{S} d_{ij,s} \qquad (3.7)$$

These distances can be used to directly calibrate experimental peaks and extract distance restraints for iteration $i+1$, by setting the calibration factor C to:

$$C = \sum_{noe's} \frac{\hat{d}^{-6}}{V}$$

(3.8)

where V is the observed NOE volume or intensity and the sum runs over all experimentally measured NOEs for which the corresponding $\hat{d}$ is smaller than a cutoff of 6 Å. The observed distance $d^{obs}$ is calculated as:

$$d^{obs} = \left(CV\right)^{-1/6}$$

(3.9)

The calibration will improve from iteration to iteration with the improving structure. Alternatively, a calibration factor can be calculated from reference NOEs with known distance or from a simulated NOESY spectrum, obtained from the mixing time and the relaxation matrix derived from $\hat{d}$. A spin diffusion correction term can be included in the calibration to improve the accuracy of the distance restraints [52]. For iteration 0 a template structure is used for calibration and assignment [53].

In every iteration, ARIA calibrates and assigns the NOESY spectra, merges the restraint lists and calculates a structure ensemble using CNS [50]. Using then the distance restraints, the program decides which peaks can be considered real NOEs and which are noise or wrongly assigned. The violation analysis is made by comparison of the distance bounds obtained in a set of calculation with the ones found in the structure ensemble; a restraint is violated if the distance found in the specific structure lies outside the bounds for more than a user defined violation tolerance. Restraints that are violated in the majority of the structures and with a large upper-bound violation are inconsistent or stem from a wrong assignment: they should then be inspected manually.

ARIA takes into account the chirality of methylene and methyl protons of the isopropyl groups of valine or leucine, for which it is often impossible to have a specific assignment for each proton. ARIA compensates this lack of information by testing both alternatives during structure calculation, and for each calculated structure it gives the energetically preferred option.

Finally ARIA provides a protocol for the refinement of the structures in a solvent, usually $H_2O$.

The ARIA package also provides several programs that calculate quality indices for the output structures and permit to judge the reliability of the results [54-56].

### 3.3.2  Structure validation

Knowing the structural details of a macromolecule is fundamental for drug design, homology modelling and structural genomics. A calculated structure must therefore satisfy specific quality parameters, to ensure local and global good definition. In this study PROCHECK and WHATIF were used for structure validation.

#### 3.3.2.1  Procheck

Procheck and Procheck_NMR are validation programs useful to check crystallographic and NMR structures [57]. They read a .pdb file, with the spatial coordinates of the molecule of interest. The PROCHECK suite comprises five different programs run in succession. The first program, CLEAN.F, produces a clean-up version of the input structure, ensuring that all the atoms have been correctly labelled in accordance with the IUPAC naming convention. The program also checks that the right L/D stereochemical labels have been assigned to the residues. The second program, SECSTR.F, is responsible for making the secondary structure assignments on a residue-by-residue basis. NB.C, the third program, identifies all nonbonded interactions, ANGLE.F checks the angle values and  PPLOT.F produces all the plots and the residues listing [58].

The checks carried out by Procheck make use of ideal bond lengths and bond angles derived from a comprehensive analysis of small molecule structures. The quality level is assessed with G-factors, that are log-odds scores based on the observed distribution of the stereochemical parameters in an internal protein database.

Table 3-2 provides a summary of the quality checks carried out by the program.

**Table 3-2: Summary of the stereochemical parameters that are checked by the program PROCHECK. The first column lists the parameters, in the second column a brief description of non obvious parameters is provided together with the comparison values. Ticks on the third column indicate the presence of a graphical output.**

| Parameters | Notes | Graphical output |
|---|---|---|
| Main chain bond lengths | They must be below 0.5 Å from the reference value | ✓ |
| Main chain bond angles | They must be below 10° from reference value | ✓ |
| Planarity of aromatic rings | They must have a RMS within 0.03 Å from the best fit plane | ✓ |
| Planarity of end groups (Arg, Asn, Asp, Gln, Glu) | They must have a RMS within 0.02 Å from the best fit plane | ✓ |
| Ramachandran: psi-phy distribution | Regions of the Ramachandran are divided in: 1) Core; 2) Allowed; 3) Generous; 4) Disallowed | ✓ |
| Chi1-Chi2 distribution | | ✗ |
| Chi1 torsion angles | | ✗ |
| Chi3-Chi4 torsion angles | | ✗ |
| Omega torsion angles | | ✗ |
| Chirality check | Measure of the $C_\alpha$ tetrahedral distortion. The expected value is 33.9° | ✗ |
| Non bonded interactions | Bad contacts when two non hydrogen bonded atoms are closer than 2.6 Å | ✗ |
| Main chain hydrogen bonds | Hydrogen bonds energy. Ideal value= -2.0 Kcal/mol | ✗ |
| Disulphide bonds | Ideal S-S separation=2.0 Å. Ideal Chi3 torsional angle value, defined by S-S bridge: -85.3° for left handed conformation 96.8° for right handed conformation | ✗ |
| Parameter comparison | How the structure compares to well refined structure at similar resolution | ✗ |
| Residue by residue analysis | Overview of how typical each residue in the structure is | ✓ |

### 3.3.2.2  Whatif

Whatif is a computer program that provides a flexible environment for displaying, manipulating and analysing small molecules, proteins, nucleic acids and their interactions. It contains a high number of options, for graphics, molecular comparison, drug docking, protein mutations, structure analysis. The core of the program is constituted by a buffer called '*soup*', where the molecule's coordinates are stored together with related parameters such as matrices, file names, pointers to databases, standard datasets. External programs are interfaced, for instance the PSQ/PIR for access to protein sequence database, GROMOS for molecular dynamics and energy minimization, MODEL for design small molecules. The input file must be a .pdb file; in the case of a NMR structure, ensembles of more molecules can be entered as well. The user is then prompted to choose an action from a large menu. Stereochemical controls similar to those carried out by Procheck can be entered with the option FULCHK (fullcheck). The quality of each parameter is evaluated in terms of Z scores. In the output the structure-Z scores such as packing quality, the Ramachandran plot appearance, the chi1-chi2 rotamer normality and the backbone conformation are listed. Whatif also calculates RMS-Z scores referred to local geometry on bond lengths, bond angles, omega angle restraints, side chain planarity, improper dihedral angle distribution. The parameters that need further checking are highlighted.

Given a Gaussian distribution of the parameter X, the Z score is defined as:

$$Z - score = \frac{X - average}{\sigma}$$

(3.10)

where σ is the standard deviation.

When X is calculated for different residues, the RMS-Z score is a better measure of the quality of X. For n residues, the RMS-Z score is defined as:

$$RMSZ - score = \sqrt{\frac{\sum_n Z^2}{n}}$$

(3.11)

Since the reference RMS-Z score is 1, values of RMS-Z score closer to 1 indicate a better quality.

The overall quality is checked against a dataset of well refined X-ray structures that is periodically updated, while for the local geometry quality assessment the Cambridge small molecules database (CSD) is also used [59].

## 3.4 *Structure comparison*

Structure tends to diverge more conservatively than sequence during evolution. In particular, patterns of contacts between residues are shared by related protein, even in very distant homologues. By identifying and comparing related pattern, we can distinguish variable and conserved features and therefore understand which ones are crucial to structure and function [40].

### 3.4.1 Dali

ali (Distance-matrix ALIgnment), is a powerful tool to reveal ancient evolutionary relationships by structure-structure comparison. The program can perform an all-against-all structural comparison of all protein families in the PDB or searches with a single structure [60]. A database called FSSP of pre-calculated structural neighbours is available on the Web Site of the program and contains a classification of folds. Each protein in the database occupies a specific position in a protein fold classification tree and is linked to a list of structural neighbours. The database is continuously updated.

The search with a single structure is extremely useful for NMR spectroscopists and crystallographers: at the end of a structural determination, it is important to check whether the structure has a unique fold or if it shares structural similarities with known proteins. In this latter case in fact, functional information can be derived.

When a new structure is complete, one can send its coordinates through the Web to the Dali server, which will provide a list of all the structural neighbours of the query structure in the fold space and an optimal structural alignment with each neighbour.

The search is fully automated. Structural relatedness is measured in terms of similarities of intramolecular matrices, using exclusively three dimensional coordinates. For two proteins Dali builds matrices of contact patterns and then seeks for maximal matching submatrices [40]. The distance between two structures is measured in terms of root mean square deviation (RMS deviation):

$$RMS = \sqrt{\sum_i d_i^2 \Big/ n} \qquad\qquad (3.12)$$

where $d_i$ is the distance between the $i^{th}$ pair of points after optimal fitting and $n$ is the number of points.

The statistical significance of structural similarity has been determined empirically. Hits in the database search are ranked in order of statistical significance with a cutoff set to a Z-score of 2 for any pair of structural domain. The output file also gives the RMSD between the two structures [60].

# Frataxin and Friedreich's Ataxia 4

## 4.1 *The disease*

Friedreich's ataxia (FRDA) is a severe neurodegenerative pathology, that affects the central nervous system and the myocardium. Typical clinical features are progressive gait and limb ataxia, absent lower limb reflexes, neuropathy, hypertrophic cardiomyopathy, degeneration of the spinal cord -mainly the posterior columns and spinocerebellar tracts- loss of large cells in the dorsal root ganglia, loss of large myelinated axons in the peripheral nerve, diabetes mellitus, skeletal deformities. The disease manifests itself between the 5$^{th}$ and the 15$^{th}$ year of life and has a relentless progression that leads to death [7].

It is an autosomal recessive disease, mainly spread in the Caucasian population, with a prevalence of 1:29000 but with a carrier frequency as low as 1:60 [7]. The molecular defect was identified as a deficiency of frataxin, a 210 amino acid protein encoded by the FRDA gene, that was mapped in 1988 on chromosome 9 and successfully cloned in 1996 [61]. The gene comprises seven exons (1-5a, 5b, 6). The commonest transcript arises from exons 1-5a but exon 5b may be transcribed as well by alternative splicing, resulting in a 171 amino acids protein; exon 6 is non-coding. Frataxin is extremely conserved, with the highest conservation score for amino acids 141-167, that represent the C terminus of the protein, and has no apparent sequence homology with any other proteins of known function. It is a nuclear encoded protein associated with the mitochondrial membrane, ubiquitously expressed, but with higher expression levels in heart and spinal chord, and intermediate levels in the liver, skeletal muscle, pancreas, cerebellum [61]. Mouse embryo studies have shown that frataxin expression increases between day 10-14.

Deficiency of frataxin may occasionally be traced to nonsense splice sites or missense mutations, but the primary cause is the hyperexpansion of a polymorphic GAA trinucleotide repeat situated in the first introne of the FRDA gene, which results in a marked reduction in the steady-state level of mature frataxin mRNA. While normal alleles contain between 6 and 34 GAA triplets, expanded alleles contain 67-1700 repeats [62]. The expansion is unstable with respect to its transmission from parents to offsprings: maternal transmission may result in smaller or larger alleles in the offsprings, while male transmission almost always results in a size decrease of

the repeats. The so called 'premutation' alleles are of intermediate size between the one found in a healthy person and the one found in FRDA patients, and they are very likely to sensibly increase size within one generation. The length of the GAA repeat in the smaller allele correlates with the age of onset, disease progression and cardiomyopathy, which is the principal cause of death [7].

Low levels of frataxin mRNA were found in cerebral and cerebellar cortex, skeletal muscle and limphoblasts of FRDA patients, indicating that lack of frataxin is caused by a suppressed gene expression [8]. The presence of 'sticky' DNA, that forms upon self aggregation of strands bearing the GAA triplet repeats, is thought to interfere with the correct transcription [63].

While on the 95% of the cases FRDA patients are homozygous for a GAA expansion, 5% of the affected persons are heterozygous for an expansion and a point mutation affecting the frataxin codon sequence. Fifteen point mutations have been identified so far. Eight of them are truncating mutations, mainly found in exon 1, while the others are missense mutations found in the last three exons. Three mutations are found in the ATG initiation codon and result in a mislocalisation of the protein. The second ATG in frame codon is in fact located in exon 2 and translation of this coding sequence produces a N terminal truncated protein that lacks the mitochondrial targeting peptide. The other truncating mutations and the point mutations I154F, L156P, L182H, H183R, W173G affect residues contained in the carboxy-terminal domain, and result in an almost complete loss of the protein's function. Patients heterozygous for the GAA repeat and one of these point mutations develop symptoms undistinguishable from those of homozygous. On the contrary, D122Y and G130V mutations occur at the N terminal domain and are associated with milder FRDA features. To date no cases of patients homozygous for point mutations are known [64].

## 4.2 *Frataxin adopts a new fold*

Human and yeast frataxins are synthesised on cytoplasmic ribosomes as pre-proteins with an N-terminal mitochondrial signal sequence. After importation into the mitochondria, the proteins are processed by the metallopeptidase MPP (mitochondrial processing protease) and converted into their mature form [65,66]. Upon expression in *E. Coli* human frataxin (hfra) spontaneously degrades to an N truncated form that spans residues 63-210 while the bacterial orthologue CyaY naturally lacks the N terminal import sequence [67]. The C terminal domain is highly conserved: CyaY shares 24.5% of identity with human frataxin (residues 90-210) and

26.4% with yeast frataxin, Yfh1p (residues 68-174), which in turn shares 37.8% sequence identity with hfra [68,69].

The 3D structures of human, yeast and bacterial frataxin reveal that they adopt a compact αβ sandwich fold, conserved within the family, that do not resemble any of the proteins' folds already determined [70, 21]. Both NMR and crystal structure of hfra were obtained on N-terminal truncated species (residues 92-210 and 88-210 respectively) [21,70]. Conversely, the NMR structure of the yeast frataxin includes also the dynamic 18 amino acid N terminal mitochondrial import sequence, marked by a $3_{10}$ helix between residues 12-15 [69]. The comparison of the structures of the human and bacterial form, shows that human frataxin has a four residue insertion in the loop between α1 and β1 with respect to the bacterial orthologue and that residues 180-181 are structured to form an additional β strand, while in CyaY they are unstructured [71]. The C terminal domain has different length in the three orthologues and is responsible for their different thermodynamical stability. In both hfra and CyaY, it is long enough to insert between the two helices and to pack against the hydrophobic core, thus increasing the folding stability. In Yfh1p instead, the C terminal tail is shorter, including only four residues after the end of the secondary structure, and cannot form any stabilising interactions [23].
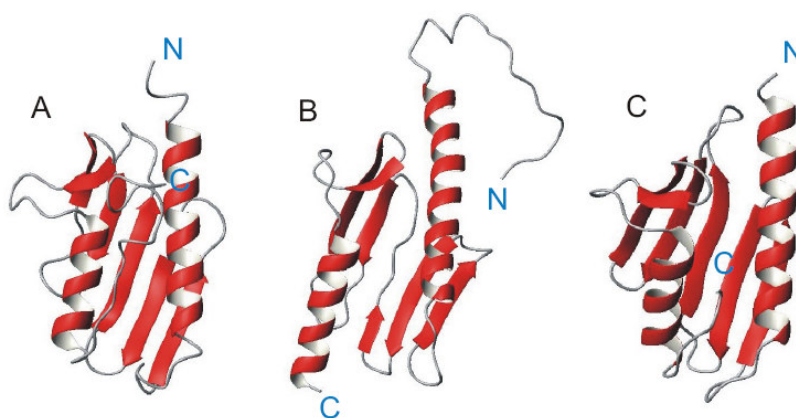


**Figure 4-1: 3D NMR structures of the human, yeast and bacterial frataxin orthologues (A, B and C). The C terminal tail can be seen in the human and bacterial proteins, being longer in human frataxin. The long unstructured N terminal mitochondrial import sequence is evident in the yeast orthologue [21,69,71].**

The three proteins have a very acidic nature: the measured isoelectric point (PI) for human, bacterial and yeast frataxin are respectively 4.98, 4.24 and 4.34. Strikingly the charge distribution is asymmetric, with a conserved superficial negative ridge on the region comprising α1 and the edge of β1. A large patch of conserved hydrophobic residues is also present on the surface.
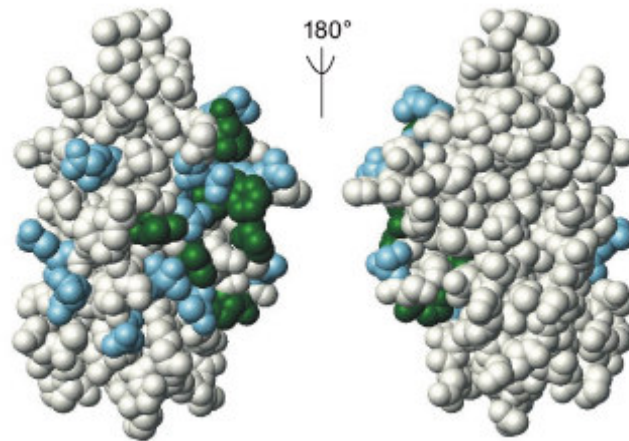
**Figure 4-2: Distribution of the conserved residues on human frataxin surface. Hydrophilic and hydrophobic residues are shown in cyan and green respectively [21,70].**

It is interesting to note that four of the missense mutations associated with FRDA disease, occur within these two conserved regions. D122Y (human frataxin numbering) is within the anionic patch, while G130V, W155R and R165C are located within the flat, conserved external surface of the large β sheet [70]. These observations suggest that both regions are functionally significant, probably as mediator of intermolecular interactions. The other conserved residues are buried to form the hydrophobic core of the proteins; missense mutations in these residues provoke the disassembly of the structure. The proteins do not contain cavities or typical iron binding motifs, but can weakly bind iron II and iron III with residues belonging to the negative conserved patch [21,68-71]. A detailed description of CyaY interaction with cations is the subject of chapter 5.

## 4.3 *Frataxin is involved in oxidative stress and iron homeostasis*

How frataxin deficiency leads to neuro- and cardio-degeneration is still unknown, but potential pathogenic mechanisms have been proposed in the past years. Accumulating evidence suggests that frataxin is involved in iron metabolism, iron-suphur cluster synthesis, oxidative phosphorilation and antioxidant defense [7]. Its remarkable conservation throughout evolution, from purple bacteria to humans, suggests that it plays a vital role. This is also supported by the observation that frataxin *knock-out* mice die *in utero* shortly after implantation [72]. The protein is important during early development of neuronal lineage of cells: reduced levels of the

protein in fact lead to an increased concentration of reactive oxygen species (ROS) in differentiating neuroectodermal cells, resulting in a higher susceptibility to apoptosis [73].

The involvement of frataxin in iron metabolism is indicated by several data based on studies of the human and yeast frataxins [15,74-77].

Iron deposits are found in the biopsied myocardial fibers from FRDA patients, together with deficient activities of iron-sulphur cluster containing proteins, such as subunits of mitochondrial respiratory complexes I, II, III and mitochondrial and cytosolic aconitase. Fibroblasts from FRDA patients show a higher content of mitochondrial iron with respect to control fibroblasts and a greater susceptibility to oxidants-induced death, phenotype that can be in part reduced by adding iron and calcium chelators as well as apoptosis inhibitors [78].

Yeast frataxin knock-out strains develop similar features. Deletion of the YFH1 gene results in a 'petite' phenotype, i.e. a severe growth deficit on fermentable carbon sources and a dramatically reduced oxygen consumption, due to the loss of mitochondrial DNA. Moreover, an abnormal increase of the mitochondrial iron concentration (10 times that of wild type cells mitochondria), high sensitivity to oxidative stress and reduction of oxidative phosphorilation were described. Similarly to what observed in human cells, in frataxin-defective yeast strains the levels of mitochondrial and cytosolic aconitase are very low and cannot be restored by addition of iron chelators in the culture media. Conversely, the same treatment normalises the oxidative respiration, suggesting that defects in the respiratory chain depend on mitochondrial iron accumulation, whereas reduced aconitase activity is directly linked to frataxin deficiency [79]. This suggests an important role of frataxin in the synthesis of the iron-sulphur clusters.

However an alternative role of the protein in activating oxidative phosphorilation and energy conversion was also postulated. In fact it was noticed that Friedreich's ataxia phenotype in humans is associated with reduced ATP-levels in skeletal muscles during exercises and that the most severely affected tissues, i.e. heart and neuronal tissues, are exclusively dependent on oxidative metabolism. In addition, mutant mice in which the frataxin gene was selectively inactivated in neuronal and cardiac tissues develop neuronal sensory defects and cardiomyopathy simultaneously to mitochondrial iron-sulphur enzyme deficiencies and show mitochondrial iron accumulation only at a later stage. From these data it appears that iron accumulation is a consequence of ATP depletion [80,81].

## 4.4 *Frataxin is involved in iron-sulphur clusters' synthesis*

One of the major problems encountered in the study of a complex molecular disease such as FRDA is to distinguish between early and late consequences of the molecular defect.

To unmask the first event caused by frataxin depletion, Lill et al. created a yeast mutant harbouring a switchable YFH1 gene, Gal-YFH1. By regulated expression of the gene, they were able to follow the time course of the typical hallmarks of the FRDA disease. They found that Yfh1p does not bind iron *in vivo* in a detectable stable way and that Gal-YFH1 cells were strongly impaired in Fe-S protein maturation, even in the absence of an excess of iron. Fe-S cluster deficiency therefore appears to be the primary event in the FRDA disease [76].

Much evidence that frataxin primarily influences iron-sulphur clusters' biosynthesis comes also from studies in human cells. Upon specific depletion of the FRDA gene product using the RNA interference technology, decrease of frataxin levels in HeLa human cells was achieved. Concomitantly, severe reduction of aconitase activity, succinate dehydrogenase and increase of the RNA binding activity of the cytosolic IRP1, indicating loss of its Fe-S cofactor, were observed. More important, disturbances in the mitochondrial iron content were not detected within the first three weeks of frataxin deficiency, thus confirming that iron accumulation is a secondary effect of the impaired Fe-S cluster synthesis [15].

In addition, another study showed that the gene expression of the transcripts involved in the sulphur amino acid biosynthetic pathway and in the related iron-sulphur cluster pathway are frataxin-dependent in human FRDA fibroblasts, limphoblasts and in mutated neuronal NT2 cell lines [82]. In support of these findings, a phylogenetic study discovered the co-occurrence of the frataxin gene with hscA and hscB, two genes that belong to the Isc operon and that encode proteins directly involved in the biosynthesis of the clusters. The analysis of the evolutionary history of the Isc proteins and frataxin, led to the observation that, while iscR, iscS, iscU and Fdx have a widespread distribution, hscA, hscB and frataxin were invented later, at the same time, and were lost in the same species [11]. Another study reported that expression of the ISU1 gene, one of the central gene in the ISC operon, increases in Yfh1p mutant cells, indicating a link between these genes [75]. As further support, a synthetic lethality screen for proteins functionally interacting with Yfh1p, i.e. a search for extragenic mutations that cause lethality on iron media in the presence of a mild Yfh1p mutation, identified Isu1p as the main interacting species [14].

A genome wide analysis in yeast had revealed that frataxin deficiency results in enhanced transcription of genes that are under the control of iron-sensing transcription factors, AFT1/AFT2, including genes involved in iron uptake [75]. Since the AFT regulatory system is induced by low cytosolic iron concentration, the deletion of frataxin results in a redistribution of iron from the cytosol to mithocondria. A role of frataxin in the efflux of iron from the mitochondria was suggested by early studies when it was observed that induction of YFH1 gene in Yfh1p deleted yeast cells led to efflux of accumulated mitochondrial iron [74]. A similar iron cycle aberration is caused by depletion of Nfs1p, a critical enzyme involved in Fe-S cluster synthesis. It must be noted that alteration of the other two mitochondrial activities, heme synthesis and respiration, does not lead to increased mitochondrial iron accumulation [77].

A defective heme synthesis was indeed observed in yeast, resulting in low cytochrome c levels. In particular it was demonstrated that the iron accumulated in mitochondria of Yfh1p depleted cells was not bio-available for the ferrochelatase, the protein that catalyses the insertion of iron into the porfirin ring. The observation that iron was present in the cells as nano-particles of ferric phosphate explained its unavailability and suggested a role of Yfh1p in mediating iron use by ferrochelatase [12]. However it was later showed that defective heme synthesis is a general secondary consequence of defects in Fe-S protein biosynthesis and that it develops after shutdown of Fe-S cluster biosynthesis [83].

Finally, it has to be noted that, even if the yeast model has been widely extended to human, yeast frataxin does not play an essential role in Fe-S synthesis as long as Isu1p is not in limiting amounts. Cells with no frataxin in fact were still able to synthesize ISC proteins, although with very low efficiency [84]. Moreover cytosolic Fe-S cluster proteins were identified in frataxin knock out yeast, thus indicating that another not yet identified protein might be involved in the cluster biosynthesis or that a direct cytosolic Fe-S clusters' synthesis pathway is present. These findings, together with the distinct iron-promoted oligomerisation processes of yeast and human frataxin (see later, paragraph 4.6), provide warnings that the yeast model might not be fully carried over to humans [16,81].

## 4.5 *Frataxin as an iron chaperone*

Although a precise role of frataxin within the iron-sulphur clusters' synthesis mechanism has not been clarified yet, large evidence indicates that the protein might be responsible for iron delivery.

In an attempt at defining the cellular role of frataxin, Gerber et al succeded in experimentally determining the interaction of Yfh1p with the central Isc-assembly complex Isu1/Nis1 in mitochondrial extracts, already proposed by genomic analysis [14,85]. The interaction of the proteins was traced with a GST pull-down and a co-immunoprecipitation assay and was significantly higher when the concentrations of ferrous ion were kept at a physiological level during the purification of the complex. Using *Gal-YFH1* cells (see previous paragraph) to modulate frataxin expression and monitoring the amount of radiolabelled $^{55}$Fe co-immunoprecipitated with Isu1, they also suggested that the presence of Yfh1p induced formation of an ISC on Isu1p.

Their findings were supported by an *in vitro* study on human frataxin, that demonstrated the binding of ferric and ferrous iron by the monomeric protein and the transfer of the ions to Isu.

Using a colorimetric assay together with fluorescence and calorimetry measurements, Yoon and Cowan demonstrated that frataxin can bind ~ 6-7 iron ions per mole, with a dissociation constant of ~ 55 and 11.7 μM for ferrous and ferric ions respectively. They showed that frataxin is stable to iron release in the absence of external ligands and that it transfers it to ISU, where a temporary Fe-S cluster is assembled (see paragraph 6.3.3). The cluster could not be detected when free iron was incubated with ISU alone under the same conditions [16].

These studies strongly suggest that holo-frataxin mediates the reconstitution of the ISC on ISU, acting as the functional iron donor (chaperone) for the initial step of the reaction pathway for mitochondrial Fe-S cluster assembly.

However the iron chaperone role of frataxin does not appear to be specific for the synthesis of the Fe-S clusters. In fact it was recently observed that the protein can interact in an iron mediated fashion with ferrochelatase and aconitase as well. The interaction with ferrochelatase had already been described by surface plasmon interaction studies, although no evidence for frataxin mediated iron delivery in heme synthesis was presented [12]. Isothermal calorimetry measurements instead, provided a measured $K_d$ ~ 17 nM for a 1:1 complex between human frataxin and dimeric ferrochelatase. No binding was observed in the absence of iron. Measured in the same conditions, the $K_d$ for the frataxin/ISU complex is ~48 μM, with a binding stoichiometry of 0.95. Ferrochelatase activity reaches a maximum in the presence of 1 molecule of frataxin per dimer at a concentration of ferrous iron of 5.8 μM, supporting the hypothesis that frataxin recruits iron and delivers it to ferrochelatase [17]. A similar iron mediated interaction was also observed with NMR [69].

The interaction with aconitase was demonstrated by Bulteau et al., who suggested that frataxin is an iron chaperone protein that protects the aconitase $[Fe_4-S_4]^{2+}$ cluster

from disassembly and promotes the enzyme's activity. They in fact observed that the aconitase inactive $[Fe_3-S_4]^{1+}$ cluster was converted into the active form $[Fe_4-S_4]^{2+}$ only in the presence of iron loaded frataxin, citrate and DTT [9].

Further support to the iron chaperone activity of frataxin comes from an *in vitro* study performed on the bacterial homologue, where it was shown that CyaY is able to weakly bind two ferrous ions per monomer ($K_d$~4µM), thus making them bio-available. The authors hypothesise an iron chaperone role for the bacterial protein similar to the one indicated for the human and yeast orthologues, where the protein prevents the participation of iron to the Fenton chemistry while keeping it available for different biosynthetic pathways [86].

The ability of frataxin to deliver iron to aconitase, ferrochelatase and ISU could be part of a wide iron homeostasis control system: in a heme deficient condition, holo-frataxin should be used mainly for heme synthesis, while in the case of iron-sulphur cluster deficiency the majority of holo-frataxin would be utilised for Fe-S cluster biosynthesis. The difference in the binding constants of frataxin with ISU and ferrochelatase would also reflect the observed phenotype: since the affinity for ISU is much weaker than the one for ferrochelatase, decrease in the frataxin levels would primarily affect ISC synthesis [17].

However the relevance of frataxin iron binding ability is still a fairly controversial question. In fact neither Yfh1p nor human frataxin were indicated to stably bind iron in mitochondrial extract and under physiological conditions, and their 3D structures reveal that they do not possess any typical iron binding motifs [67,76]. On the other hand iron induced aggregation was also reported [13,18,87].

## 4.6 *Frataxin as an iron storage protein*

The evidence that yeast and bacterial frataxin form high molecular weight aggregates in the presence of high concentrations of iron (II) and iron (III), suggest that the protein might be an iron storage protein.

It was shown that Yfh1p undergoes a stepwise aggregation process *in vitro* that leads to a 48 subunits molecule ($\alpha_{48}$) in the presence of Fe(II) and $O_2$. As the aggregation takes place, Fe(III) accumulates, until an apparent maximum loading capacity of 50 atoms per subunit in $\alpha_{48}$. The iron stored in the protein then can be transferred to the ferrochelatase, involving an intermolecular interaction. Yeast frataxin would therefore act as an iron reservoir, that makes Fe(II) ions available to different users perhaps via hydrophobic interactions mediated by the neutral conserved surface found in the protein. In this respect the oligomeric yeast protein

would act as an iron e, as demonstrated for the monomeric form [17]. It was observed that when Fe(II) is not transferred, it is oxidised and stored in a water soluble mineral form within the assembled protein. A model was proposed where the monomer might be activated by Fe(II) in the presence of $O_2$ and might form a three subunit oligomer, with negatively charged inner surface able to sequester Fe(II) from the solution. The oligomer would then catalyse the oxidation of Fe(II) to Fe(III) and promote nucleation of ferrihydrite crystallite at the negatively charged surface. If the Fe(II) concentration exceeds that of the ferroxidase sites on the protein, ferroxidation is rapidly overcome by a slower autoxidation reaction at the surface of the growing crystallite. Alignment and binding of one ferrihydrite molecule to the other would lead to the interaction of trimers which would further facilitate biomineralisation [13,18].

The behaviour of human frataxin however does not resemble the one of the yeast orthologue In fact it was observed that, despite the human protein can be isolated from *E. Coli* as a mixture of monomeric and oligomeric species, the purified monomeric species is not able to undergo self assembly *in vitro* as it is, and only minimally when treated with increasing iron concentrations. Nevertheless, recovery of a high molecular weight frataxin species upon purification, suggests that aggregation can occur *in vivo* in the presence of physiological concentration of iron and of a still unidentified assembly factor [87]. Electron microscopy of the high molecular weight purified frataxin reveals the existence of rod-shaped structures containing small electron dense iron cores, similar to those formed by iron loaded ferritin. Moreover, a high molecular weight form of frataxin was showed to complement yeast frataxin *knock out* strains. The presence of high molecular weight frataxin in mouse heart, suggests that frataxin assembly can occur not only during expression in bacteria or yeast but also when the protein is expressed at physiologic levels under native conditions [87].

Another report indicates that human frataxin can undergo an *in vitro* oligomerisation process only under strongly destabilising conditions, i.e. incubating the protein at 57°C in the presence of $Fe(NH_4)_2(SO_4)_2$ 1:20 protein:iron ratio, differently to what observed for the yeast and bacterial proteins. Moreover, detailed analysis on the effect of the experimental conditions on CyaY aggregation, shows that the phenomenon is not evident at physiological ionic strength and that it is mostly of electrostatic nature. Mutation of conserved negative residues on the surface of the protein strongly alters the aggregational properties of CyaY; in particular mutations of CyaY Glu18 and Glu33 into alanine and lysine (the corresponding residues in hfra) produce an almost complete loss of the aggregational properties, thus explaining the differences in the behaviour of the bacterial and human protein [67].

A non essential role for YFH1 oligomerisation *in vivo* is reported by Aloria et al. They showed that a mutated oligomerisation defective protein is able to complement cells lacking Yfh1 gene, rescuing growth defect, aconitase deficiency and high mitochondrial iron levels [20].

These findings suggest that the iron promoted oligomerisation process might not have a physiological relevance and that the iron storage role might be a secondary function of the protein.

# Study of the Iron Binding Site of CyaY          5

This chapter describes the ion binding experiments carried out on bacterial frataxin CyaY. Bacterial frataxin represents a suitable model system to examine by NMR because it is a well folded and thermodynamically stable protein. In addition, complete NMR assignment and structure calculation have been recently achieved in the laboratory where the described studies were conducted. Moreover a bacterial model system is advantageous because several ISC proteins from *E. Coli* are well characterised (see chapter 6). In order to better characterise the iron binding properties of the protein, NMR titrations were performed with different ions and chemical shift perturbation was used to map the most affected residues onto the structure. Other than iron (II) and iron (III), we used paramagnetic ions such as lanthanides, manganese and cobalt, both to analyse the binding specificity and selectivity and possibly to gain insights into the geometry of the binding site. In the methods part a brief summing of the characteristic of paramagnetic NMR is reported.

## 5.1 *Paramagnetic NMR*

The unpaired electrons of a paramagnetic moiety originate an additional magnetic moment that severely affects the magnetic properties of a molecule. Typical paramagnetic agents are transition metals and lanthanides, but nitroxide radicals as well have been widely used as paramagnetic spin labels [88].

It is generally advised to carefully avoid and remove species with unpaired electrons when recording NMR spectra. In fact they provide additional relaxation mechanisms that cause broadening of the resonances, resulting in a more difficult detection of the signals. As a consequence, NOE signals are reduced and the multidimensional spectra difficult to record. In addition many paramagnetic species induce variations in the chemical shift of the residues, sometimes making the assignment procedure extremely complicated. Nevertheless, there are cases when paramagnetic effects can be exploited both in a qualitative and a quantitative manner to gain structural information on the region surrounding the paramagnets. Not only, but spin labels and paramagnetic ions can also be informative of regions as far as 40Å away from the probe, thus providing important structural restraints for solving structures of entire macromolecules [89].

A paramagnetic centre affects the magnetic properties of the surrounding nuclei in both a time-averaged and a time-dependent manner. The time-averaged magnetic field severely affects the resonance frequencies of nuclei close to the paramagnetic centre causing the shift of many signals outside the classical diamagnetic window. Time-dependent fluctuations of the magnetic field resulting from the unpaired electron at the metal centre, provide relaxation mechanisms that give rise to shorter $T_1$ and $T_2$ relaxation times and, in the case of $T_2$, line broadening [90]. Moreover paramagnetic ions induce a self orienting capability on the macromolecule that hosts them, giving rise to residual dipolar couplings (RDCs), which can be exploited as structural restraints. A more complete treatment of the RDCs can be found elsewhere [89].

### 5.1.1 Paramagnetic shift

In the NMR spectrum of a paramagnetic molecule, the observed chemical shift is described by:

$$\delta_{obs} = \delta_{param} + \delta_{diam} \tag{5.1}$$

where $\delta_{obs}$, $\delta_{param}$ and $\delta_{diam}$ are the observed, the paramagnetic and the diamagnetic shift.

The paramagnetic shift is caused by the hyperfine interaction of the unpaired electrons with the nuclei. It comprises two contributions, the Fermi contact term and a dipolar, or pseudocontact, term.

The contact shift is determined by the additional magnetic field generated at the nucleus by the electric magnetic moment located at the nucleus itself. It is proportional to the Fermi contact term $A$, which results from delocalisation of the electron spin density from the metal orbitals onto the s-orbital of the ligand nuclei.

$$A = \frac{\mu_0}{3S} \hbar \gamma_I g_e \mu_B \sum_i \rho_i \tag{5.2}$$

where $g_e$ is the electron $g$ value 2.0023, $\mu_B$ is the Bohr magneton $\gamma_I$ is the proton giromagnetic ratio, $S$ indicates the electron spin, $\rho$ is the electron spin densityand $i$ are the s orbitals.

In the high field approximation, when $g_e \mu_B B_0 \gg A$, the contact shift is given by:

$$\delta_{con} = -\frac{A}{\hbar \gamma_I B_0} \langle S_z \rangle = \frac{A}{\hbar} \frac{g_e \mu_B S(S+1)}{3\gamma_I kT} \tag{5.3}$$

where $\langle S_z \rangle$ is the expectation value of $S_z$.

The contact shift is important when the paramagnetic ion is covalently bound to the ligand.

The second contribution to the paramagnetic shift is the dipolar coupling, or pseudocontact shift. It derives from the local magnetic field induced at the nucleus under study by the anisotropic magnetic moment of the paramagnetic ion, and affects both ligated and non-ligated atoms [90,91]. Applying the so called 'metal-centred dipole' approximation for a metal ion with spin S=1/2, the dipolar contribution can be expressed as:

$$\delta_{PC}^j = \frac{1}{12\pi} \left[ \Delta\chi_{ax} \left( 3\cos^2\theta - 1 \right) r^{-3} + 3/2 \, \Delta\chi_{rh} \left( \sin^2\theta \cos 2\phi \right) r^{-3} \right] \tag{5.4}$$

where $\Delta\chi_{ax} = \chi_{zz} - \frac{1}{2}\left( \chi_{xx} + \chi_{yy} \right)$ and $\Delta\chi_{rh} = \chi_{xx} - \chi_{yy}$ are respectively the axial and rhombic magnetic susceptibilities, $r$, $\theta$ and $\phi$ are the spherical polar coordinates of the nucleus $j$ relative to the principal axes of the $\underline{\underline{\chi}}$ tensor, chosen as reference axes [91]. In the case of axial symmetry, the second term vanishes because $\chi_{xx} = \chi_{yy} = \chi_\perp$ and the pseudocontact shift becomes:

$$\delta_{PC} = \frac{1}{12\pi r^3} \left( \chi_\parallel - \chi_\perp \right)\left( 3\cos^2\theta - 1 \right) \tag{5.5}$$

Pseudocontact shifts are extremely useful for structure calculation and refinement because they are intimately related to the geometrical characteristics of their environment [89,92].

### 5.1.2 Relaxation

Analogously to what happens for the chemical shift, the relaxation times of a paramagnetic molecule are influenced by the unpaired electrons, as well. Both longitudinal and transverse relaxation rates ($T_1^{-1}$ and $T_2^{-1}$) are enhanced and can be decomposed into the sum of a diamagnetic and a paramagnetic contribution. The latter may be attributed to the presence of the fluctuating magnetic field generated by the unpaired electrons of the paramagnetic centre on the nuclear spin. In order to better understand the effect of paramagnetic ions on the relaxation of a molecule, it is convenient to refer to correlation times. Since the electron-nuclear interaction is randomly modulated by electron spin relaxation, molecular tumbling and chemical exchange, the overall correlation time is the sum of the electronic and tumbling correlation times and of an exchange lifetime ($\tau_s$, $\tau_r$ and $\tau_M$).

$$\left(\tau_c\right)^{-1} = \tau_s^{-1} + \tau_r^{-1} + \tau_M^{-1} \tag{5.6}$$

Electronic correlation times fall in the range $10^{-7}$ to $10^{-13}$ s, whereas the exchange time may vary very largely, depending on the nature of the system under investigation. Rotational correlation times can be around $10^{-8}$ s for small proteins up to $10^{-6}$ s for bigger macromolecules. High resolution NMR of systems containing a paramagnetic metal ion can be recorded when the correlation time is governed by a short electronic relaxation of the ion (and in the absence of exchange broadening). As illustrated in Table 5-1, lanthanides (III) with the exception of gadolinium (III), and low spin iron (III) are particularly suitable for NMR, having relaxation times of $10^{-12}$-$10^{-13}$ s [92].

The relaxation rates are composed of contributions mainly arising from dipolar and Curie mechanisms.

Dipolar relaxation arises from through-space interactions due to the fluctuation of the electronic magnetic field. A complete treatment of the paramagnetic dipolar relaxation theory can be found in [92]. Both longitudinal and transverse dipolar relaxation rates are inversely proportional to $r^{-6}$ and can thus be used to obtain structural restraints.

$$R_{1M} = \frac{2}{15}\left(\frac{\mu_0}{4\pi}\right)^2 \frac{\gamma_I^2 g_e^2 \mu_B^2 S(S+1)}{r^6} \times \left[\frac{\tau_s}{1+\left(\omega_I - \omega_S\right)^2 \tau_s^2} + \frac{3\tau_s}{1+\omega_I^2 \tau_s^2} + \frac{6\tau_s}{1+\left(\omega_I + \omega_S\right)^2 \tau_s^2}\right] \tag{5.7}$$

$$R_{2M} = \frac{1}{15}\left(\frac{\mu_0}{4\pi}\right)^2 \frac{\gamma_I^2 g_e^2 \mu_B^2 S(S+1)}{r^6} \times \left[ \frac{4\tau_c + \dfrac{\tau_s}{1+(\omega_I-\omega_S)^2\tau_s^2} + \dfrac{3\tau_s}{1+\omega_I^2\tau_s^2} +}{\dfrac{6\tau_s}{1+(\omega_I+\omega_S)^2\tau_s^2} + \dfrac{6\tau_s}{1+\omega_S^2\tau_s^2}} \right] \qquad (5.8)$$

where $\omega_I$ is the nuclear frequency and $\omega_S$ corresponds to the transition frequency between the various electron spin levels [92].

**Table 5-1: Electronic relaxation times of some common paramagnetic metal ion and relative line broadening effects (from [92] and references therein).**

| Metal ion | Electronic configuration | S | $\tau_s$ (s) | Line broadening (Hz)[a] |
|---|---|---|---|---|
| $Ti^{3+}$ | $d^1$ | 1/2 | $10^{-10}$-$10^{-11}$ | 20-200 |
| $VO^{2+}$ | $d^1$ | 1/2 | $\sim 10^{-8}$ | 10000 |
| $V^{3+}$ | $d^2$ | 1 | $\sim 10^{-11}$ | 50 |
| $V^{2+}$ | $d^3$ | 3/2 | $\sim 10^{-9}$ | 5000 |
| $Cr^{3+}$ | $d^3$ | 3/2 | $5 \times 10^{-9}$-$5 \times 10^{-10}$ | 3000-25000 |
| $Cr^{2+}$ | $d^4$ | 2 | $10^{-11}$-$10^{-12}$ | 20-150 |
| $Mn^{3+}$ | $d^4$ | 2 | $10^{-10}$-$10^{-11}$ | 150-1500 |
| $Mn^{2+}$ | $d^5$ | 5/2 | $\sim 10^{-8}$ | 100000 |
| $Fe^{3+}$ (H.S.) | $d^5$ | 5/2 | $10^{-9}$-$10^{-11}$ | 200-1200 |
| $Fe^{3+}$ (L.S.) | $d^5$ | 1/2 | $10^{-11}$-$10^{-13}$ | 0.5-20 |
| $Fe^{2+}$ (H.S.) | $d^6$, 5-6 coord | 2 | $10^{-12}$-$10^{-13}$ | 5-20 |
|  | $d^6$, 4 coord | 2 | $\sim 10^{-11}$ | 150 |
| $Co^{2+}$ (H.S.) | $d^7$ 5-6 coord | 3/2 | $5 \times 10^{-12}$-$10^{-13}$ | 2-50 |
|  | $d^7$, 4 coord | 3/2 | $\sim 10^{-11}$ | 100 |
| $Co^{2+}$ (L.S.) | $d^7$ | 1/2 | $10^{-9}$-$10^{-10}$ | 200-1000 |
| $Ni^{2+}$ | $d^8$, 5-6 coord | 1 | $\sim 10^{-10}$ | 500 |
|  | $d^8$, 4 coord | 1 | $\sim 10^{-12}$ | 5 |
| $Cu^{2+}$ | $d^9$ | 1/2 | $(1\text{-}5) \times 10^{-9}$ | 1000-5000 |
| $Ru^{3+}$ | $d^5$ | 5/2 | $10^{-11}$-$10^{-12}$ | 2-20 |
| $Re^{3+}$ | $d^4$ | 2 | $10^{-12}$-$10^{-13}$ | 5-20 |
| $Gd^{3+}$ | $f^7$ | 7/2 | $10^{-8}$-$10^{-9}$ | 20000-200000 |
| $Ln^{3+}$ | $f^n$ |  | $10^{-12}$-$10^{-13}$ | 1-100 |

[a] For a proton at 5 Å from the metal; 500 MHz 1H resonance frequency; only dipolar relaxation estimated from (5.7) and (5.8), with $\tau_c = \tau_s$.

The Curie relaxation mechanism is a dipolar effect arising form the interaction between the nuclear spin with the thermal average of the electronic spins [90] Due to the splitting of the S levels in fact, a difference in the population of each electronic spin level will be generated, according to the Boltzmann distribution. Therefore an averaged magnetic spin moment related to the $\langle S_z \rangle$ expectation value will result. Curie terms are modulated exclusively by the rotational correlation time, and are field dependent.

$$R_{1M} = \frac{2}{5}\left(\frac{\mu_0}{4\pi}\right)^2 \frac{\gamma_I^2 g_e^2 \mu_B^2}{r^6} \langle S_z \rangle^2 \left(\frac{3\tau_r}{1+\omega_I^2\tau_r^2}\right) \qquad (5.9)$$

$$R_{2M} = \frac{1}{5}\left(\frac{\mu_0}{4\pi}\right)^2 \frac{\gamma_I^2 g_e^2 \mu_B^2}{r^6} \langle S_z \rangle^2 \left(4\tau_r + \frac{3\tau_r}{1+\omega_I^2\tau_r^2}\right) \qquad (5.10)$$

where

$$\langle S_z \rangle = -\frac{g_e \mu_B S(S+1) B_0}{3kT} \qquad (5.11)$$

The Curie spin relaxation provides the largest contribution when $\tau_e \ll \tau_c$, which is the case in proteins, and it is particularly important at high magnetic fields [89].

## 5.2 *Results*

### 5.2.1 Titration with iron (II)

We titrated a solution of CyaY with Fe(II) and for each point a [$^1$H,$^{15}$N] HSQC spectrum was recorded. We observed many shifts, but the most striking feature is the total disappearance of specific resonances without the concomitant appearance of other signals. This result could derive from the influence of the paramagnetism of iron (II) on the coordinated residues. Otherwise it could arise from the presence of an intermediate rate equilibrium between the free and the bound form; in this case in fact the line broadening caused by the dynamic process would be extremely large, preventing the detection of the peaks. However this hypothesis has to be investigated further.

Already at a 1:1 ratio of iron(II): protein, the resonances of Arg20, Asp22 and Asp23 disappear and the resonance of Asp21 is shifted; at 2:1 ratio also residues 19 and 44

disappear while resonances 24, 28, 29, 31, 32, 33 shift. At higher concentrations of iron(II) many other resonances broaden.



**Figure 5-1: superimposition of a reference [$^1$H,$^{15}$N] HSQC spectrum of CyaY (blue) with a [$^1$H,$^{15}$N] HSQC spectrum of a 1:2 mixture of CyaY:Fe$^{2+}$ (red).**
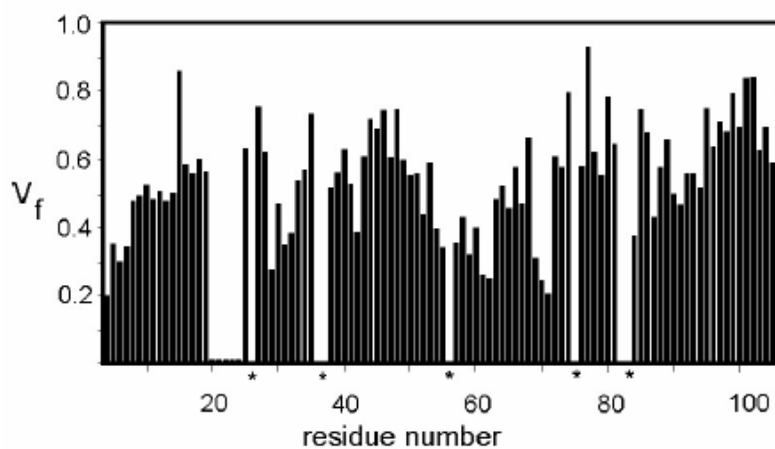


**Figure 5-2: Histogram of the fractional volume from a spectrum recorded in presence of Fe$^{2+}$ as compared to a control solution: Fe$^{2+}$ was added at 1:2 ratio. Asterisks indicate the position of resonances not present (proline), or not identified in the spectrum, or for which overlap was too severe and volume determination is unreliable.**

When we reach stoichiometric ratios of nearly 6:1 many superficial resonances are affected; this effect is probably due to the bulk paramagnetism, i.e. the presence of free paramagnetic Fe(H$_2$O)$^{2+}$ that affects indiscriminately all the exposed peaks.

## 5.2.2   Titration with iron(III)

We carried out the same titration experiment using ferric iron. In this case too, the most affected peaks at protein:iron ratio 1:1 are Arg20, Leu 21, Asp22 and Asp23. At 1:2 ratio the above resonances completely disappear together with the resonances of the amides 29, 30 and 31. At a 1:6 protein/iron ratio resonances 19-35 are completely bleached while resonances 42, 44, 104, 105 are slightly affected.



**Figure 5-3: superimposition of a reference [$^1$H,$^{15}$N] HSQC spectrum of CyaY (blue) with a [$^1$H,$^{15}$N] HSQC spectrum of a 1:6 mixture of CyaY:Fe$^{3+}$ (red).**
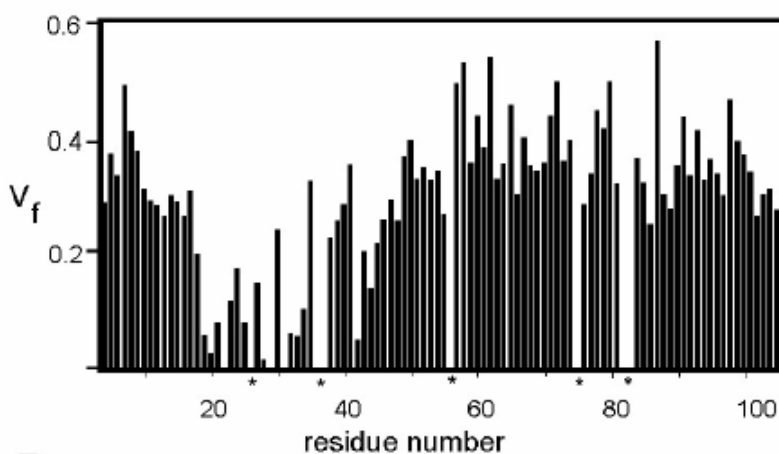


**Figure 5-4: Histogram of the fractional volume from a spectrum recorded in presence of Fe$^{3+}$ as compared to a control solution: Fe$^{3+}$ was added at 1:6 ratio. Asterisks indicate the position of resonances not present (proline), or not identified in the spectrum, or for which overlap was too severe and volume determination is unreliable.**

### 5.2.3  Titration with other ions

Titrations with cations different from Fe(II) and Fe(III) enabled us to investigate the selectivity and specificity of the binding.

### 5.2.4  Titration with $Ca^{2+}$

The choice of $Ca^{2+}$ is due to the fact that previous experiments had hinted at the possibility of a competition between it and iron for the same binding site [67,86]. Moreover, being diamagnetic, it could help in the unambiguous identification of the binding site, without the interferences of the paramagnetism.

We recorded [$^{1}$H,$^{15}$N] HSQC spectra at increasing calcium-protein ratios. With this ion as well we noticed that the first resonances to be affected by a small shift, are the amides of residues 14, 17, 22, 23, 25, 27-32, 34, 44 at a 1:1 ratio. No other resonances are affected at higher ratios; the chemical shift variation reaches a plateau at protein:$Ca^{2+}$ ratio of 1:6.
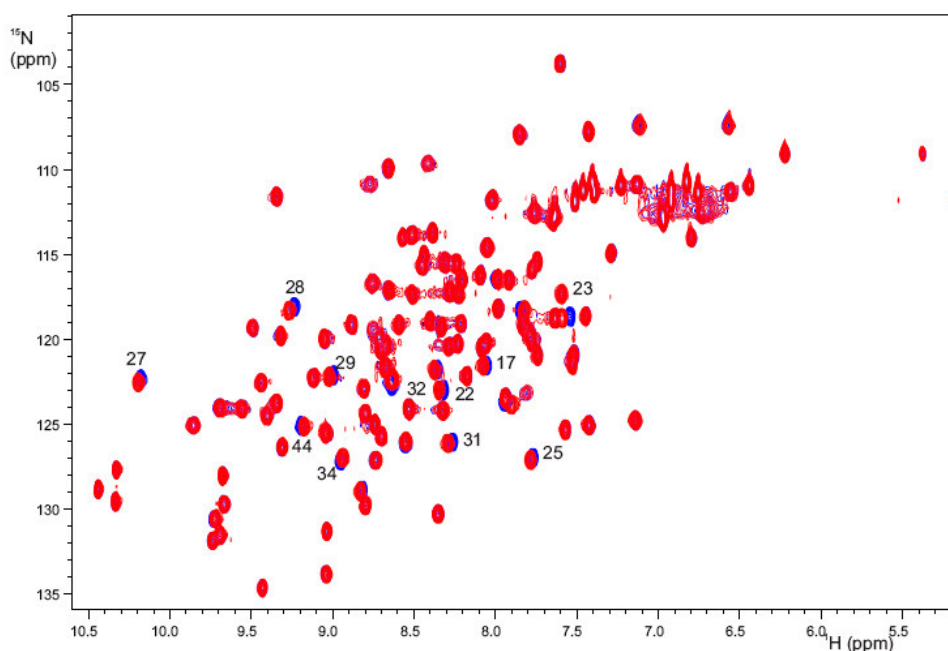


**Figure 5-5: superimposition of a reference [$^{1}$H,$^{15}$N] HSQC spectrum of CyaY (blue) with a [$^{1}$H,$^{15}$N] HSQC spectrum of a 1:1 mixture of CyaY:$Ca^{2+}$ (red).**

The presence of shifts is indicative of a small change of the chemical environment around the binding site upon $Ca^{2+}$ binding. It is probable that the bound and the free species exchange rapidly between each other; the detected NMR signal in this case

would in fact fall at a frequency which is the average between those of the signals produced by each species.

### 5.2.5   Titration with Lanthanides

Trivalent lanthanide ions generally induce a line broadening of 1-100 Hz for most nuclei distant up to 5 Å from $Ln^{3+}$, except for $Gd^{3+}$ which has a much more relevant effect, broadening the signals of 20,000-200,000 Hz [92] and $La^{3+}$ and $Lu^{3+}$, which are diamagnetic. We chose to titrate CyaY with lanthanides to enhance the paramagnetic effects around the binding site.

#### 5.2.5.1   Titration with $Eu^{3+}$

The titration with $Eu^{3+}$, an $f^6$ ion, was first followed with a TOCSY and a NOESY-HSQC spectra, to observe the effects on the side chains. The results were not satisfactory; since these spectra are very crowded, it is very difficult to follow the destiny of the peaks, in particular to say if they disappear completely or shift and overlap with other signals. However, at a 0.5:1 $Eu^{3+}$:protein ratio of, we can say with a high degree of confidence that the diagonal peaks of residues  20, 21, 22, 25, 27-32, 44, 47, 59-61, 75, 78, 90, 103, 105 and 106 disappear in both TOCSY and NOESY experiments. This is partly confirmed by the HSQC spectra: the main effect of the addition of one equivalent of $Eu^{3+}$ to CyaY is a large broadening of resonances 20-31. Resonances 14, 17-19, 32, 33, 34, 42-44, 106 shift.

At higher concentrations of lanthanide ion, an unspecific broadening starts to be particularly important.

#### 5.2.5.2   Titration with $Yb^{3+}$

$Yb^{3+}$ is a $f^{13}$ ion, which usually determines a large pseudocontact shift variation on the atoms of the binding site. However addition of $Yb^{3+}$ to CyaY does not cause any variation in the chemical shift. As previously seen with $Eu^{3+}$, a broadening of the signals up to disappearance is evident.
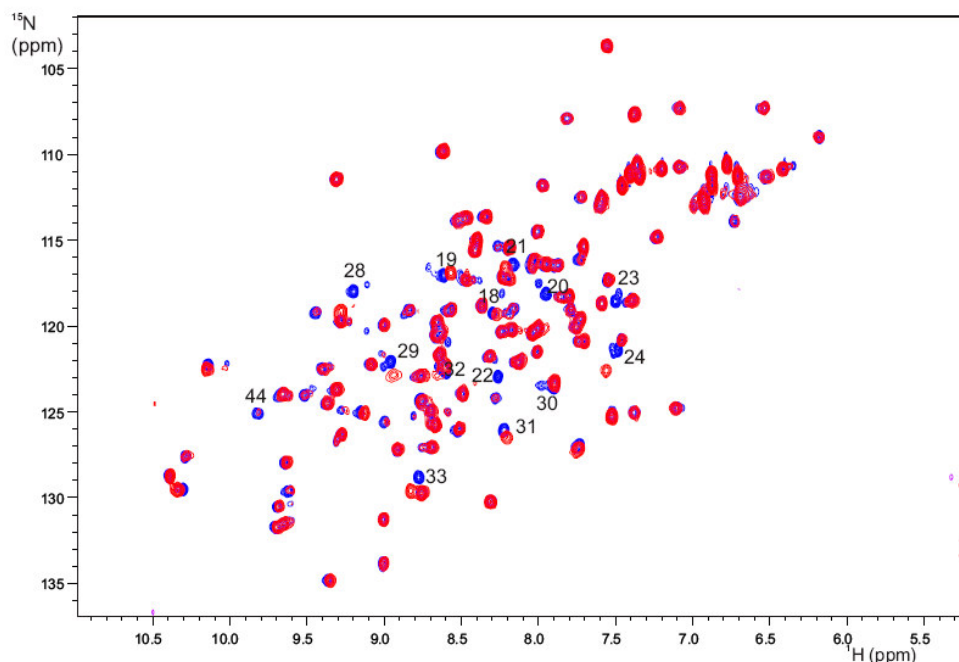
**Figure 5-6: superimposition of a reference [¹H,¹⁵N] HSQC spectrum of CyaY (blue) with a [¹H,¹⁵N] HSQC spectrum of a 1:1 mixture of CyaY:Eu³⁺ (red).**

Before recording the 2D HSQC we recorded a TOCSY and a NOESY-HSQC. With this ion as well many changes induces in the spectra cannot be easily followed. In both TOCSY and NOESY-HSQC, the diagonal peaks that disappear at ion:protein ratio 1:1 are 18-25, 27-34, 40, 44, 65, 66, 101, 103, 106. At higher Yb³⁺ concentrations all the peaks suffer from a severe line broadening that does not allow us to reach any further conclusions.

In the HSQC spectrum we cannot appreciate many changes at a 1:1 ratio. However at a Yb³⁺: CyaY ratio 2:1 we can easily say that the mainly affected resonances are 18-25, 27-29, 31, 32, 34, 44, as shown in Figure 5-7.

### 5.2.5.3    Titration with Gd³⁺

Gd³⁺ has a f ⁷ configuration, meaning that its *f* orbitals are only half occupied. Due to the spherical distribution of the electrons around its nucleus, Gd³⁺ does not cause any paramagnetic shift; nevertheless it has a very powerful effect on the relaxation rate of the close atoms, as proved by the fact that it is used in functional MRI as a relaxation agent [93] The titration with this ion was therefore carried out at very low ion:protein ratio, relying on the fact that the effects would be evident even at low concentrations of Gd³⁺. The first point was recorded at a ratio 0.05:1 ion:protein and a very small broadening effect was observed on resonances 22, 23 and 33. Already

at a ratio of 0.1:1 however peaks 19, 20-25, 27-29, 31-34, 42-46 sensibly weaker. This effect is much stronger at a 1:1 ratio, as shown in the following picture.
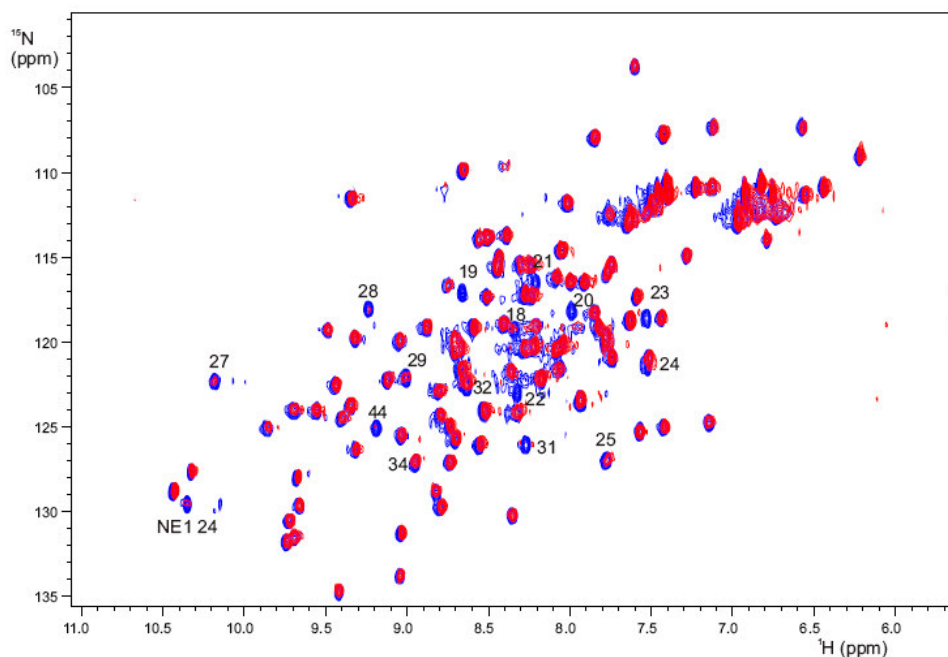


**Figure 5-7: superimposition of a reference [$^1$H,$^{15}$N] HSQC spectrum of CyaY (blue) with a [$^1$H,$^{15}$N] HSQC spectrum of a 1:2 mixture of CyaY:Yb$^{3+}$ (red).**
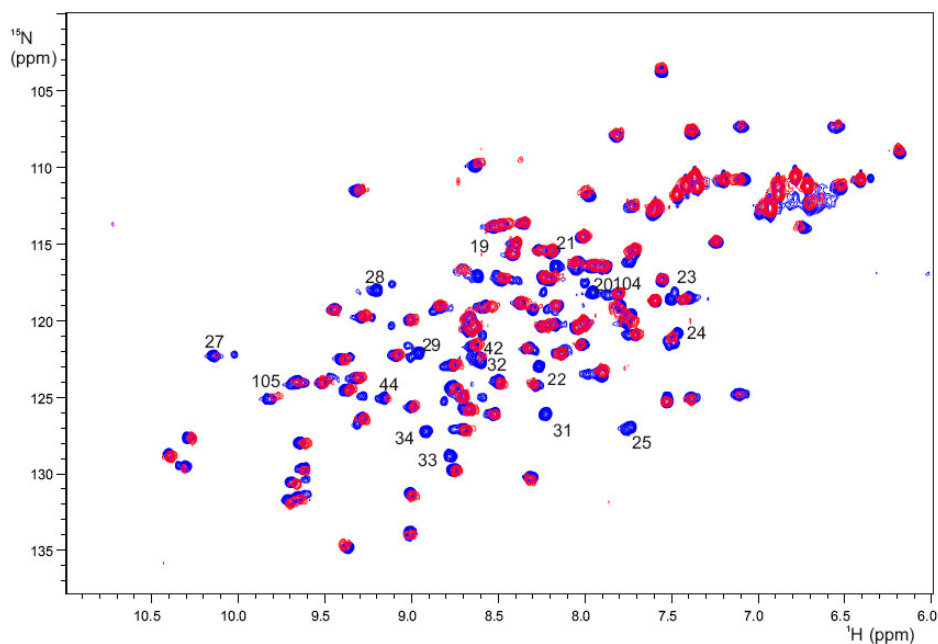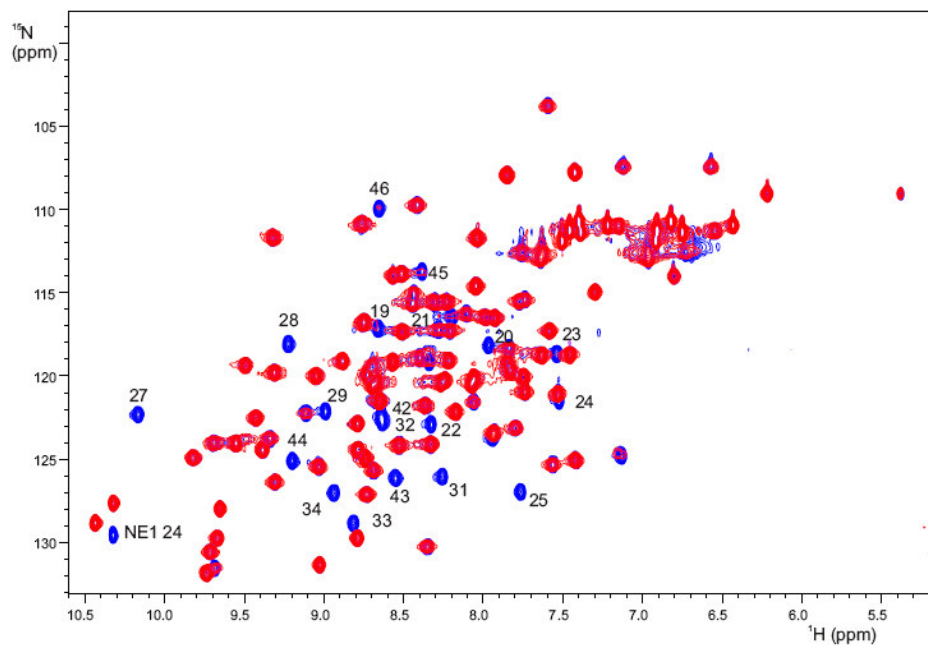


**Figure 5-8: superimposition of a reference [$^1$H,$^{15}$N] HSQC spectrum of CyaY (blue) with a [$^1$H,$^{15}$N] HSQC spectrum of a ca 1:1 mixture of CyaY:Gd$^{3+}$(red).**

80

## 5.2.5.4    Titration with $Lu^{3+}$

$Lu^{3+}$ is a diamagnetic lanthanide, having the f orbital shell full (it has a $f^{14}$ electronic configuration). We decided to perform a titration with $Lu^{3+}$ because it has an ionic radius very close to the one of $Yb^{3+}$ and because we wanted to investigate the effect of a diamagnetic ion different from $Ca^{2+}$ on the protein.

Being diamagnetic, $Lu^{3+}$ should not cause any bleaching of the peaks. However already at 0.5:1 Lu:CyaY ratio, peaks 23 and 27 disappear. Peaks 18, 19, 22, 24, 25, 28, 29, 44, 105 are gradually shifted and bleached until disappearance as the titration progresses (until a 3:1 ratio). Peaks 14, 17, 19, 20, 31, 34, 43 are only shifted. Residues 32, 42, 13, 24 are also affected, but they overlap with each other (42 with 32, 13 with 24) and we cannot precisely say which peak is affected and in which way.
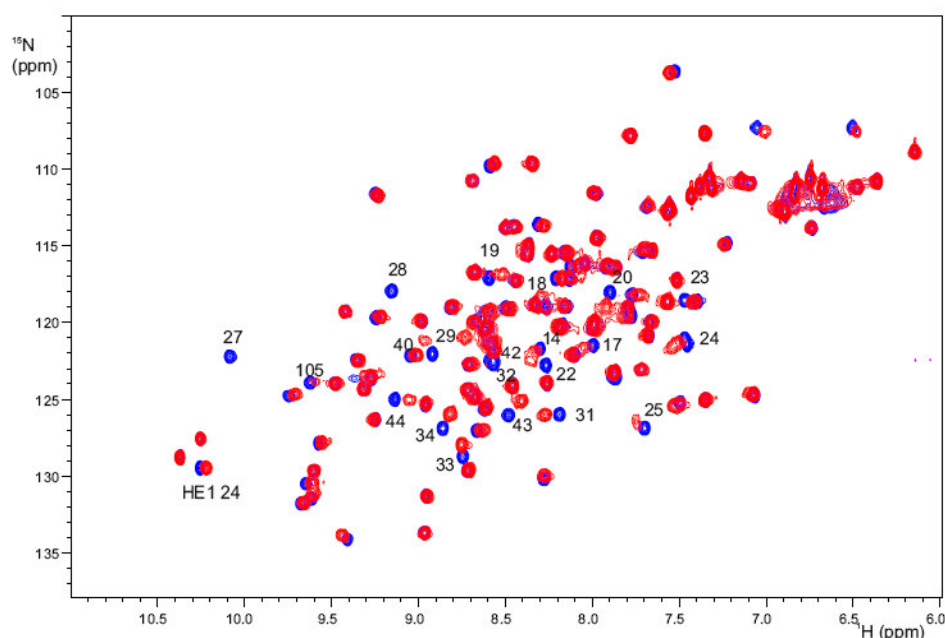


**Figure 5-9: superimposition of a reference [$^1$H,$^{15}$N] HSQC spectrum of CyaY (blue) with a [$^1$H,$^{15}$N] HSQC spectrum of a ca 1:2 mixture of CyaY:Lu$^{3+}$ (red).**

The fact that some resonances disappear without the obvious appearance of new signals in different regions of the spectrum, suggests that they are involved in an intermediate rate exchanging equilibrium between the metal-bound and the free protein. In contrast, the chemical shift variation of many other peaks suggests that they are in a regime of fast exchange.

Further analysis is needed to assess the nature of the equilibrium and the possible presence of a conformational rearrangement.

### 5.2.6  Titration with Mn$^{2+}$

Mn$^{2+}$ is a d$^5$ ion which induces a line broadening of 100,000 Hz. For this reason it is not the best ion to choose for paramagnetic NMR [92]. However, the addition of very small amounts of this metal ion can help the identification of the binding site.

Already at a Mn$^{2+}$ :CyaY  ratio of 0.1:1 the effect is quite evident: peaks 18-34, 42-44, 46, 86, 105 and 106 disappear. At higher ratios, the substantial broadening of the peaks prevents further analyses.
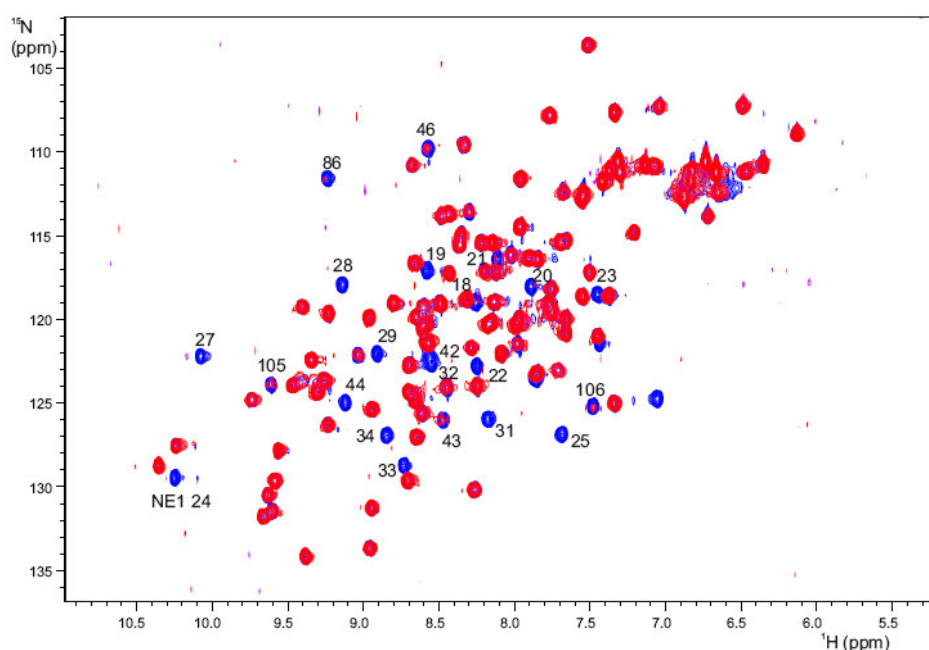


**Figure 5-10: superimposition of a reference [$^1$H,$^{15}$N] HSQC spectrum of CyaY (blue) with a [$^1$H,$^{15}$N] HSQC spectrum of a ca 1:0.1 mixture of CyaY:Mn$^{2+}$ (red).**

### 5.2.7  Titration with Co$^{2+}$

Co$^{2+}$ is a d$^7$ ion, which can be high spin or low spin, according to its coordination sphere. Co$^{2+}$ H.S. and Co$^{2+}$ L.S. induce a very narrow line broadening (2-50 Hz) and a very large one (200-1000 Hz) respectively [92].

Since we did not know the coordination of the metal, we decided to titrate CyaY with small quantities of the salt solution. We were only able to add a total of 1 equivalent, before the broadening of the peaks was too severe.

In Figure 5-11, the effect of the addition of 0.5 equivalent of Co$^{2+}$ to the protein is shown: the main affected signals are 19-35, 43-45, 95, 96, 103-106.
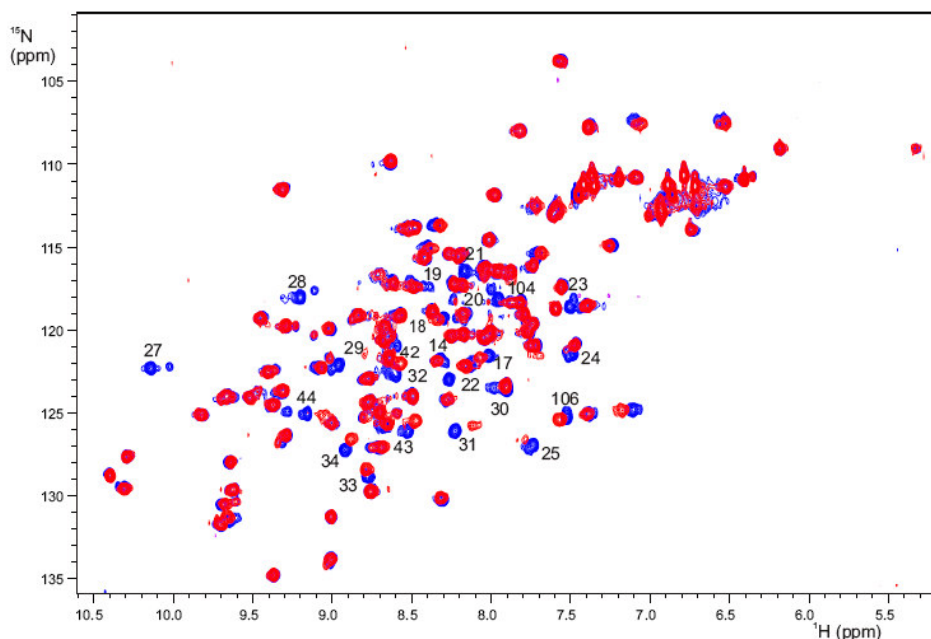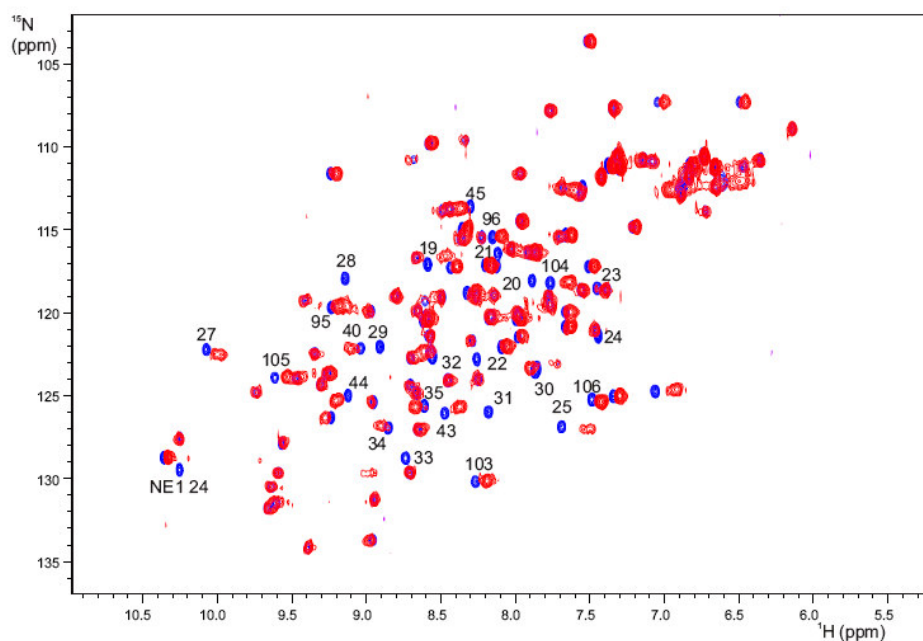
**Figure 5-11: superimposition of a reference [$^1$H,$^{15}$N] HSQC spectrum of CyaY (blue) with a [$^1$H,$^{15}$N] HSQC spectrum of a ca 1:0.5 mixture of CyaY:Co$^{2+}$ (red).**

### 5.2.8 Titration of a CyaY mutant

In order to drastically change the electrostatic properties of the protein surface we designed a mutant called CyaY_181922, where Glu 18, Glu 19 and Asp 22 were mutated to Lys, and we titrated it with iron(II) and iron(III).

Before starting the titration experiments, we recorded a NOESY and TOCSY-HSQC to assign the mutated resonances; the assignment was achieved by comparison with the correspondent wild type spectra. The mutant retains the 3D fold of the wild type, despite the drastic polarity change induced by the mutation, as previously demonstrated by far UV-CD spectroscopy [67]. This is reasonable considering that the mutated residues do not contribute to build the protein's backbone, being located on its surface.

Interestingly, the presence of up to 6 equivalents of iron(II) with respect to the protein, does not affect any peak, meaning that iron(II) binding is completely abolished.

Conversely, titration with iron(III) causes the disappearance of residues 29 and 30 and a severe weakening of amides 28 and 31 already at a 1:0.5 ratio. At 1:2 protein/iron ratio the resonances of 25, 27, 28 also broaden, while at 1:6 a

generalized broadening is observed, with the most affected residues being 24-32 and 44. Residues 20, 21, 22, 23 remain unaffected.
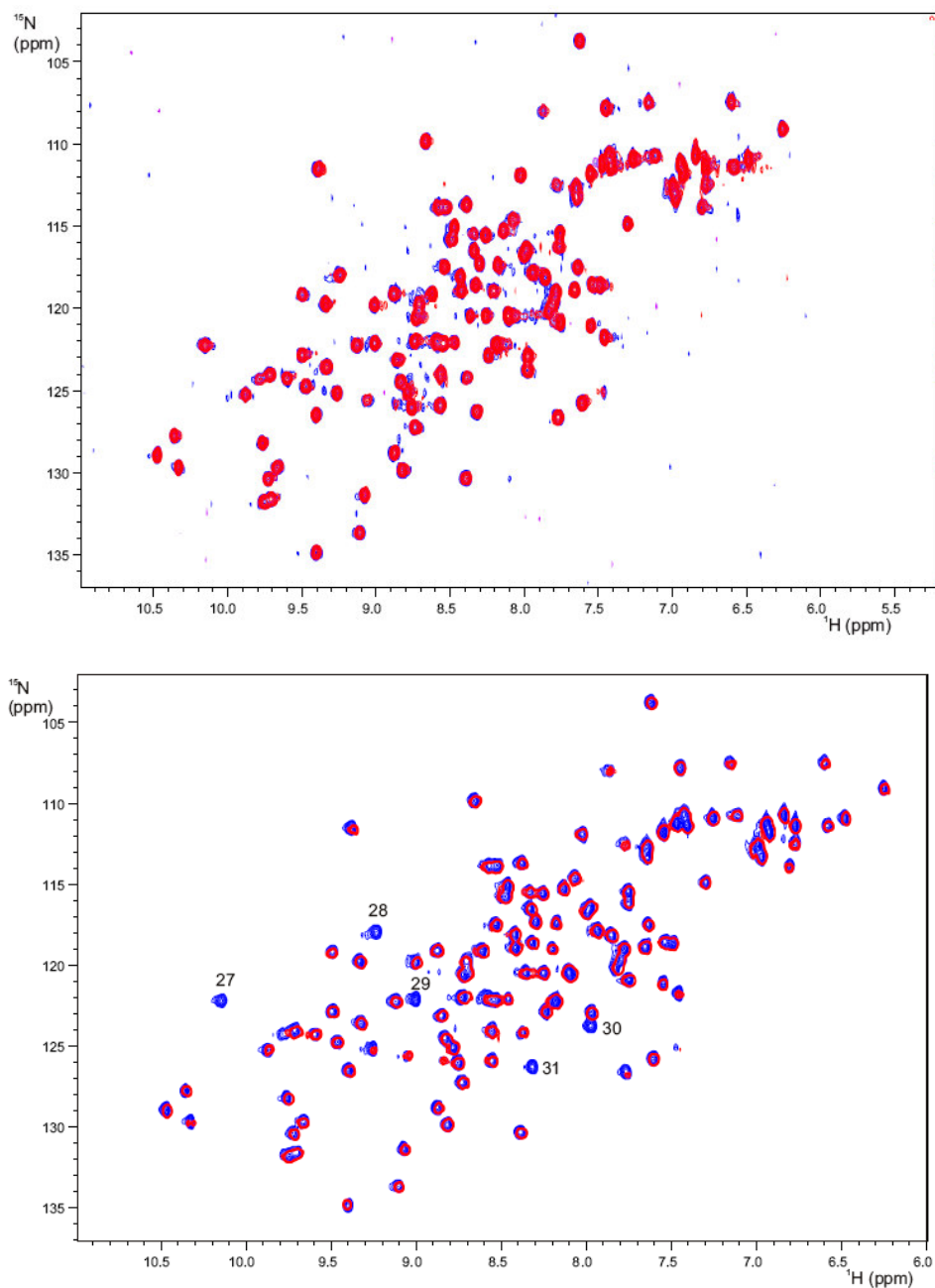


**Figure 5-12: Top: comparison of a reference [$^1$H,$^{15}$N] HSQC spectrum (blue) of CyaY181922 with a mixture of the mutant protein and Fe(II), 1:2 (red). None of the peaks is affected. Bottom: comparison of a reference HSQC [$^1$H,$^{15}$N] spectrum (blue) of CyaY181922 with a mixture of the mutant protein and Fe(III), 1:6 (red). The binding affects the secondary binding site constituted by residues 27-31.**

These experiments prove conclusively that Glu18, Glu19 and Asp22 are involved in the iron binding site; complete loss of Fe(II) binding in the mutant shows that this is the region responsible for the 1:2 protein/iron stoichiometry. On the other hand, retention of Fe(III) binding with immediate disappearance of residues 29-31 is in accord with a higher stoichiometry of this cation and indicates the hyerarchical order in which the binding sites are populated in the wild type.

## 5.3 *Discussion*

CyaY is the bacterial orthologue of frataxin, the protein involved in the severe neurodegenerative genetic disease Freidreich's ataxia (FRDA) [94]. Although the role of the protein has not been demonstrated yet, several evidences hint at its involvement in mitochondrial iron metabolism. In particular it was proposed that the protein functions as an iron chaperone in the Fe-S clusters' synthesis (see paragraph 4.5).

Despite the fact that CyaY depleted *E. Coli* cells do not show the typical FRDA phenotype [95], the high structural conservation of the protein in human and bacterial species encouraged us to carry out NMR experiments on the bacterial orthologue [67], which has a high stability and can be easily expressed [23].
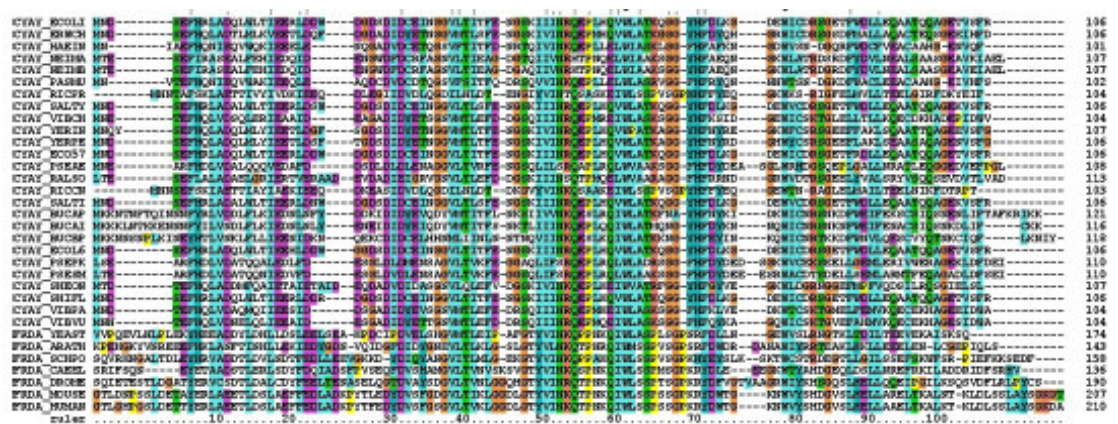


**Figure 5-13: Multiple alignment of frataxin orthologues, colour coded according to the ClustalX convention to emphasize conserved sequence features. The sequences are identified by their entry names in the SWISSPROT database [71].**

Previous evidence that CyaY binds iron has already been provided. In particular, anaerobic isothermal titration calorimetry revealed that at least two Fe(II) weakly

associate with monomeric CyaY ($K_d$=3.8μM) and that the protein can bind up to four more ferrous ions, although with too low affinities to be measured [86].

Moreover, it has been found that CyaY can also bind ferric ions, with a stoichiometry of 1:6. An excess of $Fe^{3+}$ promotes the formation of large spherical aggregates, able to sequester up to 26 ions [86].

The 3D structure of the protein does not show any typical iron binding motifs, such as conserved cysteines or histidines, or cavities. To characterise the iron binding site and to determine its specificity and selectivity, we performed NMR titrations of CyaY with Fe(II), Fe(III) and other different paramagnetic and diamagnetic ions. We observed several variations, depending on the ion used.

Paramagnetic iron (II) and iron (III) affect the amide resonances of the [$^1$H,$^{15}$N] HSQC spectrum in a different way. While Fe(II) induces both a hyperfine shift (presumably of a dipolar origin) and relaxation enhancement of the signals, iron (III) does not cause chemical shift variations but has exclusively a relaxation effect, indicating that the ion is in its HS electronic configuration, with an isotropic distribution of the electrons in its outer shell [92]. On the whole however, the two ions influence the same resonances. In both cases resonances from Arg20, Asp22, Asp23 are bleached. Amides of Ser28 and Asp29 are also severely affected by the ions; they are shifted by iron (II) and completely bleached by iron (III). A similar behaviour can be observed for Leu21, Asp27, Glu19, Trp24, Asp31, Glu33. However iron (III) has a more pronounced effect on resonances 31-34 with respect to iron (II). It can be suggested that the site centred around 19 and 23 is the first being populated, both from $Fe^{2+}$ and $Fe^{3+}$, while the 24-34 region, much weaker for $Fe^{2+}$, would account for the observed 1:6 stoichiometry for $Fe^{3+}$ binding. Residues 19-34 are located on the protein surface at the edge between the N-terminal α1 helix and β1 strand, in a region rich in conserved negatively charged residues, whose presence on the exposed surface of the protein could possibly hint at a specific functional role (Figure 5-14). This is confirmed by the titration on Cyay181922, where the 19-23 binding site is vacant, thus allowing the region around 24-31 to be immediately populated.

$Ca^{2+}$ induces only small chemical shift variations on the spectrum, mainly of residues 22, 23 and 27-31. It was shown that $Ca^{2+}$ can inhibit iron binding [67]; it is reasonable that it binds to the same site, involving the conserved aspartates 22 and 23 and 27-28 [23]. $Ca^{2+}$ influences weakly also residues 44, 30, 32, 34, 14, 17. Except for 44 and 32, the other resonances are hydrophobic residues; it is conceivable that the observed effects are due to weak interactions of the ion with the backbone oxygen. Residue 44 instead is a glutamic acid, whose sidechains can give electrostatic

interaction with $Ca^{2+}$, while 32 is a cysteine, which can interact through its sulphur atom.
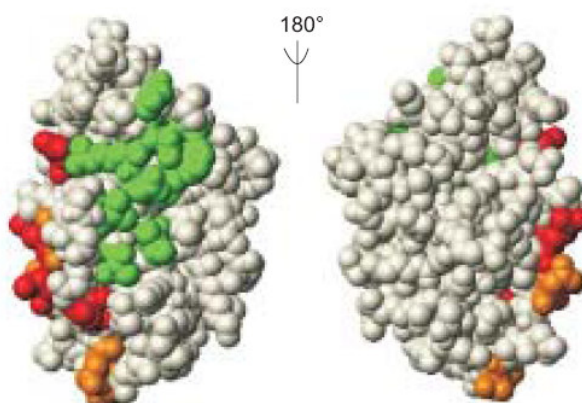


**Figure 5-14: Distribution of the conserved and semi-conseved residues onto the CyaY three-dimensional structure. They mainly cluster on one surface. Hydrophobic and polar conserved residues (Val38, Thr40, Ile50, Asn52, Arg53, Gln54, Pro56, Gln59, Trp61) are indicated in green. They form a conserved patch. The conserved and semi-conserved negatively charged residues are shown in red and orange, respectively. They all cluster on α1 and β1 [71].**

Lanthanides are often used as probes in the study of the metal binding sites in biomolecules. Their trivalent ions show paramagnetic properties, except for the first and last element of the series, $La^{3+}$ and $Lu^{3+}$. They possess 4f electrons, which are shielded by the 5s and 5p electrons and are not readily available for covalent interactions. The two main consequences of the shielding are that the interactions of the ions with the ligands are mainly electrostatic and that they show a large similarity in chemical behaviour. Nevertheless, the number of the unpaired electrons modulates the paramagnetic properties of the ions, that can thus be exploited to gain specific structural information. For example, $Gd^{3+}$ has 7 unpaired electrons distributed isotropically throughout its *4f* shell and therefore cannot produce an NMR dipolar shift. Conversely it has a fairly long electronic relaxation time and induces extensive line broadening of the resonances of the coordinated atoms. The other lanthanidic ions originate a variation of the chemical shift of the neighbouring atoms, as well as an enhanced relaxation, because of the anisotropic spatial distribution of their unpaired electrons.

We titrated CyaY with three different paramagnetic lanthanides, $Yb^{3+}$, $Eu^{3+}$ and $Gd^{3+}$, and with the diamagnetic lanthanide $Lu^{3+}$, mainly to gain qualitative information on the binding site, rather than determining its geometry: we must also keep in mind that

we do not have any information on the possible isomorphism of the protein bound to iron and to lanthanides.

According to their different paramagnetic behaviour, the three chosen lanthanides provoke different effects. As expected, one equivalent of $Gd^{3+}$ is sufficient to bleach many signals: amides 19-29, 31-34 and 42-44 disappear, while resonances from Asn45 and Gly46 are broadened but still visible. One equivalent of $Eu^{3+}$ instead, causes amides 20-25, 27-30 to disappear and induces a shift of resonances 14, 17-19, 31-34, 42-45 and 106. At 2:1 $Yb^{3+}$/CyaY ratio resonances 19-23 disappear completely, 27-29, 31 and 44 are extremely weak and 25, 32, 34 are slightly affected. The behaviour of the two ions is surprising: in fact, although in general the presence of $Yb^{3+}$ induces large pseudocontact shift, we did not notice any chemical shift variation when we titrated the protein with this ion. On the contrary, we observe it in the presence of $Eu^{3+}$. The effects that we see may be largely ascribed to the presence of an exchange process, with particular characteristics for each ion.

The effect of $Lu^{3+}$ is peculiar as well. Being diamagnetic the ion should not provoke any paramagnetic shift or relaxation. Conversely, two equivalents of the ion cause the shift and the disappearance of many signals, without the concomitant appearance of any other peak. In particular, resonances 14, 17, 19-22, 24, 25, 29, 31, 32, 34, 42-44 are shifted, residues 23, 27, 28 disappear and residues 18 and 105 are partially bleached. Here as well the variations are attributed to the presence of an exchange mechanism between the metal-bound and the free species. The residues that shift are probably involved in a fast equilibrium between two different conformations, while the peaks that disappear might participate to an intermediate exchange process. To verify this hypothesis it could be useful to measure the same spectrum at a different temperature or with a different external field strength, in order to modify the exchange regime.

Finally, we titrated the protein with two highly paramagnetic ions, $Mn^{2+}$ and $Co^{2+}$. Their effect is evident at very low ion/protein ratio. $Mn^{2+}$ is characterised by a long relaxation time and behaves in a similar way to $Gd^{3+}$, producing an efficient line broadening on atoms closer than 5 Å 0.1 equivalents of ion with respect to the protein provoke the complete bleaching of resonances 19-34, 42-46 and 105-106. Conversely, 0.5 equivalents of $Co^{2+}$ induce both relaxation enhancement and shift of the amide peaks. Resonances 19-22, 24, 25, 28, 29, 31, 33 disappear, 23 and 45 are affected by severe line broadening and resonances 27, 30, 32, 34, 35, 40, 42-44, 95, 96, 103, 104, 105 shift.

The experiments with these ions and titration of a CyaY mutant lacking the acidic residues 18, 19, 22, confirmed the presence of different binding sites. We identified a

primary binding site, constituted by residues 19-23, which are the first resonances to be affected by all the ions used. The complete bleaching of the signals observed in every experiment prevented us to draw any conclusions on the geometry of the binding site. However, it appears that both Curie relaxation and chemical exchange strongly influence the peaks' lineshapes. It could therefore be possible to reduce the broadening of the signals upon variation of the temperature or of the field strength; modulation of both effects could result in narrower lines, whose behaviour could be more easily followed.

Additional ion binding sites were identified: in particular region 27-34, similarly to what observed with $Fe^{2+}$ and $Fe^{3+}$, and region 42-44 are influenced by all the ions and the region 104-106 is influenced by $Co^{2+}$, $Mn^{2+}$, and, to a minor extent, by $Lu^{3+}$ and $Eu^{3+}$. Ser104 and Phe105 are also weakly affected by $Fe^{3+}$, although this may be a non-specific effect since they are quite distant from the main binding site. Table 5-2 and Figure 5-15 recapitulate the results of the titrations.

**Table 5-2: Effects of different ions on CyaY. The resonances mainly affected in the [$^1$H,$^{15}$N] HSQC spectrum of the protein are indicated in the right column.**

| | |
|---|---|
| $Fe^{2+}$ | E19, R20,L21, D22, D23, W24, S28, D29, I30, D31, C32, D33, E44 |
| $Fe^{3+}$ | E19, R20,L21, D22, D23, W24, D25, D27, S28, D29, D31, C32, D33, I34, T42, E44 |
| $Ca^{2+}$ | I17, R20, D22, D23, D25, D27, S28, D29, D31, C32, I34, E44 |
| $Eu^{3+}$ | I17, E18, E19, R20, L21, D22, D23, W24, D25, D27, S28, D29, I30, D31, C32, E33, I34, T42, F43, E44, R106 |
| $Yb^{3+}$ | E18, E19, R20, L21, D22, D23, W24, D25, D27, S28, D29, D31, I34, E44 |
| $Gd^{3+}$ | E19, R20, L21, D22, D23, W24, D25, D27, S28, D29, D31, E33, I34, T42, F43, E44, N45, G46 |
| $Lu^{3+}$ | I17, E18, E19, R20, D22, D23, W24, D25, D27, S28, D29, D31, C32, E33, I34, T40, T42, F43, D44, S104, F105 |
| $Co^{2+}$ | E19, R20, L21, D22, D23, W24, D25, D27, S28, D29, I30, D31, C32, E33, I34, N35, T42, F43, E44, N45, G46, A95, A96, V103, S104, F105, R106 |
| $Mn^{2+}$ | E18, E19, R20, L21, D22, D23, D25, D27, S28, D29, D31, C32, E33, I34, T42, F43, E44, G46, F86, F105, R106 |

```
                    1        10        20        30        40        50        60        70        80        90        100   106
Fe³⁺ (d⁵)   MNDSEFHRLADQLWLTIEERLDDWDGDSDIDCEINGGVLTITFENGSKIIINRQEPLHQVWLATKQGGYHFDLKGDEWICDRSGETFWDLLEQAATQQAGETVSFR
Mn²⁺ (d⁵)   MNDSEFHRLADQLWLTIEERLDDWDGDSDIDCEINGGVLTITFENGSKIIINRQEPLHQVWLATKQGGYHFDLKGDEWICDRSGETFWDLLEQAATQQAGETVSFR
Gd³⁺ (f⁷)   MNDSEFHRLADQLWLTIEERLDDWDGDSDIDCEINGGVLTITFENGSKIIINRQEPLHQVWLATKQGGYHFDLKGDEWICDRSGETFWDLLEQAATQQAGETVSFR
Fe²⁺ (d⁶)   MNDSEFHRLADQLWLTIEERLDDWDGDSDIDCEINGGVLTITFENGSKIIINRQEPLHQVWLATKQGGYHFDLKGDEWICDRSGETFWDLLEQAATQQAGETVSFR
Eu³⁺ (f⁶)   MNDSEFHRLADQLWLTIEERLDDWDGDSDIDCEINGGVLTITFENGSKIIINRQEPLHQVWLATKQGGYHFDLKGDEWICDRSGETFWDLLEQAATQQAGETVSFR
Yb³⁺ (f¹³)  MNDSEFHRLADQLWLTIEERLDDWDGDSDIDCEINGGVLTITFENGSKIIINRQEPLHQVWLATKQGGYHFDLKGDEWICDRSGETFWDLLEQAATQQAGETVSFR
Co²⁺ (d⁷)   MNDSEFHRLADQLWLTIEERLDDWDGDSDIDCEINGGVLTITFENGSKIIINRQEPLHQVWLATKQGGYHFDLKGDEWICDRSGETFWDLLEQAATQQAGETVSFR
Ca²⁺        MNDSEFHRLADQLWLTIEERLDDWDGDSDIDCEINGGVLTITFENGSKIIINRQEPLHQVWLATKQGGYHFDLKGDEWICDRSGETFWDLLEQAATQQAGETVSFR
Lu³⁺        MNDSEFHRLADQLWLTIEERLDDWDGDSDIDCEINGGVLTITFENGSKIIINRQEPLHQVWLATKQGGYHFDLKGDEWICDRSGETFWDLLEQAATQQAGETVSFR
```

**Figure 5-15: Effect of the ions on CyaY. For each ion the mainly affected residues are highlighted onto the sequence; the ion/protein ratios are the same as those previously reported in the text. The bleached residues are coloured in red, while in blue are indicated the residues affected by hyperfine shift. For the paramagnetic ions the electronic distribution is indicated in parenthesis. The first three rows group the ions which have an isotropic electronic distribution; their only effect is to enhance the relaxation of the signals.**

Weaker binding sites have been indicated to be responsible for iron-induced aggregation; the ion would form intermolecular bridges between primary and secondary iron binding sites, generating transient asymmetrically arranged dimers [86]. Presumably these interactions are significantly stabilized only at high concentrations of iron(III).

Surprisingly, the main binding site identified in this study is not constituted by histidines, cysteins or tyrosines, but only by aspartates and glutamates. The stoichiometry of two $Fe^{2+}$ or six $Fe^{3+}$ per monomer suggests the formation of one or more di-iron clusters [96]; from our data it is tempting to locate one of them in the quadrangle formed by Glu18, Glu19, Asp22 and Asp23, whose suitably reoriented side chains are at distances well compatible with a di-iron cluster [97]. However the amide of Glu18 is not affected at 1:2 protein:iron ratio, and it is not conserved in human frataxin, where it is substituted by an alanine. In other eukaryotes this residue is mutated into a hydrophobic amino acid.

An exposed binding site formed exclusively by carboxylates would be a new feature for iron chelation and would explain the modest affinity constant measured for $Fe^{2+}$ binding. It would therefore be consistent with the putative chaperone role proposed for the protein [9,13,16]. Nevertheless, the weak affinity for iron and the low ion specificity highlighted by our data question the effective biological meaning of the binding site. First of all in fact, the iron binding constants measured for CyaY and human frataxin are too low when compared with the free iron concentration in mitochondria (4.8-9.2 μM) [98] and with the high concentration of other iron chelating agents, as citrate, that would compete with the frataxin monomer.
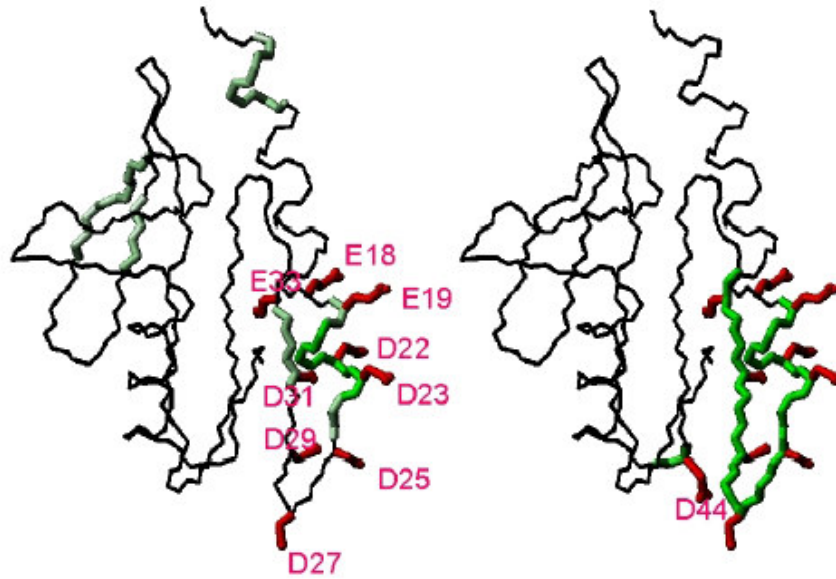
**Figure 5-16: Left: the backbone of residues disappearing and broadening upon the iron titration up to a protein:Fe$^{2+}$ ratio of 1:2 are indicated in green and pearl green, respectively. Right: the backbone of residues affected by the iron titration at 1:6 protein:Fe$^{3+}$ ratio are indicated in green. The side chain of the negatively charged residues is indicated in red.**

In addition it seems improbable that the iron bound by the protein is not substituted by other cations, present at higher concentrations in the cell, such as Mg$^{2+}$ or Na$^+$.

A possibility is that frataxin needs a partner to stabilize the interaction with iron. This would be an attractive hypothesis for at least two reasons: first, the very low affinities and poor selectivity reported for iron binding in CyaY and human frataxin could be ameliorated by the presence of a cofactor. Secondly frataxins have a highly conserved hydrophobic exposed patch, which could be the perfect interaction site for a partner protein. At present a similar cofactor has not been identified yet.

Otherwise, frataxin could bind iron as a dimer, with iron being the binding bridge between the monomers. Even if appealing, this is not likely because CyaY is able to bind iron even at a monomeric state, as shown by our NMR spectra; moreover no dimeric species can be detected in our experimental conditions, in accord with the results from previously reported studies [86]. For the short construct of human frataxin the same conclusions were reached [71].

In alternative, frataxin could have a ferritin like behaviour and form a supermolecular assembly to protect the iron in an insoluble core as the one described for ferritin. If this were the case, binding of iron could be the first step of the formation of a spherical aggregate that occurs at high iron:protein ratios [18]. However it has been

91

shown that aggregates formation occurs only at an ionic strength too low to be of physiologic relevance [67,86].

Our results suggest ad hoc mutations which can be designed to influence iron binding and/or aggregation; studying their effect *in vitro* and *in vivo* could provide valuable information on the actual role of frataxin. If its effective function is to bind iron, then mutation of the residues mainly involved in the binding should prove to be lethal for the cells. Alternatively, if the protein has a storage role, distinct mutations could be introduced to selectively affect the secondary binding sites involved in aggregation [71].

## 5.4 *Materials and Methods*

The effect of different cations on CyaY and on the CyaY_181922 mutant was assessed by titrating samples of typically 0.5 mM concentration either in 20mM Tris-HCl or 10 mM HEPES, 50 mM NaCl (pH=7-7.5) at 25 °C. The [$^1$H, $^{15}$N] HSQC spectra were recorded at 25° C on Varian Unity 500 instrument operating at 500 MHz and equipped with a 5 mm triple-resonance probe.

$Fe(NH_4)_2(SO_4)_2$, $FeCl_3$, $CaCl_2$, $CoSO_4.7H_2O$, $MnSO_4.4H_2O$, $LuCl_3$ from Sigma and $EuCl_3$, $YbCl_3$, $GdCl_3$ from Strem Chemicals were used.

Iron titrations were performed on a 0.8mM protein sample both in aerobic conditions with Fe(II) or Fe(III) and in anaerobic condition with Fe(II), preparing the samples in a glove box filled with argon.

For the different titrations the following protein:ion ratios were used:

Fe(II): 1:0.5, 1:1, 1:2, 1:3, 1:5, 1:7 and 1:8; CyaY 0.8 mM.

Fe(III): 1:0.5, 1:1, 1:2, 1:3, 1:5, 1:7 and 1:8; CyaY 0.8 mM.

$Ca^{2+}$: 1:0.5, 1:1, 1:2, 1:3, 1:5, 1:7, 1:9, 1:11, 1:18; CyaY 0.3 mM.

$Eu^{3+}$ :1:0.5, 1:1, 1:1.5, 1:2; CyaY 0.5 mM.

$Yb^{3+}$ :, 1:4; CyaY 0.5 mM.

$Lu^{3+}$ : 1:0.5, 1:1, 1:1.5, 1:2, 1:2,5 ; CyaY 0.6 mM.

$Gd^{3+}$ : 1:0.05, 1:0.1, 1:1.15, 1:1.2 ; 1:1.25 ;1:1.3; CyaY 0.58 mM.

$Co^{2+}$: 1:0.5, 1:1; CyaY 0.58 mM

$Mn^2$: 1:0.1, 1:0.2; Cyay 0.58 mM

[$^1$H,$^{15}$N] HSQC spectra were recorded for each of the protein:cation ratio typically using 80 increments for the indirect dimension.

# Iron Sulphur Clusters                6

Biological iron-sulphur clusters have been identified 40 years ago as acid-labile prosthetic groups contained within a class of electron carrier proteins called "Ferredoxins". For a recent review, see [99].

Since their discovery, they have been considered agents of electron transfer, given their versatile electronic properties. However, increasing numbers of proteins containing iron-sulphur clusters with roles which differ from the canonical one have been identified. In some cases the cluster has a role in stabilizing the structure of the protein, in others it can act as an environmental sensor, or as a modulator of gene regulation. In many proteins it is known to participate in radical formation [100]. Such functional diversity reflects the chemical versatility of iron and sulphur, a characteristic which led to the suggestion that pre-biotic [Fe-S] complexes could have played an important role in the emergence of life on earth [101].

The relevance of Fe-S clusters can be further appreciated considering that the three main processes in life (nitrogen fixation, photosynthesis and respiration) all require the presence of iron-sulphur clusters.

## 6.1 *Structure and function of Iron Sulphur Clusters*

[Fe-S] clusters are modular structures, with a remarkable facility for conversion, interconversion, ligand exchange, oxidative degradation, under both free and protein bound conditions [102].

The different types of Fe-S clusters architectures that can be found in nature are shown in Figure 6-1. The $[Fe_2(\mu2\text{-}S)_2]$ rhomb is the basic building block for the assembly of high nuclearity centres: cubane-type $[Fe_4\text{-}S_4]$ clusters, for example, can be assembled from two $[Fe_2\text{-}S_2]$ units while $[Fe_3\text{-}S_4]$ and $[Fe_7\text{-}S_8]$ clusters derive from $[Fe_4\text{-}S_4]$ via loss of one iron and cluster fusion respectively. In some [Fe-S] proteins then, Fe is not the only metal in the cluster: the nitrogenase MoFe protein, for example, contains a cluster called FeMo-cofactor, which has a [7Fe-9S-Mo] core [99].

In general, the coordination of Fe in the cluster is tetrahedral, with cysteines completing the S coordination at each Fe site. Occasionally though, aspartates, histidines, serines and backbone amides can be found at a unique Fe site.

| | | |
|---|---|---|
| | $[2Fe-2S]^{2+}$ | $S = 0$ |
| | $[2Fe-2S]^{+}$ | $S = 1/2$ or $9/2$ |
| | | |
| | $[3Fe-4S]^{+}$ | $S = 1/2$ |
| | $[3Fe-4S]^{0}$ | $S = 2$ |
| | $[3Fe-4S]^{-}$ | $S = 5/2$ |
| | $[3Fe-4S]^{2-}$ | $S = ?$ |
| | | |
| | $[4Fe-4S]^{3+}$ | $S = 1/2$ |
| | $[4Fe-4S]^{2+}$ | $S = 0$ |
| | $[4Fe-4S]^{+}$ | $S = 1/2$ or $3/2$ |
| | $[4Fe-4S]^{0}$ | $S = 4$ |
| | | |
| | $[8Fe-8S]^{5+}$ | $S = 7/2$ |
| | $[8Fe-8S]^{4+}$ | $S = 3$ or $4$ |
| | $[8Fe-7S]^{3+}$ | $S = 1/2$ or $5/2$ |
| | $[8Fe-7S]^{2+}$ | $S = 0$ |
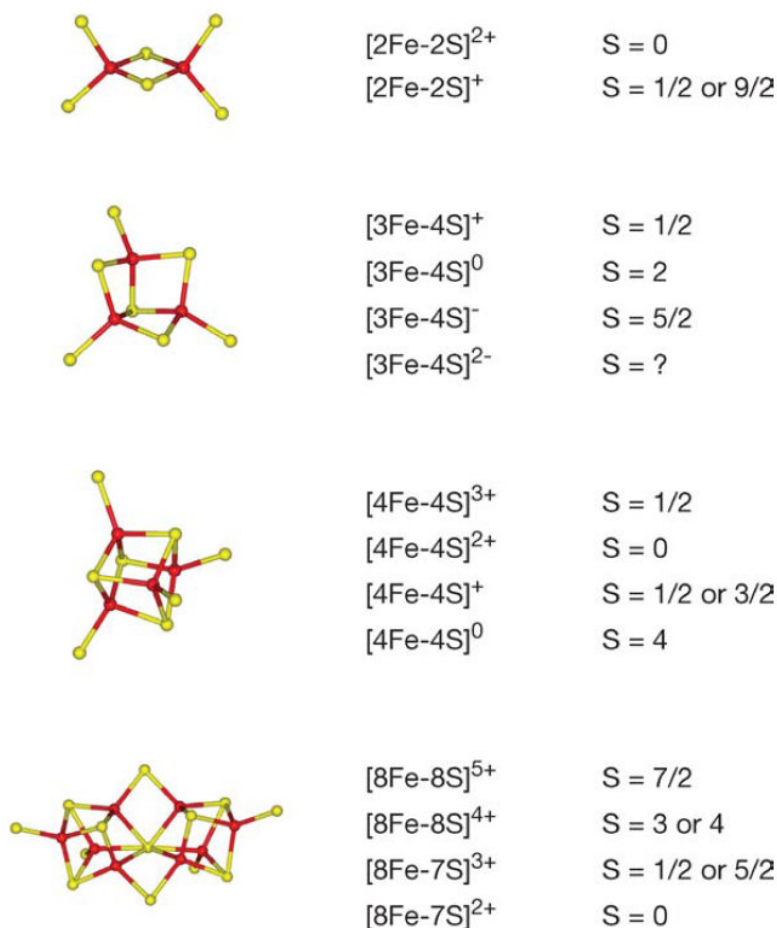
**Figure 6-1: Structures, oxidation states and spin state of crystallographic defined [Fe-S] clusters. Iron is shown in red and sulphur is shown in yellow. The spin state denoted by a question mark has not been determined yet. The [Fe$_3$-S$_4$]$^-$ cluster has only been observed as a fragment in heterometallic [M-Fe$_3$-S$_4$] clusters in which M is a divalent transition metal ion [99].**

Their exquisite structural and chemical versatility is the reason for their extreme functional flexibility. Iron sulphur clusters are involved in:

- electron transfer: uptake, release and storage of electrons. The ability to delocalise electrons over both Fe and S, makes [Fe-S] clusters ideal cofactors for mediating biological electron transport. Not only they accomplish this role in the photosynthetic and respiratory electron transport

chains, but they also define the electron transport pathways in many membrane bound and soluble redox enzymes and constitute the redox active centres in ferredoxin, the principal electrons carrier protein [103].

- catalysis: substrate binding, activation and turnover. [Fe-S] clusters constitute the substrate-binding site of many redox or non-redox enzymes. For this purpose, the typical tetrahedral S coordination around the Fe atoms is not complete. For example aconitase, the enzyme involved in the interconversion of citrateand isocitratate in the TCA cycle, adopts a non-cysteinil ligation at a unique Fe site of the cluster, to facilitate substrate binding. Here the cofactor does not show redox properties, but functions as a Lewis acid, activating the $OH^-$ group of the substrate for elimination. Interestingly, the protein interconverts its cluster from an inactive $[Fe_3\text{-}S_4]$ form to the active $[Fe_4\text{-}S_4]$ form when needed [104].

- ions storage: uptake, binding, release of metals in soluble forms. Clusters can also have an iron storage role, as found for many polyferredoxins expressed in operons that encode for enzymes such as hydrogenase, which contains multiple $[Fe_4\text{-}S_4]$ clusters [99].

- structure stabilisation: configuration of protein tertiary and/or quaternary structure. In many DNA repairing enzymes these cofactors allow the host protein to adopt its correct fold: in endonuclease III, for instance, the redox-inactive $[Fe_4\text{-}S_4]$ cluster is structurally functional because it controls the structure of a protein loop essential for DNA recognition.

- regulation of gene expression. There are now several well-characterised examples of clusters involved in transcriptional or translational regulation of gene expression in bacteria [105]. Each of them senses a different type of environmental stimulus and reacts to it via a diverse response, such as cluster assembly, conversion or redox chemistry. SoxR, for instance, senses oxidative stress through the oxidation of the $[Fe_2\text{-}S_2]^{2+}$ and thereby stimulates expression of SoxS, which in turn activates the transcription of many enzymes in the oxidative stress response [106]. FNR (Fumarate and Nitrate Reduction) is an oxygen sensor, active as a dimer. When the cluster is disassembled due to oxidative stress, the protein becomes an inactive monomer. Iron Regulating Protein IRP instead senses the cellular iron concentration. When the iron concentration is low, the apo form of the protein regulates the relative levels of transferrin and ferritin, responsible for iron upload and iron storage respectively. However when iron becomes

available again, a $[Fe_4-S_4]^{2+}$ cluster is reconstituted and, strikingly, the protein shows an aconitase activity [107].

In the table below the main functions of [Fe-S] clusters are summarised.

**Table 6-1: Functions of some biological [Fe-S] clusters [99]**

| Function | Examples | Cluster type |
|---|---|---|
| Electron transfer | Ferredoxins; redox enzymes | [2Fe-2S]; [3Fe-4S]; [4Fe-4S] |
| Coupled electron/proton transfer | Rieske protein | [2Fe-2S] |
|  | Nitrogenase | [8Fe-7S] |
| Substrate binding and activation | (de)Hydratases | [4Fe-4S] |
|  | Radical SAM enzymes | [4Fe-4S] |
|  | Acetyl-CoA synthase | Ni-Ni-[4Fe-4S], [Ni-4Fe-5S] |
|  | Sulphite reductase | [4Fe-4S]-siroheme |
| Fe or cluster storage | Ferredoxins | [4Fe-4S] |
|  | Polyferredoxins | [4Fe-4S] |
| Structural | Endonuclease III | [4Fe-4S] |
|  | MutY | [4Fe-4S] |
| Regulation of gene expression | SoxR | [2Fe-2S] |
|  | FNR | [4Fe-4S]/ [2Fe-2S] |
|  | IRP | [4Fe-4S] |
|  | IscR | [2Fe-2S] |
| Regulation of enzyme activity | Glutamine PRPP amidotransferase | [4Fe-4S] |
|  | Ferrochelatase | [2Fe-2S] |
| Disulphide reduction | Ferredoxin:thioredoxin reductase | [4Fe-4S] |
|  | Heterodisulfide reductase | [4Fe-4S |
| Sulphur donor | Biotin synthase | [2Fe-2S] |

Abbreviations used are SAM, S-adenosylmethionine; acetyl-CoA, acetyl coenzymeA; FNR, fumarate and nitrate reduction; IRP, iron-regulatory protein; IscR, iron-sulfur cluster assembly regulatory protein; PRPP, phosphoribosylpyrophosphate.

## 6.2 *Assembly of the clusters*

Considering the importance of these groups, it is not surprising the increasing interest on their origin and biosynthesis. Analogues of $[Fe_2-S_2]$ and $[Fe_4-S_4]$ have been synthesized from $Fe^{3+/2+}$, $S^{2-}$ and thiolates by spontaneous self-assembly in aprotic media [108]. Moreover the $[Fe_2-S_2]$ and $[Fe_4-S_4]$ centres in many ferredoxins and simple Fe-S proteins can be reassembled spontaneously *in vitro* by anaerobically incubating apo-proteins with $Fe^{3+/2+}$ and $S^{2-}$ in the presence of dithiothreitol or β-mercaptoethanol. *In vivo* however, the mechanism of self-assembly

cannot be a viable biochemical pathway for the Fe-S formation, but specific proteins and es are required. In fact, both the transport and the internal metabolism of iron in cells are tightly regulated, owing to the insolubility of $Fe^{3+}$ in aqueous solution at neutral pH and to the ability of $Fe^{2+}$ to interact with oxygen and generate hydroxyl radicals, resulting in damage of cellular components. In addition, $S^{2-}$ would be extremely toxic to the cell at the levels required for self-assembly of the cluster [109].

In prokaryotes at least three systems are responsible for the clusters' synthesis: the *Nif* (Nitrogen Fixation), the *Isc* (Iron-Sulphur Cluster) and the *Suf* (Sulphur mobilization) system.

The identification of macromolecules involved in the clusters' assembly first came from genetic and biochemical analyses of proteins specifically involved in nitrogen fixation. In particular, studies on nitrogenase provided many insights on the plausible mechanism of clusters' biogenesis. Nitrogenase is a complex metalloenzyme that catalyzes the nucleotide dependent reduction of dinitrogen (nitrogen fixation); it is constituted by two component proteins, the Fe protein and the MoFe protein, containing respectively a $[Fe_4-S_4]$ and a Mo-Fe mixed cluster. The primary translation products of the genes encoding the nitrogenase component proteins are not active, lacking the associated metallocluster. It was found that a specific group of genes, the *nif* operon, is responsible for the synthesis and the insertion of the co-factors onto the immature proteins. The *nif* cluster comprises ten genes: *orf6, nifU*, *nifS*, *nifV*, *orf7*, *orf8*, *nifW*, *nifZ, nifM*, *orf9* [110]. Inactivation of many of them causes impairment of either the Fe protein or the MoFe protein; however deletion of two particular genes, nifS and nifU, results in an activity loss of both components. Because the only feature shared by the Fe protein and the MoFe protein is the presence of the [Fe-S] cluster, the nifU and nifS phenotype uniquely points at defects in the early stage of nitrogenase-specific cluster assembly [110]. As it was later discovered, NifS is a pyridoxal-phosphate (PLP) containing homodimer, involved in the mobilization of S for maturation of the nitrogenase. The protein catalyses the desulphuration of L-cysteine through the formation of a cysteine-PLP ketimine adduct and the subsequent nucleophilic attack by the thiolate anion of a very conserved cysteine (Cys325) on the sulphur of the substrate cysteine (Figure 6-2) [111].
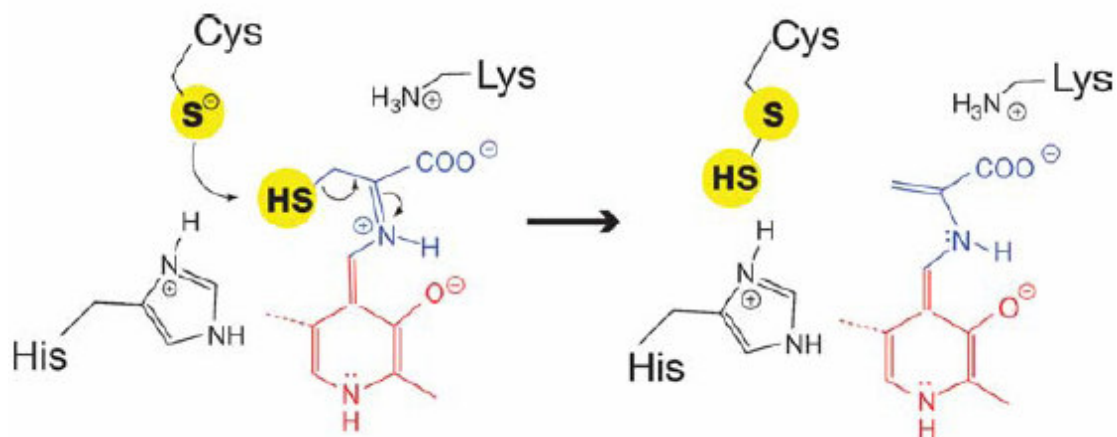
**Figure 6-2: Salient mechanistic features of the NifS/IscS class of cysteine desulphurases. An enzyme-bound persulphide is formed through nucleophilic attack by the thiolate anion of the active site-cystein located in a flexible loop on the sulphur of the cystein-PLP (blue and red respectively) adduct. The imidazole ring of a histidine residue acts as an acid-base catalyst in protonation-deprotonation steps during the catalysis [99].**

NifU is a homodimeric modular protein that contains one $[Fe_2-S_2]^{2+}$ per monomer [112] and whose function is to provide an intermediate scaffold for the assembly of the [Fe-S] cluster necessary for nitrogenase activation [113]. The permanent clusters are located in the core domain and might have a role in the formation or release of the transient ones [114].

The observation that, although very low, the activities of both Fe and FeMo proteins are not completely suppressed when NifU and NifS are absent, suggested that some other proteins, related with general [Fe-S] cluster biosynthesis, could weakly replace them [110]. The so called '*Isc* operon' was identified in *A. Vinelandii*, as well as in *E. Coli* and *H. Influenzae*, as the result of a search for homologue counterparts of nifU and nifS involved in 'housekeeping' functions [24]. The *A. Vinelandii* operon contains nine genes, namely cisE2, iscR, iscS, iscU, iscA, hscA, hscB, fdx and orf3. CisE2 is homologue to the cysE1 gene present within the *nif* operon; it is absent in *E. Coli*, and is probably transcribed separately from the rest of the cluster. Its product catalyses the production of O-acetylserine, involved in the synthesis of L-cysteine, suggesting the possibility of a mechanism for the specific targeting of L-cysteine for [Fe-S] cluster formation. IscS, IscU and IscA exhibit sequence identity respectively with NifS, NifU and orf6. The proteins encoded by the genes of the *E. Coli isc* operon will be further discussed in paragraph 6.3.

While inactivation of IscU and IscS is lethal in *A. Vinelandii*, it resulted in mutation strains which were still viable in *E. Coli*. Further investigation led to the discovery in *E. Coli* of an additional iron-sulphur cluster biosynthetic system, the *suf* operon, not encoded in *A.Vinelandii* [115]. The *suf* operon is formed by six genes, sufA, sufB, sufC, sufD, sufS, sufE. SufA and SufS are homologous to IscA and IscS; SufC is a soluble ATPase [116] while SufB and SufD do not have an assigned function, although they interact with SufC to form a ternary complex. SufE is a novel sulphur transfer protein that enhances SufS cysteine desulphurase activity by accepting sulphane sulphur [117]. The SufBCD complex in turn stimulates SufE enhancement of SufS by an unknown mechanism [118].

Differently from what happens in *A. Vinelandii*, where the *nif* operon specifically encodes for the proteins involved in nitrogenase maturation, thus having a well separated role from the *isc* operon, in *E.Coli* the distinction between *suf* and *isc* function is not as clear cut. Mutant strains where the *suf* operon was inactivated did not show any apparent phenotype under normal, nonstress growth conditions. However all the *suf* products, with the exception of SufD, were shown to be fundamental when the *isc* operon is inactivated [115]. Based on the observation that the *suf* operon is regulated by OxyR and Fur, which are optimised to sense respectively $H_2O_2$ and low iron concentrations [119], it was recently hypothesised that the physiological roles of the *suf* and *isc* operons diverge under hydrogen peroxide stress and iron starving conditions. While the *isc* operon is in charge of the clusters synthesis under healthy conditions, the *suf* operon would therefore be able to compensate for the loss of iron and sulphur availability that occurs in concomitance with oxidative stress and specifically build the Fe-S clusters in these situations [118].

## 6.3 *ISC operon of E.Coli*

Understanding the functions of the Isc proteins in prokaryotes is important to provide insights into iron homeostasis and Fe-S clusters assembly in mitochondria, with potential relevance to iron-storage diseases and the control of cellular iron uptake.

The Isc operon of the bacterium *E. Coli* comprises eight proteins, as shown in Figure 6-3. Only three of them, IscS, IscU and IscR, have been functionally characterized. The role of IscA is still controversial and at the moment there is no definitive evidence of how the gene products of hscA, hscB and fdx are involved in the cluster formation. Similarly the function of YFHJ, encoded by orf3, is not clear yet. Knock out experiments, however, have shown that deletion of orf3 and iscR do not cause alterations in the Fe-S cluster proteins synthesis, while strains with mutation in the

iscSU-hscAB-fdx genes exhibit conspicuous phenotypical consequences, almost identical to one another. The effect of inactivation of iscA is small but still appreciable on Fe-S enzymes [120].
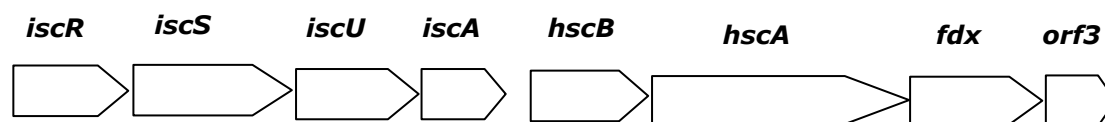


**Figure 6-3: Scheme of the genes clustered in the _E. Coli_ ISC operon.**

In the following paragraphs a brief description of what is known about the proteins encoded by the operon is provided.

### 6.3.1  IscR

Although a detailed structure of IscR does not exist, sequence homology with the bacterial transcription factor MarA (*m*ultiple *a*ntibiotic *r*esistance) has suggested Schwartz at al. a thourough investigation on the putative regulatory function of IscR [121]. They found that the protein is co-transcribed with IscSUA and that the transcription of the four proteins originates from a single promoter located upstream of IscR and is repressed by IscR itself. Because IscR contains a [$Fe_2$-$S_2$] cluster, they proposed that the regulation of the Isc proteins expression depends on the levels of apo-IscR. When all the Isc proteins are engaged in the Fe-S cluster assembly and this process becomes rate limiting, the level of [$Fe_2$-$S_2$] IscR would decrease and repression of IscRSUA synthesis would be relieved. The resulting increase in the Isc assembly proteins would subsequently lead to an increase in the rate of Fe-S cluster formation, causing levels of [$Fe_2$-$S_2$] IscR to rise, thus re-establishing the repression of the *isc* operon. In this way, [Fe-S] cluster assembly would be finely tuned and adapted to the cellular needs for synthesis or repair of Fe-S proteins.

### 6.3.2  IscS

IscS is an extremely conserved protein, composed of two identical 45kDa subunits. It uses a pyridoxal 5'-phosphate (PLP) to catalyze the reductive elimination of sulphur from L-cysteine to yield alanine and sulphur ($S^0$) or sulphide in the presence of a reducing agent. It was shown that the sulphane sulphur from the cysteine substrate

of *E. Coli* IscS is donated directly as persulphide to apo-IscU during iron-sulphur cluster formation [122]. The protein is also involved in thiamine, thionucleoside and biotin biosynthesis [123].

The crystal structure of the protein reveals that it contains 40.1% and 13.8% of α-helical and β-strand elements respectively. Each IscS monomer is composed of two domains, the larger of which is mainly involved in the formation of the dimer interface [123].
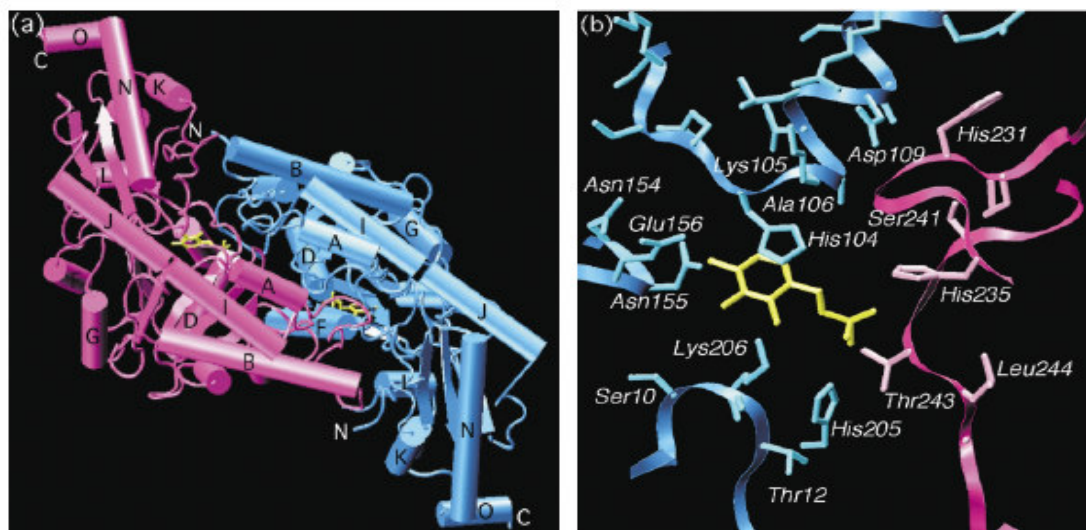


**Figure 6-4: a) Crystal structure of E. Coli IscS dimer where the monomers are colour coded pink or blue. b) The dimer interface near the active site of IscS. PLP is shown in yellow [123].**

The active site and the PLP of each monomer are 30 Å apart from one another, in highly solvated pockets close to the protein surface. The PLP molecule is anchored to the active site by formation of an internal aldimine Schiff base with Lys206 as well as a number of polar and non-polar interactions Figure 6-4-b.

Sulphur multiple transfer most likely occurs through a sequential rather than a cooperative process involving both subunits. Similarly to what happens for NifS, the first step of the reaction catalyzed by IscS is the formation of an external aldimine Schiff base between the substrate cysteine and the PLP group. This is followed by the nucleophilic attack of the bound cysteine by Cys328 to form a Cys328-persulphide and subsequent release of alanine [111]. The position of Cys328 is not clearly observable in the crystal structure but the position of the flanking residues suggests that it is located near the surface. The imidazole ring of His104 is positioned directly above the pyridine ring of PLP and seems to act as an acid-base catalyst in several protonation-deprotonation steps during catalysis (see Figure 6-2).

Cys328 plays a dual role in sulphur transfer since it participates in catalysis by attacking the substrate cysteine thiol and by transferring the bound persulphide to the acceptor protein IscU [123]. The fact that the electron density of Cys328 is not detectable and that the entire 328-333 region seems disordered may reflect the presence of a significant conformational change of the active site region upon catalysis. In fact the position of Ala327 restricts the possible location of Cys328 such that the γ-sulphur atom is >17 Å from the PLP cofactor, too great a distance for the formation of a persulphide bond to occur. However, no amino acid side chains or segments of the protein backbone project in the cleft between the thio side chain and the PLP; therefore Cys328 may move closer to the cofactor upon the rotation of the peptide backbone. In particular, residues not involved in bonding interactions with other parts of the protein may easily rotate. Computer simulations demonstrated that rotation about the Φ and Ψ angles of Gly325 allows a conformational rearrangement that brings Cys328 to within 3 Å of the PLP, thus allowing catalysis. A mutated IscS with a Val instead of Ala 327 was found to have decreased activity; even though the biochemical reasons for this were not investigated, one could think that the branched side chain of the valine residue interferes with the flexibility of the active loop [123].

### 6.3.3  IscU

*E. Coli* IscU represents one of the most conserved protein sequences in nature [124]. It is homologous to the N-terminal domain of NifU, with which it shares three conserved cysteines (Cys37, Cys63, Cys106). It was demonstrated that the protein represents a scaffold for the assembly of a transient Fe-S cluster and that it interacts with IscS to accept a sulphur atom (see paragraph 6.3.4). The origin of the iron is still unknown. Some evidence suggests that frataxin, a protein expressed at abnormally low levels in Friedreich's ataxia patients and involved in iron homeostasis [16,125], is responsible for the delivery of iron ions to IscU, but a direct interaction between the two proteins has not been found yet. According to some authors, the iron chaperone would instead be IscA [126].

Biophysical studies on IscU produced controversial results.

Mansy et al. reported that the D40A mutant of *T. maritime* (*Tm*) IscU, chosen for stability reasons, dimerizes upon cluster acquisition [127]. Besides, NMR investigations revealed that *Tm* IscU adopts a molten globule-like state: even though NMR chemical shifts and dihedral angles suggested the presence of several

elements of secondary structure, the tertiary structure appeared to be fluxional among widely different conformational arrangements [128].

Conversely, in a paper from 2004, Adinolfi et al showed that *E. Coli* IscU is a stable and well folded protein. Deconvolution of the far UV CD spectrum yielded the following percentage of secondary structure elements: 40.8% of α-helix, 13.7% of β-strand and 19.6% of turns. The thermal unfolding was 95% reversible and the excellent chemical shift dispersion of the NMR spectra suggested the presence of a tertiary structure and provided strong evidence for a stable globular fold. The fairly homogeneous distribution of the NMR relaxation times confirmed that IscU is a compact protein, without relevant differences of local flexibility. Gel filtration, ultracentrifugation and the rotational correlation time provided by NMR are consistent with the presence of a monomeric form in solution, capable to host a $[Fe_2-S_2]^{2+}$ cluster in a typical experiment of iron-sulphur cluster reconstitution [129].

The affinity of IscU for iron is also controversial: on one hand, calorimetric and fluorescence experiments showed that *Tm* IscU has a high affinity for both ferric and ferrous ions [130]. On the other hand, however, Raman spectroscopy and magnetic circular dichroism on *A. Vinelandii* IscU and NMR perturbation experiments on *E. Coli* IscU demonstrated that the protein binds iron extremely weakly, if at all [129,131].

Recently the 3D NMR structure of Zn loaded *H. Influenzae* (*Hinf*) IscU has been published. According to the authors, the calculated structure could demonstrate structural features found in the cluster loaded form [132]. The presence of Zn stabilises the protein and allows structure determination. In fact several $^1H-^{15}N$ HSQC cross-peaks disappear or sensibly broaden upon addition of the Zn chelating agent EDTA to the NMR sample. The crystal structure of Zn bound *S. pyogenes* (*Sp*) IscU also has been recently solved [133]. The two structures are very similar; both are compact monomers, characterized by a αβ tertiary fold. $Zn^{2+}$ ions were externally provided during cell growth in the first case, whereas *Sp* IscU crystallised with the Zn bound; it is not clear whether Zn binding has a biological role or it mimics an iron ion. Although a detailed description of the cluster binding site is not available, it is conceivable that the cofactor binds in the Zn binding site identified in the above cited structures. Raman studies of the *A.Vinelandii* IscU homodimer coordinating one $[Fe_2-S_2]^{2+}$ cluster had already showed that at least one non-cysteinil ligand is involved in cluster ligation [131]. Accordingly, in both the solved structures, Zn is coordinated by the three conserved cysteines. The crystal structure of *Sp* IscU revealed that the fourth ligand is represented by the OD1 of an aspartic residue (Asp42 in *Sp* IscU

numbering). In both structures the metal is solvent exposed and located close to a flexible loop that could allow the coordination of the $[Fe_2\text{-}S_2]^{2+}$ cluster, supporting the idea that the cluster binds at the zinc binding site [133].
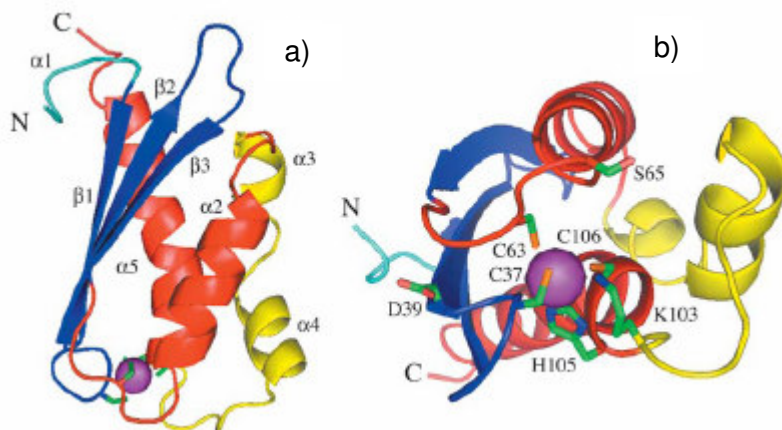


**Figure 6-5: a): Ribbon diagram representation of residues 21-128 of *Hinf* IscU. $Zn^{2+}$ is shown in purple while the side chains of the residues believed to chelate the metal ion are displayed in green. b) Ribbon diagram of the zinc binding region, with the three conserved cysteines Cys63, Cys37, Cys196 and His105 [132].**

How the cluster is assembled on IscU is still a matter of debate, also considering the uncertainties about the origin of iron. However at the moment it is believed that the cluster assembly starts with the sulphur transfer from IscS to IscU, which will be the subject of next paragraph [131].

### 6.3.4 Mechanism of the sulphur transfer from IscS to IscU

The first investigations on a putative IscU-IscS complex were subsequent to the finding that the $[Fe_2\text{-}S_2]$ cluster in IscU was formed, *in vitro*, upon addition of IscS in the presence of an excess of L-cysteine and stoichiometric amount of ferric ammonium citrate. Gel exclusion chromatography was used to show the formation of a weakly bound $\alpha_2\beta_2$ heterotetramer from a mixture of equimolar IscS and IscU dimers [131]. However, later studies demonstrated that sulphane transfer from IscS to IscU is catalysed by a heterodimeric complex, identified by mass spectrometry. Direct transfer of sulphane sulphur from a cysteine persulphide intermediate on IscS to a cysteine on IscU was demonstrated again by mass spectrometry [134]. This was suggested to be the first step on the clusters' synthesis. In fact, although in principle

104

Fe binding could initiate the clusters' assembly on IscU, there is no direct evidence of an iron binding to IscU. $Fe^{2+}$ has instead been indicated as the reducing agent for the reduction of the $S^0$ from the cysteine persulphide to the $S^{2-}$ of the cluster [135]. Alternatively, the reducing equivalents for the reaction can be provided by Ferredoxin, as discussed in paragraph 6.3.7.

In a paper published in 2002, a Japanese group showed that the IscU/IscS complex is characterised by a disulphide bond between Cys328 of IscS and Cys63 of IscU and that IscU enhances the cysteine desulphurase activity of IscS through Cys63. Moreover, they showed that IscU exists as a monomeric and a covalent dimeric form, and that either of these forms can receive sulphur form IscS [122].

Since the complex between the two proteins is formed only when L-cysteine is present, it is likely that the target of IscU is IscS with the sulphane sulphur derived from L-cysteine attached at the Cys328 active site. It was hypothesised that the Sγ of Cys63 of IscU attacks the Sγ of Cys328 of IscS that carries the sulphane sulphur. If IscU reacts as a monomer, the sulphane sulphur may be transferred to a cysteine residue close to Cys 63, which performs the attack on IscS Cys328. Alternatively, if IscU reacts as a dimer, Cys63 would be the primary site for sulphur transfer: the disulphide bridged IscU would react with persulphide-containing IscS resulting in a persulphide linkage between Cys-328 of IscS and Cys-63 of one of the two subunits of the IscU dimer, releasing the other subunit of IscU with a persulphide on Cys63. Because the two Cys328 of each IscS subunit are located on opposite side of the protein, two IscU dimers would react with one IscS dimer as exemplified in Figure 6-6. Since Cys63 appears to be fundamental for the desulphurase activity of IscS, it is conceivable that the attack is crucial for the IscU-mediated activation of IscS: probably it facilitates the release of the sulphane sulphur from Cys328 of IscS. The dissociation of IscU from IscS must be quick, in order to make Cys328 available for the next cycle of the catalysis. *In vitro* this process is facilitated by the reducing agent DTT. It is still unknown how the process takes place *in vivo*, but the involvement of the two chaperones Hsc66 and Hsc20 was postulated [122].
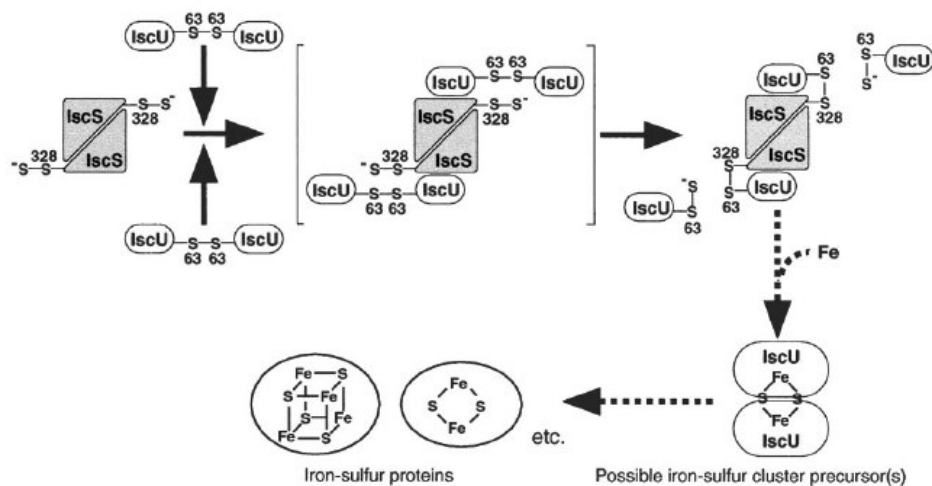
**Figure 6-6: Proposed scheme for the formation of the covalent IscS/IscU complex involving the sulphur transfer from Cys-328 of IscS to Cys-63 of IscU. The bracketed α2β4 hexamer of the IscS/IscU complex is supposed to be produced transiently in the reaction. Details of the processes indicated by dotted arrows are not known. The type of iron-sulphur clusters on IscU serving as iron-sulphur cluster precursor(s) has not been determined [122].**

### 6.3.5  IscA

*E. Coli* IscA is a fairly conserved protein whose role is still under debate. Although deletion of IscA gene is not lethal, it leads to a decrease of Fe-S proteins [120]. It was suggested that IscA is a [Fe-S] cluster scaffold, in analogy with IscU or that it is responsible for the iron delivery to the clusters [126,136-138]. The protein elutes mainly as a dimer in gel filtration experiments, even if a minor peak corresponding to a tetrameric form was also found [138].

Many biochemical studies proved that the dimeric form of the protein can coordinate a transient [$Fe_2$-$S_2$] cluster with two conserved cysteines from each subunit [138-140]. Moreover *in vitro* experiments demonstrated that holo-IscA can tightly bind ferredoxin, while apo-IscA has no affinity for it. In addition, it was also found that holo-ferredoxin was formed exclusively when holo-IscA was present, and not when the protein was incubated only with iron and sulphur. This was considered a strong indication of the involvement of IscA in the delivery of the cluster to ferredoxin [138].

Other studies, however, suggested that the biological role of the protein consists in recruiting intracellular iron and delivering it to IscU for the assembly of the clusters. A detailed iron binding analysis, carried out by Ding and co-workers, provided evidence of a tight iron binding, characterised by an apparent association constant of 3.0 X $10^{19}$ $M^{-1}$. The authors claimed that two IscA monomers form one iron binding site, in

accord with the proposed dimeric or tetrameric nature of the protein [141]. *In vitro* experiments showed that the iron centre of IscA is mobilized by L-cysteine and that, upon addition of IscS, the ion is transferred from IscA to IscU [136,137].

The X ray structure of the monomeric form of the protein reveals that it folds into a single domain that comprises α helical and β stranded elements (Figure 6-7 a). In the crystal state it was found that monomer $\alpha_1$ is closely associated with a second monomer $\alpha_2$, which exhibits minor structural differences. The two monomers interact to form a $\alpha_1\alpha_2$ dimeric form, which, upon association with a symmetry related dimer, can originate the tetrameric form reported in Figure 6-7 c). The dimers arrangement produces a basket-like structure that may serve as a location for a [Fe-S] cluster or as a binding site for proteins that interact with IscA.
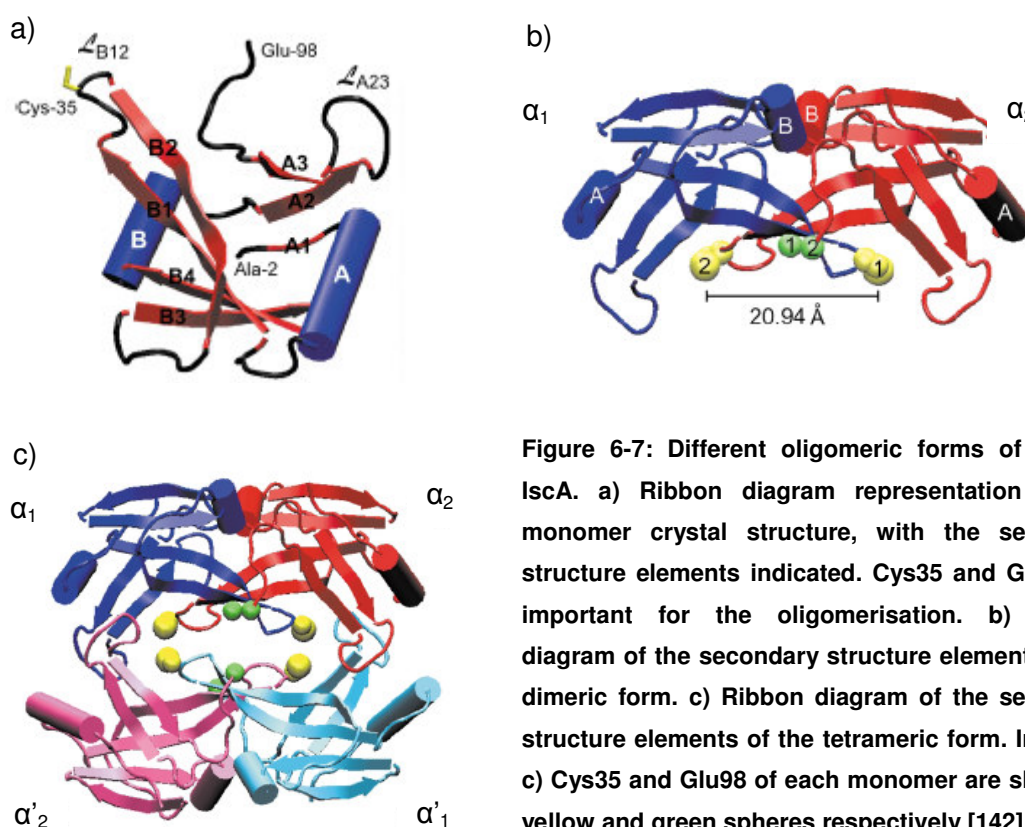


**Figure 6-7: Different oligomeric forms of *E. Coli* IscA. a) Ribbon diagram representation of the monomer crystal structure, with the secondary structure elements indicated. Cys35 and Glu98 are important for the oligomerisation. b) Ribbon diagram of the secondary structure elements of the dimeric form. c) Ribbon diagram of the secondary structure elements of the tetrameric form. In b) and c) Cys35 and Glu98 of each monomer are shown as yellow and green spheres respectively [142]**

The position of the two conserved cysteines, Cys99 and Cys101, are not well determined in the crystal structure, because of intrinsic disorder of the region. However, the position of Glu98 from each monomer in the dimeric form places restriction on the positioning of Cys99, that are too far apart to coordinate the same [FeS] cluster. On the contrary, the formation of the tetramer brings Cys99 and

Cys101 from equivalent monomers of different dimers close enough to cooperate in iron binding and/or cluster formation [142].

### 6.3.6 HscAB

Hsc66 and Hsc20, the products of hscA and hscB genes, form the second Hsp70-type chaperone system in addition to DnaK-DnaJ-GrpE in *E. Coli*.

The members of the hsp70 family are ATP-dependent molecular chaperones that play a role in protein folding, degradation of misfolded proteins, membrane trafficking, regulatory processes and in the maintenance of cellular viability upon stress [143].

Hsc66, a 66 KDa protein, exhibits 40% of sequence identity with DnaK; in particular the N terminal domain responsible for the ATPase activity of the hsc70 is well conserved, while the C-terminal domain, which is believed to function in protein recognition, contains important sequence differences. This suggests that the role of hsc66 is somehow specific for a particular process [144].

HscB is located upstream of hscA and encodes for the 20 kDa protein Hsc20. Hsc20 bears sequence similarities with the hsp40 class of proteins, also termed J proteins, that function as co-chaperones for the hsp70 proteins. The N terminal domain of Hsc20 bears the 20% of sequence similarity with DnaJ and contains the His-Pro-Asp sequence that characterises the J type proteins. However, its C terminal domain is extremely different from DnaJ C-terminus, thus suggesting a specialised role for the molecule. Both Hsc66 and Hsc20 are monomers in solution [144].

Hsp70 proteins couple ATP hydrolysis to conformational changes that modulate substrates' affinity. ATP binding leads to the formation of a *tense* state (T), with reduced substrate affinity and increased substrate exchange rates. Subsequent ATP hydrolysis to ADP and phosphate, results in a *relaxed* state (R), with increased substrate affinity and decreased substrate exchange rates. Hsc20 binds IscU and Hsc66. Strikingly, Hsc20 and IscU can synergistically stimulate Hsc66 activity > 400 fold [145,146]. In particular, binding of the co-chaperone and the substrate stimulates both the ATP hydrolysis and the conformational exchange from T to R state. Hsc20 would then dissociate from the complex, while IscU would remain trapped in the high affinity ADP-bound R state. Dissociation of the ADP molecule and subsequent binding of an ATP molecule would cause the dissociation of IscU, with restoration of the initial T state. IscU was actually found to enhance R→T conversion as an auto-regulatory mechanism to allow additional cycles of substrate binding and release [145].
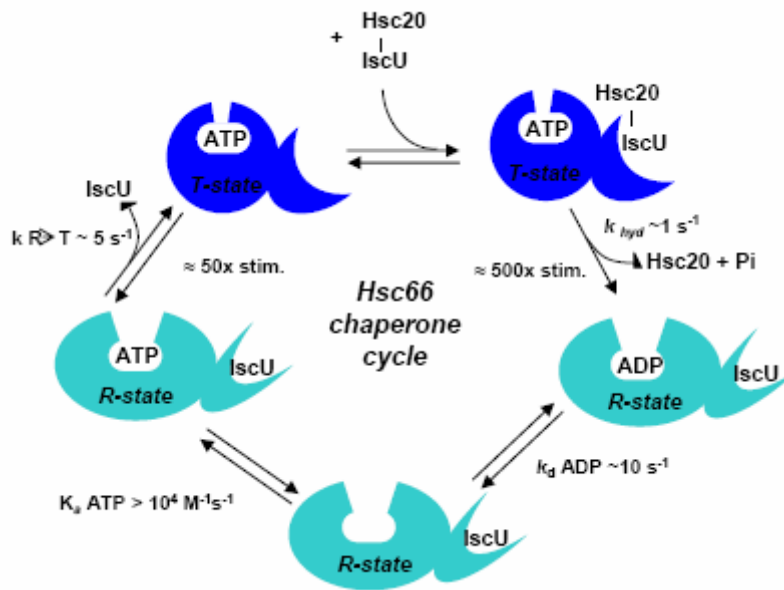
**Figure 6-8: A model for IscU binding cycle of Hsc66. Kinetic constants refer to rates in the presence of IscU and Hsc20 (adapted from [147]) .**

The region of IscU that is specifically recognised by Hsc66 is the conserved LPPVK motif that spans residues 99-103. The synthetic peptide ELPPVKIHC, corresponding to residues 98-106 of IscU, although able to stimulate Hsc66 activity does not show any synergic effect when coupled with Hsc20 and does not show any interaction with the co-chaperone. This suggests that Hsc20 may bind a region of IscU that is not present in small unstructured peptides, and that this interaction is necessary for synergistic stimulation of Hsc66 activity [148].

Although the biological meaning of these findings is still not clear, it is very likely that the Hsc66/20 system facilitates Fe-S cluster assembly on IscU or clusters transfer to acceptor proteins.

Interestingly, it was also found that the Hsc66 and Hsc20 are cold shock proteins, meaning that the expression of their genes is induced in conditions of cold shock. Although mutation of Hsc66 does not affect cold shock survival rates, thus suggesting the presence of a not yet discovered redundant system active in cold shock response, this mutation does affect the profile of many cold shock proteins. It was therefore hypothesised that Hsc66/Hsc20 act as molecular chaperones for the expression or the proper processing of specific proteins following a cold shock, in a manner analogous to the way DnaK/DnaJ act as molecular chaperones in the heat shock regulon [143].

### 6.3.7 Ferredoxin

Ferredoxin is a widespread family of electron transferring iron-sulphur cluster proteins found in every organism, from Archaeobacteria to Eukaryotes. According to the nature of their active centre they show different redox characteristics and are involved in biosynthetic and electron transfer mechanisms [149]. *E. Coli* ISC ferredoxin (Fdx) contains a $[Fe_2S_2]$ centre. Two independent works demonstrated that the cluster's transfer from IscA or IscU to the *apo* form of ferredoxin is the key step for its activation [138,150]. Genetic analyses underlined that the protein covers a crucial role within the mechanism of the clusters' biosynthesis. Bacterial strains lacking Fdx in fact, showed a markedly decreased activity of Fe-S containing proteins. Moreover, it was shown that mutation of the fdx gene caused a phenotype similar to the one provoked by mutations on the hscA and hscB genes, and that depletion of one of the three proteins resulted in loss of function of the other two [120]. The latest finding hints at a co-ordinated function of the three proteins that may work together either in a sequential way or as a complex to promote the cluster's transfer to target apo proteins. In accord with this hypothesis, co-purifications assays identified an interaction between Fdx and Hsc66. The same experiment also revealed an interaction between holo-Fdx and IscS. This may suggest a redox role of Fdx in the reduction of the sulphide ion from $S^0$ to $S^{-2}$, a fundamental step for IscS catalytic role. Consistent with this view, a unique cysteine residue is located on the molecular surface of the protein, near the cluster [151].

Alternatively, Fdx could be involved in the reduction of the iron inside the cells, to prevent precipitation of the ferric ion, or could be important for the release of the cluster from IscU, which requires DTT in *in vitro* experiments [152].

### 6.3.8 YFHJ

Very little is known about the protein YFHJ. The gene that encodes it, *orf3*, is located immediately downstream of *fdx* and it is very likely co-trascribed with the other *isc* genes. Inspection of the nucleotide sequence following *orf3* indicated the presence of a transcriptional termination signal and the next gene, which encodes a nucleotide dikinase, is situated 96 nucleotides away from it [24]. Therefore it seems that *orf3* is the last gene of the *isc* operon. Whether it encodes a protein or is silent, is still unclear. YFHJ belongs in fact to the category of the 'hypothetical proteins', whose genes have been discovered during a structural genomics project but that have not been isolated in living organisms. Although initially it was thought that YFHJ was

110

present only in prokaryotes [24], recently it has been identified also in some eukaryotic species (see paragraph 7.2.1). Its function within the operon has not been clarified yet but YFHJ knock out bacteria do not show a decrease in the concentration of the Fe-S cluster containing proteins [120].

Early functional studies on *isc* proteins suggested that YFHJ could deliver iron to Ferredoxin, which was thought to be the scaffold for the assembly of the Fe-S clusters [153]. This hypothesis has never been investigated further.

An interaction between the protein and IscS was discovered some years ago and the existence of an IscU-IscS-YFHJ complex was suggested [151].

Very recently, a crystallographic structure of the protein, similar to our NMR solution structure, was published [154]. A detailed discussion of the structural features of YFHJ and of the possible functional implications will be given in Chapter 7.

# YFHJ Structure Determination          **7**

The chapter that follows reports the structural analysis of the protein YFHJ by optical and NMR spectroscopies. The process that takes from a blurred image to a well defined picture of a protein structure will be illustrated step by step and the level of information that can be obtained at each step will be discussed. Finally, an attempt to rationalise what the details of the structure can tell us about the function will be made.

## 7.1  *Biophysical techniques*

The main technique used to carry out our study is NMR, which has been exhaustively discussed in chapter 1 and 2 and which will not be addressed once more here. Instead, I would like to describe some principles of CD spectroscopy, a technique that is widely used in biomolecular analyses and that was used in this study as well to gain preliminary information on YFHJ structure.

### 7.1.1  CD spectroscopy

CD spectroscopy is an optical technique extremely convenient to describe polypeptides' secondary and tertiary structures. Because of its extreme sensitivity to structural modifications induced by solvent or mutations, it is extremely suitable for the description of macromolecules' conformations and rearrangements.

The two informative spectral regions in a protein CD spectrum are the region below 250 nm, the so called 'far UV' region, and the region between 250 and 300 nm [155,156].

The far UV region is sensitive to the secondary structure of a protein and is dominated by the three principal electronic transitions of the amidic chromophore: the nπ* transition around 220 nm and two ππ* transitions at ~190 nm and ~140 nm, called respectively NV1 and NV2 [155].

The first transition (which is similar to the carbonylic transition in ketones and aldeides) is electric dipole prohibited but magnetic dipole permitted and produces a weak absorption, with $\varepsilon_{max}$<100. On the contrary, ππ* transitions are electrically and magnetically permitted and have transition dipole moments approximately parallel

(NV1) and perpendicular (NV2) to the C-N bond direction, as schematically depicted in Figure 7-1.

$$m_{n\pi*}$$

$$\mu NV1 \qquad O$$

$$C_{\alpha-1} \qquad \mu NV2$$

$$N \qquad C_{\alpha}$$

$$H$$
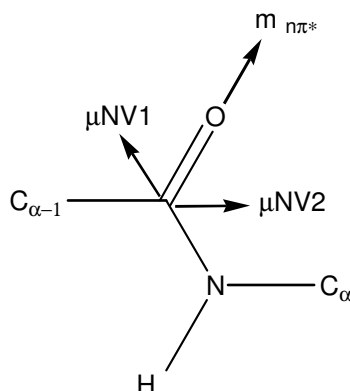
**Figure 7-1: Directions of the transition magnetic dipolar moment $n\pi^*$ and of the transition electric dipolar moment $\pi\pi^*$ in the peptidic group.**

The CD spectrum in the far UV region results from three types of interactions between the above described transitions [155]:

- Exciton coupling of the $\pi\pi^*$ from two different peptidic groups;
- Mixing of the $n\pi^*$ and $\pi\pi^*$ from the same residue, by means of the one electron mechanism.
- Mixing of the $n\pi^*$ and $\pi\pi^*$ transitions localized on two different residues, by means of $\mu$-m coupling

The bands around 190 nm are determined by the first phenomenon and have been described by Moffitt [157].

The intense CD bands observed in correspondence of the prohibited $n\pi^*$ transition arises from the phenomena described at points 2 and 3. A single residue can contribute to the spectrum by means of the coupling of the electrically permitted NV1 transition with the large magnetic dipole moment from the neighbouring residues (the $\mu$-m coupling). In comparison to simple carbonyl groups, the $n\pi^*$ and NV1 $\pi\pi^*$ transitions of peptide groups are much closer in energy. Conjugation of the lone pair of electrons on the nitrogen with the carbonyl $\pi$ electrons leads to a red shift of the NV1 transition, relative to those observed in aldehydes and ketons. The proximity in energy of the electrically allowed NV1 transition to the $n\pi^*$ transition enhances the mixing of the two types of transitions, which can occur within a single peptide group or between neighbouring residues. In the first case, mixing of the $n\pi^*$ and $\pi\pi^*$ states is described as the one-electron effect, and results from the interaction of the peptide group with the static field of the rest of the molecule. Since the peptide group itself

114

has a large permanent dipole moment, and many of the side chains in peptides also contain polar groups, the static coupling or one-electron effect can be appreciable. Mixing of the nπ* transition with the NV1 transition of a neighbouring peptide, the µ-m coupling, can also make a significant contribution to the CD of peptides [155].

Since the described interactions give rise to spectra extremely sensitive to the spatial distribution of the chromophores, the classic protein secondary structures, α helix, β sheet and random coil, can unequivocally be recognised from their CD patterns.

The strongest and most characteristic spectrum is the one of α helix, which has two negative bands of similar amplitude near 222 nm and 208 nm and a stronger positive band near 190 nm. The 222 nm band is due to the nπ* transition while the 208 nm and the 190 nm bands result from the excitonic splitting of the lowest ππ* peptide transition. Except for peptides with a large content of aromatic residues that contribute to the far UV CD spectrum, the CD spectra of α helical polypeptides is largely independent of the side chains and of the solvent. On the contrary, the CD spectra of β sheet structures do not show the same consistency, due to the large variability among the members of this structural class (parallel, anti parallel, intra- or intermolecular β sheets). More important, the β sheet structure itself is not intrinsically chiral, therefore even small distorsions can result in important variations of the spectra. General features can anyhow be identified. A negative band near 217 nm, generated by the peptide nπ* transition, a positive band near 195 nm and a negative band near 180, which arise from the excitonic splitting of the lowest ππ* peptide transition, characterise every β sheet conformation. The intensity of the bands, their ratios and their positions however, are sensibly influenced by the solvent, the side chains and other environmental factors. Finally, CD spectra of unstructured peptides show a typical strong negative band around 200 nm and a weaker band, which can be either positive or negative, at longer wavelengths. The different spectra are shown in Figure 7-2 [155,158].
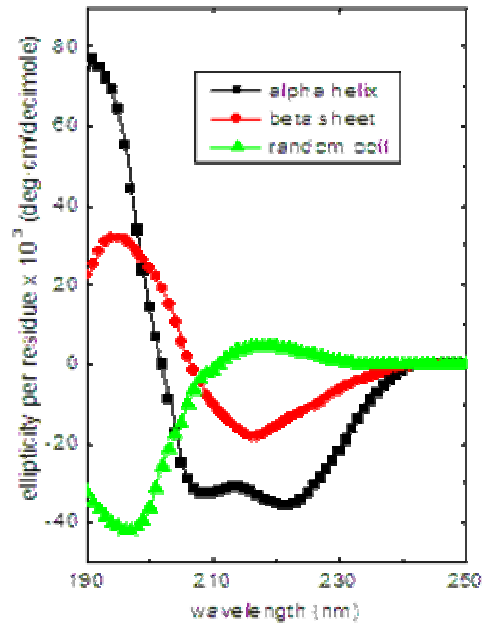
**Figure 7-2: Superimposition of the CD spectra of the three protein secondary structure elements.**

The spectrum of a complex protein is however much more complicated to interpreter. Different regions of a protein can in fact fold into distinct geometries, each contributing to the recorded dichroism. The global spectrum $C_\lambda$ can nevertheless be considered as a linear combination of single component $B_{k\lambda}$, correspondent to the spectra of the secondary structure elements that constitute the biomolecule, as obtained from model polypeptides or proteins with known structure. We have therefore:

$$C_\lambda = \Sigma f_k B_{k\lambda} \tag{7.1}$$

where $f_k$ represents the fraction of the secondary structure K in the protein. In order for such a formula to be valid, we have to assume that [155,156]:

- the contributions from the single structures are additive;
- the solution structures are similar to the ones obtained from the models (i.e. if the model was obtained from a X-ray structure, we have to assume that the solution and the solid structures are similar);
- the non peptidic chromophores from the single residues have no influence on the recorded CD;
- the tertiary structure does not contribute to the spectrum;
- the prototypic structures are perfectly reproducible

Early secondary structure computational methods made use of CD spectra of model polypeptides in specific secondary structures as the 'pure component spectra'

116

derived from a set of proteins with known secondary structures. Least squares based methods were then developed to solve the set of simultaneous equations formed by a reference set of proteins, with known structures and CD spectra, and analyse a new protein CD spectrum. Three of the most used methods are SELCON, CONTINN and Cdsstr [159-161].

The CD recorded in the region 250-300 nm is dominated by the transitions of the aromatic side chains. Disulphide and prosthetic groups may also contribute to the spectrum, when present. Disulphide contributions can generally be distinguished from those of aromatic side chains because the former are much broader [158]. Being mainly reflective of the environment of aromatic groups, near-UV CD is representative of the protein tertiary structure. The small number of aromatic residues and their characteristic bands are advantageous in detecting variations in their environment due to conformational changes in the protein. In many cases it is also possible to assign specific CD bands to individual residues. Near-UV CD has been used to monitor the effect of ligand binding and site mutagenesis and to investigate conformational transitions and protein folding mechanisms. For example, CD bands represented a useful marker for the R-T transition in hemoglobin, or for monitoring the exchange among the insulin conformers [158]. Moreover one can follow the intensity of the near-UV CD bands to identify the presence of molten globules as intermediate in a protein folding process. Finally, from the analysis of the spectrum, information on the relative positions of Phe, Tyr and Trp within the protein can be obtained. According to the generally accepted interpretation of solvent effects in fact, red-shifted signals correspond to residues in a highly polarizable environment, i.e. deeply buried, while blue-shifted bands are generated by solvent exposed side chains. [155,158].

## 7.2 *Results*

### 7.2.1 Primary sequence analysis

YFHJ is a small 66 amino acids protein, of nearly 8 kDa.

A database search using PSI-Blast with default settings and YFHJ as query shows that homologues of YFHJ occur mainly in the α-, β-, and γ-proteobacteria and in the *Cyanobacterium Gloeobacter violaceus* and *Leptospirillum ferrooxidans*. The protein is not universal to the α-, β-, and γ-proteobacteria, and has been lost many times in independent lineages like *Buchnera* and *Ricketssia prowazekii* and *Pasteurella*

*multocida*. Examination of the STRING database shows that the class of proteobacteria in which YFHJ is absent have retained the remainder of the Isc operon, indicating that YFHJ is not absolutely necessary for Fe-S cluster assembly. YFHJ also occurs in specific groups of eukaryotes, such as the Apicomplexa which include *Plasmodium falciparum*, *Plasmodium yoelli yoelli*, *Plasmodium chabaudi*, *Theileria annulata* and *Theileria parva* (Figure 7-3).
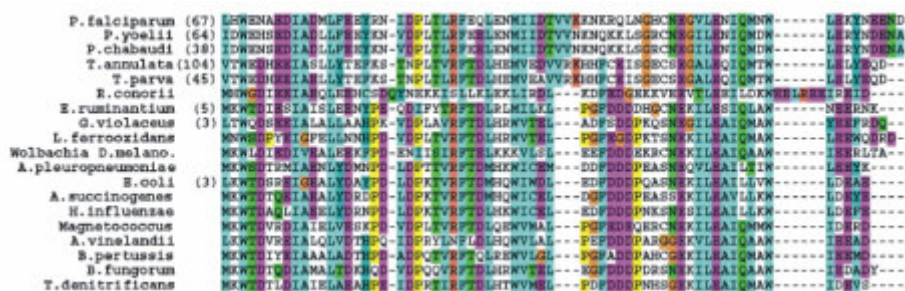


**Figure 7-3: Sequence alignment of YFHJ throughout different species, as provided by the program ClustalX [45]. The colour code depends on the physical characteristics of the residues.**

The sequence presents many conserved acidic residues and contains a high percentage of aromatic residues, 3 tryptophans, 2 tyrosines and 2 phenylalanines. Cys37 is the only cysteine present in the sequence and it is not conserved.

### 7.2.2 Aggregation state

Before starting an NMR investigation of a protein it is very important to assess its aggregation status. The narrow cross-peaks of the preliminary [$^1$H,$^{15}$N] HSQC experiment suggest the presence of a monomeric protein in solution. However we observed the shift of a few resonances during time, which could be partially reverted by the addition of the reducing agent Tris(2-carboxyethyl)phosphine hydrochloride (TCEP). To assess whether this phenomenon was due to the presence of an aggregation process, we measured the correlation time of the protein and run SDS-page and mass spectrometry experiments.

The very low value of the protein correlation time (3.9 ns, see paragraph 7.2.6 for details) is characteristic of a species in a low aggregation state in solution [3]. Moreover, when we run a polyacrilamide gel in non reducing conditions, to prevent the disruption of any possible disulphide bridges, we observed the presence of a low molecular weight protein in solution. Finally, mass spectrometry experiments measured on double labelled samples in the presence and absence of the reducing

118

agent and at different times, show similar features, consistent with the presence of a 8 kDa molecule. From these data we can conclude that YFHJ is a monomer under the experimental conditions adopted for our NMR studies.
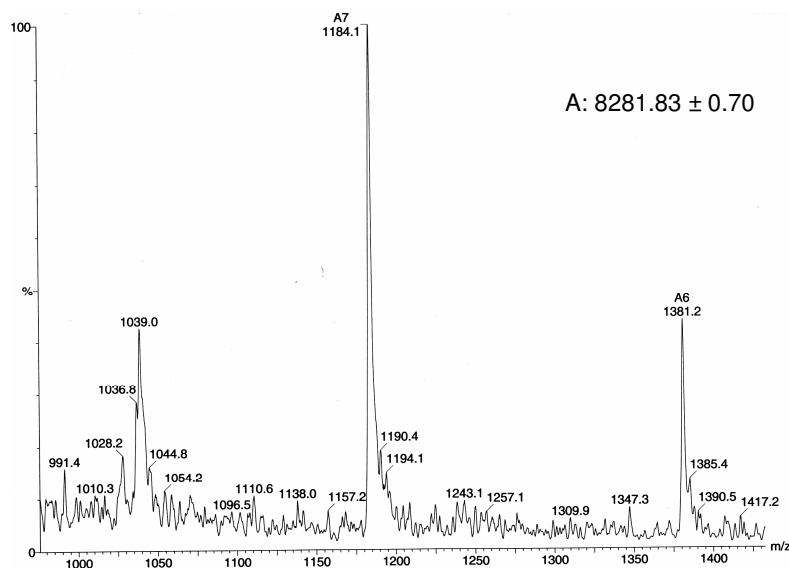


**Figure 7-4: Mass spectrum of a double labelled YFHJ sample in phosphate buffer 10 mM pH 6.5 and NaCl 50 mM.**

### 7.2.3   Preliminary conformational analysis

The conformational analysis of YFHJ was first carried out with low resolution optical techniques.

The far UV CD spectrum recorded between 190 and 270 nm, shows two negative bands, the main one centred at 224 nm and a smaller one at 207 nm. The spectrum is typical for a α helical protein (Figure 7-5).

Deconvolution of the CD spectrum to estimate the fractions of secondary structure elements was done using three different software packages, SELCON, CONTIN and CDsstr [159-161]. Best results were obtained with CONTIN, using two different protein databases, one of which included proteins with poly-proline structures ($PP_{II}$). Estimate of the secondary structures' fractions is reported in Table 7-1.
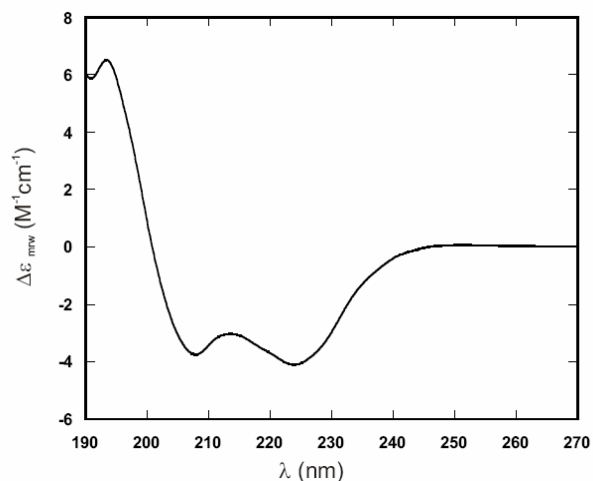
**Figure 7-5: Far UV-CD spectrum of YFHJ. [YFHJ]=203μM, in Tris-HCl pH 8, 150 mM NaCl, 20 mM β-mercapto-ethanol.**

**Table 7-1: Secondary structure elements' content as calculated by the program CONTINN. SP37A and SP43 are the two databases that were used.**

|  | α helix % | β sheet% | Random coil% | Turn% | PP2% | RMSD |
|---|---|---|---|---|---|---|
| CONTINN SP37A | 51 | 3 | 31.3 | 7.7 | 7 | 0.3 |
| CONTINN SP43 | 54 | 4 | 33 | 9 |  | 0.2 |
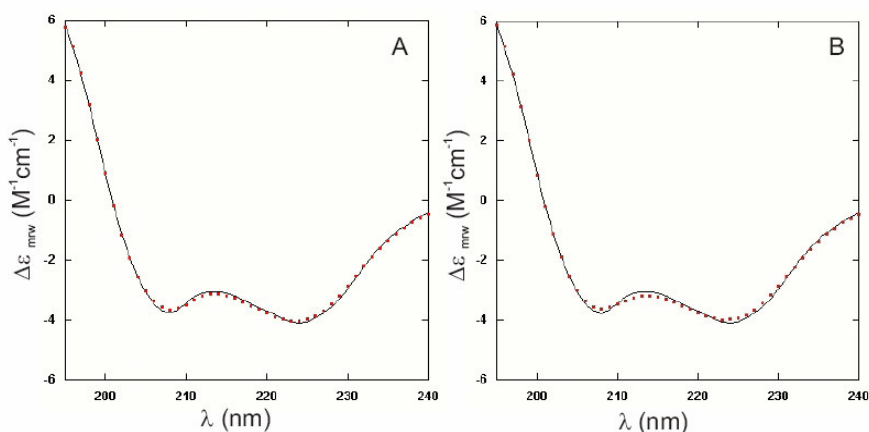


**Figure 7-6: Comparison between the experimental CD spectrum (continuous line) and the calculated spectra (dotted line) as provided by the program CONTINN using the database SP43 and SP37A, A) and B) respectively [161].**

These values however might not be accurate because of the high content of aromatic residues of the protein (see paragraph 7.1.1).

The pH dependence of the CD signal at 225 nm was studied in Tris 20mM, NaCl 150 mM, adjusting the pH with HCl and is shown in Figure 7-7 A and B. It can be seen that the protein adopts different conformations on lowering the pH. The native state is present between pH 8 and pH 6; between pH 5 and 6 the curve seems to indicate the presence of an intermediate state. Below pH 3 the protein is presumably unfolded.

The unfolding process is not reversible, i.e. the protein does not fold back to the native conformation upon addition of NaOH, as shown in Figure 7-7 C and D.
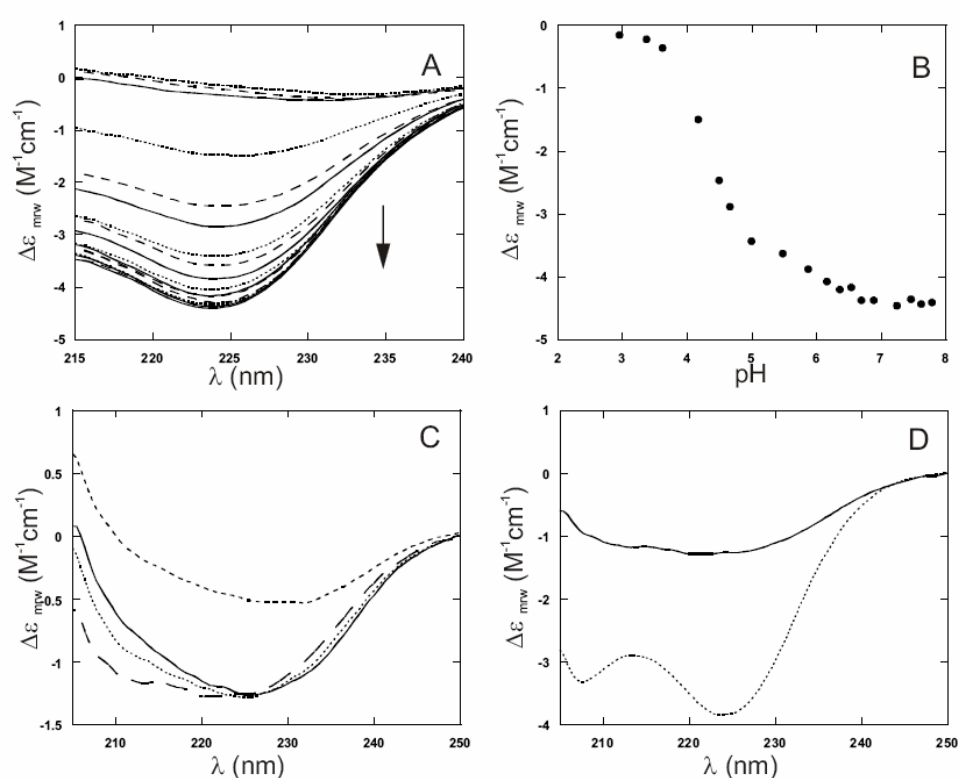


**Figure 7-7: Monitoring the effect of the pH on YFHJ by CD spectroscopy. A) Superimposition of the CD curves recorded at decreasing pH. An enlargement of the region between 215 and 240 nm is reported. The arrow indicates increasing pH. B) Dependence of the CD signal at 225 nm upon addition of HCl to the protein solution. on decreasing pH in. C) Effect of the addition of NaOH to the unfolded YFHJ followed by CD. The curves were recorded respectively at pH 3.86 (----), 4.50 (——), 5.49 (·····) and 5.88 (— —). D) Comparison of the CD curve of YFHJ at pH 5.89 in the direct titration (·····) and at pH 5.89 in the inverse titration (——). All the experiments were recorded 20 mM Tris-HCl, 150 mM NaCl at 20 °C, with an initial protein concentration of 3 µM.**

The 1D spectrum of YFHJ is shown in Figure 7-8. It is typical of a well folded and non degraded protein; the presence of ring-current shifted peaks around 0 ppm, which arises from spatial proximity of hydrophobic residues to aromatic rings, is characteristic of proteins with a well defined hydrophobic core.
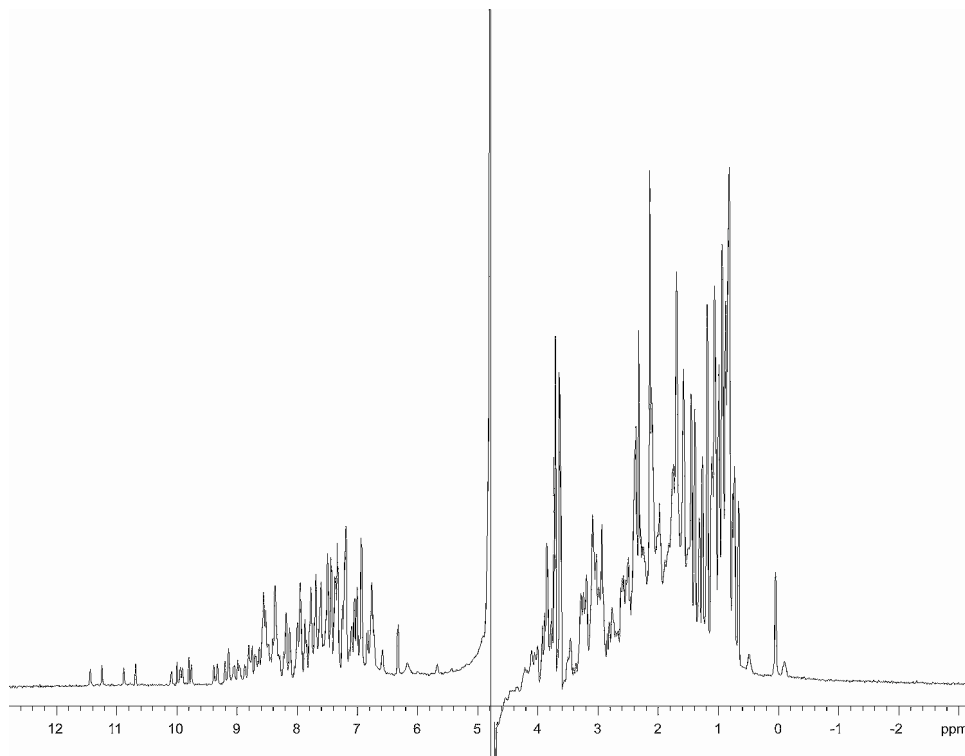


**Figure 7-8:** [1]**H monodimensional NMR spectrum of YFHJ recorded at 30 °C, at 500 MHz. [YFHJ]=0.6 mM, 10 mM phosphate buffer pH 6.5, 50 mM NaCl.**

Evidence for the presence of a well folded structure is also provided by the excellent dispersion of the peaks in the [^{1}H,^{15}N] HSQC spectrum shown in Figure 7-9.

### 7.2.4 Spectral assignment

Once we analysed the preliminary NMR spectra, we proceeded to the assignment procedure, which allowed us to uniquely link every signal of the spectra to a specific atom of the protein. The sequential assignment allows identification of the backbone connectivities alone while the side chains of every residue are separately assigned.

122

## 7.2.4.1    Sequential assignment

Complete assignment of the $^1$H, $^{15}$N, $^{13}$C backbone resonances was obtained using CBCA(CO)NH, CBCANH, HNCA, HN(CO)CA, HNCO recorded at 600 MHz. From the CBCA(CO)NH, CBCANH, and  HNCO we assigned all backbone atoms, with the exception of prolines 20, 24 and 46. These residues were initially identified from their cross peaks in the $^{15}$N NOESY-HSQC and consequently assigned by means of the HCCH-TOCSY. The presence of the typical NOESY cross-peaks between their δ protons and the α protons of the preceding residues confirmed that they are all *trans* prolines [3].
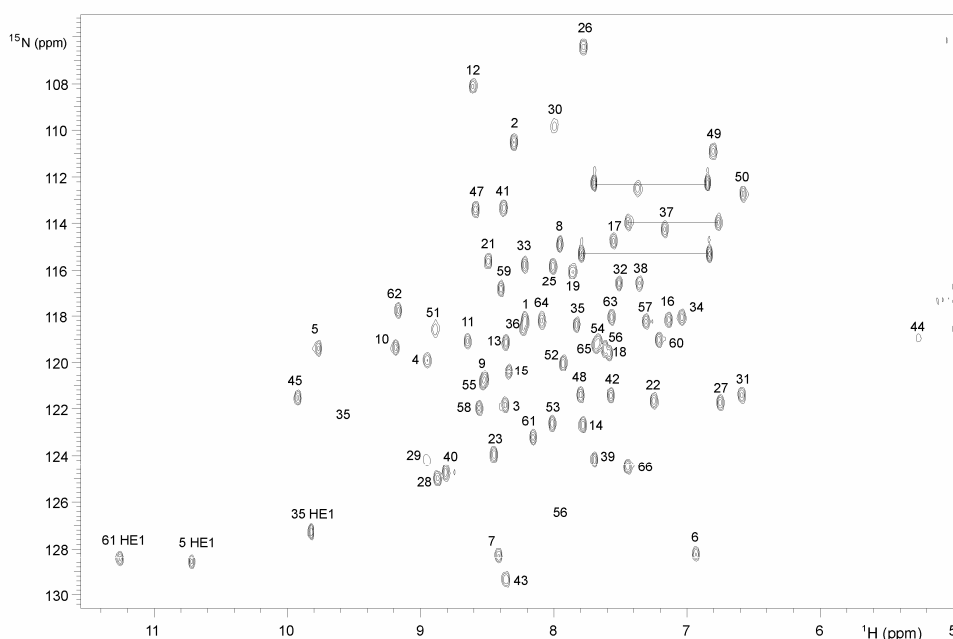


**Figure 7-9: [$^1$H,$^{15}$N] HSQC of YFHJ, recorded at 303 K, 500 MHz. [YFHJ]=0.6mM, phosphate buffer 10 mM, pH 6.5, NaCl 50mM.**

## 7.2.4.2    Side chains assignment

Proton resonances of the side chains from the aliphatic residues were assigned with $^{15}$N TOCSY-HSQC and $^{15}$N NOESY-HSQC spectra, recorded respectively with 70 ms and 100 ms mixing time at 600 MHz. Amide protons of residues 29, 30, 51 were not detected in these spectra probably because they are in exchange with the solvent or they are close to an aromatic residue. Their $H_N$ was anyway determined during the sequential assignment and can be observed, although weakly, in the [$^1$H $^{,15}$N] HSQC. Complete spin system recognition was achieved using a HCCH-TOCSY recorded in

$D_2O$ with 70 ms mixing time. Being more sensitive than the corresponding ^{15}N spectrum, the HCCH-TOCSY allows a more precise assignment of the proton side chain resonances (e.g. the δ and γ protons of Leu or Ile are collapsed in the ^{15}N TOCSY-HSQC and well resolved in the HCCH-TOCSY). This experiment was recorded in $D_2O$ to prevent covering of the signals from the α-protons by the intense water signal. With this experiment we also assigned the Cα and Cβ atoms that we could not previously assign. $[^1H,^{13}C]$ HSQC was used mainly to check the carbon assignment.

The aromatic side chains were assigned using (Hβ)Cβ(CγCδ)Hδ and (Hβ)Cβ(CγCδCε)Hε experiments [39] recorded at 600 MHz and HCCH-TOCSY specific for the aromatic region recorded with 30 ms mixing time at 500 MHz. 63 out of 66 well resolved non proline backbone amide resonances are observed in the $[^1H$ $^{15}N]$ HSQC. The NMR assignment of backbone ^{1}H, ^{13}C and ^{15}N is complete for all residues, with the exception of the C' of residues followed by prolines and those of residues 28 and 29. The assignment of side chain ^{1}H, ^{13}C and ^{15}N aliphatic and aromatic groups is almost complete. The NεHε of only one of the two expected Arg is visible in the HSQC, presumably that of Arg28, whereas the guanidinium groups are not visible. The $^{15}NH_2$ of Asn and Gln and the indoles of the Trp are assigned by means of the NOEs correlations with their side chain protons, while the $NH_{δ1}$ of His exchanges with the solvent and is not visible.

The $εCH_3$ proton of Met 32 has been identified from its chemical shift in the $[^1H, ^{13}C]$ HSQC and from its NOESY pattern in the $[^1H, ^{13}C]$ NOESY-HSQC spectra. The amides of the first two N terminal residues, Gly and Ala introduced for cloning reasons, were not assigned, though their ^{1}H and ^{13}C backbone and side chain atoms were identified in the HCCH-TOCSY and ^{13}C edited NOESY-HSQC.

The complete assignment of YFHJ was deposited at the BMRB web site and is reported in appendix A, table I.

### 7.2.5 Secondary structure prediction

Details of the local backbone geometry were obtained by an extension of the assignment process. Secondary structure elements of YFHJ were identified using a combination of data, including medium-range and long-range NOEs, $^3J_{HNHα}$ coupling constants [32], torsion angles estimated with the Talos approach [49] and secondary chemical shift analysis [47,48].

Strong correlations between $H_{Ni}$ and $H_{Ni+1}$ in the [$^1$H$^{15}$N] NOESY-HSQC are diagnostic of a α helical conformation, while correlations between $H_{Ni}$ and $H_{\alpha i+1}$ arise from residues in β sheet conformation [3]. Typical α helical NOE patterns can be distinguished for residues 5-19, 31-42, 47-66, indicating the presence of three helices. An example of the correlation pattern is given in Figure 7-10.
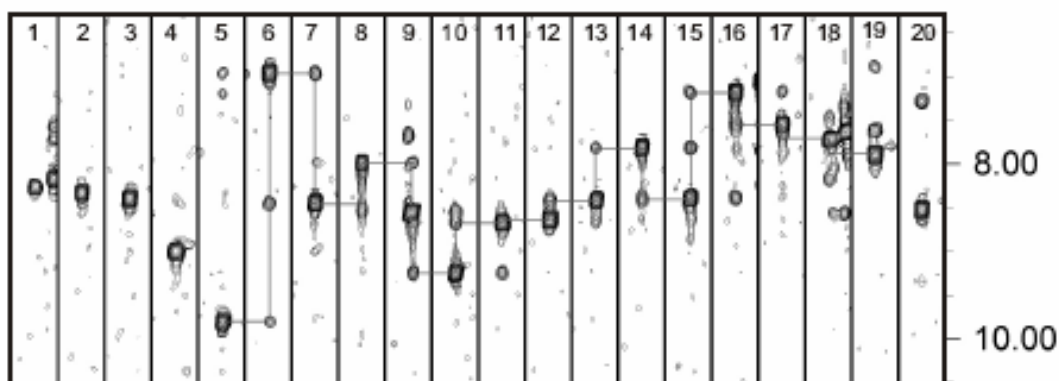


**Figure 7-10: Strips from residues 1-20 taken from a [$^1$H,$^{15}$N] NOESY-HSQC spectrum. For residues 5-19 the typical α-helical connectivities are drawn.**

Torsion angles represent additional parameters that can be exploited to gain information on the secondary structure arrangement of the residues.

The values of the φ torsion angles are related to the homonuclear three bond J (H$^N$H$^\alpha$) coupling constant by the Karplus equation. In turn, J couplings can be easily obtained from the HNHA spectrum. Here the magnetisation is transferred from the H$^N$ proton of a residue to its own H$_\alpha$. The spectrum will thus show for each H$^N$ the corresponding H$^\alpha$ cross peak. The ratio between the volumes of the diagonal peak and the cross-peak is related to the $^3$J(H$^N$H$^\alpha$) coupling constant by the equation:

$$Scross/Sdiag = -\tan^2\left(2\pi J_{HNH\alpha}\zeta\right) \tag{7.2}$$

Where $S_{cross}$ and $S_{diag}$ are the intensities of the cross peak and of the diagonal peak signal respectively, J(H$_N$H$_\alpha$) is the desired coupling constant and ξ is a fixed delay in the pulse sequence [32].

Given the shape of the Karplus curve, as many as four different conformations can yield the same J(H$_N$H$_\alpha$); the dihedral angles are therefore given as a range of values compatible with the measured coupling constant. In particular, we used for J>8 φ =-

120 ± 50, for 7<J<8 φ = -120 ± 60, 6<J<7 φ =-100 ± 80 and J<5.5 φ =-60 ± 30 [162]. Some residues did not give strong enough signals to calculate J couplings. From the J and φ values obtained for YFHJ, which are listed in appendix A, table II, we can assume that residues 8-9-11, 13-17, 32-37, 39-40, 50-66 are in a α helical conformation.

Values of torsion angles can also be predicted by TALOS on the basis of the assigned chemical shifts [49]. Accordingly, residues 8-18, 29-37, 51-60 and 62-64 are predicted in a helical conformation, while residues 2-7 and 43-46 are in an extended conformation. For all the other residues the prediction is considered unreliable. TALOS results are summarised in appendix A, table III.

Further information on the secondary structures can be gained from the secondary chemical shifts, that can be calculated using the program CSI, already described in paragraph 3.2.1, or through the more compact formula:

$$\frac{\left(4\Delta\delta H_\alpha + \Delta\delta C_\beta - \Delta\delta C_\alpha - \Delta\delta C'\right)}{n} \tag{7.3}$$

The Δs in the numerator are the differences between the chemical shift of the $H_\alpha$, $C_\beta$, $C_\alpha$, C' of each residue respect to the random coil value, while n is the number of assignments.

The advantage of the second method is that it takes into account the secondary chemical shift of every atom type at the same time and gives a weighted consensus value ΣΔδ.

The secondary chemical shifts are plotted against the residue number in Figure 7-11. It has to be noted that in the top histogram, where we used the traditional secondary chemical shifts, the negative bars indicate the presence of helical elements, whereas in the bottom one, α helices correspond to positive bars. Apart from this formal aspect, the results obtained in the two cases are very consistent with each other: the helical elements are comprised between residues 8-18, 29-37, 51-59, 61-63, while residues 43-46 are in an extended conformation.

In conclusion we can deduce the distribution of the secondary elements along the sequence from the residues' chemical shifts. The methods that we used agree on finding helices between residues 8-18, 29-37 51-59 and 62-63. However, according to TALOS residues 60 and 64 are in a helical conformation as well, while according to CSI residue 61 is helical. It is reasonable to believe that residues 51-64 form a long helix.

126

Figure 7.12 is a summarising graph correlating the primary sequence to the secondary structure elements, as predicted with the above described methods.
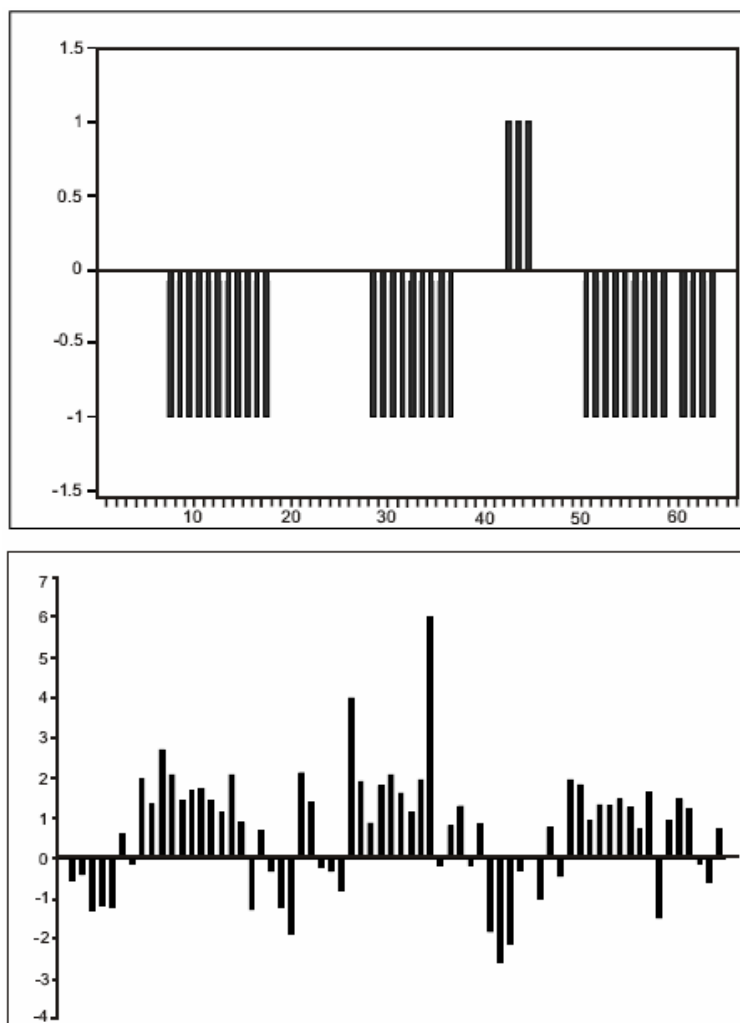


**Figure 7-11: Distribution of the secondary structure elements as computed from the secondary chemical shifts. The histogram on the top represents the consensus bars as provided by the program CSI [47]. Negative bars indicate helical elements. In the histogram on the bottom, a weighted consensus value ΣΔδ, calculated according to equation 7.2, is plotted against the amino acid sequence. Here positive bars indicate helical elements.**
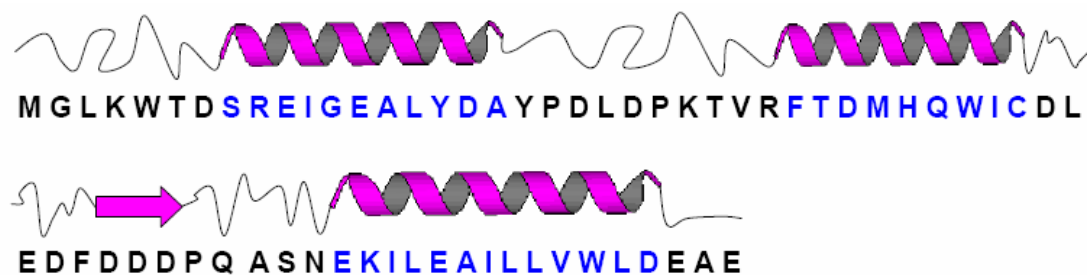


**Figure 7-12: Schematic representation of the distribution of the secondary structure elements as resulted from Talos, CSI and $^{3}J_{HN H\alpha}$ couplings (see text for details).**

### 7.2.6 Relaxation and NOEs

The dynamic behaviour of the protein's backbone was analysed measuring the ^{15}N relaxation times and the steady state NOEs, as described in the Material and Methods section.

For most residues the ^{15}N relaxation values are rather uniform, with average $T_1$=380±45 ms, $T_2$=280±45 ms, noe=0.576. Large deviations are observed for Met1 and Gly2, which have high values of $T_1$ (722 ms and 607 ms respectively) and negative values of NOE (-0.666 and -0.329). Residue 1 also has a $T_2$ of 328 ms, slightly higher than average. Small deviations can be observed for Glu66 that has $T_1$=286 ms, $T_2$=521 ms and noe=0.243, for Leu54, with a $T_1$=257 ms, for Ala18 with $T_1$=258 ms and Tyr19, $T_1$=276 ms. Overall however, the fairly constant trend measured for the dynamic parameters indicates that the protein is compact.

Assuming isotropic tumbling of the protein in solution, we applied the model free approach to calculate the global rotation correlation time $\tau_c$ [163]. The value that we calculated is 3.9 ns, consistent with a low molecular weight protein.

The $T_1$, $T_2$ and NOEs values for each residue are summarised in the graphs presented in Figure 7-13.

### 7.2.7 3D structure

The calculation of the 3D structure of the protein was accomplished using the program ARIA 1.2, described in paragraph 3.3.1 [51,54-56,164]. The method is automated and iterative. The input files used by ARIA consist of a list of peak volumes of NOESY-like experiments, the list of chemical shifts' assignments and, when available, a list of angular and hydrogen bond restraints. Proton distances are obtained from the peaks volumes, according to equation 7.4 and a suitable distance calibration.

$$r_i = r_{ref} \left( V_{ref} \middle/ V_1 \right)^{1/6}$$

(7.4)

The program performs partial assignment, calibration, violation analysis and merging of assigned data from different sources to calculate the final structure.
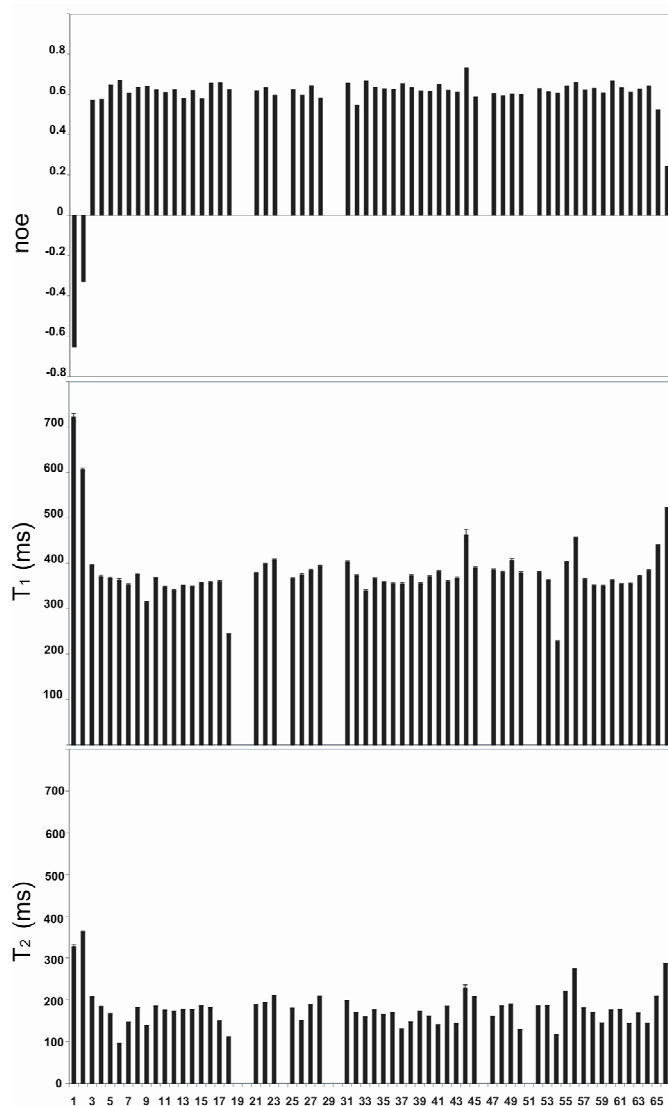
**Figure 7-13: From the top: steady state NOEs, $T_1$ and $T_2$ plotted against the amino acid sequence of YFHJ.**

We obtained the distance restraints from 3D $^{15}$N and $^{13}$C NOESY-HSQC spectra; 61 angular $\varphi$ and $\psi$ restraints were obtained from TALOS [49] and 50 $\varphi$ dihedral angles evaluated from HNHA spectra [32]. We did not use a list of hydrogen bonds. For each NOESY-like experiment, the peaks were previously peak-picked and provided as a list of cross-peak volumes identified by their frequencies. Cross peaks from $^{15}$N NOESY-HSQC spectra were almost completely manually assigned, while the correspondent $^{13}$C NOESY spectrum was only partially assigned. In particular, regions where we observed severe overlapping were left unassigned. As already described, ARIA can in fact perform ambiguous assignments when the peaks' overlap is too severe, and can take into account all the possible assignments by weighting them with a specific function during the structure calculation process.

129

A total of 8 iterations were performed in order to achieve the final refined structure. In each ARIA iteration, 20 structures were calculated by simulated annealing using the standard CNS protocol [50]. The first bundle of structures generated by iteration 0 was obtained using a high violation tolerance, 1000 Å, and an ambiguous cutoff of 1.01 ppm, so that no NOE distance was excluded from the calculation. Both parameters were progressively narrowed in the following iterations, in order to exclude restraints which were not significant for the final structure. Floating assignment for prochiral groups and correction for spin diffusion during iterative NOE assignment were applied [54]. At the end of each iteration the 7 best structures in term of lowest global energy were selected and used for assignment of additional NOEs during the following iteration. A restraint was not used in the next iteration if it was violated by more than a target value in maximally 50% of the seven structures.

In iteration 0, 1117 ambiguous and 1144 unambiguous restraints were calculated. In the final iteration the program used 1699 distance restraints, of which 1583 were unambiguous and 116 ambiguous. A total of 61 $\psi$ and 83 $\varphi$ rotation angles were used as dihedral restraints in the last iteration. In the final iteration 100 structures were calculated and the best 20 were refined in water. Table 7-2 summarises the number and nature of the restraints used in the structure calculation and the quality parameters that characterise the 20 structures. It is important to note that none of them have distance violations greater than 0.5 Å or dihedral violations greater than 5°. As shown in Figure 7-14 A, the structure ensemble is well defined. The precision of the structures, defined by the rmsd (root mean square deviation) with respect to a putative mean structure, is 0.70 ± 0.17 Å for the backbone and 1.41 ± 0.19 Å for the side chain atoms. A ribbon representation of the lowest energy structure is also reported in Figure 7-14. The program MolMol was used to visualize the structures and to calculate the rmsd [165].

The solution structure can be superimposed to the crystallographic one, released in the PDB few months before completion of the NMR calculation process [154], with an rmsd of only 1.2 Å calculated on all atoms, thus indicating the good quality of our results (Figure 7-15).
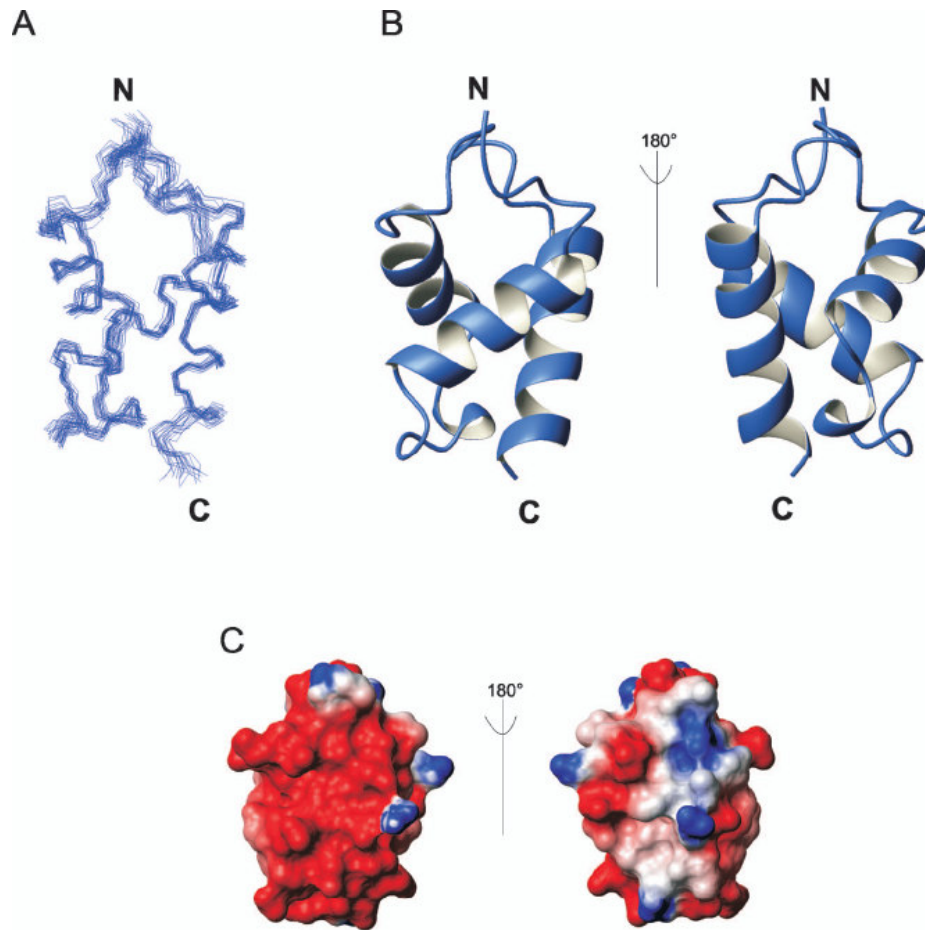
**Figure 7-14: 3D structure of the protein YFHJ. A) Bundle of the 20 lowest energy structures calculated by the program ARIA 1.2 B) Ribbon representation of the NMR structure of YFHJ C) Surface distribution of the electrostatic potential.**
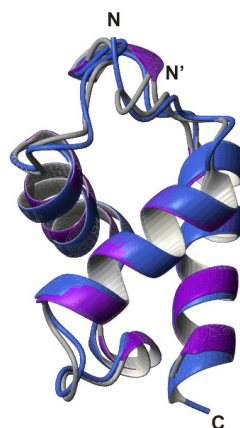


**Figure 7-15: Superimposition of the solution with the crystal structure of YFHJ, shown in blue and purple respectively. The N terminal of YFHJ's crystal structure is indicated in the picture as N'.**

### 7.2.8 Structure Validation

The last step of the structure calculation process is the structure validation, which was accomplished with the programs Whatif and Procheck [57,59].

We run a fullcheck Whatif quality control on the best 20 structures obtained after the last ARIA iteration. The packing parameters are reported in Table 7-2.

As for Procheck, the most important parameter is the Ramachandran plot. The program measures the dihedral angles of each residue in the 20 structures and plots them in the Ramachandran map. As reported in Table 7-2 there are no residues in the disallowed region of the plot, only 1.9% of the residues are in a generously allowed region, 12.5 % in additional allowed region and 85.7% in the most favoured region. The philosophy of Procheck program is illustrated in paragraph 3.3.2.1; the Ramachandran plot for the ensemble of calculated structures can be seen in appendix A.

The structure coordinates were deposited at the Protein Data Bank website with the pdb code 2BTZ.

### 7.2.9 Dali

In order to extract functional information from the spatial folding of YFHJ, we looked for structural homologues of the protein, using the program Dali, described in paragraph 3.4.1 [60].

The results showed that YFHJ is structurally related to DNA binding domain of eukaryotic proteins, containing the characteristic helix-turn-helix motif involved in DNA recognition [166].

The entire list of proteins or domains found to share the folding with YFHJ is presented in table IV, appendix A.

In Figure 7-16, the superimpositions of YFHJ with four representative domains are presented. 1iuf-A is the PDB accession code for the DNA binding domain of the centromere abp1 protein fragment [167], while 1h40 refers to the FF domain of human HYPA/FBP11 [168]. FF domains are known to bind to the carboxy-terminal-domain of RNA Polymerase II, and are involved in RNA metabolism. It is interesting to observe that the protein shares high structural similarity with the DNA binding domain of the protein Genesis (PDB:2hfh), a HNF3 homologue [169], and with the DNA binding domain of the human protein DEK (PDB:1q1v) [170]. They both are helix-turn-helix motifs belonging to the winged helix family, a class of transcription factors very important for mammalian development. To identify putative binding

regions, we aligned the structurally superimposable tracts of sequences of the considered proteins. The alignment reported in Figure 7-17, shows that the proteins have no common residues.

**Table 7-2: Structural statistics for YFHJ**

| Final NMR restraints | |
|---|---|
| Total distance restraints | 1699 |
| Unambiguous/ambiguous | 1583/116 |
| Intraresidue | 775 |
| Sequential | 281 |
| Medium (residue i to i+j, j=2-4) | 230 |
| Long-range (residue i to i+j, j>4) | 413 |
| *Dihedral angle restraints* | |
| Φ | 83 |
| ψ | 61 |

| Deviation from idealised geometry | |
|---|---|
| Bond lengths (Å) | 0.001± 0.000 |
| Bond angles (º) | 0.263 ± 0.024 |
| Improper dihedrals (º) | 0.114 ± 0.004 |

| Restraints violations | |
|---|---|
| Distance restraint violation > 0.5Å | 0 |
| Dihedral restrain violation > 5º | 0 |

| Coordinate precision (Å) with respect to the mean structure | |
|---|---|
| Backbone of structures regions | 0.70 ± 0.17 |
| Heavy atoms of structures regions | 1.41 ± 0.19 |

| Whatif quality check | |
|---|---|
| First generation packing quality | -0.189 |
| Second generation packing quality | -0.632 |
| Ramachandran plot appearance | -0.241 |
| χ1-χ2 rotamer normality | -0.368 |
| Backbone conformation | -0.076 |

| Procheck Ramachandran statistics | |
|---|---|
| Most favoured region | 85.7% |
| Additional allowed region | 12.5% |
| Generously allowed regions | 1.9% |
| Disallowed regions | 0% |
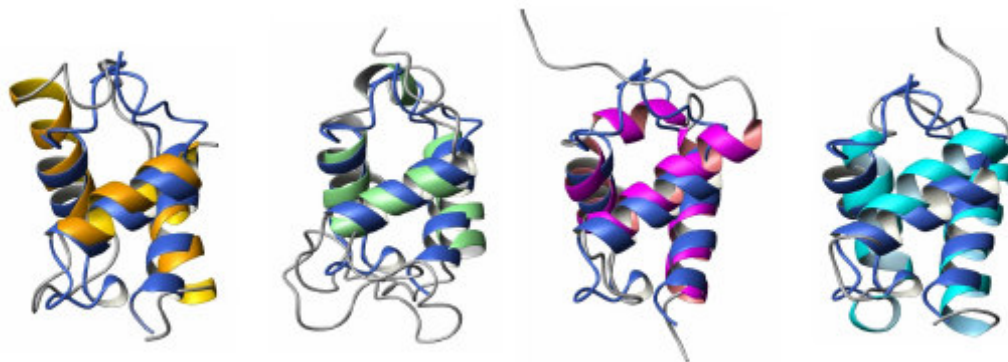
**Figure 7-16: Superimposition of YFHJ with structurally related proteins. From left to right: C-terminal domain of the DNA binding protein Apb1 (PDB code: 1iuf, RMSD=2.4Å) [167], FF-motif from human HYPA-FBP11 (PDB code: 1h40, RMSD=2.5Å) [171], protein Genesis (PDB code: 2hfh, RMSD=2.9Å) [169], C-terminal domain of human protein DEK (PDB code: 1q1v, RMSD=2.7Å) [170].**



**Figure 7-17: Sequence alignment of YFHJ with the proteins 1iuf, 1h40, 2hfh and 1q1v, as provided by Dali according to the structural superimposition.**

## 7.3 *Discussion*

The bacterial protein YFHJ is a 66 amino acids hypothetical protein whose gene has been found in the Iron Sulphur Cluster operon of *E. Coli* [24]. Its role within the iron sulphur clusters' biosynthesis mechanism has not been clarified yet.

To gain functional information, we undertook a structural study of the protein by means of CD spectroscopy and NMR.

According to NMR, mass spectrometry and ultracentrifugation data, YFHJ is a monomer in solution. In our first NMR [$^1$H, $^{15}$N] HSQC spectra we noticed a slight rearrangement of the peaks with time, which lasted a couple of hours. Initially we attributed this phenomenon to the formation of a dimeric species. Upon addition of the reducing agent TCEP in fact, we saw that some peaks were shifted back to their

134

initial positions. However mass spectrometry analyses on samples of different age, with and without TCEP, did not detect any dimeric species, nor did polyacrilamide gel run in non reducing conditions. Moreover ultracentrifugation measurements on the protein carried out by Dr V. Rybin at European Molecular Biology Laboratory in Heidelberg were consistent with the presence of a monomeric protein. It is possible that the initial rearrangement that we detected by NMR derives from small variations of the solution conditions during time, which affect particularly sensitive residues.

Apart from what I have just described, YFHJ is a stable protein. It can be easily frozen and once it has been thawed it can survive in the fridge for a period of several months, although the doubly labelled protein used for NMR studies has an intrinsic lower stability.

The thermal unfolding curves, measured with CD spectroscopy in different solution conditions (see Materials and Methods section for details), show that YFHJ undergoes a reversible unfolding process with temperature. The melting point was calculated in the range 47-55°C. Conversely, the unfolding process induced by decreasing the pH of the solution is not reversible.

Preliminary CD and NMR spectra are consistent with a well folded protein, mainly helical.

The 3D NMR structure of YFHJ confirms that the protein comprises three helices, $\alpha_1$ from residue Ser8 to Ala18, $\alpha_2$ from residue Phe29 to Asp38, $\alpha_3$ from residue Glu51 to Ala65. $\alpha_1$ and $\alpha_3$ are antiparallel to each other while the second helix is at almost 90° respect to the others (Figure 7-14). The three helices form a compact core which is flanked by long unstructured regions: loop 1 between $\alpha_1$ and $\alpha_2$ is ten residues long, while loop 2 between $\alpha_2$ and $\alpha_3$ is formed by twelve residues.

The protein has an unusual high content of Trp, 3 out of 66. Trp35 stacks against Tyr19 and their aromatic rings are relatively accessible to the solvent while Trp61 and Trp5 are buried and contribute to the hydrophobic core. However the $H_\delta 1$ of Trp5 is exposed. Trp61 gives NOEs with Leu15, Leu22, Met32, Pro24, Tyr16, Leu58, Tyr19, Val27, while Trp5 interacts with Ile11, Ile36, Ile53, Leu54, Ile57. The relative position of helix $\alpha_1$ and $\alpha_3$ is confirmed by a large number of NOEs between Ser8 and Ile57, Arg9 and Val60, Gly12 and Val60 and Gly12 and Ile57, Tyr16 and Glu64. Contacts between $\alpha_2$ and $\alpha_1$ and between $\alpha_2$ and $\alpha_3$ are provided by NOEs between Trp35 and Tyr19 and between Met32 and Leu54 respectively.

From the observation of the structure, we can now explain some peculiarities which we encountered during the assignment process. The amide proton of Asp44 points directly towards the aromatic ring of Phe42 and it is spatially close to the indole of Trp5, thus explaining its particularly high field shifted frequency. The same

explanation can justify the absence of the amide signal of Phe29 in the [$^1$H, $^{15}$N] HSQC; the amide proton in fact is right above the phenyl ring and it is therefore heavily influenced by the ring current shift. The $H_N$ signal of Tyr30 and Glu51 are very weak, possibly because of an exchange phenomenon or because they are influenced by the chemical shift anisotropy generated by the ring current of Phe29, which is spatially close to both residues. The only Cys present in the protein is partially exposed, thus possibly forming disulphide bridges. However, since it is not conserved, it is unlikely that it plays an important functional role.

The structure is compact and the measure of the $^{15}$N relaxation times and steady state NOEs confirmed that only the very end of the N terminus represents a dynamic region. The C terminus is in fact embedded into a rigid helical element and the loops themselves are not flexible.

The analysis of the conserved residues can give information on the relative importance of each amino acid in the protein. Conserved non polar residues are normally found in the internal region of the protein, where they have a structural role, being part of its hydrophobic core. Trp61, Ile11, Leu15, Leu54, Gly12, Met32, Ile36, Ile57 belong to this class of residues. While Trp61, Ile11, Leu54 are extremely conserved, Gly12, Met32, Ile36, Ile57 are only partially conserved, being substituted in some organisms by other hydrophobic residues, such as Ala, Leu, Val. Trp5, another well conserved residue, contributes as well to building the internal architecture but its aromatic ring is partially exposed. The presence on the surface of hydrophobic conserved residues such as Leu3, Leu39, Phe42, Phe29 and Val27 is more puzzling. The distribution of the above mentioned amino acids does not define a hydrophobic surface, which could have represented an interaction site with other proteins. However, having resisted to evolutionary pressure, these residues must be important for the protein's function. Moreover, several conserved and semi-conserved charged residues are exposed: Asp7, Asp44, Glu51, Glu55, amongst the extremely well conserved residues and Asp17, Asp21, Asp23, His33, Asp43, Asp45 and Asp63 amongst the semi-conserved amino acids are superficial, presumably covering an important functional role (Figure 7-18).

The high content of negatively charged residues determines the very low protein's PI (3.7). The electrostatic potential of the protein plotted onto its surface shows that the negative charges are mainly distributed on one side of the protein, while on the opposite side a neutral/basic region can be observed (Figure 7-14). A possibility is that both surfaces are functionally relevant. Binding of YFHJ with a basic protein might activate its function, which in turn may be accomplished through the neutral/basic cleft.
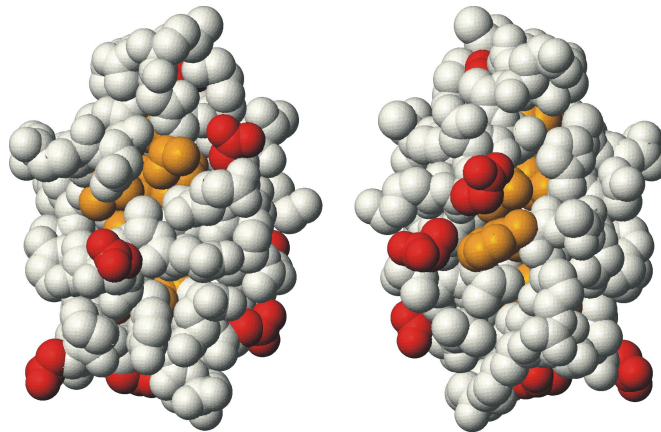
**Figure 7-18: Distribution of the conserved and semi-conserved residues on the surface of YFHJ. Acidic and hydrophobic residues are displayed in red and orange respectively.**

Similarities amongst the 3D structures of two or more proteins often indicate that the proteins share a common function. This is the principle that inspires the program Dali, a software package that searches the PDB for protein structurally related to the query in order to obtain functional information. In many cases however, the results provided by the program can be misleading. A Dali search with YFHJ as query for example, indicates that the protein has a fold remarkably similar to that of DNA binding domains, in particular to the Helix-Turn-Helix motif of larger proteins. Because of the acidic nature of the protein, though, it seems extremely unlikely that it can encounter a DNA molecule *in vivo* as it is, since the repulsion energy between the negative phosphate backbone of the nucleic acid and the negatively charged surface of YFHJ would be too high to overcome.

Most likely the observed similarity reflects an adaptation of the HTH fold of YFHJ to a new function, rather than a direct role in DNA binding. The poor sequence alignment of the structurally superimposable tracts shown in Figure 7-17 seems to support this hypothesis as well.

## 7.4 *Materials and Methods*

### 7.4.1 Expression and purification of YFHJ

Doubly and singly labelled YFHJ were kindly prepared by John McCormick, from National Institute of Medical Research in London.

### 7.4.2 CD measurements and fluorescence

The far-UV CD spectrum was recorded on a Jasco J-715 spectropolarimeter at 20°C, using a round quartz cuvette of 0.01 cm path length. The measurement was carried out in 20 mM Tris-HCl, pH 8, NaCl 150 mM, β-mercapto-ethanol 20 mM, using a protein concentration of 203 μM.

Direct pH titration was carried out upon addition of HCl 1-0.1 M to a sample initially 3 μM dissolved in 20 mM Tris-HCl, pH 8 and 150 mM NaCl, using a quartz cuvette of 1 cm path length. We recorded 18 points, from pH 7.82 to 2.95. The reverse titration was carried out adding NaOH 0.1-1 M to a protein solution at pH 3.95. In this case we recorded only 4 points.

For each addition we recorded the CD spectrum between 250 and 205 nm and monitored the signal at 225 nm. Spectra are averages of four scans at a scan rate of 20 nm/min; each spectrum was baseline corrected by subtraction of the buffer spectrum and smoothed.

### 7.4.3 NMR data acquisition and processing

Spectra for NMR assignment and structure determination of YFHJ were acquired using $^{15}$N and $^{13}$C uniformly labelled samples at 0.6 mM concentration in 90% $H_2O$ and 10% $D_2O$ in 10 mM phosphate buffer (pH 6.5), 50mM NaCl, 0.05% sodium azide.

All the spectra were recorded at 30°C on Varian Unity, Unity-plus, and Inova instruments operating at 500, 600 and 800 MHz proton frequencies equipped with 5 mm triple-resonance probes. The HNCA, HNCOCA and HNCO were recorded on a Bruker Avance spectrometer operating at 600 MHz, equipped with a 5 mm triple-resonance cryoprobe. The WATERGATE sequence was used for water suppression [37]. Two-dimensional $^1$H-$^1$H TOCSY and NOESY and heteronuclear $^{15}$N TOCSY and NOESY-HSQC [3,172], $^{13}$C HSQC, HNCA [173], HN(CO)CA [174], HNCO [175], CBCA(CO)NH [176], CBCANH [177], HCCH-TOCSY [178], $^{15}$N-TOCSY-HSQC,

HNHA [32] were acquired for YFHJ assignment. Typical mixing times for NOESY and TOCSY experiments were 100 and 70 ms respectively.

The spectra were processed using NMRPipe [179] and analyzed using the XEASY software [180]. FIDs were typically processed applying a Gaussian window function and zero filled to twice the size of the data.

### 7.4.4 Relaxation measurements

$T_1$, $T_2$ and NOEs measurements were performed on a 0.6 mM $^{15}$N labelled sample dissolved in 90% $H_2O$ and 10% $D_2O$, in 10mM phosphate buffer, 50 mM NaCl, 0.05% sodium azide, pH 6.5. All the experiments were conducted on a Varian Unity instrument operating at 500 MHz of proton frequency. Both $R_1$ ($T_1^{-1}$) data and $R_2$ ($T_2^{-1}$) data were acquired with 10 relaxation delays (10, 100, 200, 300, 400, 500, 600, 700, 800, 200 ms and 10, 20, 35, 50, 65, 80, 100, 125, 150, 20 ms respectively). Peak height intensities were measured with the automatic routine in the NMRPipe/NMRDraw [179] software and the $^{15}$N $T_1$ and $T_2$ relaxation times were obtained by fitting the data to a two parameter single exponential decay function. $\tau_c$ values were obtained from the experimental $T_1$ and $T_2$ data using the Szabo Model-Free approach [163]. Residues with $T_1$ and $T_2$ values that differ more than a standard deviation from the mean value were not used for $\tau_c$ calculation.

Experimental steady states NOE values were determined from the peak intensity ratios of amide signals obtained by recording interleaved 2D Watergate $^1$H-$^{15}$N HSQC spectra with and without saturation delay of 4 s and a repetition delay of 4.2 s [181].

### 7.4.5 Structure calculation

Distance restraints were obtained from three dimensional $^1$H-$^{15}$N NOESY-HSQC and $^1$H-$^{13}$C NOESY-HSQC and two dimensional NOESY spectra recorded with 100 ms mixing time. $^1$H-$^{15}$N NOESY-HSQC was recorded at 600 MHz while $^1$H-$^{13}$C NOESY-HSQC was recorded at 800 MHz.

Angular restraints were obtained from TALOS [49] and HNHA spectra [32]. From the latter spectra we obtained the $^3J_{HNH\alpha}$ coupling constants, which gave a range of compatible torsional angles. In particular for J>8 φ =-120 ± 50, for 7<J<8 φ =-120 ± 60, 6<J<7 φ =-100 ± 80 and J<5.5 φ =-60 ± 30 [162]. Structure calculations were performed using the ARIA program, version 1.2 [51,54-56,164] and validation of the final structure was carried out with the programs Whatif and Procheck-NMR [57,59].

# Functional Investigation of YFHJ                    8

This chapter describes the functional analysis carried out on YFHJ. Classical NMR chamical perturbation analyses allowed the investigation of YFHJ's interaction with the proteins IscS and CyaY and with iron (II) and iron (III). Mass spectrometry and fluorescence techniques were used to further characterise the complex formed by YFHJ and the desulphurase IscS.

The first part of the chapter shortly describes the methods which were used, with the exclusion of NMR, which was described in chapter 1 and 2, while the results are discussed later on.

## 8.1 *Biomolecular techniques*

Here after I shall describe all the methodologies used for the production and characterisation of YFHJ.

### 8.1.1   Expression of recombinant proteins

Most of the proteins used for structural biology studies derive from recombinant DNA, i.e. DNA molecules which are formed by a combination of heterologous DNA sequences. The gene that codes for the protein of interest is cloned from a specific organism and inserted into a host cell, normally a bacterial cell, via a suitable molecular vector (transformation).

Bacterial plasmids are generally used as vectors because they can easily enter the bacterial cell. They are circular molecules of DNA separate from the bacterial chromosome but replicated with the same machinery that replicates chromosomal DNA. They normally contain only few genes, including an antibiotic resistance gene, and have a single origin of replication [182].

The transformation of the plasmid into an opportunely treated host cell (competent cell) is usually carried out with a heat shock treatment. *E. Coli* is the most commonly used host organism while different kinds of competent cells are commercially available.

After transformation, cells are grown in a nutrient medium at 37°C, in the presence of an antibiotic which guarantees that only the transformed cells, resistant to the

antibiotic, survive. Several dilution steps are made to ensure that the cells have enough nutrition to grow. When they reach their mid log phase, characterised by an optical density (O.D.) at 600 nm of 0.6-0.8, they are induced with isopropyl beta-thiogalactoside (IPTG).

IPTG is an inductor of the lactose (lac) operon that is not metabolised by the cells. The lac operon contains three structural genes, a regulator gene, an operator site and a promoter. The structural genes express β-galactosidase, a galactosidase permease and a thiogalactosidase transacetylase, necessary to metabolise lactose. Although in normal conditions bacteria do not use this sugar as a carbon source, when a high concentration of lactose is added to the cells the lac operon is activated.
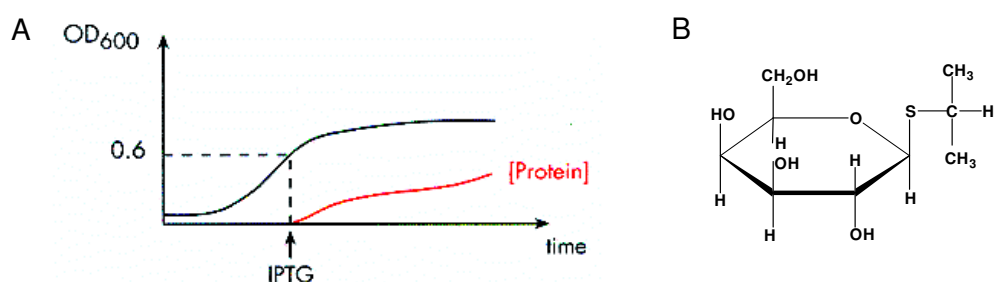


**Figure 8-1: Induction of bacterial cells. A) Diagram showing the log phase of the cell growth, when the IPTG is added to the cell culture. B) Molecular formula of IPTG.**

Lactose itself binds to the regulator gene preventing the expression of the lac repressor. IPTG is structurally reminiscent of lactose and plays a similar activator role when added to the cells.

Since the plasmids used as expression vectors contain a modified lac promoter where the three structural genes are replaced by the gene expressing the desired protein, addition of IPTG will induce over-expression of the protein. Subsequently the cells will be lysed with a proper buffer and the protein collected and extensively purified.

### 8.1.2 Purification of recombinant proteins

Recombinant proteins are generally produced with a specific tag, which can be a short tail of histidines (6-His tag) or a Glutathion-S-Transferase fused protein (GST-tag), specifically introduced to allow the purification step. Proteins purification, in fact, is often achieved by chromatography, exploiting the affinity of the tag for the stationary phase of the column.

In immobilised metal ion affinity chromatography (IMAC) [183], interactions between an electron donating group on the protein surface and a metal ion with one or more accessible coordination sites, permit the binding of the macromolecule to the column. Histidine is the amino acid with the highest affinity for metal ions; application of a 6-His tag either at the N or at the C-terminus of a protein therefore, constitutes a powerful tool for protein recovery using IMAC. The most widely used affinity columns for 6-His tagged proteins are the $Ni^{2+}$-NTA columns, where the stationary phase is constituted by a $Ni^{2+}$-nitrilotriacetic matrix.

The selectivity of the metal for the protein can be improved by modifying the composition of the mobile phase. In fact, a buffer characterised by a weakly alkaline pH and a high ionic strength, favours retention of the histidine residues and, simultaneously, decreases non specific ionic interactions between the matrix and the protein. To disrupt the binding of dispersed histidine residues in non tagged proteins which might be present in the cell lysate, a buffer with a low concentration of the histidine displacer imidazole is used. Accordingly, a high concentrated imidazole solution buffer elutes the protein from the column [183]. Details of the purification procedure can be found in the experimental section (paragraph 8.5.1).

An alternative purification method is adopted when the recombinant protein is obtained as a fusion protein with GST. In this case the macromolecule is purified exploiting the affinity of GST for glutathione immobilized to a Sepharose matrix.

The 6-His and the GST tags can be cleaved using the Tobacco Etch Virus (TEV) protease, that recognises a specific sequence of amino acids generally inserted during the cloning between the purification tag and the protein.

Finally the protein is further purified using a gel-filtration chromatography and the collected fractions can be analysed with a SDS-Page gel to assess their purity.

## 8.2 *Biophysical techniques*

### 8.2.1 Electrospray Ionisation Mass Spectrometry

Electrospray Ionisation Mass Spectrometry (ESI-MS) has recently emerged as a very useful technique for studying large biomolecules as well as non-covalent biomolecular complexes [184]. The measurement of non–covalent protein complexes is a challenge for several reasons: primarily they tend to dissociate easily, secondly they acquire a high mass to charge (m/z) ratio which represents an instrumental problem and finally non specific adducts can form during the measurements, causing a bias on experimental results.

Mass spectrometry measurements are carried out in the gas-phase on ionised analytes. The most common system for the study of macromolecular complexes is the Electrospray Ionisation mass spectrometry coupled to the Time of Flight analysis (ESI/ToF mass spectrometry). ESI is a very mild ionisation method that produces intact singly or multicharged ions in vacuo from large and complex species in solution [185]. In electrospray, small charged droplets are initially formed at the tip of a capillary because of the action of an applied electric field on the solution at the capillary tip (1-4 kV). To get the sample to spray, adjustment of the capillary voltage and of its distance from the ionisation cone is often necessary. A weak backing flow of inert gas can be used to push the sample through the capillary. Heat or dry gas, typically nitrogen, is then applied to the charged droplets to evaporate the solvent. The evaporation process causes the radius of the droplets to decrease and since the charge remains unchanged, Coulomb forces overcome the surface tension of the liquid at some critical radius, leading to the fission of the droplets into even smaller droplets. Repeated evaporation and fission steps result in the formation of very small charge droplets, which are the precursors of the gas phase ions. Since this evaporation process takes place at atmospheric pressure, it does not cause significant fragmentation of the analyte ions in the gas phase [185]. Once the droplets are formed, they are directed to the ToF chamber through a cone extractor. A quadrupole or a hexapole is interfaced between the ESI and the ToF system, to convert the continuous ion beam that derives from the ionisation source into a pulsed ion beam necessary for ToF analyses. These are based on measuring the times at which a set of ions, accelerated by the same amount of energy, reach the detector. Data are presented as a relative population of the various charged states of the molecule separated according to their m/z ratio. The spectrum obtained with ESI is referred to as the multiple charged spectrum. The molecular mass of the molecule can be calculated from the multiple charged peaks by software algorithms included with the instrument software. Two recent advances in the ESI-ToF technique, the nanoflow ESI and the quadrupole ToF instrument operating up to 22000 m/z, made it even more appealing for biomolecular applications [184,185].

Strikingly, non-covalent macromolecular complexes do not benefit from high vacuum conditions within the spectrometer. It was verified that under standard pressure conditions (2 mbar) no ions from non-covalent complexes can be detected. Upon increasing of the pressure a continuous ion current is instead detected. It was suggested that larger ions may acquire substantial kinetic energy when they are sprayed, affecting the transfer and orthogonal extraction into the ToF region. Increasing the pressure may enhance the collisional dampening and cool the ions;

moreover it could induce a more efficient desolvation, thus being effective in the enhanced detection of ions with high m/z values [185].

## 8.2.2 Fluorescence Polarization

Some of the functional analyses performed on YFHJ were based on fluorescence intensity and polarisation measurements. In the following paragraph a brief description of the polarisation phenomenon is provided.

Fluorescence anisotropy experiments have been widely used to quantify protein denaturation, protein-ligand associations and rotational rates of proteins. Upon excitation with polarised light, the emission from fluorescent samples is polarised as well. Fluorophores are in fact selected according to their orientation with respect to the direction of the polarised excitation [186]. In practice, polarisation or anisotropy measurements reveal the average angular displacement of the fluorophore between the absorption and subsequent emission of a photon. The recorded values therefore are determined by the rate and extent of the rotational diffusion during the lifetime of the excited state. This, in turn, depends on solvent viscosity and on the size and shape of the diffusing molecules.

In the figure below the measurement of fluorescent polarisation is illustrated.
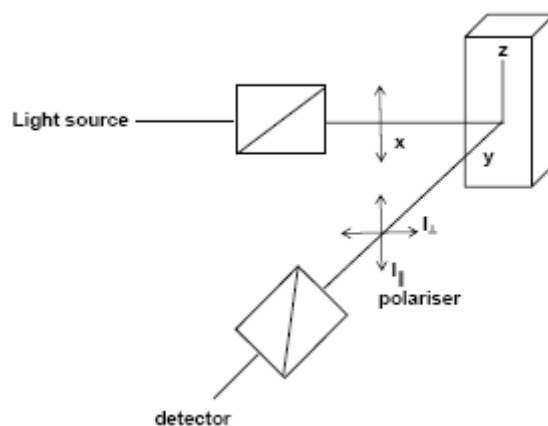


**Figure 8-2: Schematic representation of the fluorescence polarisation measurement [186].**

If the sample is excited with vertically polarised light, the electric vector of the excitation is parallel to the z axis. The intensity of the emission is measured through a polariser, which can be oriented parallel or perpendicular to the direction of the

polarised excitation. In the first case the measured intensity will be called $I_{\parallel}$ while in the second case will be $I_{\perp}$. The polarisation P and the anisotropy r will then be defined as:

$$P = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + I_{\perp}} \qquad (8.1)$$

$$r = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + 2I_{\perp}} \qquad (8.2)$$

In order to absorb light polarised along a certain axis, for example the z axis, it is not necessary that a fluorophore is precisely aligned along it. The probability of absorption is in fact proportional to cos2θ, where θ is the angle that the absorption dipole makes with the z axis. Excitation with polarised light will therefore result in a population of excited fluorophores symmetrically distributed around the z axis (photoselection), weighted with the probability that the fluorophores orient at θ degrees with respect to z. For a random distribution, the number of molecules at an angle between θ and θ + dθ is proportional to sinθdθ (proportional to the surface area on a sphere within the angles θ and θ + dθ). Hence the maximum selection that can be obtained upon excitation of an isotropic solution is given by the probability distribution below.

$$f(\theta)d\theta = \cos^2\theta \sin\theta d\theta \qquad (8.3)$$

It has been demonstrated that the maximum value of anisotropy that can be observed when the absorption and emission dipoles are co-linear is 0.4, in the absence of depolarising processes [186].

### 8.2.3  Measurements of fluorescence anisotropies

In this paragraph I will describe only the so called 'L-Format or Single Channel Method' of measurement, that is the one that was used for the performed studies. A detailed description of the second method, the 'T Format or Two Channel Method', can be found  in [186].

In the L-Format method the excitation and emission polarisers are alternatively rotated in the horizontal and vertical plane. Excitation and emission wavelengths are

146

selected with monochromators that partially polarise the light. As a result, rotation of the polarisers in the horizontal and vertical plane yields different intensities of the incident and emitted light. The measured intensities therefore do not correspond to the desired parallel and perpendicular intensities. To get unbiased results from the detection system one has to measure a G factor, which is the ratio between the sensitivities of the emission channel for vertically and horizontally polarised components. We have:

$$I_{VV} = kS_V I_\parallel \tag{8.4}$$

$$I_{VH} = kS_H I_\perp \tag{8.5}$$

$$\frac{I_{VV}}{I_{VH}} = \frac{kS_V I_\parallel}{kS_H I_\perp} = G\frac{I_\parallel}{I_\perp} \tag{8.6}$$

$$G = \frac{S_V}{S_H} \tag{8.7}$$

Where the subscripts indicate respectively the orientation of the exciting and emitting polariser and k is a proportionality factor that accounts for quantum yield of the fluorophore and other instrumental factors.

G can easily be evaluated using horizontally polarised excitation. In fact when this is measured both horizontal and vertical polarised components are proportional to $I_\perp$, being both oriented perpendicular to the excitation plane. Hence any differences between $I_{HH}$ and $I_{HV}$ must be due to the properties of the detection system.

$$\frac{I_{HV}}{I_{HH}} = \frac{S_V I_\perp}{S_H I_\perp} = \frac{S_V}{S_H} = G \tag{8.8}$$

Possible variations in the intensity of the excitation light upon rotation of the polariser from the vertical to the horizontal plane are constant factors in the numerator and the denominator, and therefore they nullify each other.

When the G factor is known the ratio $I_\parallel / I_\perp$ can be calculated from equation (8.6) and from it the value of polarisation and anisotropy can be evaluated [186].

## 8.3  *Results*

### 8.3.1  Investigation of the YFHJ/IscS complex

#### *8.3.1.1    Titration with IscS*

At present, the most relevant clue on the role of YFHJ derive from the observation that it interacts with IscS, the protein which provides sulphur for the iron-sulphur cluster biosynthesis.

The interaction was initially identified by Tokumoto et al. in 2002 [151]. A GST pull down assay confirmed the presence of an interaction, subsequently traced through NMR chemical shift perturbation experiments.

Upon titration of a $^{15}$N labelled sample of YFHJ with unlabelled IscS, we observed that even at very low IscS concentration (10:1 YFHJ/IscS  ratio), amides 3, 10, 11, 14, 29, 30, 51, 62, 65 are broadened and that residues 13, 17, 33, 35, 61, 63 are affected by a small shift. 64 is both broadened and appreciably shifted. An increased concentration of IscS affects peaks 18, 44, 63, 66 as well (ratio 6:1 YFHJ/IscS). Peaks 17 and 64 are broadened almost beyond the detection limit.  At 4:1 YFHJ:IscS ratio the spectrum of YFHJ is seriously affected, the signals being extremely weak.

 A spectrum of a 3:1 mixture of YFHJ and IscS acquired overnight confirms that residues 3, 10, 11, 13-15, 17, 18, 61-66 are the most affected by the binding, since they have a very large line broadening even after many hours of acquisition, thus excluding that their intensities are depressed by a poor signal to noise ratio. Broadening of residues 29, 30, 51 is probably independent from the presence of IscS, since these residues are weak in the reference spectrum as well and were already shown to be particularly sensitive to solvent variations.

In the [$^{1}$H,$^{15}$N] HSQC of a mixture 2:1 YFHJ/IscS one can see that the amines from the side chain of Gln47 and Asn50 and the indolic proton of Trp35 are still reasonably sharp. The side chain of Gln34 is not visible because it has a higher degree of rigidity probably due to the formation of a saline bridge with His33.

#### *8.3.1.2    Investigation of the stoichiometry of YFHJ/IscS with mass spectrometry*

Mass spectrometry has become an important tool for investigating protein-protein complexes. In particular, recent advances in ESI nanospray techniques have allowed the investigation of the stoichiometry and the $K_d$ of covalent complexes.

In collaboration with Helena Hernandez, from the Cambridge University Chemical Laboratory, we performed mass spectrometry on various mixtures of YFHJ and IscS. From the peak pattern of the mass spectrum, it is possible to evaluate the molecular weight of the species present in the gas phase, and therefore the stoichiometry of the complex.
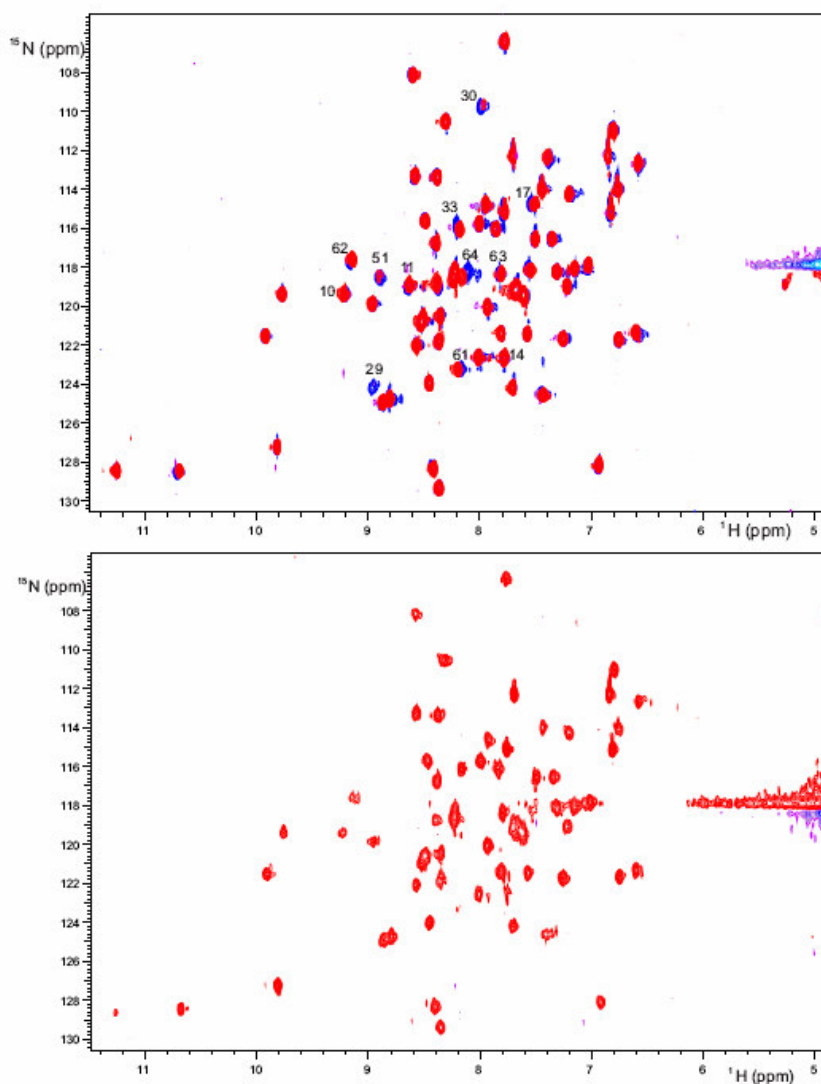


**Figure 8-3: Effect of IscS on YFHJ. The picture shows on the top the superimposition of the [$^1$H,$^{15}$N] HSQC spectra of YFHJ alone (blue) and in the presence of 0.1 equivalents of IscS (red). On the bottom the [$^1$H,$^{15}$N] HSQC of a 1:6 YFHJ:IscS mixture is reported. [YFHJ]=0.4mM, in Tris-HCl 10 mM, pH 6.5 and NaCl 50 mM.**

One of the major challenges in analysing a protein complex with mass spectrometry is to choose the appropriate sample and instrumental conditions. For non-covalent interaction measurements, the cone and extractor voltages and the pressures in the source and, when used, in the quadrupole/hexapole regions must be carefully

calibrated, as they control the energy and number of collisions a protein complex undergoes on its transit from the source to the ToF analyser.

In Figure 8-4-A, a spectrum of a 2.2 YFHJ/IscS (monomer) mixture in 100 mM ammonium acetate buffer at pH 7 is reported. Ammonium acetate was chosen because of its highly volatility, important to guarantee an optimal desolvation of the sample.

The peak pattern indicates the presence of a 2:2 YFHJ/IscS complex, characterised by MW=98906±22.20, and of a smaller concentration of the 1:2 complex, with MW= 106823.14±23.70. The peaks from the free proteins are absent, suggesting that the complex is specific.

However, the same mixture dissolved in 500 mM ammonium acetate gives a different spectrum (Figure 8-4, bottom). Here we can identify a 1:2 YFHJ/IscS complex (MW=98994±31.97) and a high concentration of free IscS (MW=91105.84±12.42). This indicates that the ionic strength of the solution has an influence on the complex formation.
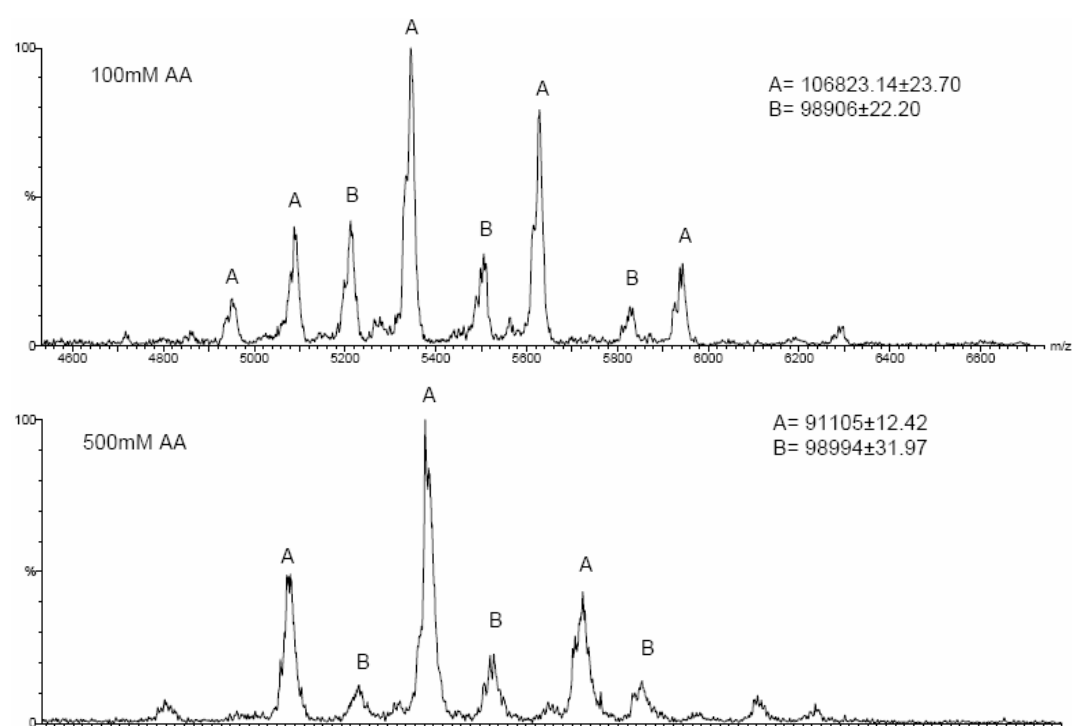


**Figure 8-4: Top: ESI-ToF mass spectrum of a 2:2 YFHJ/IscS (monomer) mixture, in 100 mM ammonium acetate buffer. Peaks A and B correspond respectively to a 2:2 YFHJ/IscS and to a 1:2 YFHJ/IscS complex. Bottom: ESI-ToF mass spectrum of a 2:2 YFHJ/IscS (monomer) mixture in 500 mM ammonium acetate. Peaks A and B correspond respectively to dimeric IscS and to a 1:2 YFHJ/IscS complex. Pressure reading in the hexapole region=1.1 e$^{-6}$ mbar, pressure at the ion source= 6.1 mbar.**

### 8.3.1.3 *Investigation of the stoichiometry and $K_d$ of YFHJ/IscS complex with fluorescence methods.*

To determine the binding constant of the complex and to have an independent measurement of its stoichiometry we performed fluorescence anisotropy titrations.

We labelled YFHJ with an extrinsic probe covalently conjugated to the partially exposed Cys37, and monitored the variation of the fluorescence polarisation upon addition of IscS. The presence of tryptophans in both proteins prevented us to use the intrinsic fluorescence, because it would have produced misleading results.

In order to carry out the experiments in the best conditions, we tried different probes.

First we used MIANS, a maleimidylanilino dye containing a naphthalene-sulphonic acid moiety. However the fluorescence polarisation measured for MIANS-YFHJ was too high to correspond to a monomeric protein. It is possible that the presence of the probe influenced YFHJ's conformation and induced an aggregation process.

We then labelled YFHJ with AlexaFluor 488, characterised by a tetrafluorophenyl moiety. In this case we measured a polarisation value lower than the expected value, suggesting that, in our experimental conditions, the probe was reorienting independently from YFHJ. As a consequence it could not be used to obtain information on the molecular weight of the species in solution.

Finally, labelling with AlexaFluor 532, that contains a succinimidyl ester moiety (Figure 8-5), resulted in a labelled protein YFHJ* with a polarisation compatible with a monomeric species.
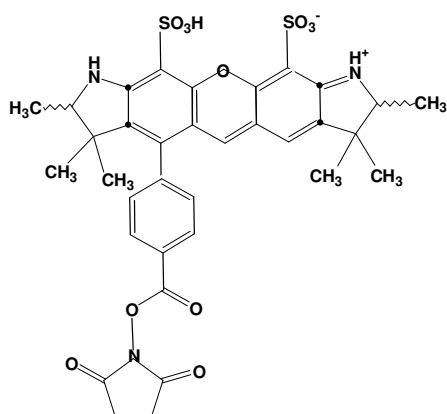


**Figure 8-5: The figure represents the fluorophore AlexaFluor 532 carboxylic acid, succinimidyl ester.**

The titration was performed at 20°C using a solution of IscS containing the same amount of YFHJ* that was present in the titrated solution. In this way we could neglect the effect of dilution on the fluorescence signal.

Upon addition of IscS, we observed a significant increase in the intensity of the fluorescence and in the polarisation value, consistent with the formation of a high molecular weight complex with a slow rotation rate in solution. We recorded three

different sets of values to account for experimental errors. Assuming a 1:1 stoichiometry with IscS monomer, we could fit the intensity titration curves with a $K_d$ value of 9.9 ± 0.8 µM. Competition titrations in which labelled YFHJ was displaced from the complex by addition of unlabelled protein gave a $K_d$ for the latter of 14.3 ± 0.75 µM. However the same curves could be reasonably fitted supposing a stoichiometry 1:1 with IscS dimer; in this case we obtained a $K_d$ of 1.37 ± 0.23 µM and of 2.2 µM ± 0.75 for the labelled and unlabelled complex respectively.

The polarisation data were converted into fractional saturation ([YFHJ/IscS]/[YFHJ total]) to take into account the variation in the total intensity fluorescence. A correcting factor of 1.77 ± 0.03 was calculated. We measured a polarisation value for the free labelled YFHJ of 0.185 and we estimated a value of 0.345 for the complex polarisation. The calculated $K_d$ were in accord with those calculated using the fluorescence intensity, 12.4 ± 2.3 µM and 1.68 ± 0.19 µM, assuming a 1:1 stoichiometry with IscS monomer and dimer respectively. The titration curves are reported in Figure 8-6.

### 8.3.2  Titrations with iron (II) and iron (III)

The location of YFHJ gene in the Fe-S cluster operon suggested to titrate the protein with iron (II) and iron (III). Moreover preliminary gel filtration experiments showed that the protein has a tendency to aggregate in the presence of iron.

Addition of iron (II) causes observable effects on the [$^1$H,$^{15}$N] HSQC spectrum of YFHJ already at low ion:protein ratio. In fact 0.5 equivalents of iron on the solution bleach completely the amides form residues 17 and 64. Signals from residues 10, 13, 16, 18, 19, 39, 41, 65, 66 are also affected in both their intensity and their position in the spectrum. Residues 11, 14, 15, 21, 22, 61, 63 experience a small shift. At higher concentrations of iron (II), the chemical shift variation and the line broadening of the same amides are enhanced. At a 2:1 iron:protein ratio signals from Glu66, Leu39 and Ala18 are too broad to be observed; Asp41 and Ala65 have further shifted and broadened, Gly2, Leu3, Lys4, His33, Leu62, Met32, Trp35, Asp38, Phe42 are slightly shifted. At 2.5:1 iron(II)/YFHJ ratio, we noticed partial oxidation of iron (II) to iron (III), as evidenced by the pale yellow colour of the solution. We recorded additional titration points but a colloidal species started to precipitate, preventing a good shimming of the sample, and eventually provoking the precipitation of the protein.

The effect of iron (III) is smaller.
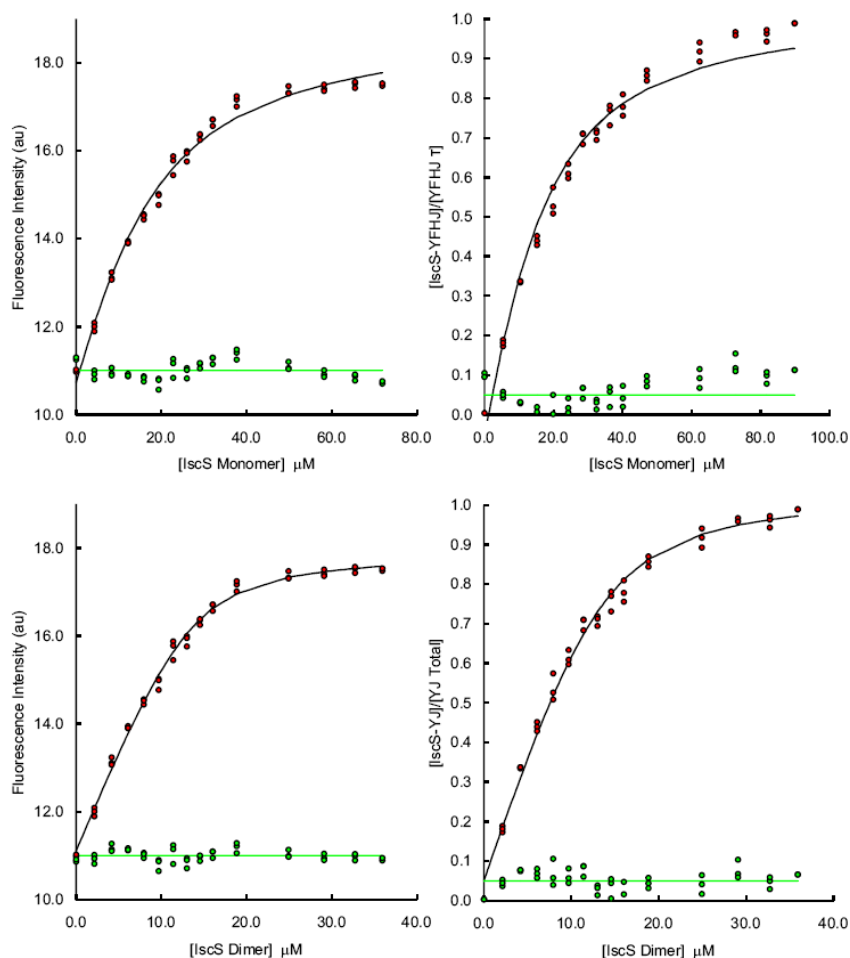
**Figure 8-6: Fluorescence measurements of the affinity constant for the YFHJ/IscS complex. On the top and bottom panel fluorescence intensity on the left and fractional saturation on the right are plotted against the concentration of monomeric and dimeric IscS respectively. In green the deviation between the experimental and the theoretical points is shown. The concentration of YFHJ\* was kept constant during the experiment at 3.9 μM: aliquots of a solution containing IscS 222 μM and YFHJ\* 3.9 μM in Tris-HCl 20 mM, pH 8, 150 mM NaCl, 20 mM β-mercapto-ethanol were added.**

In the presence of 0.5 equivalents of iron (III) few peaks experience a shift and no peak disappears. However amides from residues 64 and 66 are broadened and those of residue 30 and 33 are slightly shifted. Resonances of amide 29, 30, 51 also weaken, but these residues were already weak in the reference spectrum.

At an equimolar iron:protein ratio we observe a general broadening of the peaks, with a complete bleaching of amides 64 and 66 and a pronounced effect on amides 13, 16, 17, 63, 65. Protons 10, 15, 33 are weakly shifted. At an ion:protein ratio of 2:1 the effects are enhanced; signals from residue 4, 11, 37 experience a weak shift. Moreover, we noted the appearance of a peak around 113-7.1 ppm. At 2.5:1

resonances 62, 32, 34, 35, 39 slightly shift while 10, 15, 33, 34 shift further; the amides of Asp41 and Tyr16 sensibly broaden and the $H^N$ from residues 17, 36, 65, 66 disappear.

Further addition of iron (III) provokes the formation of a colloidal species and a general decrease of the peaks' intensities.
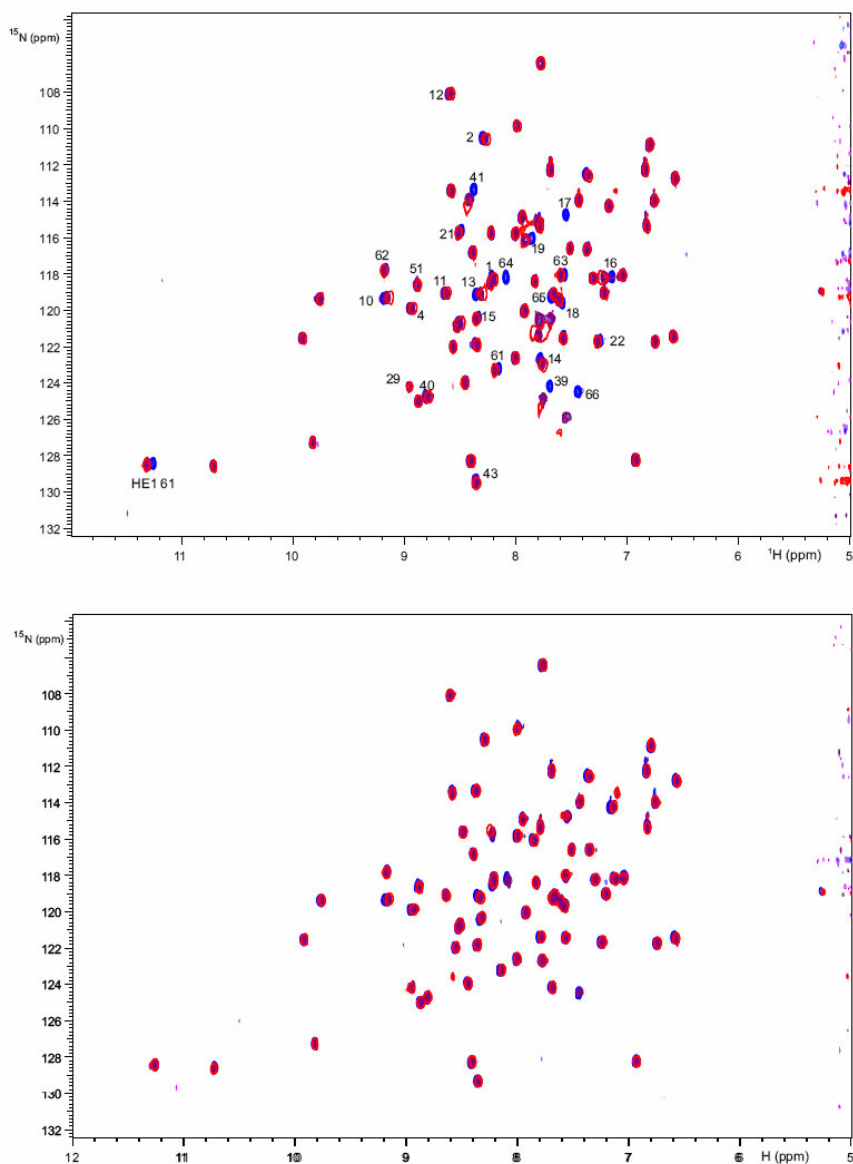


**Figure 8-7: Effect of Fe(II) and Fe(III) on YFHJ. Superimposition of the [$^1$H,$^{15}$N] HSQC spectra of a YFHJ in the absence and in the presence of 0.5 and 1 equivalent of A) iron (II) and B) iron (III) (blue, purple and red respectively). [YFHJ]=0.3 mM in Tris 10 mM, pH 6.5, 50 mM NaCl.**

### 8.3.3   Titrations with CyaY

CyaY, the bacterial homologue of human frataxin, is essential for iron homeostasis and many studies have suggested its involvement in the biosynthesis of the iron-sulphur clusters, in particular in iron delivery to IscU [16,79,125,187].

As part of a wider project aimed at discovering the link between the disease caused by lack of frataxin, Friedreich's ataxia, and Fe-S clusters biosynthesis, we decided to investigate a possible interaction between YFHJ and CyaY. Furthermore, to study a hypothetical role of YFHJ in iron-binding and possibly in iron-delivery, we analyzed the behaviour of the iron (III) loaded protein in the presence of increasing concentration of CyaY.

First, we carried out an NMR titration to follow the effect of unlabelled CyaY on a labelled sample of YFHJ. We added only two aliquots of titrating protein, as we were not interested in a stoichiometric titration and because we did not notice large variations in the spectrum.

At 0.2:1 CyaY/YFHJ ratio, we observe the usual broadening of peaks 29, 30, 51 and a small shift of resonances 10, 17, 33, 64. At 0.4:1 ratio we do not notice any further changes.

We then added a total of 100 μl of CyaY 0.6 mM to a solution of YFHJ/iron (III) formally 1:3 (YFHJ 0.306 mM), after centrifugation. As previously discussed, the presence of 3 equivalents of Fe (III) causes some variations on the spectrum. Interestingly some of these variations are abolished by the presence of CyaY: in particular we can observe that the amide groups of the residues that were displaced (mainly 4, 10, 13, 15, 33, 34, 35, 62) are shifted back towards their original position. However the peaks which were bleached are still not detectable.


### 8.3.4   Genomic links

A search for genomic correlations was made by Dr. Huynen, from the Radboud University Nijmegen Medical Centre. He examined whether the gene that encodes for YFHJ tends to occur with specific other genes of the Isc operon amongst the sequenced genomes, a technique that was successfully used to predict a role for frataxin in Fe-S cluster assembly [11]. All of the genes of the Isc operon have a wider distribution among sequenced genomes with respect to YFHJ. IscS occurs for instance in 91 sequenced α, β and γ proteobacteria, 34 of which do not have a homologue for YFHJ. The reverse, the presence of YFHJ without IscS, never occurs (Table 8-1). YFHJ shows the strongest dependence with the presence of IscS, IscU,

HscA and HscB and to a minor degree with the presence of Fdx, IscA and IscR. Several examples of the presence of the other Isc operon proteins in the absence of YFHJ are consistent with the experimental evidence that YFHJ is not essential for Fe-S cluster assembly [120].

The analysis was extended to CyaY. This gene has the weakest link with YFHJ, since either sequence occurs regularly without the other. However in species *P. falciparum*, *P. chabaudi*, *P. yoelli yoelli*, *T. parva* and *T. annulata*, which are the only sequenced eukaryotes that have a detectable orthologue of YFHJ, CyaY and YFHJ appear to replace each other. With the exception of *T. annulata*, these are also the only so far sequenced eukaryotes in which homologues of CyaY could not be detected.

**Table 8-1: Occurence of orthologues of *yfhj* with other members of the Isc operon and with CyaY in α-, β- and γ-bacterial genomes. In the first row is indicated the frequency of co-occurrence of the *yfhj* gene with those indicated in the respective colums. The second row corresponds to the frequency at which the indicated gene occurs without *yfhj* while the third row represents the frequency at which *yfhj* occurs without the indicated gene. The presence of *yfhj* clearly correlates better with the gene of the Isc operon than with that of CyaY. Within the Isc operon, the link with IscS is the strongest, although the differences with the other genes are small.**

|                   | *iscR* | *iscS* | *iscU* | *iscA* | *hscA* | *hscB* | *Fdx* | *CyaY* |
|-------------------|--------|--------|--------|--------|--------|--------|-------|--------|
| **Co-occurence**  | 52     | 57     | 57     | 53     | 56     | 57     | 53    | 46     |
| **No *yfhj***     | 37     | 34     | 27     | 53     | 28     | 30     | 35    | 26     |
| ***yfhj* alone**  | 5      | 0      | 1      | 4      | 1      | 1      | 4     | 11     |

## 8.4 *Discussion*

Preliminary work undertaken by Tokumoto et al [151] described the complex network that links the isc proteins and, based on co-purification and yeast two-hybrid screens, demonstrated that YFHJ interacts with the desulphurase IscS. They also reported an interaction of YFHJ with IscU, but these results were not supported by two-hybrid screen assays, thus suggesting that the interaction is mediated by other factors.

We have confirmed the presence of an interaction between YFHJ and IscS by NMR and studied the macromolecular complex with mass spectrometry and fluorescence.

The disappearance of the amide peaks of YFHJ from the [$^1$H, $^{15}$N] HSQC spectrum upon addition of the 90 kDa IscS provides evidence of the formation of a large complex, characterised by a long tumbling time and therefore not visible with a standard NMR technique. We were able to follow the bleaching of the peaks upon slow titration of YFHJ with IscS. Although many peaks were already weak at low IscS/YFHJ ratio, we could identify the most affected ones and map them on the protein surface (Figure 8-8). Mutagenesis experiments could clarify the relative importance of each amino acid.

We do not know which IscS residues might be involved in the binding, but it was shown that a C328A mutation in IscS did not give rise to differences in the interaction with YFHJ, thus ruling out the direct involvement of Cys328 in the binding with YFHJ [151].
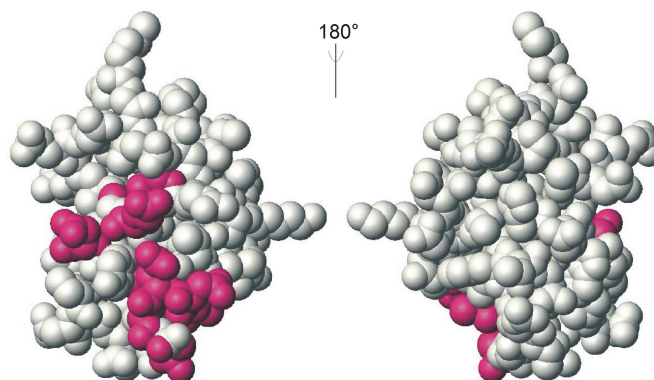


**Figure 8-8: YFHJ's surface with the most affected residues upon addition of IscS (3, 10, 13, 14, 11, 17, 18, 62-66). They form a surface on one side of the protein, thus suggesting that they constitute a binding region.**

Nano-ESI mass spectrometry performed on various mixtures of YFHJ/IscS in 100 mM ammonium acetate buffer, suggested that a 1:2 and a 1:1 YFHJ/IscS (monomer) species can form. The 1:1 complex seems to be the most abundant species, but the experiments do not provide a definitive answer on the stoichiometry of the complex.

At higher ionic strengths peaks from unspecific binding are visible, thus it is conceivable that the interaction between the two proteins has an electrostatic component.

Measurements of fluorescence did not provide a definitive answer on the complex' stoichiometry either. In fact they are in accord with the presence of both a 1:1 and a 1:2 (YFHJ/IscS monomer) complex in solution. However, it must be considered that the crystal structure of IscS is a pseudo symmetric dimer. Therefore, unless a structural rearrangement of the two IscS monomers takes place upon formation of

the complex, it would be reasonable to expect a 1:1 stoichiometry of the complex, with each monomer binding one YFHJ molecule.

A dissociation constant of 9.9± 0.8 µM was obtained from fluorescence polarisation measurements, supposing that one YFHJ molecule binds to one IscS monomer. The value of the $K_d$ is not extremely strong, thus justifying the variability of the obtained results. Further analyses must be conducted to reach a conclusive result.

A preliminary gel filtration analysis had pointed out an oligomerisation process promoted by Fe(II) under aerobic conditions. Accordingly, NMR chemical shift perturbation experiments suggested that the protein is able to weakly bind both iron (II) and iron (III) in the solution state.

Iron (II) affects the HSQC peaks of YFHJ already at low stoichiometries (0.5:1, ion/protein), causing bleaching and shifting of many signals. Upon mapping the affected residues onto the surface of the protein, we could identify an interaction site in the area determined by residues Glu13, Asp17, Glu64, Glu66. Tyr16 is strongly influenced by iron (II) as well, and might be involved in the binding. However, many other peaks are influenced by the ion, in particular Leu39, Asp41 and Tyr19. Asp41 might be directly involved in chelating iron, together with Glu10, Glu13. Surprisingly, non polar residues such as Ala14, Leu3, Leu15 are slightly influenced at higher stoichiometries. It is conceivable that the effects of iron in these cases are echoes of the electrostatic interactions between the cation and the negative residues that surround them. This hypothesis is supported by the observation that iron (II) specifically influences residues from the 'acidic' side of the protein.

The effects of iron (III) on the HSQC spectrum of YFHJ are not as important as the one produced by iron (II). The presence of the ion, in fact, is evident at higher concentrations and does not cause a dramatic shift of the peaks. The peak cluster at ~ 119 and 7.4 ppm, where the amides from residues 54, 56, 65 and 18 resonate, is paradigmatic of this difference. While in the presence of 0.5 equivalents of iron (II) amides 65 and 18 are sensibly shifted, when one equivalent of iron (III) is added one can see only a slight broadening of amide 65. Nevertheless, iron (III) bleaches the signals from Asp17, Asp64, Ala 65, Glu66 and, slightly, Glu63 and induces a small shift of resonances from Glu19, Glu13, Leu15, similarly to iron (II). A specific effect of iron (III) is the shift of His33 and Cys37.

Here as well, the side chains of Asp17, Glu13, Glu64 and Glu66 could chelate an iron (III) ion, supposing that the side chain of Glu66 could rotate. His33 and Cys37, particularly suitable for iron ligation, are properly directed to chelate the ion, but two coordination sites would be vacant.

This analysis indicates that YFHJ can bind both iron (II) and iron (III) and that many residues are affected by the metal ions. The effects produced by iron (III) are consistent with the presence of the cation in its high spin (H.S.) configuration: H. S. iron (III) in fact, does not have an anisotropic distribution of its $d^6$ electrons in the outer shell and therefore does not cause pseudo-contact shift. On the other hand displacement of the signals is expected when iron (II) is added; the presence of many shift variations moreover, may hint at the presence of a fast exchange equilibrium between different conformations. For example, it could be hypothesised that the protein undergoes a structural rearrangement that interests the region comprised by residues 39-41 upon iron chelation.

At iron/protein stoichiometric ratios up to ~3:1, the HSQC spectra of YFHJ are consistent with the presence of a monomeric species. However, at higher iron (II) and iron (III) concentrations, the appearance of a colloidal species and the general broadening of the resonances indicates the presence of a precipitation process, which could involve formation of high molecular weight aggregates, as suggested by gel-filtration analysis (Adinolfi, personal communication). The phenomenon however could be a non specific aggregation process induced by the increase of the ionic strength of the solution, i.e. a *salting out* effect. Measuring the relaxation parameters and the light scattering of the iron loaded form of YFHJ could clarify the nature of the species on solution.

Both iron (II) and iron (III) mainly affect Glu13, Asp17, Glu64, Glu66. Among these residues, Glu13 is the only one to be partially conserved. It would be interesting to measure the relative affinity of iron (II) and iron (III) to this binding site with competition assays. Moreover, titrations with different cations could be informative on the specificity and selectivity of the binding site. It is noteworthy to observe that the surface involved in IscS binding is the same as the one implicated in iron binding (Figure 8-9).

This finding strongly suggests that YFHJ could be a molecular adaptor, which mediates the interaction of IscS with a yet unidentified partner, through an iron mediated mechanism. An interesting experiment would be to evaluate the stoichiometry and binding constant of the IscS/YFHJ complex in the presence of iron (II) and iron (III) to determine whether the metal ion can effectively favour the formation of the complex.

The physico-chemical properties of YFHJ and its behaviour in the presence of iron are strongly reminiscent of bacterial frataxin. In fact, both proteins are highly acidic, weakly bind iron (II) and iron (III) with glutamates and aspartates, and undergo an oligomerisation process promoted by Fe(II).
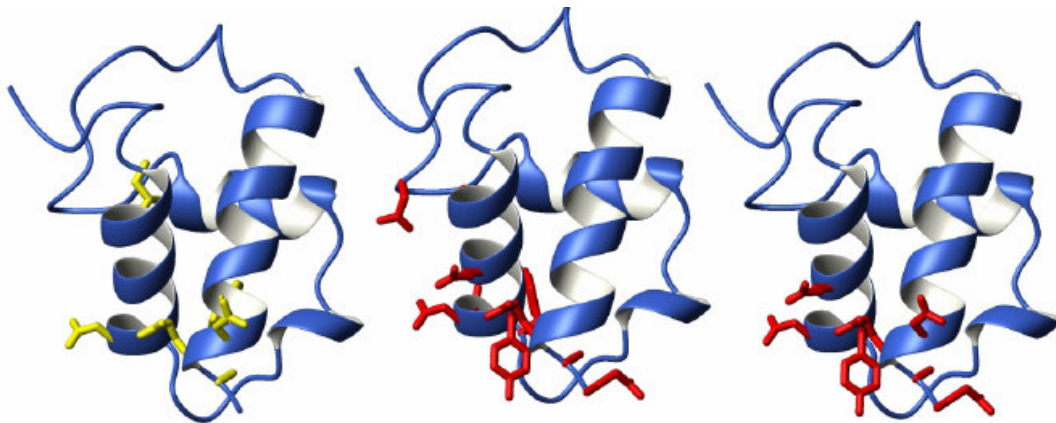
**Figure 8-9: Mapping the interactions with IscS and iron cations onto the structure of YfhJ. From left to right, ribbon representation of YfhJ indicating the side chains of the residues most affected by IscS, $Fe^{2+}$ and $Fe^{3+}$.**

The HSQC spectrum of $^{15}$N labelled YFHJ does not show many changes upon the addition of unlabelled Cyay, except for small chemical shift variations of Glu10, Asp17, Glu33 and Glu64, that can be ascribed to small modification of the solvent composition. This suggests that the two proteins do not interact with each other. The effect of CyaY on a sample containing YFHJ and ~ 3 equivalents of iron (III) is more interesting. CyaY appears to be able to bind the iron (III) which is unspecifically bound to YFHJ. This would explain the fact that many resonances which were affected by the ion, shift back to their original position when CyaY is added, while the amides from Asp17, Glu13, Glu64 and Glu66 do not reappear in the spectrum. CyaY was shown to have a weak affinity for iron (II) and iron (III), as described in chapter 5 ($K_d$=55 µM for iron (III) [16]); this appears to be sufficient to compete with YFHJ superficial residues for the ion binding but it is not strong enough to capture the ion from the chelating residues of the binding site. We would expect the same behaviour with iron (II), but further experiments are needed to confirm the data.

Genomic analysis of the co-occurrence of the two proteins in the sequenced genomes revealed that YFHJ and CyaY are complementary in eukaryotes, although the analysis was performed on the only 5 eukaryotic genomes where YFHJ could be detected. This result and the similarity of the two proteins suggest that YFHJ could substitute frataxin in some primitive eukaryotic species. This would be an extremely appealing hypothesis, but would not explain the redundancy of the two proteins observed in bacteria. It might be possible that the two proteins have different function in prokaryotes and eukaryotes. Interestingly, *E. Coli* strains lacking CyaY gene do not show any of the classical phenotypes of FRDA cells, such as mitochondrial iron

accumulation, defective oxidative phosphorilation and oxidative damage [22]. Could YFHJ effectively compensate for CyaY absence? An experiment where both genes are knocked out could answer this question and increase the understanding of both proteins.

## 8.5 *Materials and Methods*

### 8.5.1  Protein expression and purification

Unlabelled YFHJ and IscS proteins for fluorescence and mass spectrometry measurements were cloned into a pET derived plasmid vector as His tagged glutathione-S-transferase (GST) fusion proteins comprising a Tobacco Etch Virus (TEV) protease cleavage site that leaves a short N terminal AlaGly tag. The proteins were over-expressed in *E. Coli* BL21(DE3) cells. The insertion of the plasmid into the cells was achieved by a thermal shock: 5 µg of plasmid were mixed with 50 µl of cells and put on ice for 15 minutes. Then the mixture was put at 42 °C for one minute, to help the insertion of the plasmid, and immediately placed back on ice. After the transformation, the cells were spread over Petri dishes and incubated overnight at 37°C. An isolated bacterial colony was then picked with a sterile tip and diluted into 5 ml of Luria Bertani (LB) broth in the presence of 150 mg of kanamycin. The cells were grown for 8 h at 37°C and were subsequently ten fold diluted and shaked overnight at 37°C. After a 20 fold dilution in LB, the cells were grown until they reached $OD_{600}$=0.6-0.7 and induced with 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG) at 37°C. After four hours the cells were centrifuged at 4°C and 4800 rpm for 25 minutes and the pellet re-suspended with the lysis buffer whose composition is given in appendix B. Cells were lysed by vortexing and sonication and the insoluble components were removed by centrifugation. The soluble proteins were then purified using $Ni^{2+}$-NTA chromatography (QUIAGEN), following a standard procedure. The columns were first equilibrated with binding buffer and then the protein solution was loaded. The columns were gently rotated for 30 minutes, to guarantee binding of the His tagged protein to the resin. Subsequent washings first with a buffer containing low salt and low imidazole concentrations and then with one at a higher salt concentration, allowed removal of the impurities and breaking of the non specific bindings to the resin. Finally the elution buffer containing a 300 mM imidazole was used to recover the desired protein. The 6-His tag was removed using TEV protease, which was added to the recovered proteins before a

dialysis step, necessary to remove the high salt and high imidazole concentrations. After the dyalisis, a second purification on a Ni-NTA column allowed the separation of the proteins from the His-GST and from the TEV protease, this latter containing a non-cleavable 6His-sequence. The composition of the buffers used for the chromatographic purification and for the dialysis are given in appendix B.

The untagged polypeptide was isolated by gel-filtration chromatography on a (1.6 x 60 cm) HR Superdex-75 from Pharmacia, eluted at a flow rate of 0.4 ml/min with 20 mM Tris-HCl, pH 8.0, containing 150 mM NaCl buffer and 20 mM β-mercapto ethanol. The effluent was monitored by recording the absorbance at 280 nm. The column was calibrated using the low molecular mass protein standard mixture (Pharmacia). The purity of the proteins was evaluated by 12% polyacrilammide gel-electrophoresis in the presence of sodium dodecylsulfate (SDS-PAGE). The proteins were concentrated using a Vivaspin (VivaScience) centrifugation tube to a final concentrations of 230 μM and 60 μM for YFHJ and IscS respectively. Protein concentrations were determined spectrophotometrically by measuring the absorbance at 280 nm, using $\varepsilon_{280}= 19630 M^{-1}cm^{-1}$ and $\varepsilon_{280}=39760 M^{-1}cm^{-1}$ for YFHJ and monomeric IscS respectively.

### 8.5.2  NMR titrations

NMR titrations were performed by recording [$^1$H,$^{15}$N] HSQC spectra at 30°C on a Varian Utility instrument operating at 500 MHz proton frequency. $^{15}$N labelled samples of YFHJ (0.3-0.4mM) in 10mM Tris-HCl buffer at pH 6.5, 50mM NaCl, 0.05% $NaN_3$ were used. Aliquots of 30 μl of an IscS stock solution at 0.23 mM dissolved in Tris-HCl (pH 7.0) were added to reach YFHJ/IscS ratios of 10:1, 6:1, 4:1, 3.5:1; 3:1; 2.9:1, 2.6:1, 2.4:1, 2.3:1 2.1:1. Titrations with Fe(II) and Fe(III) were carried out using concentrated stock solutions of $Fe(NH_4)_2(SO_4)_2.6H_2O$ and $FeCl_3.6H_2O$ (Sigma) at 25mM and 38mM respectively. 0.5 equivalents of ion solution were added up to a protein:ion ratio of 6:1 and 3:1 for the Fe(II) and Fe(III) titrations respectively. Titration of YFHJ with CyaY was performed by adding two 50 μl aliquots of unlabelled CyaY from a solution 0.6 mM, dissolved in 20mM Tris-HCl, 50mM NaCl at pH 7.0. Titration of Fe(III) loaded YFHJ with CyaY was carried out using an unlabelled 0.6 mM CyaY solution up to 3:1 CyaY/YFHJ-Fe(III) ratio. Unlabelled CyaY was expressed and purified by John McCormick.

### 8.5.3 Mass spectrometry

ESI-ToF spectra were recorded on a LCD mass spectrometer, consisting of a nanoflow source, a hexapole ion guide and a ToF analiser. The instrument was modified to allow the admission of argon gas in the hexapole region of the spectrometer.

Different YFHJ/IscS mixtures were prepared from standard solutions of YFHJ 11.5 µM and IscS 60 µM in 500 mM or 100 mM ammonium acetate buffer at pH 7, in presence of 1 mM dithiothreitol (DTT). Aliquots of 1-2 µl of protein solution were placed in a nanoflow needle and sprayed into the spectrometer. Mass spectrometry conditions were set with a capillary voltage of 1700 volts, a cone voltage of 200 volts and an extractor voltage of 20 volts. The pressure applied at the ion source was 6.1 mbar, while a pressure of 1.1 e$^{-6}$ mbar was read in the exapole region.

### 8.5.4 Fluorescence measurements

Labelling of YFHJ with the fluorescent probe AlexaFluor 532 dye (Invitrogen) was obtained following standard procedure. 1 ml of YFHJ 22 µM in 20 mM Tris-HCl buffer (pH 7) containing 150 mM NaCl was reacted for one hour in the dark with 75 µl of AlexaFluor 532 1.23 mM dissolved in DMSO. It is important to have a big excess of the probe to ensure a complete 1:1 reaction, still maintaining the volume of DMSO in the solution in the range of 5-10% respect to the total volume, to prevent denaturation of the protein.

The labelled protein (YFHJ*) was purified from the free fluorophore using a size exclusion chromatographic column (PD10 column, Amersham Biosciences).

0.550 ml of protein solution was loaded in the column, which was subsequently washed with 3.55 ml of freṣh buffer. The first 2.25 ml of buffer were discarded, while the remaining 1.3 ml, which contained the desired product, were collected. After solvent evaporation, we obtained a concentration of 79.6 µM of YFHJ*, as measured from the absorbance of AlexaFluor 532 at 532 nm ($\varepsilon$=75000).

Dissociation constants were determined at 20°C in 20 mM Tris-HCl buffer (pH 8) containing 150 mM NaCl and 20 mM β-mercapto ethanol using two different approaches. On the one hand, we used direct fluorescence intensity (and polarisation) titrations of YFHJ labelled with the Alexa Fluor 532 fluorescence probe, using a SPEX FluoroMax fluorimeter with excitation at 465 nm and emission at 570 nm. On the other hand, dissociation constants for unlabelled YFHJ were determined using a fluorescence competition assay in which the non-fluorescent protein was

used to displace YFHJ* from its complex with IscS. The fluorescence titration was carried out adding 10-20 µl aliquots of a solution containing IscS 207 µM and YFHJ* 15.9 µM to a 15.9 µM YFHJ* solution. The presence of YFHJ* into the titrating solution allowed us to keep the concentration of YFHJ* constant over the experiment. The competition assay was carried out adding YFHJ 230 µM to a solution containing IscS 69 µM and YFHJ* 3.9 µM.

The direct fluorometric titrations of labelled YFHJ* with IscS were fit to the following equation:

$$F_{obs} = F_{YFHJ*}\big[YFHJ*\big] + F_{IscS/YFHJ*}\big[IscS/YFHJ*\big] \qquad (8.9)$$

where the F values are the molar fluorescence intensities. A value for the dissociation constant ($K_{d(YFHJ*)}$) was obtained from a non-linear least squares fit of this equation with concentrations calculated by solving:

$$\big[IscS/YFHJ*\big]^2 - Kd_{(YFHJ*)} + YFHJ*_T + IscS_T\big[IscS/YFHJ*\big] + YFHJ*_T IscS_T \qquad (8.10)$$

where the subscript T denotes total concentrations. We also included correction factors ($X_{YFHJ*}$ and $X_{IscS}$) in the fitting equation to account for errors in the actual concentrations (e.g., actual concentration of IscS = $IscS_T X_{IscS}$) .

In the displacement assays a value for the dissociation constant of unlabelled ($K_{d(YFHJ)}$) was obtained from a non-linear least squares fit to equation (8.10) with the concentrations calculated by solving:

$$\big(IscS\big)^3 + \big(-IscS + Kd_{(YFHJ)} + YFHJ*_T + YFHJ_T\big)\big(IscS\big)^2 +$$
$$\big(-IscS_T Kd_{(YFHJ)} - IscS_T Kd_{(YFHJ*)} + Kd_{(YFHJ)} Kd_{(YFHJ*)} + YFHJ*_T Kd_{(YFHJ)}\big)\big(IscS\big) \qquad (8.11)$$
$$-IscS_T Kd_{(YFHJ)} Kd_{(YFHJ*)} = 0$$

The value of $K_{d(YFHJ*)}$ was kept constant in the analysis at the value determined from the direct titration.

In the case of the polarisation, measurements of the polarisation values were converted to fractional saturation (assuming the formation of a 1:1 complex with both monomeric or dimeric IscS) by correcting for the observed intensity change. In fact for a general fluorescent compound L that binds P to form the complex PL (stoichiometry 1:1), we have:

164

$$P_{obs} = \frac{P_{PL}F_{PL}[PL] + P_L F_L [L]}{F_{PL}[PL] + F_L [L]}$$

(8.12)

The variation in fluorescence intensity is given by:

$$R = \frac{F_{PL}}{F_L}$$

(8.13)

If we define the fractional saturation as:

$$f_{sat} = \frac{[PL]}{[L_{tot}]}$$

(8.14)

we have that

$$\frac{[L]}{[Ltot]} = 1 - f_{sat}$$

(8.15)

Upon substitution of expressions (8.13), (8.14) and (8.15) into equation (8.12), we obtain:

$$f_{sat} = \frac{P_{obs} - P_L}{P_{PL}R - P_L - P_{obs}R + P_{obs}}$$

(8.16)

We obtained fractional saturations from equation (8.16) using for the free YFHJ* a value of polarisation of 0.185 $(P_L)$ and for the complex an estimated polarisation of 0.345 $(P_{PL})$. R in our case was 1.77 ± 0.03.

# CONCLUSIONS

Understanding the processes that lead to iron homeostasis in cells and tissues is one of the big challenges in modern biomedicine. Large evidence links iron metabolism to neuronal function and neurodegeneration: Parkinson's and Alzheimer's diseases are both associated with iron accumulation in the brain and Friedreich's ataxia with mitochondrial iron overload, especially in heart and neuronal tissues. Many types of anemia are caused by lack of iron in the tissues, and other less well known pathologies, such as neuroferritinopathy, are related to an unbalance in the iron metabolism. The urgent need of new treatments for iron deficiency or iron overload disorders explains the importance of the studies in this area [4].

Two of the proteins involved in the complex mechanism of iron regulation were the objects of this thesis: CyaY the bacterial orthologue of frataxin, whose deficiency is responsible for Friedreich's ataxia [94], and YFHJ, a hypothetical protein whose gene is located in the Iron Sulphur Cluster operon of many prokaryotes and some eukaryotes, which is thought to be genetically linked to CyaY [24].

To investigate the iron binding site of CyaY we performed NMR chemical shift perturbation analyses. We observed that both iron (II) and iron (III) bind to the same surface. Their effect is particularly evident, even at a low ion/CyaY ratio, for a group of four acidic residues that cluster on the negative conserved patch located between $\alpha 1$ and $\beta 1$, identified as the main iron binding site. Our data are consistent with the presence of secondary weaker binding sites as well, in accord with previous reports. We carried out similar experiments on a mutated form of the protein, where the residues from the main binding site were substituted by basic amino acids. We observed that binding of iron (II) was completely abolished, while binding of iron (III) was retained at the secondary sites. It has been reported that yeast and bacterial frataxin aggregation is triggered by the presence of iron [19,67]; however, at the iron concentrations used in this study and under the experimental conditions employed, CyaY is mainly monomeric.

It's worth noting that a binding site formed exclusively by carboxylates is a new feature for iron binding proteins: iron in fact is generally chelated by histidines and cysteines or hosted in *ad hoc* cavities. This could explain the protein's low iron affinity measured *in vitro* [16,188].

The specificity of the binding sites is also very low: we carried out NMR titrations with many metal ions and we observed that, apart from the intrinsic differences due to the peculiar electronic properties of the single ions, the effects on the mainly affected resonances are similar.

The low iron affinity and specificity suggest that under our experimental conditions iron binding to CyaY alone cannot have biological significance. A third co-factor, probably a protein, is likely to be required *in vivo* if iron binding and delivery is the main function of the protein. The presence of a highly conserved hydrophobic patch on the surface of the protein supports this hypothesis [71].

Our results provide the structural basis for discriminating between the current working hypotheses on frataxin's role. The identification of the residues involved in iron binding is in fact important for future mutagenesis experiments, in that their deletion should cause severe impairment in the iron homeostasis of the host cells if iron binding is the main function of the protein.

In the second part of the thesis, the structure of YFHJ from *E. Coli* has been discussed, and functional studies have been carried out.

Using traditional solution NMR experiments, the 3D fold of YFHJ was determined: the macromolecule is an all helical protein, consisting of three helices connected by long loops. Despite its folding similarity with transcriptional factors, its highly acidic character suggests a different role within the Fe-S cluster biosynthesis mechanism.

Genomic analyses have identified the co-occurrence of YFHJ with IscS, a 90 kDa homodimer, in all the genomes containing the open reading frame of YFHJ. The interaction of the two proteins was already observed in a two-hybrid screen assay [151], and the direct evidence of a complex formation was provided in this study by NMR. The two proteins form a very large complex in solution, as inferred by the disappearance of the HSQC signals. We followed the titration of labelled YFHJ with unlabelled IscS, mapping the most affected residues on the surface of YFHJ. We attempted at characterising the stoichiometry and affinity binding of the complex with mass spectrometry and fluorescence, but we did not obtain conclusive results, our data being compatible with a 1:1 or a 1:2 complex. Nevertheless, supposing that one molecule of YFHJ binds to one monomeric subunit of IscS, for symmetry reasons, we obtained a $K_d$ of 9.9±0.8 μM using the fluorescence polarisation technique.

We also carried out chemical shift perturbation analyses on YFHJ using iron (II) and iron (III). As with CyaY, both ions influence the same residues in slightly different manners. It is interesting to note that the residues affected by iron correspond to those first influenced by IscS. This suggests that YFHJ and IscS interact in an iron-mediated manner.

168

Although we could not give definitive answers on the biological function of the protein in the cells, our work provides an important basis for further studies. For example, NMR investigations can be undertaken to pinpoint the exact role of the metal ion, in both its oxidation states, in the interaction between YFHJ and IscS and mutational analyses could explain the relative importance of the affected residues on the complex formation. Moreover, it will be interesting to analyse the possible influence of YFHJ on the IscS-IscU complex formation and on the enzymatic activity of IscS, together with the role played by iron in both cases. Finally, the observation of some common features between CyaY and YFHJ, i.e. the behaviour in the presence of iron and the highly acidic surface potential, and their complementary genetic distribution in inferior Eukaryotes, allows us to formulate the fascinating hypothesis that frataxin might substitute the function of YFHJ in higher organisms. This is only a speculation at the moment, but surely it deserves further investigations, for the deep impact that it would have on the understanding of Fe-S clusters' synthesis and of Friedreich's ataxia.

# Bibliography

1. Jahnke W, Widmer H: **Protein NMR in biomedical research**. *Cellular and Molecular Life Sciences (CMLS)* 2004, **61**:580-599.

2. Bonvin AM, Boelens R, Kaptein R: **NMR analysis of protein interactions**. *Current Opinion in Chemical Biology* 2005, **9**:501-508.

3. Wüthrich K: *NMR of proteins and nucleic acids*: John Wiley & sons; 1986.

4. Hentze MW, Muckenthaler MU, Andrews NC: **Balancing acts: molecular control of mammalian iron metabolism**. *Cell* 2004, **117**:286-297.

5. Kiley PJ, Beinert H: **The role of Fe-S proteins in sensing and regulation in bacteria**. *Curr Opin Microbiol* 2003, **6**:181-185.

6. Puccio H: **Recent advances in the molecular pathogenesis of Friedreich ataxia**. *Hum. Mol. Genet.* 2000, **9**:887-892.

7. Delatycki MB, Williamson R, Forrest SM: **Friedreich ataxia: an overview**. *J Med Genet* 2000, **37**:1-8.

8. Campuzano V, Montermini L, Lutz Y, Cova L, Hindelang C, Jiralerspong S, Trottier Y, Kish S, Faucheux B, Trouillas P, et al.: **Frataxin is reduced in Friedreich ataxia patients and is associated with mitochondrial membranes**. *Hum. Mol. Genet.* 1997, **6**:1771-1780.

9. Bulteau AL, O'Neill HA, Kennedy MC, Ikeda-Saito M, Isaya G, Szweda LI: **Frataxin acts as an iron chaperone protein to modulate mitochondrial aconitase activity**. *Science* 2004, **305**:242-245.

10. Gerber J, Lill R: **Biogenesis of iron-sulfur proteins in eukaryotes: components, mechanism and pathology**. *Mitochondrion* 2002, **2**:71-86.

11. Huynen MA, Snel B, Bork P, Gibson TJ: **The phylogenetic distribution of frataxin indicates a role in iron-sulfur cluster protein assembly**. *Hum. Mol. Genet.* 2001, **10**:2463-2468.

12. Lesuisse E, Santos R, Matzanke BF, Knight SAB, Camadro J-M, Dancis A: **Iron use for haeme synthesis is under control of the yeast frataxin homologue (Yfh1)**. *Hum. Mol. Genet.* 2003, **12**:879-889.

13. Park S, Gakh O, O'Neill HA, Mangravita A, Nichol H, Ferreira GC, Isaya G: **Yeast Frataxin Sequentially Chaperones and Stores Iron by Coupling Protein Assembly with Iron Oxidation**. *J. Biol. Chem.* 2003, **278**:31340-31351.

14. Ramazzotti A, Vanmansart V, Foury F: **Mitochondrial functional interactions between frataxin and Isu1p, the iron-sulfur cluster scaffold protein, in Saccharomyces cerevisiae**. *FEBS Lett* 2004, **557**:215-220.

15. Stehling O, Elsasser H-P, Bruckel B, Muhlenhoff U, Lill R: **Iron-sulfur protein maturation in human cells: evidence for a function of frataxin**. *Hum. Mol. Genet.* 2004, **13**:3007-3015.

16. Yoon T, Cowan JA: **Iron-sulfur cluster biosynthesis. Characterization of frataxin as an iron donor for assembly of [2Fe-2S] clusters in ISU-type proteins**. *J Am Chem Soc* 2003, **125**:6078-6084.

17. Yoon T, Cowan JA: **Frataxin-mediated iron delivery to ferrochelatase in the final step of heme biosynthesis**. *J Biol Chem* 2004, **279**:25943-25946.

18. Park S, Gakh O, Mooney SM, Isaya G: **The Ferroxidase Activity of Yeast Frataxin**. *J. Biol. Chem.* 2002, **277**:38589-38595.

19. Adamec J, Rusnak F, Owen WG, Naylor S, Benson LM, Gacy AM, Isaya G: **Iron-dependent self-assembly of recombinant yeast frataxin: implications for Friedreich ataxia**. *Am J Hum Genet* 2000, **67**:549-562.

20. Aloria K, Schilke B, Andrew A, Craig EA: **Iron-induced oligomerization of yeast frataxin homologue Yfh1 is dispensable in vivo**. *EMBO Reports* 2004, **5**:1096-1101.

21. Musco G, Stier G, Kolmerer B, Adinolfi S, Martin S, Frenkiel T, Gibson T, Pastore A: **Towards a structural understanding of Friedreich's ataxia: the solution structure of frataxin**. *Structure* 2000, **8**:695-707.

22. Li DS, Ohshima K, Jiralerspong S, Bojanowski MW, Pandolfo M: **Knock-out of the cyaY gene in Escherichia coli does not affect cellular iron content and sensitivity to oxidants**. *FEBS Letters* 1999, **456**:13-16.

23. Adinolfi S, Nair M, Politou A, Bayer E, Martin S, Temussi P, Pastore A: **The factors governing the thermal stability of frataxin orthologues: how to increase a protein's stability**. *Biochemistry* 2004, **43**:6511-6518.

24. Zheng L, Cash VL, Flint DH, Dean DR: **Assembly of iron-sulfur clusters. Identification of an iscSUA-hscBA-fdx gene cluster from Azotobacter vinelandii**. *J Biol Chem* 1998, **273**:13264-13272.

25. Gardner KH, Kay LE: **The use of 2H, 13C, 15N multidimensional NMR to study the structure and dynamics of proteins**. *Annual Review of Biophysics and Biomolecular Structure* 1998, **27**:357-406.

26. Pervushin K, Riek R, Wider G, Wüthrich K: **Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates**

an avenue to NMR structures of very largebiological macromolecules in solution. *PNAS* 1997, **94**:12366-12371.

27. Palmer AG: **NMR characterization of the dynamics of biomacromolecules**. *Chem. Rev.* 2004, **104**:3623-3640.

28. Levitt MH: *Spin Dynamics : Basics of Nuclear Magnetic Resonance*. Chichester: John Wiley & sons, LTD; 2005.

29. Cerf C: **NMR spectroscopy: from quantum mechanics to protein spectra**. *Concepts in Magnetic Resonance* 1997, **9**:17-41.

30. Cavanagh J, Fairbrother WJ, Palmer AG, Skelton NJ: *Protein NMR spectroscopy. Principle and practice*. New York: Academic Press; 1996.

31. Sattler M, Schleucher J, Griesinger C: **Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients**. *Progress in Nuclear Magnetic Resonance Spectroscopy* 1999, **34**:93-158.

32. Vuister GW, Bax A: **Quantitative J correlation: a new approach for measuring homonuclear three-bond J(HNHa) coupling constants in 15N-enriched proteins**. *Journal of the American Chemical Society* 1993, **115**:7772-7777.

33. Evans JNS: *Biomolecular NMR spectroscopy*. Oxford: Oxford University Press; 1995.

34. Keeler, J: **Understanding NMR spectroscopy** on World Wide Web URL: http://www-keeler.ch.cam.ac.uk

35. Grzesiek S: **Notes on relaxation and dynamics**. In *EMBO practical course on NMR*; *Heidelberg*: 2003.

36. Keeler J, Clowes RT, Davis AL, Laue ED: **Pulsed-field gradients: Theory and practice**. In *Methods in Enzymology*, edn Volume 239. Edited by: Academic Press; 1994:145-207.

37. Piotto M, Saudek V, Sklenar V: **Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions**. *Journal of biomolecular NMR* 1992, **2**:661-665.

38. Mandal PK, A. M: **A comprehensive discussion of HSQC and HMQC pulse sequences**. *Concepts in Magnetic Resonance* 2004, **20A(1)**:1-23.

39. Yamazaki T, Forman-Kay JD, Kay LE: **Two-dimensional NMR experiments for correlating $^{13}$Cβ and $^{1}$H δ/ε chemical shifts of aromatic residues in $^{13}$C-labeled proteins via scalar couplings**. *Journal of the American Chemical Society* 1993, **115**:11054-11055.

40. Lesk AM: *Introduction to bionformatics*: Oxford University Press; 2002.

41. Boxevanis AD, Oulette BFF (Ed): *A practical guide to analysis of genes and proteins*: John Wiley and Sons; 1998.

42. Dayhoff MO, Schwartz, RM, Orcutt, BC: *A model of evolutionary change in proteins*. Edited by Dayhoff MO. Washington: National Biomedical Research Foundation; 1978.

43. Altschul SF, Gish W, Miller W, Meyers EW, Lipman DJ: **Basic Local Alignment Search Tool**. *Journal of Molecular Biology* 1990, **215**:403-410.

44. Higgins DG, Thompson JD, Gibson TJ: **Using CLUSTAL for multiple sequence alignments**. In *Methods in Enzymology*, edn Volume 266. Edited by: Academic Press; 1996:383-402.

45. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG: **The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools**. *Nucleic Acids Res* 1997, **25**:4876-4882.

46. Wishart DS, Sykes BD: **The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data**. *J Biomol NMR* 1994, **4**:171-180.

47. Wishart DS, Sykes BD: **Chemical shifts as a tool for structure determination**. *Methods Enzymol* 1994, **239**:363-392.

48. Wishart DS, Sykes BD, Richards FM: **The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy**. *Biochemistry* 1992, **31**:1647-1651.

49. Cornilescu G, Delaglio F, Bax A: **Protein backbone angle restraints from searching a database for chemical shift and sequence homology**. *J Biomol NMR* 1999, **13**:289-302.

50. Brunger AT, Adams PD, Clore GM, L. DW, Gros P, Grosse-Kunstleve RW, Jiang J-S, Kuszewski JJ, Nilges M, Pannu NS, et al.: **Crystallography & NMR System: A New Software Suite for Macromolecular Structure Determination**. *Acta Cryst.* 1998, **54D**:905-921.

51. Nilges M, Macias MJ, O'Donoghue SI, Oschkinat H: **Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from beta-spectrin**. *J Mol Biol* 1997, **269**:408-422.

52. Linge JP, Habeck M, Rieping W, Nilges M: **Correction of spin diffusion during iterative automated NOE assignment**. *Journal of Magnetic Resonance* 2004, **167**:334-342.

53. Linge JP, O'Donoghue SI, Nilges M: **Automated assignment of ambiguous nuclear overhauser effects with ARIA**. *Methods in Enzymology* 2001, **339**:71-90.

54. Linge JP, Habeck M, Rieping W, Nilges M: **ARIA: automated NOE assignment and NMR structure calculation**. *Bioinformatics* 2003, **19**:315-316.

55. Nilges M: **Calculation of protein structures with ambiguous distance restraints. Automated assignment of ambiguous NOE crosspeaks and disulphide connectivities**. *J Mol Biol* 1995, **245**:645-660.

56. Nilges M: **Structure calculation from NMR data**. *Curr Opin Struct Biol* 1996, **6**:617-623.

57. Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM: **AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR**. *Journal of biomolecular NMR* 1996, **8**:477-486.

58. Laskowski RA, MacArthur, MW, Moss, DS, Thornton, JM: **PROCHECK: a program to check the stereochemical quality of protein structure**. *J. Appl. Cryst.* 1993, **26**:283-291.

59. Vriend G: **WHAT IF: a molecular modelling and drug design program**. *J. Mol. Graph* 1990, **8**:52-56.

60. Holm L, Sander C: **Dali: a network tool for protein structure comparison**. *Trends Biochem Sci* 1995, **20**:478-480.

61. Campunzano V, Montermini, L., Molto, MD, et al: **Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion**. *Science* 1996, **271**:1423-1427.

62. Epplen C, Epplen JrT, Frank G, Miterski B, Santos EJM, Schols L: **Differential stability of the $(GAA)_n$ tract in the Friedreich ataxia (STM7) gene**. *Human Genetics* 1997, **99**:834-836.

63. Sakamoto N, Chastain PD, Parniewski P, Ohshima K, Pandolfo M, Griffith JD, Wells RD: **Sticky DNA: Self-Association Properties of Long GAA-TTC Repeats in R.R.Y Triplex Structures from Friedreich's Ataxia**. *Molecular Cell* 1999, **3**:465-475.

64. Cossee M, Durr A, Schmitt, M et al: **Friedreich's ataxia: point mutations and clinical presentation of compound heterozygotes**. *Ann Neurol* 1999, **45**:200-206.

65. Branda S, Yang Z, Chew A, Isaya G: **Mitochondrial intermediate peptidase and the yeast frataxin homolog together maintain mitochondrial iron homeostasis in Saccharomyces cerevisiae**. *Hum. Mol. Genet.* 1999, **8**:1099-1110.

66. Koutnikova H, Campuzano V, Koenig M: **Maturation of wild-type and mutated frataxin by the mitochondrial processing peptidase**. *Hum. Mol. Genet.* 1998, **7**:1485-1489.

67. Adinolfi S, Trifuoggi M, Politou AS, Martin S, Pastore A: **A structural approach to understanding the iron-binding properties of phylogenetically different frataxins**. *Hum. Mol. Genet.* 2002, **11**:1865-1877.

68. Cho S-J, Lee MG, Yang JK, Lee JY, Song HK, Suh SW: **Crystal structure of Escherichia coli CyaY protein reveals a previously unidentified fold for the evolutionarily conserved frataxin family**. *PNAS* 2000, **97**:8932-8937.

69. He Y, Alam SL, Proteasa SV, Zhang Y, Lesuisse E, Dancis A, Stemmler TL: **Yeast frataxin solution structure, iron binding, and ferrochelatase interaction**. *Biochemistry* 2004, **43**:16254-16262.

70. Dhe-Paganon S, Shigeta R, Chi Y-I, Ristow M, Shoelson SE: **Crystal Structure of Human Frataxin**. *J. Biol. Chem.* 2000, **275**:30753-30756.

71. Nair M, Adinolfi S, Pastore C, Kelly G, Temussi P, Pastore A: **Solution structure of the bacterial frataxin ortholog, CyaY: mapping the iron binding sites**. *Structure (Camb)* 2004, **12**:2037-2048.

72. Cossee M, Puccio H, Gansmuller A, Koutnikova H, Dierich A, LeMeur M, Fischbeck K, Dolle P, Koenig M: **Inactivation of the Friedreich ataxia mouse gene leads to early embryonic lethality without iron accumulation**. *Hum. Mol. Genet.* 2000, **9**:1219-1226.

73. Santos MM, Ohshima K, Pandolfo M: **Frataxin deficiency enhances apoptosis in cells differentiating into neuroectoderm**. *Hum. Mol. Genet.* 2001, **10**:1935-1944.

74. Radisky DC, Babcock MC, Kaplan J: **The Yeast Frataxin Homologue Mediates Mitochondrial Iron Efflux. EVIDENCE FOR A MITOCHONDRIAL IRON CYCLE**. *J. Biol. Chem.* 1999, **274**:4497-4499.

75. Foury F, Talibi D: **Mitochondrial Control of Iron Homeostasis. A GENOME WIDE ANALYSIS OF GENE EXPRESSION IN A YEAST FRATAXIN-DEFICIENT STRAIN**. *J. Biol. Chem.* 2001, **276**:7762-7768.

76. Muhlenhoff U, Richhardt N, Ristow M, Kispal G, Lill R: **The yeast frataxin homolog Yfh1p plays a specific role in the maturation of cellular Fe/S proteins**. *Hum. Mol. Genet.* 2002, **11**:2025-2036.

77. Chen OS, Hemenway S, Kaplan J: **Inhibition of Fe-S cluster biosynthesis decreases mitochondrial iron export: Evidence that Yfh1p affects Fe-S cluster synthesis**. *PNAS* 2002, **99**:12321-12326.

78. Wong A, Yang J, Cavadini P, Gellera C, Lonnerdal B, Taroni F, Cortopassi G: **The Friedreich's ataxia mutation confers cellular sensitivity to oxidant stress which is rescued by chelators of iron and calcium and inhibitors of apoptosis**. *Hum. Mol. Genet.* 1999, **8**:425-430.

79. Foury F: **Low iron concentration and aconitase deficiency in a yeast frataxin homologue deficient strain**. *FEBS Lett* 1999, **456**:281-284.

80. Lodi R, Cooper JM, Bradley JL, Manners D, Styles P, Taylor DJ, Schapira AHV: **Deficit of in vivo mitochondrial ATP production in patients with Friedreich ataxia**. *PNAS* 1999, **96**:11492-11495.

81. Ristow M, Pfister MF, Yee AJ, Schubert M, Michael L, Zhang C-Y, Ueki K, Michael MD, II, Lowell BB, Kahn CR: **Frataxin activates mitochondrial energy conversion and oxidative phosphorylation**. *PNAS* 2000, **97**:12239-12243.

82. Tan G, Napoli E, Taroni F, Cortopassi G: **Decreased expression of genes involved in sulfur amino acid metabolism in frataxin-deficient cells**. *Hum. Mol. Genet.* 2003, **12**:1699-1711.

83. Lange H, Muhlenhoff U, Denzel M, Kispal G, Lill R: **The Heme Synthesis Defect of Mutants Impaired in Mitochondrial Iron-Sulfur Protein Biogenesis Is Caused by Reversible Inhibition of Ferrochelatase**. *J. Biol. Chem.* 2004, **279**:29101-29108.

84. Duby G, Foury F, Ramazzotti A, Herrmann J, Lutz T: **A non-essential function for yeast frataxin in iron-sulfur cluster assembly**. *Hum. Mol. Genet.* 2002, **11**:2635-2643.

85. Gerber J, Muhlenhoff U, Lill R: **An interaction between frataxin and Isu1/Nfs1 that is crucial for Fe/S cluster synthesis on Isu1**. *EMBO Reports* 2003, **4**:906-911.

86. Bou-Abdallah F, Adinolfi S, Pastore A, Laue TM, Dennis Chasteen N, Nair M, Politou A, Bayer E, Martin S, Temussi P, et al.: **Iron binding and oxidation kinetics in frataxin CyaY of Escherichia coli. The factors governing the thermal stability of frataxin orthologues: how to increase a protein's stability**. *J Mol Biol* 2004, **341**:605-615.

87. Cavadini P, O'Neill HA, Benada O, Isaya G: **Assembly and iron-binding properties of human frataxin, the protein deficient in Friedreich ataxia**. *Hum. Mol. Genet.* 2002, **11**:217-227.

88. Hustedt EJ, Beth AH: **Nitroxide spin-spin interactions: applications to protein structure and dynamics**. *Annu Rev Biophys Biomol Struct.* 1999, **28**:129-153.

89. Bertini I, Luchinat C, Parigi G, Pierattelli R: **NMR spectroscopy of paramagnetic metalloproteins**. *Chembiochem.* 2005, **6**:1536-1549.

90. Geraldes CF, Luchinat C: **Lanthanides as shift agents and relaxation agents in the structure of proteins and nucleic acids**. In *Metal Ions in Biological Systems*. Edited by: Marcel Dekker AG; 2003:513-588. vol 40.]

91. Ubbink M, Worral JAR, Canters GW, Groenen EJJ, Huber M: **Paramagnetic resonance of biological metal centers**. *Annu. Rev. Biophys. Biomol. Struct.* 2002, **31**:393-422.

92. Bertini I, Luchinat, C.: *Coor. Chem. Rev.*, vol 150; 1996.

93. Evans CH: *Biochemistry of the Lanthanides*. New York: Plenum Press; 1990.

94. Campuzano V, Montermini L, Molto MD, Pianese L, Cossee M, Cavalcanti F, Monros E, Rodius F, Duclos F, Monticelli A, et al.: **Friedreich's Ataxia: Autosomal Recessive Disease Caused by an Intronic GAA Triplet Repeat Expansion**. *Science* 1996, **271**:1423-1427.

95. Pandolfo M: **Frataxin deficiency and mitochondrial dysfunction**. *Mitochondrion* 2002, **2**:87-93.

96. Harrison PM, Arosio P: **The ferritins: molecular properties, iron storage function and cellular regulation**. *Biochim Biophys Acta.* 1996, **1275**:161-203.

97. Lange SJ, Que L, Jr.: **Oxygen activating nonheme iron enzymes**. *Curr Opin Chem Biol.* 1998, **2**:159-172.

98. Petrat F, de Groot H, Rauen U: **Subcellular distribution of chelatable iron: a laser scanning microscopic study in isolated hepatocytes and liver endothelial cells**. *Biochem J.* 2001, **356**:61-69.

99. Johnson D, Dean DR: **Structure, Function, and Formation of Biological Iron-Sulfur Clusters**. *Annu Rev Biochem* 2004.

100. Frazzon J, Dean DR: **Formation of iron-sulfur clusters in bacteria: an emerging field in bioinorganic chemistry**. *Curr Opin Chem Biol* 2003, **7**:166-173.

101. Huber C, Wachtershauser G: **Activated acetic acid by carbon fixation on (Fe,Ni)S under primordial conditions**. *Science* 1997, **276**:245-247.

102. Beinert H, Holm RH, Munck E: **Iron-sulfur clusters: nature's modular, multipurpose structures**. *Science* 1997, **277**:653-659.

103. Glaser T, Hedman B, Hodgson KO, Solomon EI: **Ligand K-Edge X-ray Absorption Spectroscopy: A Direct Probe of Ligand-Metal Covalency**. *Acc. Chem. Res.* 2000, **33**:859-868.

104. Beinert H, Kennedy MC, Stout CD: **Aconitase as Iron-Sulfur Protein, Enzyme, and Iron-Regulatory Protein**. *Chem. Rev.* 1996, **96**:2335-2374.

105. Beinert H, Kiley PJ: **Fe-S proteins in sensing and regulatory functions**. *Curr Opin Chem Biol* 1999, **3**:152-157.

106. Demple B, Ding H, Jorgensen M: **Escherichia coli SoxR protein: Sensor/transducer of oxidative stress and nitric oxide** *Methods in Enzymology* 2002, **348**:355-364.

107. Flint DH, Allen RM: **Iron-Sulfur Proteins with Nonredox Functions**. *Chem Rev* 1996, **96**:2315-2334.

108. Malkin R, Rabinowitz JC: *Biochem Biophys Res Commun* 1966, **23**:822-827.

109. Johnson MK: **Iron-sulfur proteins: new roles for old clusters**. *Curr Opin Chem Biol* 1998, **2**:173-181.

110. Jacobson MR, Cash VL, Weiss MC, Laird NF, Newton WE, Dean DR: **Biochemical and genetic analysis of the nifUSVWZM cluster from Azotobacter vinelandii**. *Mol Gen Genet* 1989, **219**:49-57.

111. Zheng L, White RH, Cash VL, Dean DR: **Mechanism for the desulfurization of L-cysteine catalyzed by the nifS gene product**. *Biochemistry* 1994, **33**:4714-4720.

112. Fu W, Jack RF, Morgan TV, Dean DR, Johnson MK: **nifU gene product from Azotobacter vinelandii is a homodimer that contains two identical [2Fe-2S] clusters**. *Biochemistry* 1994, **33**:13455-13463.

113. Yuvaniyama P, Agar JN, Cash VL, Johnson MK, Dean DR: **NifS-directed assembly of a transient [2Fe-2S] cluster within the NifU protein**. *Proc Natl Acad Sci U S A* 2000, **97**:599-604.

114. Dos Santos PC, Smith AD, Frazzon J, Cash VL, Johnson MK, Dean DR: **Iron-sulfur cluster assembly: NifU-directed activation of the nitrogenase Fe protein**. *J Biol Chem* 2004, **279**:19705-19711.

115. Takahashi Y, Tokumoto U: **A third bacterial system for the assembly of iron-sulfur clusters with homologs in archaea and plastids**. *J Biol Chem* 2002, **277**:28380-28383.

116. Rangachari K, Davis CT, Eccleston JF, Hirst EM, Saldanha JW, Strath M, Wilson RJ: **SufC hydrolyzes ATP and interacts with SufB from Thermotoga maritima**. *FEBS Lett* 2002, **514**:225-228.

117. Loiseau L, Ollagnier-de-Choudens S, Nachin L, Fontecave M, Barras F: **Biogenesis of Fe-S cluster by the bacterial Suf system: SufS and SufE form a new type of cysteine desulfurase**. *J Biol Chem* 2003, **278**:38352-38359.

118. Outten FW, Wood MJ, Munoz FM, Storz G: **The SufE protein and the SufBCD complex enhance SufS cysteine desulfurase activity as part of a sulfur transfer pathway for Fe-S cluster assembly in Escherichia coli**. *J Biol Chem* 2003, **278**:45713-45719.

119. Zheng M, Wang X, Templeton LJ, Smulski DR, LaRossa RA, Storz G: **DNA microarray-mediated transcriptional profiling of the Escherichia coli response to hydrogen peroxide**. *J Bacteriol* 2001, **183**:4562-4570.

120. Tokumoto U, Takahashi Y: **Genetic analysis of the isc operon in Escherichia coli involved in the biogenesis of cellular iron-sulfur proteins**. *J Biochem (Tokyo)* 2001, **130**:63-71.

121. Schwartz CJ, Giel JL, Patschkowski T, Luther C, Ruzicka FJ, Beinert H, Kiley PJ: **IscR, an Fe-S cluster-containing transcription factor, represses expression of Escherichia coli genes encoding Fe-S cluster assembly proteins**. *Proc Natl Acad Sci U S A* 2001, **98**:14895-14900.

122. Kato S, Mihara H, Kurihara T, Takahashi Y, Tokumoto U, Yoshimura T, Esaki N: **Cys-328 of IscS and Cys-63 of IscU are the sites of disulfide bridge formation in a covalently bound IscS/IscU complex: implications for the mechanism of iron-sulfur cluster assembly**. *Proc Natl Acad Sci U S A* 2002, **99**:5948-5952.

123. Cupp-Vickery JR, Urbina H, Vickery LE: **Crystal structure of IscS, a cysteine desulfurase from Escherichia coli**. *J Mol Biol* 2003, **330**:1049-1059.

124. Hwang DM, Dempsey A, Tan KT, Liew CC: **A modular domain of NifU, a nitrogen fixation cluster protein, is highly conserved in evolution**. *J Mol Evol* 1996, **43**:536-540.

125. Campuzano V, Montermini L, Lutz Y, Cova L, Hindelang C, Jiralerspong S, Trottier Y, Kish SJ, Faucheux B, Trouillas P, et al.: **Frataxin is reduced in Friedreich ataxia patients and is associated with mitochondrial membranes**. *Hum Mol Genet* 1997, **6**:1771-1780.

126. Ding H, Clark RJ, Ding B: **IscA mediates iron delivery for assembly of iron-sulfur clusters in IscU under the limited accessible free iron conditions**. *J Biol Chem* 2004, **279**:37499-37504.

127. Mansy SS, Wu G, Surerus KK, Cowan JA: **Iron-sulfur cluster biosynthesis. Thermatoga maritima IscU is a structured iron-sulfur cluster assembly protein**. *J Biol Chem* 2002, **277**:21397-21404.

128. Bertini I, Cowan JA, Del Bianco C, Luchinat C, Mansy SS: **Thermotoga maritima IscU. Structural characterization and dynamics of a new class of metallochaperone**. *J Mol Biol* 2003, **331**:907-924.

129. Adinolfi S, Rizzo F, Masino L, Nair M, Martin SR, Pastore A, Temussi PA: **Bacterial IscU is a well folded and functional single domain protein**. *Eur J Biochem* 2004, **271**:2093-2100.

130. Nuth M, Yoon T, Cowan JA: **Iron-sulfur cluster biosynthesis: characterization of iron nucleation sites for assembly of the [2Fe-2S]2+ cluster core in IscU proteins**. *J Am Chem Soc* 2002, **124**:8774-8775.

131. Agar JN, Krebs C, Frazzon J, Huynh BH, Dean DR, Johnson MK: **IscU as a scaffold for iron-sulfur cluster biosynthesis: sequential assembly of [2Fe-2S] and [4Fe-4S] clusters in IscU**. *Biochemistry* 2000, **39**:7856-7862.

132. Ramelot TA, Cort JR, Goldsmith-Fischman S, Kornhaber GJ, Xiao R, Shastry R, Acton TB, Honig B, Montelione GT, Kennedy MA: **Solution NMR structure of the iron-sulfur cluster assembly protein U (IscU) with zinc bound at the active site**. *J Mol Biol* 2004, **344**:567-583.

133. Liu J, Oganesyan N, Shin DH, Jancarik J, Yokota H, Kim R, Kim SH: **Structural characterization of an iron-sulfur cluster assembly protein IscU in a zinc-bound form**. *Proteins* 2005, **59**:875-881.

134. Smith AD, Agar JN, Johnson KA, Frazzon J, Amster IJ, Dean DR, Johnson MK: **Sulfur transfer from IscS to IscU: the first step in iron-sulfur cluster biosynthesis**. *J Am Chem Soc* 2001, **123**:11103-11104.

135. Krebs C, Agar JN, Smith AD, Frazzon J, Dean DR, Huynh BH, Johnson MK: **IscA, an alternate scaffold for Fe-S cluster biosynthesis**. *Biochemistry* 2001, **40**:14069-14080.

136. Ding H, Clark RJ: **Characterization of iron binding in IscA, an ancient iron-sulphur cluster assembly protein**. *Biochem J* 2004, **379**:433-440.

137. Ding B, Smith ES, Ding H: **Mobilization of the iron center in IscA for the iron-sulfur cluster assembly in IscU**. *Biochem J* 2005.

138. Ollagnier-de-Choudens S, Mattioli T, Takahashi Y, Fontecave M: **Iron-sulfur cluster assembly: characterization of IscA and evidence for a specific and functional complex with ferredoxin**. *J Biol Chem* 2001, **276**:22604-22607.

139. Wu G, Mansy SS, Hemann C, Hille R, Surerus KK, Cowan JA: **Iron-sulfur cluster biosynthesis: characterization of Schizosaccharomyces pombe Isa1**. *J Biol Inorg Chem* 2002, **7**:526-532.

140. Wollenberg M, Berndt C, Bill E, Schwenn JD, Seidler A: **A dimer of the FeS cluster biosynthesis protein IscA from cyanobacteria binds a [2Fe2S] cluster between two protomers and transfers it to [2Fe2S] and [4Fe4S] apo proteins**. *Eur J Biochem* 2003, **270**:1662-1671.

141. Bilder PW, Ding H, Newcomer ME: **Crystal structure of the ancient, Fe-S scaffold IscA reveals a novel protein fold**. *Biochemistry* 2004, **43**:133-139.

142. Cupp-Vickery JR, Silberg JJ, Ta DT, Vickery LE: **Crystal structure of IscA, an iron-sulfur cluster assembly protein from Escherichia coli**. *J Mol Biol* 2004, **338**:127-137.

143. Lelivelt MJ, Kawula TH: **Hsc66, an Hsp70 homolog in Escherichia coli, is induced by cold shock but not by heat shock**. *J Bacteriol* 1995, **177**:4900-4907.

144. Vickery LE, Silberg JJ, Ta DT: **Hsc66 and Hsc20, a new heat shock cognate molecular chaperone system from Escherichia coli**. *Protein Sci* 1997, **6**:1047-1056.

145. Silberg JJ, Vickery LE: **Kinetic characterization of the ATPase cycle of the molecular chaperone Hsc66 from Escherichia coli**. *J Biol Chem* 2000, **275**:7779-7786.

146. Silberg JJ, Hoff KG, Tapley TL, Vickery LE: **The Fe/S assembly protein IscU behaves as a substrate for the molecular chaperone Hsc66 from Escherichia coli**. *J Biol Chem* 2001, **276**:1696-1700.

147. Silberg JJ, Tapley TL, Hoff KG, Vickery LE: **Regulation of the HscA ATPase reaction cycle by the co-chaperone HscB and the iron-sulfur cluster assembly protein IscU**. *J Biol Chem* 2004, **279**:53924-53931.

148. Hoff KG, Ta DT, Tapley TL, Silberg JJ, Vickery LE: **Hsc66 substrate specificity is directed toward a discrete region of the iron-sulfur cluster template protein IscU**. *J Biol Chem* 2002, **277**:27353-27359.

149. Schiffler B, Bureik M, Reinle W, Muller EC, Hannemann F, Bernhardt R: **The adrenodoxin-like ferredoxin of Schizosaccharomyces pombe mitochondria**. *J Inorg Biochem* 2004, **98**:1229-1237.

150. Wu SP, Cowan JA: **Iron-sulfur cluster biosynthesis. A comparative kinetic analysis of native and Cys-substituted ISA-mediated [2Fe-2S]2+ cluster transfer to an apoferredoxin target**. *Biochemistry* 2003, **42**:5784-5791.

151. Tokumoto U, Nomura S, Minami Y, Mihara H, Kato S, Kurihara T, Esaki N, Kanazawa H, Matsubara H, Takahashi Y: **Network of protein-protein interactions among iron-sulfur cluster assembly proteins in Escherichia coli**. *J Biochem (Tokyo)* 2002, **131**:713-719.

152. Lange H, Kaut A, Kispal G, Lill R: **A mitochondrial ferredoxin is essential for biogenesis of cellular iron-sulfur proteins**. *Proc Natl Acad Sci U S A* 2000, **97**:1050-1055.

153. Takahashi Y, Nakamura M: **Functional assignment of the ORF2-iscS-iscU-iscA-hscB-hscA-fdx-ORF3 gene cluster in the assembly of Fe-S clusters in Escherichia Coli**. *J. Biochemistry (Tokio)* 1999, **126**:917-926.

154. Shimomura Y, Takahashi Y, Kakuta Y, Fukuyama K: **Crystal structure of Escherichia coli YfhJ protein, a member of the ISC machinery involved in assembly of iron-sulfur clusters**. *Proteins* 2005, **60**:566-569.

155. Berova N, Nakanishi K, Woody RW: *Circular Dichroism Principles and Applications*. New York: Wiley-VHC; 2000.

156. Johnson WC: **Protein Secondary Structure and Circular Dichroism: A Practical Guide**. *Proteins: Structure, Function, and Genetics* 1990, **7**:205-214.

157. Moffitt W: *J.Chem.Phys.* 1956, **25**:467-478.

158. Woody RW: **Circular dichroism**. *Methods Enzymol* 1995, **246**:34-71.

159. Johnson WC: **Analyzing protein circular dichroism spectra for accurate secondary structures**. *Proteins: Structure, Function, and Genetics* 1999, **35**:307-312.

160. Sreerama N, Venyaminov S, Woody R: **Estimation of the number of alpha-helical and beta-strand segments in proteins using circular dichroism spectroscopy**. *Protein Sci* 1999, **8**:370-380.

161. Provencher SW, Glockner, J.: **Estimation of globular protein secondary structure from circular dichroism**. *Biochemistry* 1981, **20**:33-37.

162. Zhang W, Smithgall TE, Gmeiner WH: **Three-dimensional structure of the Hck SH2 domain in solution**. *Journal of Biomolecular NMR* 1997, **10**:263-272.

163. Lipari G, Szabo A: **Nuclear magnetic resonance relaxation in nucleic acid fragments: models for internal motion**. *Biochemistry* 1981, **20**:6250-6256.

164. Nilges M: **Ambiguous distance data in the calculation of NMR structures**. *Fold Des* 1997, **2**:S53-57.

165. Koradi R, Billeter M, Wüthrich K: **MOLMOL: a program for display and analysis of macromolecular structures**. *J Mol Graph* 1996, **14**:51-55, 29-32.

166. Pabo CO, Sauer RT: **Protein-DNA recognition**. *Annu Rev Biochem* 1984, **53**:293-321.

167. Kikuchi J, Iwahara J, Kigawa T, Murakami Y, Okazaki T, Yokoyama S: **Solution structure determination of the two DNA-binding domains in the Schizosaccharomyces pombe Abp1 protein by a combination of dipolar coupling and diffusion anisotropy restraints**. *J Biomol NMR* 2002, **22**:333-347.

168. Allen M, Friedler A, Schon O, Bycroft M: **The Structure of an FF Domain from Human HYPA/FBP11**. *Journal of Molecular Biology* 2002, **323**:411-416.

169. Marsden I, Jin C, Liao X: **Structural changes in the region directly adjacent to the DNA-binding helix highlight a possible mechanism to explain the observed changes in the sequence-specific binding of winged helix proteins**. *J Mol Biol* 1998, **278**:293-299.

170. Devany M, Kotharu NP, Matsuo H: **Solution NMR structure of the C-terminal domain of the human protein DEK**. *Protein Sci* 2004, **13**:2252-2259. Epub 2004 Jul 2256.

171. Carty SM, Goldstrohm AC, Sune C, Garcia-Blanco MA, Greenleaf AL: **Protein-interaction modules that organize nuclear function: FF domains of CA150 bind**

the phosphoCTD of RNA polymerase II. *Proc Natl Acad Sci U S A* 2000, **97**:9015-9020.

172. Zhang O, Kay LE, Olivier JP, Forman-Kay JD: **Backbone 1H and 15N resonance assignments of the N-terminal SH3 domain of drk in folded and unfolded states using enhanced-sensitivity pulsed field gradient NMR techniques**. *Journal of Biomolecular NMR* 1994, **4**:845-858.

173. Meissner A, Sørensen OW: **A Sequential HNCA NMR Pulse Sequence for Protein Backbone Assignment**. *Journal of Magnetic Resonance* 2001, **150**:100-104.

174. Bax A, Ikura M: **An efficient 3D NMR technique for correlating the proton and 15N backbone amide resonances with the alpha-carbon of the preceding residue in uniformly 15N/13C enriched proteins**. *Journal of biomolecular NMR* 1991, **1**:99-104.

175. Muhandiram DR, Kay LE: **Gradient-enhanced triple-resonance three-dimensional NMR experiments with improved sensitivity**. *Journal of Magnetic Resonance, Series B* 1994, **103**:203-216.

176. Grzesiek S, Ikura M, Clore GM, Gronenborn AM, Bax A: **A 3D triple-resonance NMR technique for qualitative measurement of carbonyl-Hb J couplings in isotopically enriched proteins**. *Journal of Magnetic Resonance (1969-1992)* 1992, **96**:215-221.

177. Grzesiek S, Bax A: **An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins**. *Journal of Magnetic Resonance (1969-1992)* 1992, **99**:201-207.

178. Kay LE, Xu GY, Singer AU, Muhandiram DR, Forman-Kay JD: **A gradient-enhanced HCCH-TOCSY experiment for recording side-chain proton and carbon-13 correlations in water samples of proteins**. *Journal of Magnetic Resonance, Series B* 1993, **101**:333-337.

179. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A: **NMRPipe: a multidimensional spectral processing system based on UNIX pipes**. *J Biomol NMR* 1995, **6**:277-293.

180. Bartels C, Xia T-h, Billeter M, Guentert P, Wüthrich K: **The program XEASY for computer-supported NMR spectral analysis of biological macromolecules**. *Journal of Biomolecular NMR* 1995, **6**:1-10.

181. Farrow NA, Zhang O, Forman-Kay JD, Kay LE: **A heteronuclear correlation experiment for simultaneous determination of $^{15}$N longitudinal decay and chemical exchange rates of systems in slow equilibrium**. *Journal of biomolecular NMR* 1994, **4**:727-734.

182. Old RW, Primrose SB: *Principles of gene manipulation. An introduction to genetic engineering* edn Fifth edition: Blackwell Science; 1998.

183. Ueda EKM, Gout PW, Morganti L: **Current and prospective applications of metal ion-protein binding**. *Journal of Chromatography A* 2003, **988**:1-23.

184. Veenstra TD: **Electrospray ionization mass spectrometry in the study of biomolecular non-covalent interactions**. *Biophysical Chemistry* 1999, **79**:63-79.

185. Heck AJR, van der Heuvel, Robert, H. H.: **Investigation of intact protein complexes by mass spectrometry**. *MAss Spectrometry Reviews* 2004, **23**:368-389.

186. Lakowicz JR: *Principles of Fluorescence Spectroscopy* edn Second. New York and London: Plenum Press; 1999.

187. Cavadini P, Gellera C, Patel PI, Isaya G: **Human frataxin maintains mitochondrial iron homeostasis in Saccharomyces cerevisiae**. *Hum Mol Genet* 2000, **9**:2523-2530.

188. Bou-Abdallah F, Adinolfi S, Pastore A, Laue TM, Dennis Chasteen N: **Iron binding and oxidation kinetics in frataxin CyaY of Escherichia coli**. *J Mol Biol* 2004, **341**:605-615.

# Appendix A

In this appendix some pictures and tables referring to chapter 7 and 8 are reported.
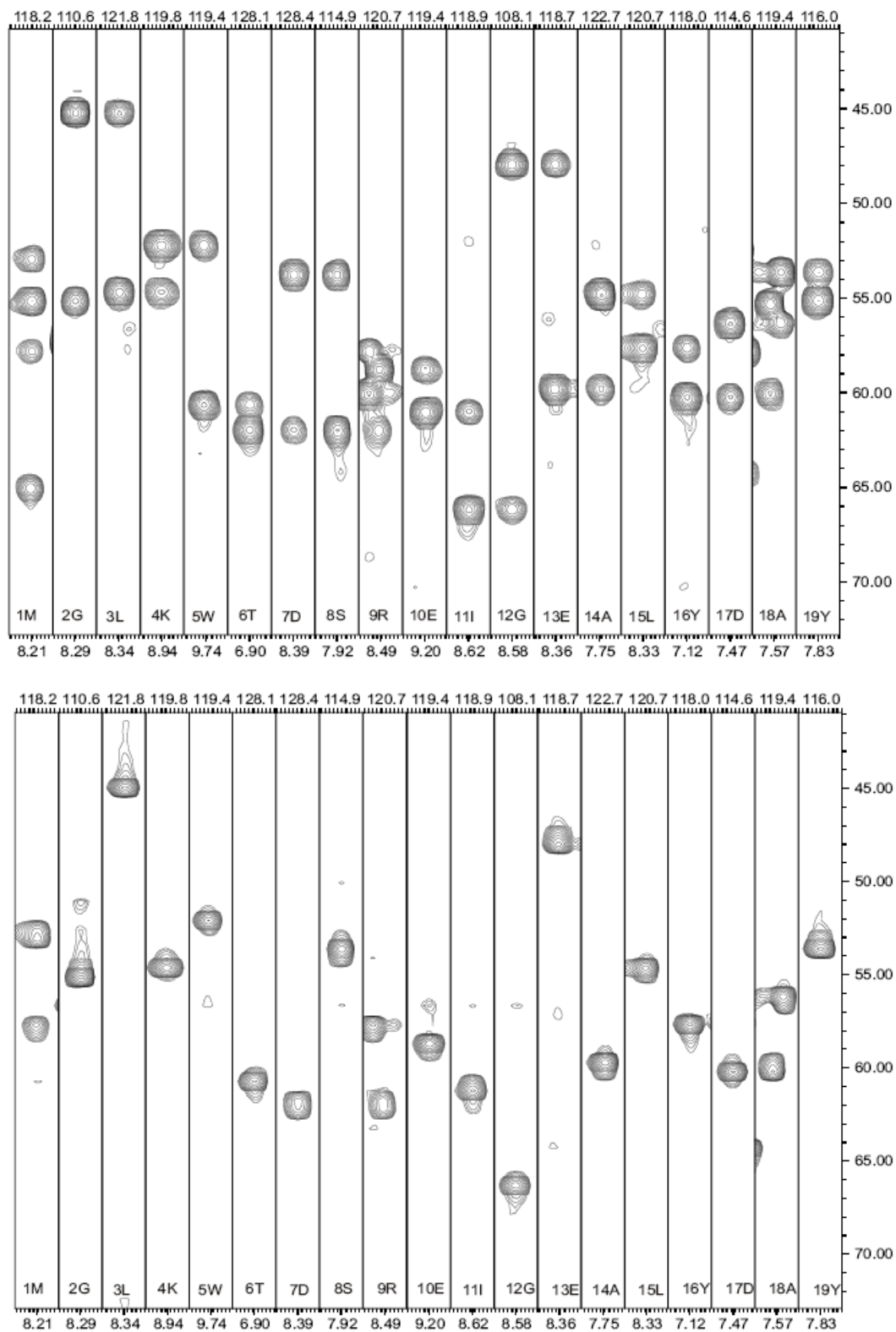


**Figure A-1: Strips of the first 19 residues taken from HNCA (top) and HNCOCA (bottom) spectra.**
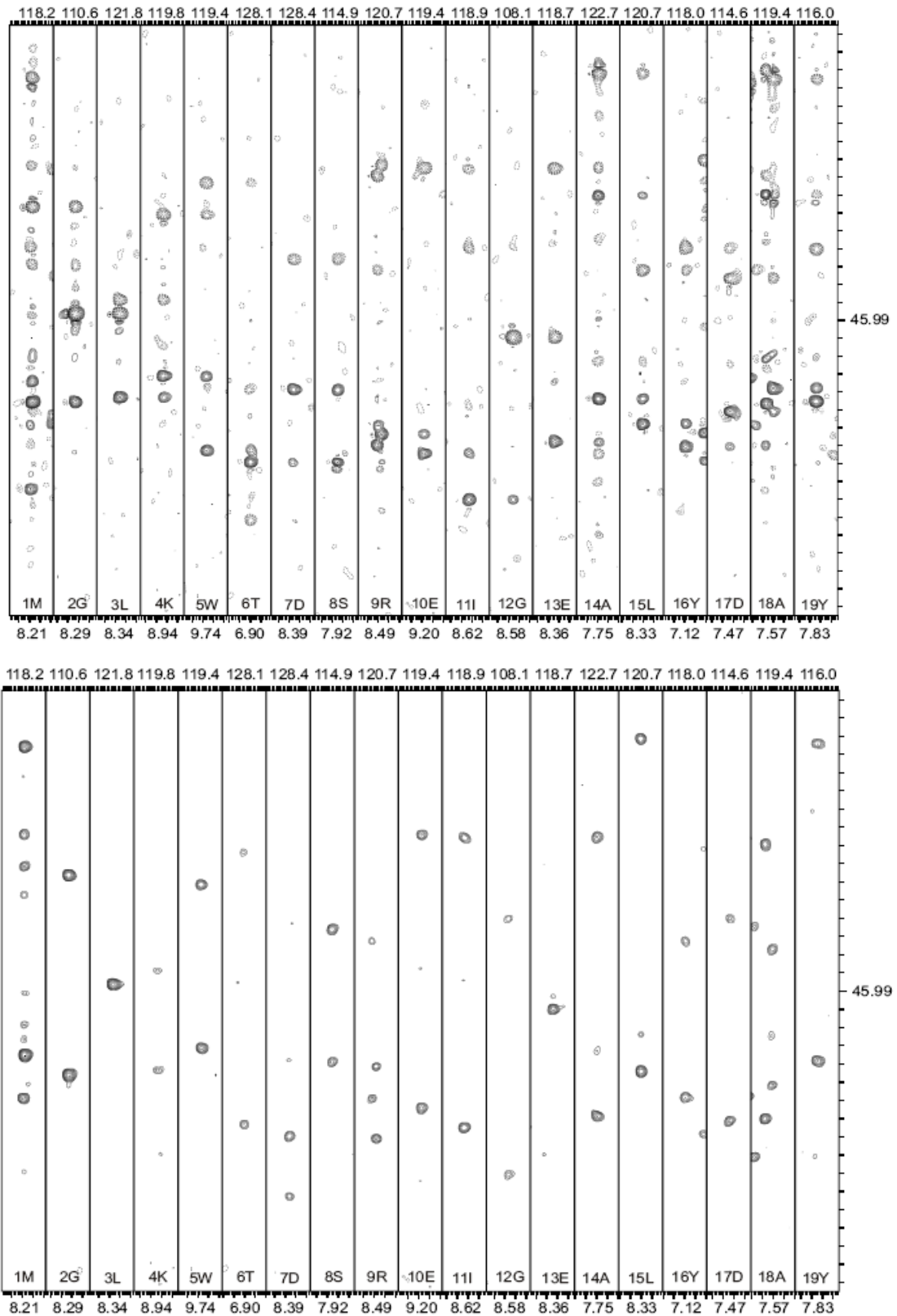
**Figure A-2: Strips of the first 19 residues taken from CBCANH (top) and CBCACONH (bottom) spectra.**
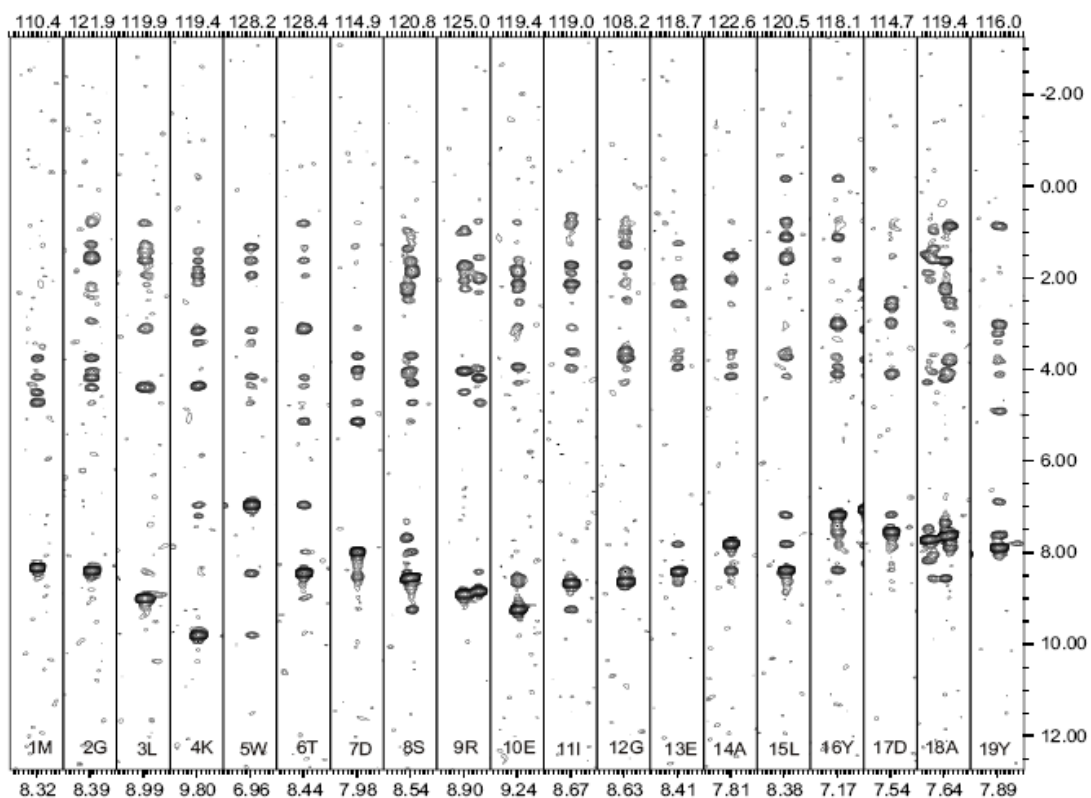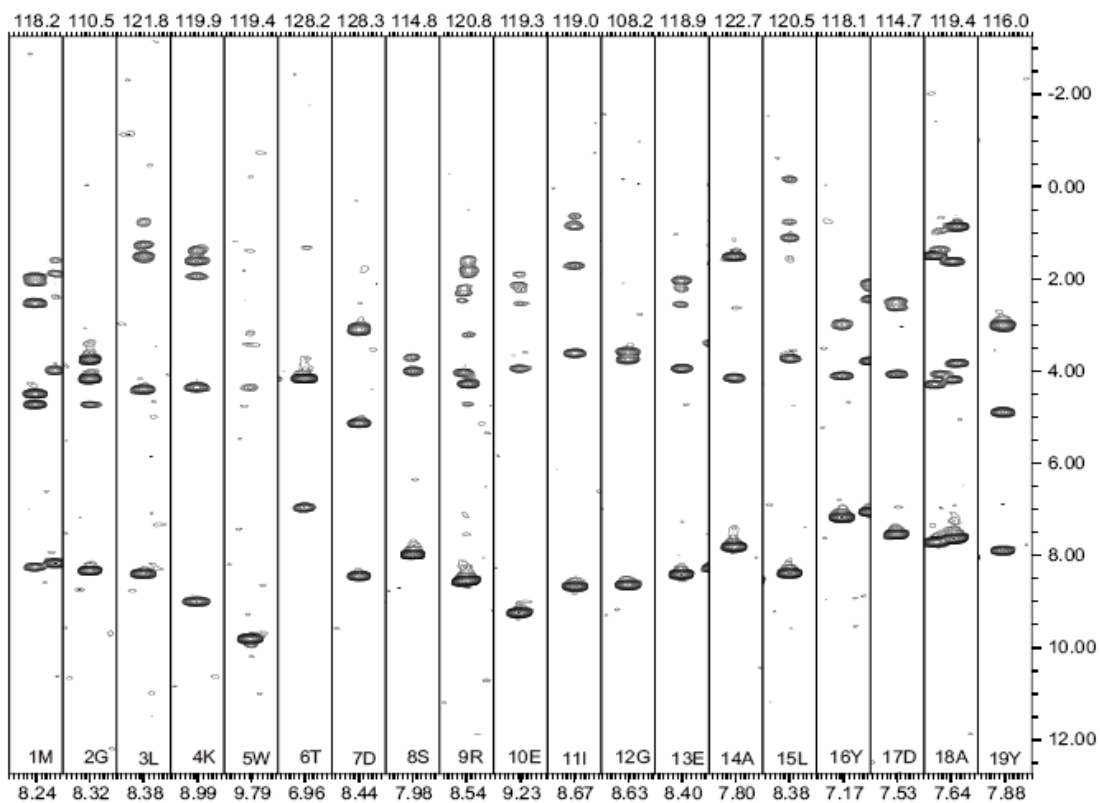
**Figure A-3: Strips of the first 19 residues taken from TOCSY-HSQC (top) and NOESY-HSQC (bottom) spectra.**

**Table I: Complete chemical shift assignment of the protein YFHJ**

| Residue type | Atom type* | Atom assignment (ppm) | Error (ppm) |
|---|---|---|---|
| MET 1 | H | 8.248 | 0.003 |
| | HA | 4.507 | 0 |
| | HB2 | 2.092 | 0.002 |
| | HB3 | 1.978 | 0.011 |
| | HG2 | 2.569 | 0.001 |
| | HG3 | 2.532 | 0.001 |
| | HE1 | 2.094 | 0 |
| | HE2 | 2.094 | 0 |
| | HE3 | 2.094 | 0 |
| | C | 175.816 | 0 |
| | CA | 55.232 | 0 |
| | CB | 33.763 | 0 |
| | CG | 31.742 | 0 |
| | CE | 16.913 | 0 |
| | N | 118.291 | 0 |
| GLY 2 | H | 8.322 | 0 |
| | HA2 | 4.179 | 0.002 |
| | HA3 | 3.768 | 0.001 |
| | C | 173.533 | 0 |
| | CA | 45.448 | 0 |
| | N | 110.647 | 0 |
| LEU 3 | H | 8.396 | 0.006 |
| | HA | 4.413 | 0.001 |
| | HB2 | 1.524 | 0.002 |
| | HB3 | 1.282 | 0 |
| | HG | 1.61 | 0.005 |
| | HD11 | 0.773 | 0.004 |
| | HD12 | 0.773 | 0.004 |
| | HD13 | 0.773 | 0.004 |
| | HD21 | 0.836 | 0.002 |
| | HD22 | 0.836 | 0.002 |
| | HD23 | 0.836 | 0.002 |
| | C | 175.161 | 0 |
| | CA | 54.689 | 0 |
| | CB | 43.818 | 0 |
| | CG | 26.425 | 0 |
| | CD1 | 25.881 | 0 |
| | CD2 | 23.164 | 0 |
| | N | 121.888 | 0 |
| LYS 4 | H | 8.998 | 0.002 |
| | HA | 4.367 | 0.001 |

| Residue type | Atom type* | Atom assignment (ppm) | Error (ppm) |
|---|---|---|---|
| | HB2 | 1.961 | 0.002 |
| | HB3 | 1.635 | 0.001 |
| | HG2 | 1.412 | 0.004 |
| | HG3 | 1.385 | 0 |
| | HD2 | 1.479 | 0.001 |
| | HD3 | 1.443 | 0.001 |
| | HE2 | 2.934 | 0.011 |
| | HE3 | 2.879 | 0.006 |
| | C | 178.322 | 0 |
| | CA | 52.243 | 0 |
| | CB | 34.306 | 0 |
| | CG | 24.522 | 0 |
| | CD | 27.784 | 0 |
| | CE | 42.459 | 0 |
| | N | 119.866 | 0.004 |
| TRP 5 | H | 9.805 | 0.004 |
| | HA | 4.369 | 0.005 |
| | HB2 | 3.422 | 0.004 |
| | HB3 | 3.171 | 0 |
| | HD1 | 7.204 | 0.002 |
| | HE1 | 10.725 | 0.001 |
| | HE3 | 7.464 | 0.001 |
| | HZ2 | 7.338 | 0.001 |
| | HZ3 | 6.892 | 0.002 |
| | HH2 | 6.675 | 0.004 |
| | C | 176.327 | 0 |
| | CA | 60.667 | 0 |
| | CB | 30.773 | 0 |
| | CD1 | 128.359 | 0 |
| | CE3 | 120.391 | 0 |
| | CZ2 | 115.469 | 0 |
| | CZ3 | 122.031 | 0 |
| | CH2 | 124.572 | 0 |
| | N | 119.409 | 0.016 |
| | NE1 | 128.419 | 0.015 |
| THR 6 | H | 6.972 | 0.002 |
| | HA | 4.177 | 0.001 |
| | HB | 4.625 | 0.003 |
| | HG21 | 1.339 | 0.002 |
| | HG22 | 1.339 | 0.002 |
| | HG23 | 1.339 | 0.002 |

| Residue | Atom | Shift | Error |
|---|---|---|---|
| | C | 176.167 | 0 |
| | CA | 61.754 | 0 |
| | CB | 68.549 | 0 |
| | CG2 | 22.348 | 0 |
| | N | 128.183 | 0 |
| ASP 7 | H | 8.448 | 0.003 |
| | HA | 5.146 | 0.001 |
| | HB2 | 3.144 | 0 |
| | HB3 | 3.073 | 0.002 |
| | C | 176.247 | 0 |
| | CA | 53.873 | 0 |
| | CB | 39.198 | 0 |
| | N | 128.408 | 0 |
| SER 8 | H | 7.983 | 0.002 |
| | HA | 4.005 | 0.003 |
| | HB2 | 4.051 | 0.003 |
| | HB3 | 3.716 | 0.002 |
| | C | 176.534 | 0 |
| | CA | 62.026 | 0 |
| | CB | 62.298 | 0 |
| | N | 114.918 | 0 |
| ARG 9 | H | 8.542 | 0 |
| | HA | 4.293 | 0 |
| | HB2 | 1.921 | 0 |
| | HB3 | 1.823 | 0.002 |
| | HG2 | 1.664 | 0.002 |
| | HG3 | 1.579 | 0.008 |
| | HD2 | 3.243 | 0 |
| | HD3 | 3.231 | 0.001 |
| | C | 177.46 | 0 |
| | CA | 58.765 | 0 |
| | CB | 28.871 | 0 |
| | CG | 26.697 | 0 |
| | CD | 42.731 | 0 |
| | N | 120.759 | 0.014 |
| GLU 10 | H | 9.239 | 0 |
| | HA | 3.961 | 0.001 |
| | HB2 | 2.146 | 0.004 |
| | HB3 | 1.916 | 0.004 |
| | HG2 | 2.558 | 0 |
| | HG3 | 2.273 | 0.001 |
| | C | 180.86 | 0 |
| | CA | 60.939 | 0 |
| | CB | 29.414 | 0 |
| | CG | 37.567 | 0 |
| | N | 119.415 | 0 |
| ILE 11 | H | 8.662 | 0.007 |
| | HA | 3.628 | 0.001 |

| Residue | Atom | Shift | Error |
|---|---|---|---|
| | HB | 1.733 | 0.002 |
| | HG12 | 2.163 | 0.002 |
| | HG13 | 0.661 | 0.002 |
| | HG21 | 0.894 | 0.003 |
| | HG22 | 0.894 | 0.003 |
| | HG23 | 0.894 | 0.003 |
| | HD11 | 0.818 | 0.005 |
| | HD12 | 0.818 | 0.005 |
| | HD13 | 0.818 | 0.005 |
| | C | 177.492 | 0 |
| | CA | 66.374 | 0 |
| | CB | 38.111 | 0 |
| | CG1 | 31.045 | 0 |
| | CG2 | 20.174 | 0 |
| | CD1 | 13.924 | 0 |
| | N | 118.966 | 0.003 |
| GLY 12 | H | 8.637 | 0.001 |
| | HA2 | 3.611 | 0.001 |
| | HA3 | 3.763 | 0.001 |
| | C | 175.241 | 0 |
| | CA | 48.166 | 0 |
| | N | 108.168 | 0.015 |
| GLU 13 | H | 8.416 | 0.001 |
| | HA | 3.949 | 0.001 |
| | HB2 | 2.054 | 0.001 |
| | HB3 | 2.054 | 0.001 |
| | HG2 | 2.574 | 0.004 |
| | HG3 | 2.216 | 0.003 |
| | C | 178.801 | 0 |
| | CA | 59.852 | 0 |
| | CB | 29.142 | 0 |
| | CG | 36.48 | 0 |
| | N | 118.965 | 0 |
| ALA 14 | H | 7.814 | 0 |
| | HA | 4.165 | 0.001 |
| | HB1 | 1.539 | 0.002 |
| | HB2 | 1.539 | 0.002 |
| | HB3 | 1.539 | 0.002 |
| | C | 181.275 | 0 |
| | CA | 54.96 | 0 |
| | CB | 18.272 | 0 |
| | N | 122.787 | 0 |
| LEU 15 | H | 8.385 | 0.009 |
| | HA | 3.74 | 0.003 |
| | HB2 | 1.129 | 0.001 |
| | HB3 | -0.153 | 0.003 |
| | HG | 0.768 | 0.001 |
| | HD11 | 1.63 | 0.005 |

| | HD12 | 1.63 | 0.005 | | HB3 | 3.036 | 0.018 |
|---|---|---|---|---|---|---|---|
| | HD13 | 1.63 | 0.005 | | HD1 | 6.884 | 0.001 |
| | HD21 | 0.795 | 0.003 | | HD2 | 6.884 | 0.001 |
| | HD22 | 0.795 | 0.003 | | HE1 | 6.279 | 0.002 |
| | HD23 | 0.795 | 0.003 | | HE2 | 6.279 | 0.002 |
| | C | 177.173 | 0 | | CA | 55.232 | 0 |
| | CA | 57.678 | 0 | | CB | 38.383 | 0 |
| | CB | 40.557 | 0 | | CD1 | 133.749 | 0 |
| | CG | 27.512 | 0 | | CD2 | 133.749 | 0 |
| | CD1 | 26.425 | 0 | | CE1 | 116.875 | 0 |
| | CD2 | 22.62 | 0 | | CE2 | 116.875 | 0 |
| | N | 120.535 | 0.011 | | N | 116.04 | 0.009 |
| TYR 16 | H | 7.171 | 0.002 | PRO 20 | HA | 4.409 | 0.001 |
| | HA | 4.121 | 0 | | HB2 | 2.077 | 0 |
| | HB2 | 3.051 | 0.004 | | HB3 | 1.852 | 0 |
| | HB3 | 2.981 | 0.004 | | HG2 | 1.771 | 0 |
| | HD1 | 7.444 | 0.006 | | HG3 | 1.89 | 0.001 |
| | HD2 | 7.444 | 0.006 | | HD2 | 3.408 | 0.002 |
| | HE1 | 7.005 | 0.004 | | HD3 | 3.241 | 0.001 |
| | HE2 | 7.005 | 0.004 | | C | 177.412 | 0 |
| | C | 176.806 | 0 | | CA | 65.016 | 0 |
| | CA | 60.396 | 0 | | CB | 31.86 | 0 |
| | CB | 38.111 | 0 | | CG | 26.968 | 0 |
| | CD1 | 133.281 | 0 | | CD | 50.069 | 0 |
| | CD2 | 133.281 | 0 | ASP 21 | H | 8.512 | 0.002 |
| | CE1 | 118.516 | 0 | | HA | 4.77 | 0.006 |
| | CE2 | 118.516 | 0 | | HB2 | 2.643 | 0 |
| | N | 118.062 | 0.012 | | HB3 | 2.78 | 0 |
| ASP 17 | H | 7.539 | 0.001 | | C | 176.055 | 0 |
| | HA | 4.083 | 0 | | CA | 53.798 | 0 |
| | HB2 | 2.633 | 0 | | CB | 40.828 | 0 |
| | HB3 | 2.503 | 0 | | N | 115.597 | 0.009 |
| | C | 177.971 | 0 | LEU 22 | H | 7.285 | 0 |
| | CA | 56.319 | 0 | | HA | 4.363 | 0.001 |
| | CB | 41.683 | 0 | | HB2 | 1.734 | 0.003 |
| | N | 114.697 | 0.008 | | HB3 | 1.542 | 0.002 |
| ALA 18 | H | 7.622 | 0 | | HG | 1.453 | 0.003 |
| | HA | 3.852 | 0 | | HD11 | 0.613 | 0.002 |
| | HB1 | 0.886 | 0.001 | | HD12 | 0.613 | 0.002 |
| | HB2 | 0.886 | 0.001 | | HD13 | 0.613 | 0.002 |
| | HB3 | 0.886 | 0.001 | | HD21 | 0.683 | 0.002 |
| | C | 177.684 | 0 | | HD22 | 0.683 | 0.002 |
| | CA | 53.798 | 0 | | HD23 | 0.683 | 0.002 |
| | CB | 18.815 | 0 | | C | 175.736 | 0 |
| | N | 119.415 | 0 | | CA | 54.145 | 0 |
| TYR 19 | H | 7.889 | 0.008 | | CB | 43.818 | 0 |
| | HA | 4.906 | 0.007 | | CG | 26.968 | 0 |
| | HB2 | 3 | 0.001 | | CD1 | 25.066 | 0 |

192

| | CD2 | 24.522 | 0 | | N | 106.375 | 0.001 |
|---|---|---|---|---|---|---|---|
| | N | 121.658 | 0.014 | **VAL 27** | H | 6.794 | 0 |
| **ASP 23** | H | 8.483 | 0 | | HA | 4.047 | 0 |
| | HA | 4.903 | 0.001 | | HB | 2.071 | 0.002 |
| | HB2 | 2.853 | 0.004 | | HG11 | 1.013 | 0.001 |
| | HB3 | 2.696 | 0 | | HG12 | 1.013 | 0.001 |
| | CA | 50.34 | 0 | | HG13 | 1.013 | 0.001 |
| | CB | 41.644 | 0 | | HG21 | 0.938 | 0.001 |
| | N | 123.917 | 0.015 | | HG22 | 0.938 | 0.001 |
| **PRO 24** | HA | 3.46 | 0.003 | | HG23 | 0.938 | 0.001 |
| | HB2 | 1.209 | 0.001 | | C | 175.577 | 0 |
| | HB3 | 0.444 | 0.001 | | CA | 62.57 | 0 |
| | HG2 | 1.475 | 0.002 | | CB | 33.219 | 0 |
| | HG3 | 0.925 | 0.002 | | CG1 | 22.62 | 0 |
| | HD2 | 3.875 | 0.002 | | CG2 | 22.077 | 0 |
| | HD3 | 3.814 | 0 | | N | 121.667 | 0.009 |
| | C | 176.726 | 0 | **ARG 28** | H | 8.907 | 0.003 |
| | CA | 63.657 | 0 | | HA | 4.499 | 0.004 |
| | CB | 30.501 | 0 | | HB2 | 1.928 | 0.001 |
| | CG | 26.697 | 0 | | HB3 | 1.752 | 0.002 |
| | CD | 50.884 | 0 | | HG2 | 1.789 | 0.007 |
| **LYS 25** | H | 8.038 | 0.001 | | HG3 | 1.698 | 0.004 |
| | HA | 3.966 | 0.002 | | HD2 | 3.247 | 0 |
| | HB2 | 1.853 | 0.005 | | HD3 | 3.189 | 0.001 |
| | HB3 | 1.778 | 0.001 | | CA | 54.417 | 0 |
| | HG2 | 1.421 | 0.003 | | CB | 31.317 | 0 |
| | HG3 | 1.421 | 0.003 | | CG | 26.968 | 0 |
| | HD2 | 1.781 | 0.001 | | CD | 43.546 | 0 |
| | HD3 | 1.721 | 0.004 | | N | 125.035 | 0 |
| | HE2 | 3.04 | 0.003 | **PHE 29** | H | 8.983 | 0 |
| | HE3 | 3.04 | 0.003 | | HA | 4.418 | 0 |
| | C | 177.811 | 0 | | HB2 | 3.197 | 0 |
| | CA | 58.765 | 0 | | HB3 | 3.197 | 0 |
| | CB | 31.588 | 0 | | HD1 | 7.285 | 0.005 |
| | CG | 25.61 | 0 | | HD2 | 7.285 | 0.005 |
| | CD | 28.871 | 0 | | HE1 | 7.39 | 0.006 |
| | CE | 41.1 | 0 | | HE2 | 7.39 | 0.006 |
| | N | 115.813 | 0.02 | | HZ | 7.305 | 0.002 |
| **THR 26** | H | 7.81 | 0.002 | | CA | 60.124 | 0 |
| | HA | 4.425 | 0.001 | | CB | 37.295 | 0 |
| | HB | 4.477 | 0 | | CD1 | 130.468 | 0 |
| | HG21 | 1.135 | 0.001 | | CD2 | 130.468 | 0 |
| | HG22 | 1.135 | 0.001 | | CE1 | 131.874 | 0 |
| | HG23 | 1.135 | 0.001 | | CE2 | 131.874 | 0 |
| | C | 174.715 | 0 | | CZ | 130 | 0 |
| | CA | 60.939 | 0 | | N | 124.186 | 0 |
| | CB | 69.907 | 0 | **THR 30** | H | 8.002 | 0 |
| | CG2 | 21.533 | 0 | | HA | 3.958 | 0.002 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | HB | 4.095 | 0.001 | | HG2 | 2.452 | 0.002 |
| | HG21 | 1.216 | 0.002 | | HG3 | 2.452 | 0.002 |
| | HG22 | 1.216 | 0.002 | | HE21 | 7.807 | 0 |
| | HG23 | 1.216 | 0.002 | | HE22 | 6.843 | 0 |
| | C | 176.63 | 0 | | C | 177.444 | 0 |
| | CA | 65.016 | 0 | | CA | 58.765 | 0 |
| | CB | 68.005 | 0 | | CB | 28.055 | 0 |
| | CG2 | 22.077 | 0 | | CG | 33.491 | 0 |
| | N | 109.709 | 0 | | N | 118.066 | 0 |
| ASP 31 | H | 6.634 | 0.007 | | NE2 | 115.352 | 0.036 |
| | HA | 4.14 | 0 | TRP 35 | H | 7.84 | 0.004 |
| | HB2 | 2.249 | 0.004 | | HA | 4.669 | 0.002 |
| | HB3 | 1.8 | 0.003 | | HB2 | 3.282 | 0.002 |
| | C | 175.529 | 0 | | HB3 | 3.076 | 0.003 |
| | CA | 56.047 | 0 | | HD1 | 7.146 | 0.014 |
| | CB | 40.557 | 0 | | HE1 | 9.84 | 0.001 |
| | N | 121.438 | 0 | | HE3 | 7.173 | 0 |
| MET 32 | H | 7.539 | 0.003 | | HZ2 | 7.321 | 0.002 |
| | HA | 3.631 | 0.001 | | HZ3 | 7.04 | 0 |
| | HB2 | 2.187 | 0.002 | | HH2 | 7.168 | 0.002 |
| | HB3 | 2.067 | 0.002 | | C | 180.094 | 0 |
| | HG2 | 2.803 | 0.007 | | CA | 57.678 | 0 |
| | HG3 | 2.7 | 0.006 | | CB | 28.871 | 0 |
| | HE1 | 2.278 | 0 | | CD1 | 124.609 | 0 |
| | HE2 | 2.278 | 0 | | CE3 | 118.75 | 0 |
| | HE3 | 2.278 | 0 | | CZ2 | 114.063 | 0 |
| | C | 176.916 | 0 | | CZ3 | 121.328 | 0 |
| | CA | 59.037 | 0 | | CH2 | 124.609 | 0 |
| | CB | 34.034 | 0 | | N | 118.516 | 0 |
| | CG | 32.675 | 0 | | NE1 | 127.284 | 0 |
| | CE | 18 | 0 | ILE 36 | H | 8.271 | 0.01 |
| | N | 116.497 | 0.016 | | HA | 3.406 | 0.005 |
| HIS 33 | H | 8.225 | 0 | | HB | 1.69 | 0.004 |
| | HA | 3.99 | 0.001 | | HG12 | 2.094 | 0.003 |
| | HB2 | 3.134 | 0.001 | | HG13 | 0.721 | 0.005 |
| | HB3 | 3.134 | 0.001 | | HG21 | 1.144 | 0.003 |
| | HD2 | 6.945 | 0.003 | | HG22 | 1.144 | 0.003 |
| | HE1 | 7.867 | 0.006 | | HG23 | 1.144 | 0.003 |
| | C | 176.231 | 0 | | HD11 | -0.005 | 0.002 |
| | CA | 61.754 | 0 | | HD12 | -0.005 | 0.002 |
| | CB | 31.045 | 0 | | HD13 | -0.005 | 0.002 |
| | CD2 | 118.75 | 0 | | C | 176.79 | 0 |
| | CE1 | 138.905 | 0 | | CA | 65.016 | 0 |
| | N | 116.043 | 0 | | CB | 38.111 | 0 |
| GLN 34 | H | 7.055 | 0 | | CG1 | 30.501 | 0 |
| | HA | 3.796 | 0 | | CG2 | 16.913 | 0 |
| | HB2 | 2.218 | 0 | | CD1 | 18 | 0 |
| | HB3 | 2.111 | 0 | | N | 118.74 | 0 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **CYS 37** | **H** | 7.238 | 0.002 | | **HB3** | 2.421 | 0 |
| | **HA** | 3.155 | 0.001 | | **C** | 175.704 | 0 |
| | **HB2** | 2.885 | 0.005 | | **CA** | 54.42 | 0 |
| | **HB3** | 2.698 | 0.002 | | **CB** | 40.557 | 0 |
| | **C** | 175.497 | 0 | | **N** | 113.347 | 0.007 |
| | **CA** | 62.298 | 0 | **PHE 42** | **H** | 7.602 | 0 |
| | **CB** | 26.425 | 0 | | **HA** | 4.044 | 0.002 |
| | **N** | 114.245 | 0.003 | | **HB2** | 3.039 | 0.002 |
| **ASP 38** | **H** | 7.385 | 0 | | **HB3** | 2.96 | 0.004 |
| | **HA** | 4.698 | 0 | | **HD1** | 6.708 | 0.008 |
| | **HB2** | 2.881 | 0 | | **HD2** | 6.708 | 0.008 |
| | **HB3** | 2.724 | 0 | | **HE1** | 6.116 | 0.007 |
| | **C** | 177.077 | 0 | | **HE2** | 6.116 | 0.007 |
| | **CA** | 54.109 | 0 | | **HZ** | 5.617 | 0.001 |
| | **CB** | 41.768 | 0 | | **C** | 175.353 | 0 |
| | **N** | 116.499 | 0.016 | | **CA** | 60.667 | 0 |
| **LEU 39** | **H** | 7.739 | 0.003 | | **CB** | 40.285 | 0 |
| | **HA** | 4.197 | 0.003 | | **CD1** | 131.64 | 0 |
| | **HB2** | 2.067 | 0.007 | | **CD2** | 131.64 | 0 |
| | **HB3** | 1.573 | 0.005 | | **CE1** | 130.703 | 0 |
| | **HG** | 2.359 | 0.005 | | **CE2** | 130.703 | 0 |
| | **HD11** | 0.829 | 0 | | **CZ** | 129.531 | 0 |
| | **HD12** | 0.829 | 0 | | **N** | 121.434 | 0.011 |
| | **HD13** | 0.829 | 0 | **ASP 43** | **H** | 8.385 | 0.005 |
| | **HD21** | 0.791 | 0.009 | | **HA** | 4.678 | 0 |
| | **HD22** | 0.791 | 0.009 | | **HB2** | 2.279 | 0.004 |
| | **HD23** | 0.791 | 0.009 | | **HB3** | 2.175 | 0.004 |
| | **C** | 179.535 | 0 | | **C** | 172.655 | 0 |
| | **CA** | 55.776 | 0 | | **CA** | 52.786 | 0 |
| | **CB** | 41.372 | 0 | | **CB** | 42.459 | 0 |
| | **CG** | 25.881 | 0 | | **N** | 129.309 | 0.003 |
| | **CD1** | 26.697 | 0 | **ASP 44** | **H** | 5.295 | 0.001 |
| | **CD2** | 22.348 | 0 | | **HA** | 4.67 | 0.001 |
| | **N** | 124.141 | 0.015 | | **HB2** | 1.824 | 0.002 |
| **GLU 40** | **H** | 8.845 | 0.005 | | **HB3** | 2.125 | 0.002 |
| | **HA** | 4.011 | 0.001 | | **C** | 174.204 | 0 |
| | **HB2** | 2.068 | 0.002 | | **CA** | 50.612 | 0 |
| | **HB3** | 1.968 | 0.004 | | **CB** | 45.448 | 0 |
| | **HG2** | 2.32 | 0.003 | | **N** | 118.965 | 0 |
| | **HG3** | 2.32 | 0.003 | **ASP 45** | **H** | 9.951 | 0.003 |
| | **C** | 177.604 | 0 | | **HA** | 5.02 | 0 |
| | **CA** | 59.309 | 0 | | **HB2** | 3.052 | 0.001 |
| | **CB** | 29.686 | 0 | | **HB3** | 2.877 | 0.002 |
| | **CG** | 36.208 | 0 | | **CA** | 51.699 | 0 |
| | **N** | 124.811 | 0 | | **CB** | 42.187 | 0 |
| **ASP 41** | **H** | 8.409 | 0 | | **N** | 121.663 | 0 |
| | **HA** | 4.721 | 0 | **PRO 46** | **HA** | 4.87 | 0.001 |
| | **HB2** | 2.955 | 0.002 | | **HB2** | 2.915 | 0.007 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | HB3 | 2.355 | 0.007 | | CA | 52.514 | 0 |
| | HG2 | 2.348 | 0.002 | | CB | 39.47 | 0 |
| | HG3 | 2.348 | 0.002 | | N | 112.668 | 0.003 |
| | HD2 | 4.64 | 0.002 | | ND2 | 114.019 | 0.009 |
| | HD3 | 4.232 | 0.001 | GLU 51 | H | 8.915 | 0 |
| | C | 177.524 | 0 | | HA | 3.711 | 0.002 |
| | CA | 65.831 | 0 | | HB2 | 2.032 | 0.005 |
| | CB | 33.763 | 0 | | HB3 | 2.032 | 0.005 |
| | CG | 28.014 | 0 | | HG2 | 2.239 | 0 |
| | CD | 51.699 | 0 | | HG3 | 2.239 | 0 |
| GLN 47 | H | 8.612 | 0 | | C | 177.588 | 0 |
| | HA | 4.482 | 0 | | CA | 60.939 | 0 |
| | HB2 | 2.372 | 0 | | CB | 29.142 | 0 |
| | HB3 | 2.165 | 0 | | CG | 37.023 | 0 |
| | HG2 | 2.552 | 0 | | N | 118.49 | 0 |
| | HG3 | 2.437 | 0 | LYS 52 | H | 7.967 | 0 |
| | HE21 | 7.72 | 0 | | HA | 4.148 | 0.002 |
| | HE22 | 6.875 | 0 | | HB2 | 1.923 | 0.003 |
| | C | 176.518 | 0 | | HB3 | 1.859 | 0.002 |
| | CA | 56.047 | 0 | | HG2 | 1.565 | 0 |
| | CB | 28.636 | 0 | | HG3 | 1.486 | 0 |
| | CG | 34.578 | 0 | | HD2 | 1.704 | 0.002 |
| | N | 113.343 | 0.009 | | HD3 | 1.704 | 0.002 |
| | NE2 | 112.301 | 0.008 | | HE2 | 2.993 | 0 |
| ALA 48 | H | 7.84 | 0.003 | | HE3 | 2.993 | 0 |
| | HA | 4.49 | 0.002 | | C | 180.318 | 0 |
| | HB1 | 1.651 | 0.005 | | CA | 58.765 | 0 |
| | HB2 | 1.651 | 0.005 | | CB | 31.317 | 0 |
| | HB3 | 1.651 | 0.005 | | CG | 24.908 | 0 |
| | C | 176.215 | 0 | | CD | 28.327 | 0 |
| | CA | 51.971 | 0 | | CE | 42.187 | 0 |
| | CB | 19.359 | 0 | | N | 120.004 | 0.045 |
| | N | 121.438 | 0 | ILE 53 | H | 8.052 | 0.003 |
| SER 49 | H | 6.841 | 0.006 | | HA | 3.845 | 0.002 |
| | HA | 3.139 | 0.005 | | HB | 1.634 | 0.006 |
| | HB2 | 2.747 | 0.001 | | HG12 | 1.674 | 0.002 |
| | HB3 | 2.05 | 0.003 | | HG13 | 1.201 | 0.004 |
| | C | 172.703 | 0 | | HG21 | 1.02 | 0.003 |
| | CA | 57.678 | 0 | | HG22 | 1.02 | 0.003 |
| | CB | 62.842 | 0 | | HG23 | 1.02 | 0.003 |
| | N | 111.097 | 0 | | HD11 | 0.785 | 0.004 |
| ASN 50 | H | 6.62 | 0.004 | | HD12 | 0.785 | 0.004 |
| | HA | 4.497 | 0.002 | | HD13 | 0.785 | 0.004 |
| | HB2 | 2.99 | 0.001 | | C | 177.029 | 0 |
| | HB3 | 2.99 | 0.001 | | CA | 63.929 | 0 |
| | HD21 | 7.475 | 0 | | CB | 39.198 | 0 |
| | HD22 | 6.795 | 0.001 | | CG1 | 29.958 | 0 |
| | C | 174.81 | 0 | | CG2 | 18.815 | 0 |

| | | | |
|---|---|---|---|
| | CD1 | 13.108 | 0 |
| | N | 122.787 | 0 |
| LEU 54 | H | 7.681 | 0.017 |
| | HA | 4.081 | 0 |
| | HB2 | 1.376 | 0.013 |
| | HB3 | 0.979 | 0.003 |
| | HG | 1.625 | 0.005 |
| | HD11 | 0.715 | 0.003 |
| | HD12 | 0.715 | 0.003 |
| | HD13 | 0.715 | 0.003 |
| | HD21 | 0.868 | 0.002 |
| | HD22 | 0.868 | 0.002 |
| | HD23 | 0.868 | 0.002 |
| | C | 178.562 | 0 |
| | CA | 57.678 | 0 |
| | CB | 40.557 | 0 |
| | CG | 26.968 | 0 |
| | CD1 | 26.697 | 0 |
| | CD2 | 23.435 | 0 |
| | N | 119.192 | 0.005 |
| GLU 55 | H | 8.561 | 0.01 |
| | HA | 4.05 | 0.002 |
| | HB2 | 2.318 | 0.005 |
| | HB3 | 2.201 | 0 |
| | HG2 | 2.489 | 0.002 |
| | HG3 | 2.339 | 0.002 |
| | C | 177.604 | 0 |
| | CA | 59.852 | 0 |
| | CB | 29.958 | 0 |
| | CG | 36.752 | 0 |
| | N | 120.764 | 0 |
| ALA 56 | H | 7.635 | 0.002 |
| | HA | 4.201 | 0.002 |
| | HB1 | 1.638 | 0.003 |
| | HB2 | 1.638 | 0.003 |
| | HB3 | 1.638 | 0.003 |
| | C | 180.717 | 0 |
| | CA | 55.504 | 0 |
| | CB | 18.272 | 0 |
| | N | 119.408 | 0.016 |
| ILE 57 | H | 7.36 | 0 |
| | HA | 3.771 | 0.004 |
| | HB | 2.065 | 0.001 |
| | HG12 | 2.208 | 0.004 |
| | HG13 | 1.29 | 0.004 |
| | HG21 | 1.02 | 0.001 |
| | HG22 | 1.02 | 0.001 |
| | HG23 | 1.02 | 0.001 |

| | | | |
|---|---|---|---|
| | HD11 | 1.346 | 0.001 |
| | HD12 | 1.346 | 0.001 |
| | HD13 | 1.346 | 0.001 |
| | C | 177.412 | 0 |
| | CA | 65.287 | 0 |
| | CB | 39.47 | 0 |
| | CG1 | 30.773 | 0 |
| | CG2 | 19.087 | 0 |
| | CD1 | 15.554 | 0 |
| | N | 118.291 | 0 |
| LEU 58 | H | 8.606 | 0.009 |
| | HA | 4.41 | 0.003 |
| | HB2 | 2.296 | 0.006 |
| | HB3 | 1.678 | 0.001 |
| | HG | 1.73 | 0.003 |
| | HD11 | 1.05 | 0.004 |
| | HD12 | 1.05 | 0.004 |
| | HD13 | 1.05 | 0.004 |
| | HD21 | 1.073 | 0.003 |
| | HD22 | 1.073 | 0.003 |
| | HD23 | 1.073 | 0.003 |
| | C | 178.29 | 0 |
| | CA | 58.222 | 0 |
| | CB | 42.187 | 0 |
| | CG | 27.24 | 0 |
| | CD1 | 26.153 | 0 |
| | CD2 | 23.435 | 0 |
| | N | 58.147 | 0 |
| LEU 59 | H | 8.424 | 0.002 |
| | HA | 4.13 | 0.002 |
| | HB2 | 1.961 | 0.003 |
| | HB3 | 1.64 | 0 |
| | HG | 1.996 | 0.001 |
| | HD11 | 0.992 | 0.004 |
| | HD12 | 0.992 | 0.004 |
| | HD13 | 0.992 | 0.004 |
| | HD21 | 0.953 | 0.001 |
| | HD22 | 0.953 | 0.001 |
| | HD23 | 0.953 | 0.001 |
| | C | 180.078 | 0 |
| | CA | 58.493 | 0 |
| | CB | 41.644 | 0 |
| | CG | 27.082 | 0 |
| | CD1 | 25.338 | 0 |
| | CD2 | 23.164 | 0 |
| | N | 116.721 | 0.013 |
| VAL 60 | H | 7.257 | 0.005 |
| | HA | 3.878 | 0 |

| Residue | Atom | Value | Error | Residue | Atom | Value | Error |
|---|---|---|---|---|---|---|---|
|  | HB | 2.48 | 0.002 |  | CA | 58.493 | 0 |
|  | HG11 | 1.136 | 0.001 |  | CB | 42.187 | 0 |
|  | HG12 | 1.136 | 0.001 |  | CG | 27.512 | 0 |
|  | HG13 | 1.136 | 0.001 |  | CD1 | 26.425 | 0 |
|  | HG21 | 1.267 | 0.004 |  | CD2 | 24.251 | 0 |
|  | HG22 | 1.267 | 0.004 |  | N | 117.619 | 0.01 |
|  | HG23 | 1.267 | 0.004 | ASP 63 | H | 7.572 | 0 |
|  | C | 177.795 | 0 |  | HA | 4.523 | 0 |
|  | CA | 66.103 | 0 |  | HB2 | 2.891 | 0 |
|  | CB | 31.86 | 0 |  | HB3 | 2.721 | 0 |
|  | CG1 | 21.805 | 0 |  | C | 178.482 | 0 |
|  | CG2 | 21.533 | 0 |  | CA | 56.591 | 0 |
|  | N | 118.965 | 0 |  | CB | 41.224 | 0 |
| TRP 61 | H | 8.228 | 0.008 |  | N | 118.065 | 0.004 |
|  | HA | 4.497 | 0.002 | GLU 64 | H | 8.186 | 0.002 |
|  | HB2 | 3.804 | 0.001 |  | HA | 3.991 | 0.011 |
|  | HB3 | 3.339 | 0.002 |  | HB2 | 1.903 | 0.002 |
|  | HD1 | 7.579 | 0.011 |  | HB3 | 1.609 | 0.005 |
|  | HE1 | 11.28 | 0.002 |  | HG2 | 2.417 | 0.001 |
|  | HE3 | 7.557 | 0 |  | HG3 | 2.077 | 0 |
|  | HZ2 | 7.657 | 0.001 |  | C | 178.833 | 0 |
|  | HZ3 | 7.149 | 0 |  | CA | 57.406 | 0 |
|  | HH2 | 7.14 | 0 |  | CB | 29.686 | 0 |
|  | C | 178.785 | 0 |  | CG | 35.665 | 0 |
|  | CA | 59.037 | 0 |  | N | 118.519 | 0.008 |
|  | CB | 30.501 | 0 | ALA 65 | H | 7.724 | 0.002 |
|  | CD1 | 124.843 | 0 |  | HA | 4.288 | 0.006 |
|  | CE3 | 120.859 | 0 |  | HB1 | 1.519 | 0.003 |
|  | CZ2 | 45.448 | 0 |  | HB2 | 1.519 | 0.003 |
|  | CZ3 | 121.094 | 0 |  | HB3 | 1.519 | 0.003 |
|  | CH2 | 124.3 | 0 |  | C | 176.981 | 0 |
|  | N | 123.235 | 0.006 |  | CA | 52.514 | 0 |
|  | NE1 | 128.414 | 0.008 |  | CB | 19.902 | 0 |
| LEU 62 | H | 9.179 | 0.004 |  | N | 119.19 | 0 |
|  | HA | 3.832 | 0.001 | GLU 66 | H | 7.461 | 0 |
|  | HB2 | 1.947 | 0 |  | HA | 4.061 | 0.003 |
|  | HB3 | 1.538 | 0 |  | HB2 | 2.061 | 0.003 |
|  | HG | 1.643 | 0.005 |  | HB3 | 2.061 | 0.003 |
|  | HD11 | 0.894 | 0.011 |  | HG2 | 2.33 | 0.01 |
|  | HD12 | 0.894 | 0.011 |  | HG3 | 2.33 | 0.01 |
|  | HD13 | 0.894 | 0.011 |  | CA | 58.765 | 0 |
|  | HD21 | 0.763 | 0.002 |  | CB | 30.773 | 0 |
|  | HD22 | 0.763 | 0.002 |  | CG | 37.023 | 0 |
|  | HD23 | 0.763 | 0.002 |  | N | 124.586 | 0 |
|  | C | 178.977 | 0 |  |  |  |  |

\* The atoms are named after the Xeasy nomenclature. The errors were evaluated by the program Xeasy during the assignment process.

**Table II: Values $^3J_{HNH\alpha}$ couplings as calculated from the HNHA spectrum and the corresponding Φ angle ranges.**

| Residue # | $^3J_{HNH\alpha}$ | Φ(°) | Δφ(°) |
|---|---|---|---|
| 1 | 8.78 | -120 | 50 |
| 3 | 7.89 | -120 | 60 |
| 6 | 7.96 | -120 | 60 |
| 7 | 8.16 | -120 | 50 |
| 8 | 2.78 | -60 | 30 |
| 9 | 4.19 | 60 | 30 |
| 11 | 6.58 | -100 | 80 |
| 13 | 4.99 | -60 | 30 |
| 14 | 5.51 | -60 | 30 |
| 15 | 5.64 | -60 | 30 |
| 16 | 4.43 | -60 | 30 |
| 17 | 5.08 | -60 | 30 |
| 18 | 6.73 | -100 | 80 |
| 19 | 10.67 | -120 | 50 |
| 21 | 3.44 | -60 | 30 |
| 22 | 8.65 | -120 | 50 |
| 23 | 7.21 | -120 | 60 |
| 26 | 6.25 | -100 | 80 |
| 27 | 4.26 | -60 | 30 |
| 31 | 7.4 | -120 | 30 |
| 32 | 7.51 | -120 | 60 |
| 33 | 4.12 | -60 | 30 |
| 34 | 2.81 | -60 | 30 |
| 35 | 6.13 | -100 | 80 |
| 36 | 4.47 | -60 | 30 |
| 37 | 5.29 | -60 | 30 |
| 38 | 4.15 | -60 | 30 |

| Residue # | $^3J_{HNH\alpha}$ | Φ(°) | Δφ(°) |
|---|---|---|---|
| 39 | 8.58 | -120 | 50 |
| 40 | 5.2 | -60 | 30 |
| 41 | 2.98 | -60 | 30 |
| 42 | 7.31 | -120 | 60 |
| 43 | 6.33 | -100 | 80 |
| 44 | 10.07 | -120 | 50 |
| 45 | 9.97 | -120 | 50 |
| 46 | 11.85 | -120 | 50 |
| 47 | 8.74 | -120 | 50 |
| 48 | 8.82 | -120 | 50 |
| 49 | 7.77 | -120 | 60 |
| 50 | 5.89 | -100 | 80 |
| 52 | 6.09 | -100 | 80 |
| 53 | 5.14 | -60 | 30 |
| 54 | 6.69 | -100 | 80 |
| 55 | 3.61 | -60 | 30 |
| 56 | 6.54 | -100 | 80 |
| 57 | 6.67 | -100 | 80 |
| 58 | 4.22 | -60 | 30 |
| 59 | 3.95 | -60 | 30 |
| 60 | 6.77 | -100 | 80 |
| 61 | 4.69 | -60 | 30 |
| 63 | 5.61 | -60 | 30 |
| 64 | 6.48 | -100 | 80 |
| 65 | 3.87 | -60 | 30 |
| 66 | 2.57 | -60 | 30 |

**Table A-III: Talos prediction of the φ and ψ values of YFHJ's residues.**

| Residue # | Residue Type | ψ(°) | φ(°) | Δψ | Δφ | Dist | Count | Class |
|---|---|---|---|---|---|---|---|---|
| 1 | M | 9999 | 9999 | 0 | 0 | 0 | 0 | None |
| 2 | G | 81.32 | 15.33 | 14.69 | 6.05 | 19.42 | 4 | New |
| 3 | L | -94.39 | 138.57 | 34.17 | 18.55 | 27.58 | 9 | Good |
| 4 | K | -102.41 | 166.63 | 29.13 | 13.06 | 26.35 | 10 | Good |
| 5 | W | -73.74 | 124.3 | 15.7 | 22.9 | 33.99 | 9 | Good |
| 6 | T | -82.98 | 132.47 | 15.55 | 11.21 | 36.51 | 9 | Good |
| 7 | D | -98.67 | 151.61 | 20.81 | 34.32 | 35.05 | 9 | Good |
| 8 | S | -60.28 | -34.23 | 5.79 | 7.7 | 19.7 | 9 | Good |
| 9 | R | -63.84 | -42.83 | 8.22 | 6.53 | 10.15 | 10 | Good |
| 10 | E | -68.31 | -37.63 | 6.08 | 6.18 | 10.08 | 10 | Good |
| 11 | I | -68.3 | -41.67 | 6.19 | 4.95 | 17.86 | 10 | Good |
| 12 | G | -63.42 | -39.63 | 4.72 | 3.43 | 21.66 | 10 | Good |
| 13 | E | -64.94 | -40.61 | 4.26 | 10.34 | 17.18 | 10 | Good |
| 14 | A | -64.04 | -40.95 | 5.55 | 6.66 | 10.29 | 10 | Good |

| 15 | L | -64.55 | -41.24 | 6.81 | 5.76 | 13.21 | 10 | Good |
| 16 | Y | -63.89 | -42.92 | 9.4 | 6.79 | 15.08 | 10 | Good |
| 17 | D | -69.09 | -38.49 | 9.88 | 4.81 | 25.94 | 10 | Good |
| 18 | A | -77.27 | -26.01 | 10.64 | 20.39 | 39.56 | 10 | Good |
| 19 | Y | -113.42 | 128.9 | 27.58 | 41.17 | 57.69 | 6 | New |
| 20 | P | -60.88 | -21.55 | 6.27 | 9.19 | 38.03 | 9 | Good |
| 21 | D | -95.46 | 3.55 | 11.65 | 9.59 | 29.12 | 7 | New |
| 22 | L | -90.08 | 142.6 | 22.89 | 16.42 | 24.14 | 6 | New |
| 23 | D | -87.69 | 136.85 | 18.09 | 19.48 | 43.08 | 9 | Good |
| 24 | P | -58.49 | -29.95 | 5.35 | 12.09 | 47.08 | 7 | New |
| 25 | K | -64.78 | -26.99 | 7.28 | 17.24 | 34.76 | 10 | Good |
| 26 | T | -106.52 | -5.15 | 14.26 | 14.58 | 13.83 | 10 | Good |
| 27 | V | -105.58 | 117.86 | 22.36 | 17.29 | 22.55 | 6 | New |
| 28 | R | -98.38 | 131.87 | 26.71 | 19.22 | 40.19 | 9 | New |
| 29 | F | -57.6 | -35.43 | 5.02 | 11.15 | 54.2 | 9 | Good |
| 30 | T | -64.5 | -30.61 | 6.18 | 14.71 | 44.8 | 10 | Good |
| 31 | D | -66.43 | -38.84 | 6.04 | 7.1 | 25.64 | 10 | Good |
| 32 | M | -61.68 | -44.25 | 5.69 | 4.87 | 19.17 | 10 | Good |
| 33 | H | -61.43 | -44.97 | 4.24 | 3.55 | 15.73 | 10 | Good |
| 34 | Q | -60.94 | -39.35 | 3.47 | 5.82 | 16.7 | 10 | Good |
| 35 | W | -71.26 | -37.46 | 13.28 | 8.96 | 16.84 | 9 | New |
| 36 | I | -63.14 | -44.95 | 4.74 | 4.56 | 22.81 | 10 | Good |
| 37 | C | -61.69 | -35.85 | 6.84 | 11.8 | 22.83 | 10 | Good |
| 38 | D | -85.66 | -24.24 | 17.28 | 23.6 | 25.1 | 8 | New |
| 39 | L | -97.35 | 141.41 | 38.64 | 34.77 | 16.21 | 5 | New |
| 40 | E | -65.22 | -28.95 | 9.39 | 11.14 | 16.99 | 10 | Good |
| 41 | D | -94.86 | 4.07 | 15.17 | 11.48 | 18.65 | 10 | Good |
| 42 | F | -76.18 | 128.09 | 21.78 | 8.55 | 26.28 | 10 | Good |
| 43 | D | -116.29 | 134.35 | 31.74 | 10.47 | 33.49 | 9 | Good |
| 44 | D | -127.88 | 154.92 | 17.33 | 11.22 | 36.38 | 10 | Good |
| 45 | D | -97.32 | 112.75 | 24 | 21.99 | 53.03 | 10 | Good |
| 46 | P | -59.55 | -23.63 | 7.52 | 12.01 | 45.08 | 9 | Good |
| 47 | Q | -89.51 | 0.7 | 13.23 | 17.73 | 30.98 | 9 | Good |
| 48 | A | -101.88 | 155.11 | 25.2 | 12.46 | 26.33 | 6 | New |
| 49 | S | 56.8 | 43.6 | 3.04 | 10.16 | 26.55 | 6 | New |
| 50 | N | -110.56 | 153.3 | 47.41 | 23.02 | 37.77 | 5 | New |
| 51 | E | -56.49 | -40.1 | 5.29 | 7.4 | 29.86 | 10 | Good |
| 52 | K | -64.18 | -39.09 | 5.14 | 8.45 | 19.85 | 10 | Good |
| 53 | I | -65.37 | -45.49 | 6.62 | 3.23 | 12.17 | 10 | Good |
| 54 | L | -65.45 | -39.81 | 4.47 | 4.57 | 13.75 | 10 | Good |
| 55 | E | -61.61 | -45.6 | 2.22 | 5.93 | 10.06 | 10 | Good |
| 56 | A | -65.87 | -37.86 | 3.11 | 3.09 | 9.72 | 10 | Good |
| 57 | I | -67.2 | -44.14 | 2.92 | 3.87 | 13.38 | 10 | Good |
| 58 | L | -61.26 | -40.49 | 5.94 | 7.07 | 13.26 | 10 | Good |
| 59 | L | -64.96 | -38.24 | 3.1 | 4.53 | 10.96 | 10 | Good |
| 60 | V | -65.53 | -37.48 | 5.68 | 10.29 | 28.81 | 10 | Good |
| 61 | W | -91.93 | -16.71 | 17.6 | 26.22 | 31.1 | 6 | New |
| 62 | L | -57.07 | -40.48 | 5.42 | 5.9 | 22.09 | 10 | Good |

| 63 | **D** | -62.74 | -40.32 | 7.32 | 4.58 | 9.73 | 10 | Good |
|----|-------|--------|--------|------|------|-------|----|------|
| 64 | **E** | -66.85 | -28.39 | 6.89 | 8.5 | 12.93 | 10 | Good |
| 65 | **A** | -101.68 | -1.41 | 15.72 | 19.73 | 27.15 | 8 | New |
| 66 | **E** | 9999 | 9999 | 0 | 0 | 0 | 0 | None |

**Table IV: Dali output file with YFHJ's structurally related proteins, Z factors and RMSDs.**

| PDB code | Z factor | RMSD (Å) | Class and function |
|----------|----------|----------|--------------------|
| **1uj8-A** | 12.5 | 1.2 | YFHJ Crystal structure |
| **1iuf** | 3.8 | 2.4 | DNA binding protein centromere abp1 protein fragment |
| **1h40** | 3.5 | 2.5 | |
| **1mix-A** | 3.2 | 2.6 | Structural protein talin fragment (gallus gallus) chi |
| **1ef1-A** | 3.1 | 3.0 | Membrane protein moesin fragment (*H. Sapiens*) |
| **1d3y-A** | 3.0 | 2.6 | Isomerase DNA topoisomerase |
| **1a6q** | 3.0 | 3.0 | Hydrolase phosphatase 2c (*H. Sapiens*) |
| **1r5i-D** | 2.9 | 2.8 | Immune system hla class |
| **2hfh** | 2.8 | 2.9 | HFN-3 homologue genesis fragment |
| **1urj-A** | 2.8 | 2.9 | DNA binding protein |
| **1ldj-A** | 2.8 | 2.9 | Ligase cullin homologue 1 fragment |
| **1ldd-A** | 2.8 | 2.7 | Ligase anaphase promoting complex |
| **1khz-A** | 2.8 | 2.9 | Transferase phosphofruttokinase |
| **1lar-A** | 2.7 | 2.8 | Hydrolase lar fragment Mutant (*H. Sapiens*) |
| **1abv** | 2.7 | 2.9 | ATP synthesis delta subunit of the f1f0-ATP synthase |
| **1e9r-A** | 2.5 | 3.0 | Coupling protein conjugal tranfer protein |
| **1cja-A** | 2.5 | 2.7 | Transferase actin-fragmin kinase fragment |
| **1avc** | 2.5 | 3.0 | Calcium/phospholipid-binding protein |

| | | | |
|---|---|---|---|
| **1qnf** | 2.4 | 2.9 | DNA repair photolyase |
| **1q1v-A** | 2.4 | 2.7 | DNA binding protein dek protein fragment (*H. Sapiens*) |
| **1p2f-A** | 2.4 | 2.6 | Transcription response regulator |
| **1n25-A** | 2.4 | 2.6 | Virus/viral protein large t antigen fragment (simian) |
| **1ji8-A** | 2.4 | 2.8 | Oxidoreductase dissimilatory siroheme-sulphite reductase |
| **1irx-A** | 2.4 | 3.4 | Ligase lysil-t-RNA synthase (*pirococcus horikoshii*) |
| **6pax-A** | 2.3 | 3.3 | Gene regulation/DNA homeobox protein pax-6 |
| **1w5s-A** | 2.3 | 3.1 | DNA replication initiationorc2 (hypothetical protein) |
| **1u6g-C** | 2.3 | 3.5 | |
| **1p8c-A** | 2.3 | 3.3 | Structural genomics. Unknown function |
| **1oi2-A** | 2.3 | 2.1 | Kinase hypothetical protein ycgt |
| **1ltl-A** | 2.3 | 3.4 | Replication DNA replication initiator |
| **1knc-A** | 2.3 | 2.8 | Electron transport ahpd protein (*mycobacterium tuberculosis*) |
| **1h6k-A** | 2.3 | 3.2 | Nuclear protein cbp80 fragment |
| **1s9h-A** | 2.2 | 3.2 | Replication rep 40 protein |
| **1lj8-A** | 2.2 | 2.7 | Oxidoreductase mannitol dehydrogenase |
| **1ecm-A** | 2.2 | 2.9 | Chorismate mutase |
| **1e2x-A** | 2.2 | 2.8 | Transcriptional regulation fatty acid metabolism regulator |
| **1xfh-A** | 2.1 | 3.9 | Transport protein proton glutamate symport protein |
| **1p41-A** | 2.1 | 2.6 | |
| **1cez-A** | 2.1 | 3.5 | Protein/DNA bacteriphage t7 RNA polymerase |
| **1pbv** | 2.0 | 3.1 | Exchange factor arno fragment |
| **1hs6-A** | 2.9 | 3.0 | Hydrolase leukotriene a-4 hydrolase (*H. Sapiens*) |

Procheck Ramachandran plot
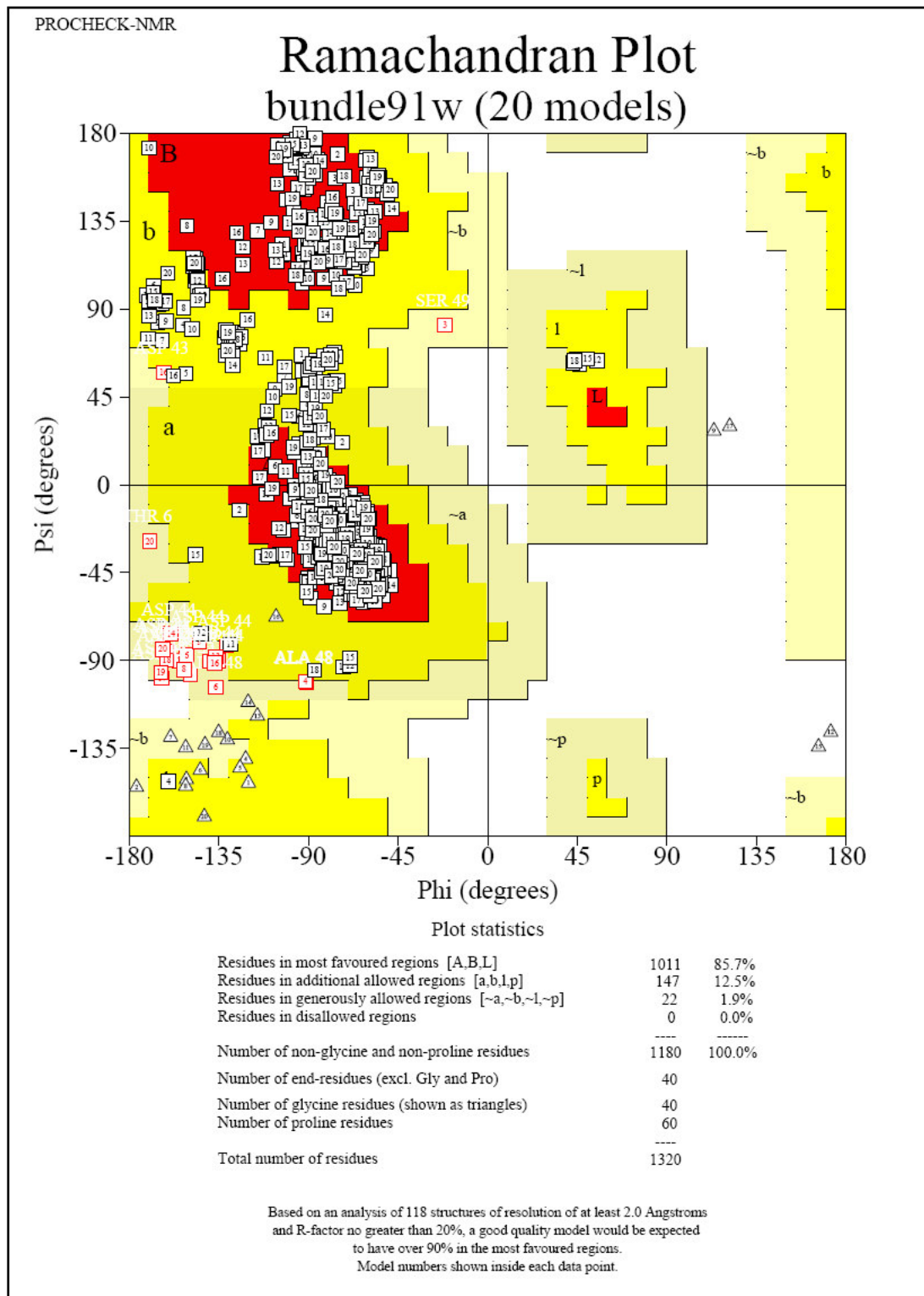
# Appendix B

Buffers' composition for the purification of unlabelled YFHJ

**Lysis buffer**

| | |
|---|---|
| 20 mM | Tris Hcl pH 8 |
| 150 mM | NaCl |
| 10 mM | Imidazole |
| 20 mM | β-mercapto ethanol |
| 0.2% | IGEPAL (detergent) |

pefac block (antiproteases)

DNAasi I

Lysozime

**Equilibrating buffer**

| | |
|---|---|
| 20 mM | Tris Hcl pH 8 |
| 150 mM | NaCl |
| 10 mM | Imidazole |
| 20 mM | β-mercapto ethanol |
| 0.2% | IGEPAL (detergent) |

**Wash buffer**

| | |
|---|---|
| 20 mM | Tris Hcl pH 8 |
| 150 mM | NaCl |
| 10 mM | Imidazole |
| 20 mM | β-mercapto ethanol |

**High salt buffer**

| | |
|---|---|
| 20 mM | Tris Hcl pH 8 |
| 1 M | NaCl |
| 10 mM | Imidazole |
| 20 mM | β-mercapto ethanol |

**Elution buffer**

| | |
|---|---|
| 20 mM | Tris Hcl pH 8 |
| 1 M | NaCl |
| 300 mM | Imidazole |
| 20 mM | β-mercapto ethanol |

**Dialysis buffer & gel chromatography buffer**

| | |
|---|---|
| 20 mM | Tris.HCl pH 8 |
| 150 mM | NaCl |
| 20 mM | β-mercaptoethanol |

# Index of Names