Filipe Miguel dos Santos Freire

Licenciado em Bioquímica

INTEGRATED STUDY BY NMR AND X-RAY CRYSTALLOGRAPHY ON THE ANALYSIS OF THE MOLECULAR INTERACTIONS IN HEME-BINDING PROTEINS

Dissertação para obtenção do Grau de Doutor em Bioquímica,
Especialidade Bioquímica Estrutural

Orientadora: Doutora Maria dos Anjos Macedo
Professora Auxiliar, FCT/UNL
Co-Orientadora: Doutora Maria João Romão
Professora Catedrática, FCT/UNL
Co-Orientador: Doutor Brian James Goodfellow
Professor Auxiliar, UA

Juri

Presidente: Professor Doutor José Paulo Barbosa Mota
Arguentes: Professor Doutor Carlos Frederico de Gusmão Campos Geraldes
Doutora Sandra de Macedo Ribeiro

Vogais: Doutor Jean-Marc Moulis
Doutor Shabir Husein Najmudin
Doutora Ana Luísa Moreira de Carvalho

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Setembro 2012

Filipe Miguel dos Santos Freire

INTEGRATED STUDY BY NMR AND X-RAY CRYSTALLOGRAPHY ON THE ANALYSIS OF THE MOLECULAR INTERACTIONS IN HEME-BINDING PROTEINS

DISSERTAÇÃO APRESENTADA PARA A OBTENÇÃO DO GRAU DE DOUTOR EM BIOQUÍMICA, ESPECIALIDADE BIOQUÍMICA ESTRUTURAL, PELA FACULDADE DE CIÊNCIAS E TECNOLOGIA, UNIVERSIDADE NOVA DE LISBOA

Caparica, 21 de Setembro de 2012

‘Integrated study by NMR and X-ray Crystallography on the analysis of the molecular interactions in heme-binding proteins’

AGRADECIMENTOS

(Acknowledgments)

'If you can't explain it simply, you don't understand it well enough.'

Albert Einstein

Os resultados apresentados nesta tese de doutoramento só foram alcançados devido ao esforço e ajuda de várias pessoas e instituições. Por isso gostaria de agradecer:

Aos meus orientadores de doutoramento, Professora Anjos Macedo, Professora Maria João Romão e Professor Brian Goodfellow por me terem dado a oportunidade de ter trabalhado com eles e pelo apoio prestado ao longo dos últimos 6 anos.

Ao Jorge, pela passagem de testemunho no meu projecto de doutoramento. À Susana, 'companheira' neste projecto.

À Ana Luísa por ter sido, para além de uma excelente colega de laboratório, a minha 'professora' de cristalografia, por me ter ajudado na planificação e execução das experiências de cristalografia, pela visita guiada ao ESRF na minha primeira ida, ainda aluno de licenciatura e sobretudo pelo esforço que dispendeu para me ajudar na (difícil) determinação da estrutura tri-dimensional da proteína SOUL.

À Teresa, por ter sido a minha primeira 'orientadora', no trabalho que desenvolvi no estágio de Licenciatura e por ser uma pessoa muito optimista, bem disposta, sempre disponível.

Ao Zé, pelas discussões científicas, pelos valiosos conselhos, cristalográficos mas também computacionais; pela amizade e disponibilidade, sobretudo nos momentos mais complicados.

A todos os membros do grupo de cristalografia que, desde 2005, me ajudaram nesta etapa da minha vida. Um agradecimento especial à Joana, Cecília, por todo o apoio laboratorial e ao Shabir, *an 'old chap' always available to teach me crystallography, to revise this thesis, to pay me a beer, to talk about football, tennis and even cricket.*

Ao Aldino, pela amizade mas também pela paciência para com as minha dúvidas elementares na preparação das experiências de RMN e na análise das mesmas.
Ao Eurico e ao Ângelo pela ajuda durante as experiências no '600 MHz'.

Ao Marino, que começou como meu 'orientando' com uma bolsa de iniciação à investigação e que agora é aluno de doutoramento, um agradecimento pelas longas trocas de palavras sobre futebol, e não só (ai os desabafos que ele ouviu da minha parte!). Ao David, antigo membro do grupo de cristalografia, que se tornou num amigo e parceiro de ténis e futebol, obrigado pelas longas conversas sobre tudo e mais alguma coisa, quase sempre sem grande conteúdo mas sempre com muito boa disposição.

To Jean-Marc Moulis for giving me the unique opportunity to work in his laboratory for approximately three weeks and learn so many things about protein cloning, overexpression and purification.

À Fundação para a Ciência e a Tecnologia, pelo apoio financeiro (SFRH/BD/30239/2006 e PTDC/QUI/64203/2006).

Ao meus grandes amigos, Nuno, Tiago, Catarina e Irina. Ao Nuno e ao Tiago pela grande amizade que nos uniu desde o início dos tempos de faculdade, pelas grandes discussões que tivemos durante a elaboração de relatórios, trabalhos e apresentações, pelos artigos que me enviaram directamente dos EUA, pelas grandes jogatanas de 'Sueca' nas viagens de comboio (o mítico comboio das 7:11, que até deu origem a um blog) e na esplanada do C5. À Catarina pelo apoio constante durante o meu doutoramento, pelas longas conversas no laboratório, pela companhia nas viagens a sincrotrões e congressos. À Irina por ter sido uma presença constante, sempre com palavras de apoio e incentivo.

À pessoa que preencheu a minha vida nos últimos 12 anos, que esteve sempre presente, nos bons mas sobretudo nos maus momentos, nos momentos de desmotivação e pessimismo; obrigado por me perceberes sempre, mesmo quando não digo nada! Caminharemos sempre juntos porque temos 'laços inquebráveis'!

Aos meus pais, por tudo o que sou, por tudo o que alcancei, por tudo o que me dão, todos os dias, a cada instante, a cada gesto, OBRIGADO!

'Underneath this smile lies everything, all my hopes and anger, pride and shame'

Mike McCready

RESUMO

O grupo hemo é essencial a todos os organismos aeróbios, uma vez que está envolvido em diversos processos biológicos. Devido à sua toxicidade e elevada insolubilidade, há proteínas que se ligam transientemente a esta molécula, garantindo o seu posterior transporte e inserção em proteínas hémicas. Nesta dissertação é efectuado um estudo estrutural sobre três proteínas de ligação ao hemo, a proteína humana SOUL (hSOUL) e as proteínas p22HBP de murganho e humana.

No capítulo 1 é efectuada uma introdução ao grupo hemo e à sua importância biológica e sistematizada a informação disponível sobre as proteínas acima referidas. São também apresentados os princípios básicos das principais técnicas que foram utilizadas na caracterização estrutural destas proteínas: Cristalografia de raios-X e Ressonância Magnética Nuclear (RMN).

Para os estudos descritos neste trabalho foi necessário obter as proteínas com elevado grau de pureza e em quantidades significativas. Para tal, foi necessário efectuar a clonagem da proteína hSOUL e optimizar a sua sobre-expressão e purificação – capítulo 2.

A proteína hSOUL apresenta uma estrutura global bastante semelhante à proteína p22HBP de murganho. No capítulo 3 é apresentada a determinação da estrutura da proteína hSOUL por Cristalografia de raios-X, descrita a respectiva estrutura tri-dimensional e discutidas as possíveis implicações funcionais da mesma.

Para compreender a interacção do grupo hemo à proteína hSOUL foram efectuados vários estudos recorrendo às técnicas de RMN, Extinção de Fluorescência e Espectroscopia de Visível. Os resultados obtidos, e apresentados no capítulo 4, indicam que, a existir, a ligação hemo-hSOUL deverá ser uma ligação não específica.

Na capítulo 5 são descritas as diversas experiências de cristalização das proteínas p22HBP humana e de murganho, com o intuito de determinar a sua estrutura tri-dimensional em complexo com o grupo hemo e, deste modo, compreender a interacção hemo-p22HBP.

No capítulo 6 são apresentadas as principais conclusões respeitantes ao trabalho que foi desenvolvido e que se encontra descrito nesta dissertação. São também incluídas algumas perspectivas futuras.

Palavras chave: Proteínas de ligação ao hemo, Proteínas com domínio BH3, Cristalografia de raios-X, Ressonância Magnética Nuclear, Extinção de Fluorescência

ABSTRACT

Heme is essential to all aerobic organisms, as it is involved in several biological processes. Due to its toxicity and high insolubility, several proteins transiently bind this molecule, ensuring the transport and insertion into heme proteins. A structural study of three heme-binding proteins, human SOUL protein and human and murine p22HBP is described in this dissertation.

In chapter 1, an introduction to heme and its biological importance is performed and all the information related to the proteins previously mentioned is summarized. The basic principles of X-ray Crystallography and Nuclear Magnetic Resonance (NMR), the techniques used in the structural characterization of these proteins, are described.

For the studies described in this thesis large amounts of pure protein are required. For this reason, hSOUL protein was cloned, and the overexpression and purification of hSOUL optimized – chapter 2.

The overall structure of hSOUL is very similar to murine p22HBP solution structure. hSOUL protein structure determined by X-ray Crystallography is described in chapter 3 and the possible biological consequences are discussed.

Understanding the heme interaction with hSOUL was an important objective of this work. For that, NMR, Fluorescence Quenching and Visible Spectroscopy studies were performed. The results obtained, and shown in chapter 4, indicate that the interaction, if it exists, is non-specific.

The several experiments to crystallize human and murine p22HBP, in order to solve their three-dimensional structure in complex with heme and therefore understand heme-p22HBP interaction, are described in chapter 5.

The main conclusions from the present work are drawn in chapter 6 together with the future perspectives.

Keywords: Heme-binding proteins, BH3-only proteins, Biomolecular Crystallography, Nuclear Magnetic Resonance, Fluorescence Quenching

TABLE OF CONTENTS

FIGURES INDEX

TABLES INDEX

EQUATIONS INDEX

ABBREVIATIONS AND SYMBOLS

ABC	(ATP)-binding cassette
ABCG2	(ATP)-binding cassette sub-family G member 2
ADP	Anisotropic displacement parameter
ALA	5-aminolevulic acid
ALAD	Aminolevulic acid dehydratase
ALAS	Aminolevulic acid synthase
ALAS 1	Aminolevulic acid synthase 1
ALAS 2	Aminolevulic acid synthase 2
Bcl-2	B-cell lymphoma-2
BCRP	Breast Cancer Resistance Protein
bp	Base pair
cAMP	3'-5'-cyclic adenosine monophosphate
CBD	Chitin binding domain
cDNA	Complementary desoxiribonucleic acid
CO	Carbon monoxide
COPRO'GEN III	Coproporphyrinogen III
COSY	Correlation Spectroscopy
CPI	Coproporphyrin I
CPIII	Coproporphyrin III
CPO	Coproporphyrinogen oxidase
Cryo-EM	Cryo-Electron Microscopy
CSA	Chemical shift anisotropy
CSI	Chemical shift index
DCs	Dendric cells
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EDTA	Ethylenediamine tetraacetic acid
EMBL	The European Molecular Biology Laboratory
EPR	Electron Paramagnetic Resonance
ESR	Electron Spin Resonance
ESRF	European Synchrotron Radiation Facility
FECH	Ferrochelatase
FID	Free induction decay
FLVCR	Feline Leukaemic Virus Receptor
FOM	Figure of merit
FPR	Formyl peptide receptor
FPRL	FPR-like receptor
FPRL 1	FPR-like receptor 1
FPRL 2	FPR-like receptor 2
FQ	Fluorescence Quenching
GFP	Green fluorescence protein
Grx	Glutaredoxin
Grx3	Glutaredoxin 3
Grx4	Glutaredoxin 4
GST	Glutathione S-transferase
Hb A	Human normal adult hemoglobin

HbCO A	Carbonmonoxy-hemoglobin A
HBP23	Heme binding protein 23
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
p22HBP	Heme-binding protein 1
HMB	Hydroxymethylbilane
HO	Heme oxygenase
HO-1	Heme oxygenase 1
HO-2	Heme oxygenase 2
HMQC	Heteronuclear Multiple Quantum Coherence
HSQC	Heteronuclear Single Quantum Coherence
ICP-AES	Inductively Coupled Plasma-Atomic Emission Spectroscopy
IMAC	Immobilized metal affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IRP	Iron-regulatory protein
LB	Luria Broth
MAD	Multiple-wavelength Anomalous Diffraction
MES	2-(*N*-morpholino)ethanesulfonic acid
MIR	Multiple Isomorphous Replacement
MIRAS	Multiple Isomorphous Replacement with Anomalous Signal
mMP	Mitochondrial membrane potential
MOPS	3-(N-morpholino)propanesulfonic acid
mPT	Mitochondrial permeability transition
mPTP	Mitochondrial permeability transition pore
MR	Molecular Replacement
mRNA	Messenger ribonucleic acid
NCS	Non-crystallographic symmetry
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Effect Spectroscopy
OD	Optical density
ORF	Open-reading frame
PAGE	Polyacrilamide gel electrophoresis
PEG	Polyethylene glycol
PBD	Peripheral blood cells
PBG	Porphobilinogen
PBGD	Porphobilinogen deaminase
PCR	Polymerase chain reaction
PDB	Protein data bank
PPIX	Protoporphyrin IX
PPO	Protoporphyrinogen oxidase
PROTO'GEN IX	Protoporphyrinogen IX
Prx	Peroxiredoxin
Prx I	Peroxiredoxin I
RA	Rheumatic arthritis
RCS	Ring current shift
RDC	Residual dipolar coupling
rmsd	Root-mean square deviation
ROE	Rotating frame NOE

ROS	Reactive oxygen species
rpm	Rotations *per* minute
SAD	Single-wavelength Anomalous Dispersion
SEC	Size exclusion chromatography
SIR	Single Isomorphous Replacement
SIRAS	Single Isomorphous Replacement with Anomalous Signal
SLC	Solute carrier
SOUL	Heme-binding protein 2
SPR	Surface Plasmon Resonance
STD	Saturation Transfer Diference
TB	Terrific Broth
Tfr1	Transferrin-receptor 1
TM	Transmembrane domain
TNF	Tumor necrosis factor
TOCSY	Total Correlation Spectroscopy
Tris	Tris-(hydroxymethyl)-aminomethane
TROSY	Transverse Relaxation-Optimized Spectroscopy
TrR1	Transferrin-receptor 1
URO3S	Uroporphyrinogen III synthase
UROD	Uroporphyrinogen Decarboxylase
URO'GEN III	Uroporphyrinogen III
UV	Ultraviolet
VDAC	Voltage-dependent anion channel
$\Delta\psi$	Mitochondrial membrane potential
K_d	Dissociation constant
k_{on}	Association rate constant
k_{off}	Dissociation rate constant

CHAPTER 1

INTRODUCTION

CONTENTS

1.1. HEME

Heme or iron protoporphyrin IX (PPIX) is a prosthetic group consisting of an iron atom in the center of a large heterocyclic organic ring, porphyrin – figure 1.1.

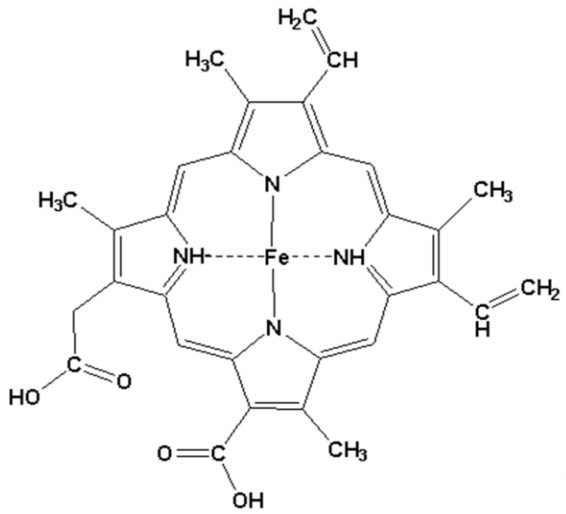


Figure 1.1 *Heme (adapted from http://met.fzu.edu.cn/cai/shenghua/resource/biochem/ch07/heme.htm).* Iron protoporphyrin IX representation.

Heme plays a vital role in many biological processes such as O_2 transport by hemoglobin, O_2 storage by myoglobin, electron transfer by cytochromes and activation of the O-O bond by P450 enzymes and peroxidases. It is, therefore, an essential molecule to all aerobic organisms.

In erythroid cells, heme synthesis regulation is mediated by erythroid-specific transcription factors and by the bioavailability of Fe in the form of Fe/S clusters. However, in non-erythroid cells, this pathway is regulated by heme-mediated feedback inhibition.

Heme is synthesized in all nucleated cells and the cellular levels of heme are tightly regulated by enzymatic synthesis and degradation processes [1].

The main catabolic pathway of heme is catalyzed by Heme oxygenase - HO (E.C. 1.14.99.3), which is itself a heme protein. The reaction catalyzed by HO leads to the production of carbon monoxide (CO) and biliverdin and to the concomitant iron release [2].

1.1.1. HEME BIOSYNTHESIS AND HOMEOSTASIS AND IRON HOMEOSTASIS IN MAMMALS

In mammals, heme synthesis (figure 1.2) occurs mainly in developing erythroid cells (approximately 85%), hepatocytes and muscle cells, and can be divided in four main stages – pyrrole formation, tetrapyrrole macrocycle formation, modification of the acetate and propionate side chains and insertion of iron into protoporphyrin IX.

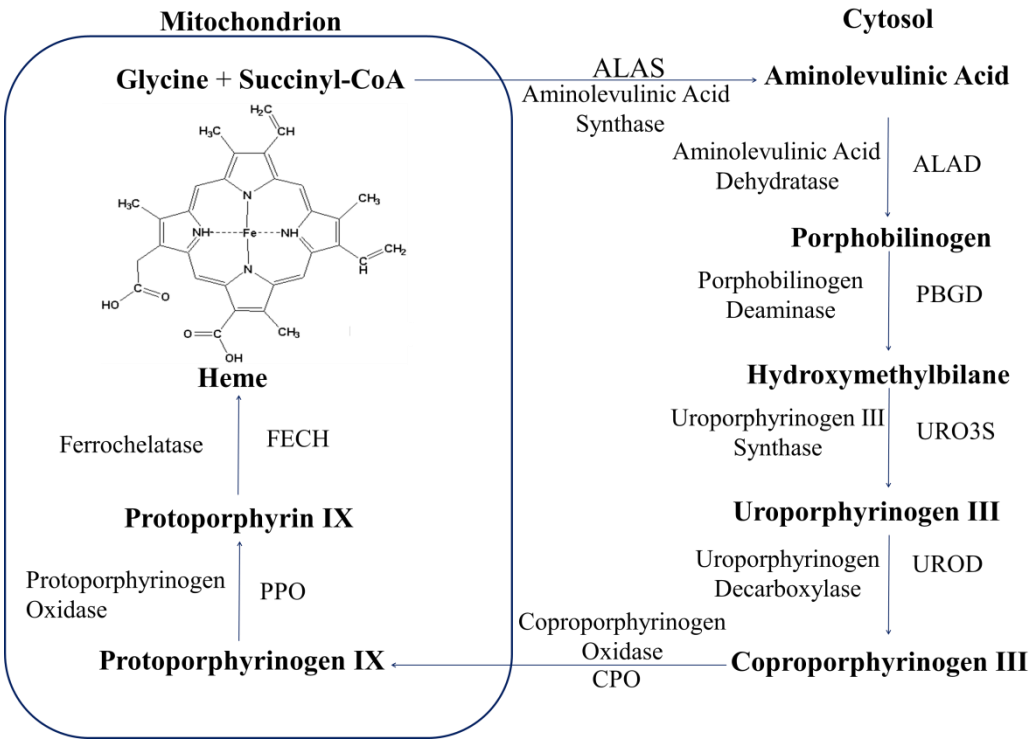


Figure 1.2 *Heme biosynthesis pathway (adapted from [3]).* Mammalian heme biosynthesis pathway scheme. The process occurs in both mitochondria (blue rectangle) and cell cytosol.

The first step of this pathway takes place in the mitochondria and consists of the condensation of glycine and succinyl-CoA to form 5-aminolevulinic acid (ALA), a reaction catalyzed by 5-aminolevulinic acid synthase - ALAS (EC 2.3.1.37) and is the rate limiting reaction of this metabolic pathway. ALA is the earliest common precursor of heme biosynthesis in all species, including eukaryotes, archaea and bacteria. Aminolevulinic acid synthase exists in two isoforms, ALAS 1 (ubiquitously expressed) and ALAS 2 (only expressed by erythroid precursors), with each isoform having the same reaction mechanism but different regulation. Aminolevulinic acid is then converted to porphobilinogen (PBG), which gives hydroxymethylbilane (HMB), a polymer of 4 molecules of PBG, in a reaction catalyzed by porphobilinogen deaminase (EC 2.5.1.61). This unstable tetrapyrrole is then converted into uroporphyrinogen III (URO'GEN III) by uroporphyrinogen III synthase (EC 4.2.1.75). URO'GEN III represents a branch point for the pathways of heme synthesis, and also for chlorophyll and corrins synthesis.

In the heme biosynthesis pathway, URO'GEN III is then converted to coproporphyrinogen III (COPRO'GEN III) by sequential removal of the 4 carboxylic groups of the acetic acid side chains, which is followed by sequential oxidative decarboxylation of the propionate groups to vinyl groups of the pyrrole rings A and B, forming protoporphyrinogen IX (PROTO'GEN IX). Inside the mitochondria PROTO'GEN IX is oxidized to protoporphyrin IX. In the inner surface

of the inner membrane of the mitochondria, iron is inserted into the protoporphyrin IX by ferrochelatase (EC 4.99.1.1) [3].
Once synthesized, heme is immediately exported to the outside of the mitochondria into the cytosol and endoplasmic reticulum and associates with apo-hemoproteins.

Heme can reversibly bind oxygen and is highly reactive to various compounds due to its capacity of exchanging electrons and activating oxygen. The reactivity of heme with oxygen can lead to toxic effects as this interaction can cause the destruction of the porphyrin rings by hydroxyl radicals. On the other hand, due to its insolubility, heme is thought to chelate transiently to amino acids, peptides or proteins while it is transported, (see chapter *1.1.2* for more details) [4].

An important regulator of heme intracellular level is heme oxygenase (enzyme with two isoforms, HO-1 and HO-2), a rate-limiting enzyme of the heme catabolic pathway. It is proposed that heme, once degraded by HO-1, is exported out of the cell by the iron-efflux protein, ferroportin [5-7]. HO-1 is an inducible isoenzyme whose expression is up-regulated by its substrate and oxidative stress. Induction of HO-1 by hemin is associated to the increased level of translated and hybridizable HO-1 mRNA. HO-2 is thought to be a constitutive enzyme, which is expressed under homeostatic conditions. Both isoforms, HO-1 and HO-2 are ubiquitously expressed and catalytically active [8-10].

Due to the biological importance of heme, iron homeostasis must also be a very tightly regulated process. Most of the iron in the organism is targeted to the mitochondria, where it is readily used for heme biosynthesis in erythroid cells but also for the synthesis of iron-sulfur clusters. However, metabolism of mitochondrial iron can be independent of heme biosynthesis since the uptake of iron by mitochondria is observed even when heme synthesis is stopped. The trafficking of iron to mitochondria is still a not well understood process. However, two conserved cytosolic glutaredoxins, Grx3 and Grx4, have been identified as very important proteins in iron intracellular sensing and trafficking. One clear fact is that the solute carrier (SLC) mitoferrin (also known as SLC25A37), located in the inner mitochondrial membrane, is required for the entrance of iron into the mitochondria [11].
In mammals, iron itself regulates its import, storage and utilization through iron-regulatory protein (IRP)/iron-response element system. Two proteins play a vital role in these processes: transferrin and ferritin. Circulating transferrin is responsible for the transport and delivery of iron that is released into the plasma, mainly from intestinal enterocytes or reticuloendothelial macrophages, to all tissues except those that are separated from the blood by endothelial cells, that form a physical barrier (like, for instance, the brain or the eye) [11]. By endocytosis, iron is

taken up into the cells by transferrin-receptor 1 (TfR1). Ferritin is an important protein in iron regulation as it can store and detoxify the excess intracellular iron in the cytosol. The protein has two subunits, H (heavy) and L (light), and folds in a shell-like structure, providing the capacity of Fe^{3+} storage in the form of ferric oxy-hydroxide phosphate.

1.1.2. HEME TRAFFICKING

Heme is highly insoluble and can be very toxic to the cells. Therefore, it is present in the intracellular environment in very low concentrations, usually less than 10^{-9} M and a wide range of proteins have been proposed to transiently bind and transport this molecule.

The mechanisms of intracellular heme channeling have not yet been totally elucidated, but a number of proteins have been characterized based on their ability to bind heme. For example, different isoforms of glutathione S-transferase (GST) have been considered as intracellular carriers of heme to the endoplasmic reticulum. A number of other heme binding proteins have been characterized that can serve as heme carriers although the role they play is still unknown. Among them, Heme Binding Protein 23, HBP23 (or peroxiredoxin I, prx I) although binding heme, has as main function protecting cells from oxidative stress. In addition, the HBP23/Prx I is proposed to be an antioxidant and tumor suppressor, though, the relation to heme binding activity is not known [12]. Other peroxiredoxins (Prxs, EC 1.11.1.15) can also bind heme, although they possess peroxidase activity, which is heme independent. Heme-binding protein 1, or p22HBP, is a cytosolic protein with high binding affinity for heme but its role in heme trafficking remains unknown [13]. It has been proposed that this protein can act as a heme transporter or even as a heme buffer protein. More recently, the protein was found to be part of a complex (complex III) that is involved in hemoglobin biosynthesis [14].

Recently, two heme exporters from maturing erythroid cells were reported: Feline Leukaemic Virus Receptor (FLVCR) and heme-efflux protein ABCG2 (also known as BCRP – Breast Cancer Resistance Protein).

FLVCR was cloned from a human T-lymphocyte cDNA libray and was shown to be essential in erythropoiesis, a process by which red blood cells (erythrocytes) are produced. This process is stimulated when a decrease in O_2 levels in circulation is detected by the kidneys, which then secrete the hormone erythropoietin. This hormone stimulates proliferation and differentiation of red blood cells precursors, which activates increased erythropoiesis in the hemopoietic tissues, ultimately producing red blood cells. FLVCR transporter is upregulated when heme synthesis increases and down-regulated when globin synthesis increases.

ABCG2, also known as BCRP/MXR/ABCP, is a member of the adenosine triphosphate (ATP)-binding cassette (ABC) transporter superfamily. This transporter expression is observed in the liver, kidney and intestines and is proposed to be involved in chemotherapeutic resistance. The overexpression of this transporter in cell lines confers resistance against chemotherapeutic

drugs. Besides this function, ABCG2 is also proposed to be involved in PPIX transport; the accumulation of PPIX in red blood cells of Abcg2$^{-/-}$ mice (mice where the gene of ABCG2 transporter is not expressed) and the up-regulation of ABCG2 during erythroid maturation (cells treated with DMSO in order to induce erythroid differentiation show high levels of ABCG2 mRNA in comparison to cells where erythroid differentiation was not induced) suggests that ABCG2 may be important in decreasing the cellular levels of PPIX. Additional experiments were designed, showing that the overexpression of the ABCG2 transporter lead to the decrease in the level of both exogenous and endogenous PPIX [15].

Both classified as 'stress proteins', FLVCR and ABCG2 seem to function similarly by getting rid of the excess of toxic heme or porphyrins during early and later stages of hematopoiesis – formation of blood cellular components, derived from hematopoietic stem cells. The activity of these two proteins may act as a supplement for HO-1 activity in bone marrow, where the requirement for oxygen by HO-1 might partially limit heme degradation in the physiological hypoxic conditions of the marrow [16].

More recently, another heme-binding protein was isolated from chicken, and was designated by heme-binding protein 2 or, more commonly, SOUL. The protein was classified as a heme-binding protein and grouped together with heme-binding protein 1 in the SOUL/HBP heme-binding family of proteins.

1.2. SOUL/HBP FAMILY OF HEME-BINDING PROTEINS

Using two-tissue subtractive hybridization technique together with database searches, Zylka *et al* detected in domestic chicken *Gallus gallus* high levels of expression of a novel gene in retina and pineal gland, that was designated as chicken *Soul* (*ckSoul)* [17]. Analysis of the resulting protein sequence allowed the identification of two similar proteins in mammals - murine p22HBP (mHBP) and murine SOUL (mSOUL). These three proteins were grouped in a new evolutionary conserved protein family – SOUL/HBP heme binding family of proteins, which also include human SOUL (hSOUL) and human p22HBP (hHBP). SOUL/HBP proteins are ubiquitous in nature, with bacterial, archaeal and eukaryotic representatives. Regarding the heme binding mechanism of these proteins, Taketani *et al* suggested the presence of a hydrophobic domain in murine p22HBP important in heme binding [13]. Sequence analysis shows that human p22HBP and ckSOUL also present this hydrophobic domain, which is not present in mSOUL and hSOUL. They have instead 3 charged amino acids in this region [17]. Further NMR and molecular dynamics studies brought important insights about the mechanism of heme binding in murine p22HBP and human p22HBP (discussed in detail in chapter *1.2.2*) [18, 19]. In this thesis, structural information on hSOUL was obtained and heme binding to the protein was studied. Until now, no clear function aside from heme-binding is attributed to SOUL/HBP family members.

1.2.1. HEME-BINDING PROTEIN 2, SOUL

In 1999, *Zylka et al,* by a combination of two-tissues suppression hybridization technique and database searches discovered a putative heme-binding protein (*hebp2*), SOUL, in chicken's retina and pineal gland (Northern blot studies showed high transcript levels of this protein in the pineal gland, an organ conjectured by Rene Descartes as the location of human soul; as a consequence, the protein was henceforth designated as SOUL protein) [17]. As the two mentioned tissues studied are involved in circadian clock mechanism, SOUL protein was immediately associated with this biological process. However, on further studies no detection of mSOUL in the suprachiasmatic nucleus (region in the brain's midline, responsible for controlling circadian rhythms) was observed, suggesting that the protein is not important for generating circadian rhythms. SOUL is mainly localized in the cytosol with much lower concentration in the mitochondria. No presence in the nucleus has been detected [17].

As mentioned previously, SOUL has been identified in human, murine and chicken tissues. Expressed Sequence Tags (EST) database ('collection of short single-read transcript sequences from GenBank) search revealed genes with sequence similarity in rice, tobacco and *Arabidopsis thaliana*.

Sato *et al*, in 2004, performed the first biochemical studies to characterize the SOUL protein. Studies on the heme binding properties of SOUL revealed that the protein specifically binds one heme *per* monomer; moreover, gel filtration analysis showed that *apo* SOUL exists as a dimeric protein becoming hexameric upon heme-binding. Analyzing the optical absorption spectrum of Fe(II) SOUL, three peaks could be observed at 422, 527 and 558 nm corresponding to a 6-coordinate low-spin heme with a 5-coordinate high-spin heme. The Soret peak of Fe(II)-CO complex of SOUL located at 418 nm rather than at 450 nm suggests the proximal ligand to be an Histidine (His 42, the only histidine present in the sequence of mouse and human SOUL). These results were confirmed by Raman spectroscopy studies. In addition, H42A mutant of mouse SOUL was constructed and reconstitution of this protein mutant with heme shows a much smaller Soret peak at 395 nm, which suggests a weak nonspecific heme binding to the protein [20].

Analyzing the sequence of SOUL - figure 1.3, the protein does not have any heme-binding motif [21, 22], but some hydrophobic amino acid segments which may be responsible for the heme binding, can be found [22].

In 2006, the solution structure of murine p22HBP was determined (the first structure from SOUL/HBP family of proteins). This protein presents 27 % sequence identity to hSOUL [18].

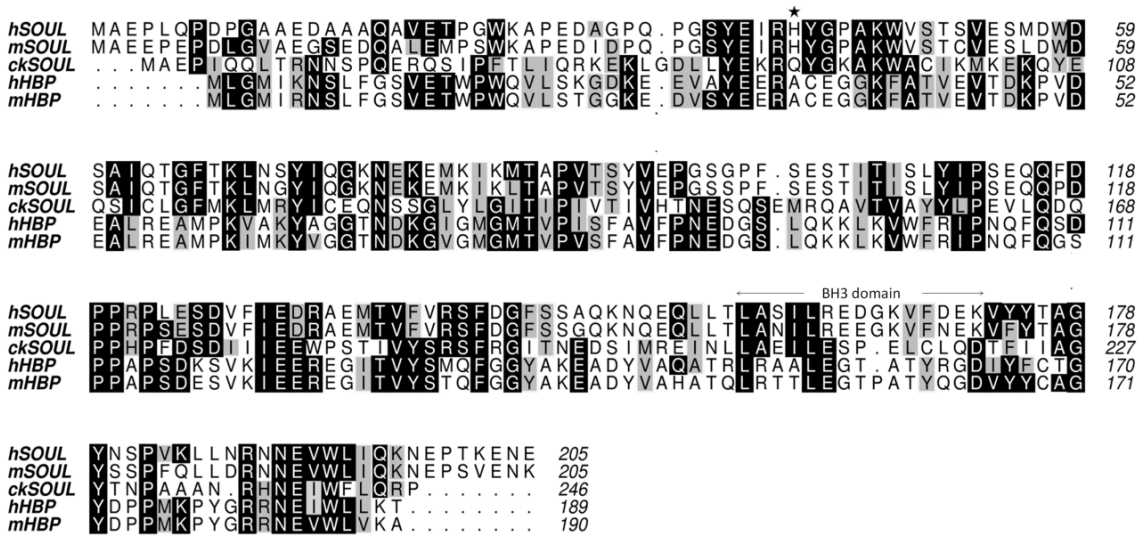


Figure 1.3 *Multi-sequence alignment of human (hSOUL), murine (mSOUL) and chicken (ckSOUL) heme-binding protein 2 with human (hHBP) and murine (mHBP) heme-binding protein 1 using ClustalW [23].* hSOUL His42, possible axial ligand of Fe (III) heme and the BH3 domain are indicated.

Chlamydomonas reinhardtii, a biflagellate unicellular alga, has been used as a model organism to study the photosynthesis process, as alga and high plants have a highly conserved photosynthetic apparatus. A proteomic analysis of the eyespot apparatus of *Chlamydomonas reinhardtii* showed the presence of different proteins including enzymes involved in the metabolism of carotenoids and fatty acids, some protein kinases and a SOUL heme-binding

protein that was classified as a putative signaling component [24, 25]. No further studies were performed with this protein.

Human SOUL is a 23 kDa heme-binding protein, belonging to the SOUL/HBP heme-binding family of proteins. hSOUL was first identified as a placental protein and designated PP23, and is expressed in a wide range of human organs, such as heart, kidneys, lungs, stomach and others [26]. Recently, the protein was also identified in human amniotic fluid [27].

The role of heme-containing proteins in cell death and survival is a well characterized event. These proteins can be involved in the formation of reactive oxygen species (ROS) inducing direct oxidative damage, and also in the induction of mitochondrial permeability transition (MPT) [28]. This process is directly implicated in both apoptotic and necrotic cell death [29, 30]. In recent studies, Szigeti *et al* suggested a new function for SOUL protein: the protein may be involved in necrotic cell death by permeabilizing both the inner and outer membrane of mitochondria [31]. In the study mentioned before, SOUL was overexpressed in NIH3T3 cells. In order to determine if SOUL affected ROS production, NIH3T3 cells were treated with H_2O_2 in order to make it easier to detect ROS formation. It was observed that SOUL did not increase the cellular levels of ROS. In order to understand how SOUL can induce cell death, MPT was induced in isolated mitochondria by applying a percoll gradient; SOUL itself did not induce mitochondrial swelling, however in the presence of a low calcium concentration (30 μM), mitochondria swelling was induced, a process that was inhibited by cyclosporine A. The mitochondrial membrane potential ($\Delta\Psi$) is directly related to MPT in a way that during MPT, a decrease in $\Delta\Psi$ can be observed. Again, SOUL, in the presence of 30 μM Ca^{2+} induced a significant decrease in $\Delta\Psi$.

Necrotic and apoptotic cell death events depend on the nature and intensity of the stimulus. MPT can, also depending on the nature and intensity of the stimulus, provoke apoptotic or necrotic cell death. Mock-transfected and SOUL overexpressing NIH3T3 cells were treated with 1M calcimycin, also known as A23187 (divalent cation ionophore, that allows ions to cross cell membranes) or 50 μM etoposide (anti-cancer agent that inhibits topoisomerase, causing DNA strands to break) for 24 hours in order to understand how SOUL is involved in one or both of these cellular events.

All the results show that SOUL can provoke both necrotic and apoptotic cell death. Moreover, in the presence of a high, but still physiological concentration of Ca^{2+}, SOUL could induce MPT and that effect could be inhibited by CsA, a specific inhibitor of MPT.

In a recent study, HeLa cells overexpressing SOUL protein were used to test the protein's response to oxidative stress conditions. To do so, the cells were treated with 300 μM hydrogen

peroxide (H_2O_2) for 24 hours. As a control, HeLa cells without SOUL overexpression were used. Immunoblot assays show that, in these oxidative stress conditions, SOUL concentration in mitochondria increases. Thus, the relocation of the protein from the cytosol into the mitochondria is proposed. In the same work, heme-binding protein 2 was shown to contain a very similar sequence region to the BH3-domain of some Bcl-2 family members. BH-3 domain proteins have been suggested to play a vital role in triggering mitochondrial-mediated apoptosis [32].

SOUL was therefore postulated to be a BH3 domain containing protein; in the (human) protein, the BH3 domain comprises residues Leu158 to Lys172 (LASILREDGKVFDEK). In order to understand if SOUL was in fact a pro-apoptotic protein, the protein was cloned without 9 amino acids (LREDGKVFD) comprised within the BH3-domain (ΔBH3-SOUL). ΔBH3-SOUL and SOUL overexpressing cells were treated with different concentrations of hydrogen peroxide from 0 to 500 μM. The results show that both cells, without SOUL and with ΔBH3-SOUL, are not sensitized to hydrogen peroxide, which leads to the conclusion that without its putative BH3 domain, SOUL cannot sensitize cells to H_2O_2 induced cell death.

It is stated that the action of BH3-domain proteins can be counter attacked by antiapoptotic proteins such as Bcl-2 and/or Bcl-x_L. To understand if SOUL interacts with any of these proteins, NIH3T3 cells were co-transfected with the full length of the open reading frame of Bcl-2 or Bcl-x_L and empty pcDNA or full-length SOUL cDNA containing pcDNA vector. In all cases, different concentrations of hydrogen peroxide (0 – 500 μM) were used. When only Bcl-2 or Bcl-x_L was overexpressed, the cells survived in all the concentrations of hydrogen peroxide used. Again, when SOUL was overexpressed, cell death occurred. When both, Bcl-2 or Bcl-x_L and SOUL were overexpressed the survival of the cells was similar to that of mock-transfected cells, which suggests that the anti-apoptotic proteins, Bcl-2 or Bcl-x_L, counter attacks SOUL activity. The same experiment was performed with ΔBH3-SOUL-overexpressing cells instead of SOUL-expressing cells; in these experiments no significant differences were observed between these cells and mock-transfected cells, which strengthens the case for the crucial importance of the BH3 domain in the cell death activity of SOUL.

At the molecular level, SOUL is proposed to facilitate cell death because of its interaction with the mitochondrial permeability transition (mPT) complex. Suppression of cyclophilin D, an integral membrane protein of the mitochondrial permeability transition pore (mMPTP), and overexpression of Bcl-2 or Bcl-x_L prevented the MMP-decreasing effect of SOUL, upon hydrogen peroxide (oxidative stress conditions). This is due to the fact that Bcl-2 analogues can inhibit mPT by binding VDAC and other components of the complex – figure 1.4.

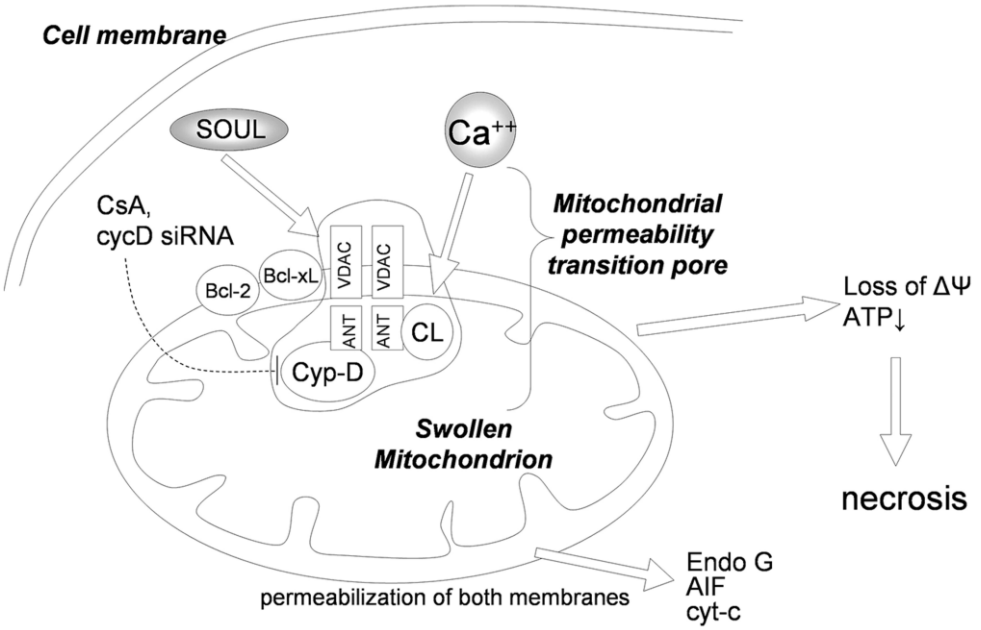


Figure 1.4 *Diagram of the proposed mechanism of hSOUL protein (from [32]).* In the presence of specific Ca^{2+} concentrations, SOUL protein induces permeability transition leading to the loss of the mitochondrial membrane potential. Bcl-2 or Bcl-xL prevents this process, indicating the direct effect of SOUL on the mitochondrial permeability transition pore (*mPTP*).

In summary, SOUL is proposed to promote the permeabilization of the inner and outer membranes of mitochondria in oxidative stress conditions; this can be reversed by removing the BH3 domain.

Genomic studies identified a novel susceptibility locus for rheumatoid arthritis (RA) (rs6920220). This locus lies close to TNFAIP3, a negative regulator of NF-kB in response to TNF (tumor necrosis factor, a cytokine involved in inflammatory response) stimulation. Heme-binding protein 2 (HEBP2) gene is also close to this region, which raised the question whether SOUL is involved in rheumatoid arthritis. To test this hypothesis, Peripheral Blood Cells (PBD) of rheumatoid arthritis patients were used to determine the expression level of mRNA of SOUL gene compared to healthy people. The results show an increase of approximately 2.5 folds of SOUL expression in RA patients [33]. No further studies were published regarding the role of SOUL protein in this disease.

Recently, studies were performed in order to deeper understand the possible pro-apoptotic function of hSOUL by studying the possible binding of the protein to anti-apoptotic protein Bcl-x_L. For that, the BH3 domain identified in hSOUL was synthetized and co-crystalized with the anti-apoptotic protein Bcl-x_L. The complex structure was solved by X-ray Crystallography to 1.95 Å showing the interaction between the SOUL BH3 domain peptide and Bcl-x_L. In addition, 2D ^{15}N-HSQC spectrum shows chemical shift changes upon SOUL BH3 domain peptide

addition to Bcl-x_L. These results suggest an interaction between hSOUL and anti-apoptotic proteins from Bcl-2 family of proteins, namely Bcl-x_L. However, data from Surface Plasmon Resonance (SPR) using intact SOUL protein with Bcl-x_L showed no interaction between the two proteins [34].

1.2.2. HEME-BINDING PROTEIN 1, P22HBP

In 1998, Taketani *et al* isolated a novel heme-binding protein from mouse liver cytosol. The protein, with approximately 22 kDa, was designated p22HBP – p22, heme-binding protein. The protein initially isolated was purified with a hemin-bound Sepharose column, hence the designation heme-binding protein. The cDNA of the protein was afterwards isolated as a 1003 base pairs (bp) insert with an open-reading frame (ORF) of 630 bp, encoding a soluble 190 amino acids protein, 21063 Da [35]. Protein sequence analysis showed no heme-binding motif, however, positions 73 to 82 consisted of a hydrophobic pocket that could be involved in heme binding [21].

RNA blots of the expression of p22HBP in mouse tissues showed an extremely high level of protein expression in the liver. Since heme metabolism is higher in the liver p22HBP can be considered an important protein in heme metabolism [35]. More recently, it was demonstrated that p22HBP is highly expressed in hematopoietic tissues like fetal liver and bone marrow [36].
p22HBP was subsequently cloned in pEGFP-N1 system in order to overexpress the protein fused with the green fluorescence protein (GFP) allowing the determination of the cellular localization of the protein. The results clearly indicate that the protein is localized throughout the cytosol but not in the nucleus [1].
A p22HBP/GST fusion protein construct was used to study the heme binding to the protein and its oligomerization state. Upon incubation with hemin *in vitro*, the UV-visible spectrum shows a Soret band at 408 nm due to the hemin-protein complex formation. Gel filtration analysis showed that the protein exists as a monomer, binding one heme molecule *per* monomer [13].
A recent proteomic study revealed that p22HBP belongs to one of the complexes (complex III) involved in hemoglobin metabolism, namely, in hemoglobin synthesis. The high concentration of p22HBP in the cytosol indicates that this protein is a positive regulator of heme biosynthesis in erythroid cells [37] .

For an adequate immune response, one of the key initial steps is the chemotaxis of dendritic cells (DCs) and monocytes. For that, formyl peptide receptor (FPR) and FPR-like receptor (FPR) 1, two G-coupled proteins, and FPR-like receptor (FPR) 2 play an important role in host defense mechanisms against bacterial infection and in inflammatory response regulation. The FPR receptors are highly expressed in phagocytic cells such as neutrophils, monocytes and

dendritic cells and the 3 receptors mentioned previously have different mechanism of action [14]. FPR is a high affinity receptor for formyl peptides whereas FPR1 is thought to be a promiscuous receptor that can be activated by several ligands. FPR2 does not respond to formyl peptides and has low affinity to several of the FPR1 agonists.

F2L is an acetylated amino-terminal peptide derived from the cleavage of the human heme-binding protein 1 (residues 1-21) and shows high affinity and high selectivity to FPLR2, binding and activating this receptor in the low nanomolar range. This peptide, when bound to FPRL2, activates the cascades stimulated by chemo-attractants which leads to the intracellular calcium release, inhibition of cAMP production and activation of MAP kinases ERK 1/2 [38].
On the other hand, F2L inhibits FPR and FPRL1 mediated signaling; F2L interacts with both receptors and inhibits calcium signaling and superoxide generation induced by the activation of FPR and FPRL1. This means that besides the function as a heme-binding protein, p22HBP may also be involved in cell death and infection due to this acetylated N-terminal fragment.

p22HBP is part of an evolutionary conserved family of heme-binding proteins together with the SOUL protein. The overall protein structure consists of a 9 stranded twisted β-barrel flanked by two α-helices. Analyzing the structure of murine p22HBP from residue 29 to residue 190, a pseudo-2-fold symmetry can be found with two repeats comprising β-β-α-β-β motif, resulting from a gene duplication event [39].
Structural similarity searches revealed SbmC from *Escherichia coli* (*E. coli*) [39], the C-terminal of Rob, an *E. coli* transcriptional factor [40] and the C-terminal of the multi-binding domain of transcription activator BmrR from *B. subtilis* [41, 42] as similar proteins. It is quite interesting to notice that the binding site of both p22HBP and BmrR is very similar: both proteins bind small hydrophobic ligands and the location of the binding site is very similar [36].
p22HBP was found to bind others porphyrins besides hemin, where its role in the cell is unknown [18, 36]. ^{15}N-labelled p22HBP 2D ^{1}H, ^{15}N TROSY spectra were acquired to follow chemical shift changes upon hemin and PPIX addition. At equimolar ratios the NMR spectra show the resonances of the protein-hemin complex, consistent with the formation of a 1:1 high affinity complex [18]. Moreover, in PPIX-murine p22HBP, twice the number of shifted peaks is observed suggesting that two orientations are possible for PPIX.

So far, only the structure of murine p22HBP, in the *apo* form, from the SOUL/HBP family of proteins has been determined. Since human p22HBP has 86% sequence identity to murine HBP one can assume that they both present a very similar overall fold.
In order to increase our understanding of the binding of heme to murine p22HBP and human p22HBP, molecular modeling (docking of tetrapyrroles, namely, hemin, PPIX, coproporphyrin I

(CPI) and coproporphyrin III (CPIII), to these proteins) and unconstrained molecular mechanics/dynamics studies were performed. For murine p22HBP, the NMR solution structure was used [18]. Regarding human p22HBP, no structural information is yet available. However, the protein has 86% sequence identity to murine p22HBP and its structure was predicted using the software Modeller [43].

Docking studies of human and murine p22HBP with hemin and PPIX indicate that the orientations of both tetrapyrroles is similar, involving mainly electrostatic interactions with lysine 177 (176 in human p22HBP) and arginine 56 implicated in the stabilization of the propionate groups of the tetrapyrroles – figure 1.5. Molecular mechanics/dynamics studies show that in both murine and human p22HBP the hemin ring is buried inside the hydrophobic pocket as described by Dias *et al* and flanked by α_A and β_8-β_9 [18]. This hydrophobic pocket is constituted mainly by non-polar residues creating a solvent exposed hydrophobic region, which is highly conserved in these two proteins, human p22HBP and murine p22HBP. For murine p22HBP-PPIX complex, ring current shift (RCS) data was used along with molecular modeling data to understand PPIX binding to the protein. In this case, the propionate group of PPIX is stabilized by lysine 64 from α_A helix. In addition, and confirming the results from NMR data, the PPIX is thought to bind to the protein acquiring two different orientations. As mentioned previously, PPIX interacts preferentially *via* a propionate group with lysine 64. A second possible orientation involves an electrostatic interaction with arginine 56.

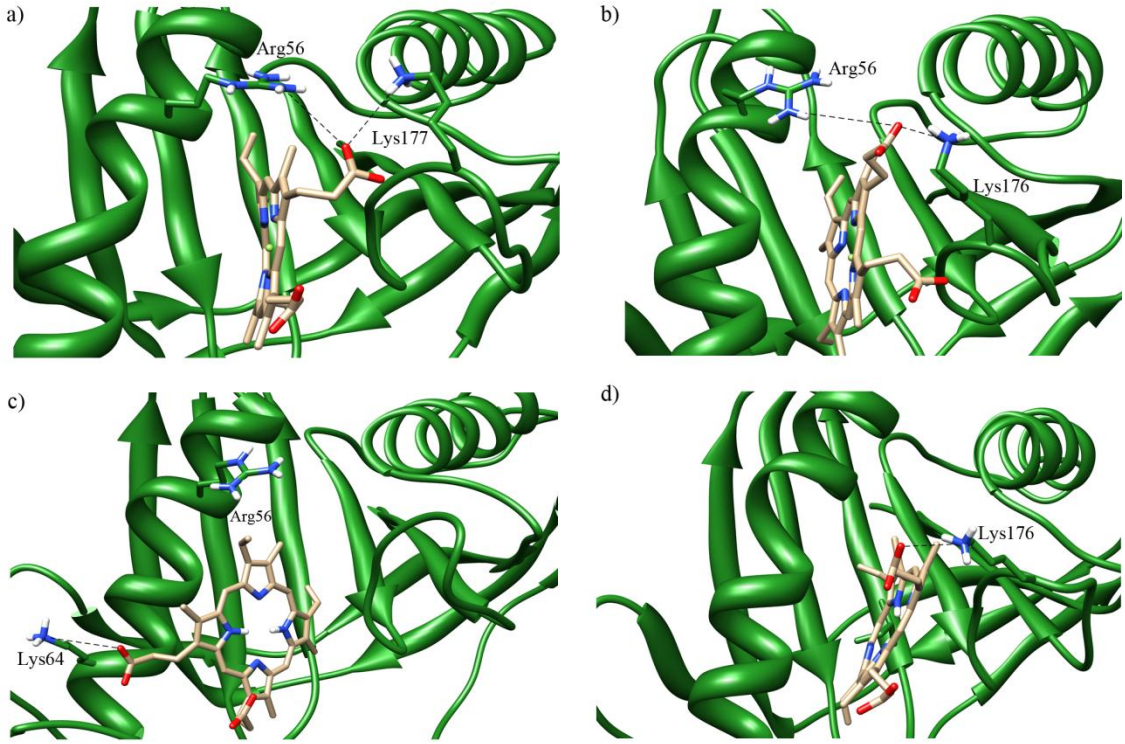


Figure 1.5 *Representative structures of human and murine p22HBP complexed with PPIX and hemin (from [19]), (a) murine p22HBP + hemin, (b) human p22HBP + hemin, (c) murine p22HBP + PPIX, (d) human p22HBP+PPIX.* The protein is rendered in cartoon with key side chain residues rendered in sticks, with the corresponding residue name.

Molecular dynamics studies indicate that the loop between β_8 and β_9 displays different mobility upon hemin/PPIX binding which indicates that this loop may be important in the heme-binding mechanism [19].

It is clear that, unlike in SOUL, where the heme is proposed to bind *via* His42, the binding of heme in p22HBP occurs at a hydrophobic pocket close to helix α_A and β-sheets β_8-β_9. Optical spectroscopy and EPR data show that iron coordination does not change upon hemin binding to murine p22HBP, which reinforces the assumption that heme binds to p22HBP and SOUL differentially.

SOUL/HBP proteins have also been identified in plants. In *Arabidopsis thaliana*, a typical model for plant biology studies, 6 homologous genes to the p22HBP/SOUL family have been identified. One of the genes was later identified as a pseudogene. From the remaining five, two contained an N-terminal amino acid sequence functioning as signal peptides to organelles. The protein products of the remaining genes were designated as cHBP1, cHBP2 and cHBP3. All cHBP show same conservation in the hydrophobic pocket, involved in heme binding to p22HBP, whereas the histidine residue proposed to act as axial ligand in SOUL protein is not conserved. cHBP1 is expressed in the leaves, whereas cHBP2 is mainly expressed in roots.

cHBP3 shows a lower level of expression on both. Further studies were performed only with cHBP1 and cHBP2, as it was not possible to overexpress recombinant cHBP3 protein in *E. coli*. Fluorescence Quenching experiments were performed with cHBP1 and cHBP2 with hemin, protoporphyrin IX (K_d values shown in table 4.1) and Mg-protoporphyrin IX dimethyl ester. The results show no significant changes in the determined K_d values independent of the presence or absence of a metal ion, so cHBPs do not bind heme *via* an axial ligand. Furthermore, in cHBP1, the two histidine residues were mutated to alanine. Heme addition shows heme-binding to the protein confirming that the histidine residues are not involved in the binding mechanism. Electron Spin Resonance spectroscopy (ESR) studies showed that cHBP bind high-spin type heme, which is in accordance with the previously published results for p22HBP. In summary, cHBP1 and cHBP2 can reversibly bind heme, protoporphyrin IX and Mg-protoporphyrin IX dimethyl ester, so these proteins are possible tetrapyrroles carrier proteins in the cytosol [44].

Although structural information on both murine p22HBP (solution structure by NMR) and hSOUL (X-ray Crystallography structure) is now available, a lot of questions still remain to be answered in the characterization of the members of SOUL/HBP family of proteins; solving the solution structure of murine p22HBP was a first step to elucidate the heme binding to p22HBP proteins; NMR data lead also to the mapping of the possible heme binding site of murine p22HBP. *In silico* data, not only confirmed the binding pocket for heme in murine p22HBP, but also showed the possible way of heme binding in human p22HBP. However, it is not clear if the transport/buffering of heme in the cytosol is the only/main function of p22HBP.

1.3. BCL-2 FAMILY OF PROTEINS IN CELL APOPTOSIS

Programmed cell death – apoptosis, plays a vital role in normal development, tissue homeostasis and removal of damaged/infected cells.

During the apoptosis process, condensation of the nucleus and its fragmentation into small pieces are observed together with an extensive hydrolysis of nuclear DNA. Although less 'intense', the Golgi complex, endoplasmic reticulum and mitochondria also undergo fragmentation. During this process, several proteins are released from the mitochondrial intermembrane compartment, namely cytochrome c, that can trigger caspase-activating complex – figure 1.6 [45]. The permeabilization of the mitochondrial outer membrane is activated by Bax and/or Bak which are themselves activated by one or more BH3-only proteins that are proposed to act as sensors of cellular stress. This permeabilization leads to the release of mitochondrial intermembrane proteins. This pro-apoptotic activity is prevented by anti-apoptotic activity proteins, such as Bcl-x_L or Bcl-2. Recent studies suggest that Bax protein is continuously translocated into the mitochondria; in healthy cells, the anti-apoptotic Bcl-x_L binds to Bax and is retranslocated into the cytoplasm. The retranslocation to the mitochondria requires a conformational change that is proposed to be carried out by BH3-only proteins [46]. The binding of Bcl-x_L to Bax is in agreement with previous studies that suggested that BH3 containing proteins such as Bak and Bax bind to the hydrophobic groove on the surface of Bcl-2 or Bcl-x_L.

All the proteins mentioned above belong to the B-cell lymphoma-2 (Bcl-2) family of proteins which are essential, as further discussed in chapters *1.3.1* and *1.3.2.*, to the cell apoptosis process.

Once cell apoptosis starts, cells tend to detach from their neighboring cells and lose contact with the extracellular matrix, which facilitates their future removal by phagocytes. The generation of binding sites for these phagocytes and the release of chemoattractants constitute the last step of the apoptosis process [45].

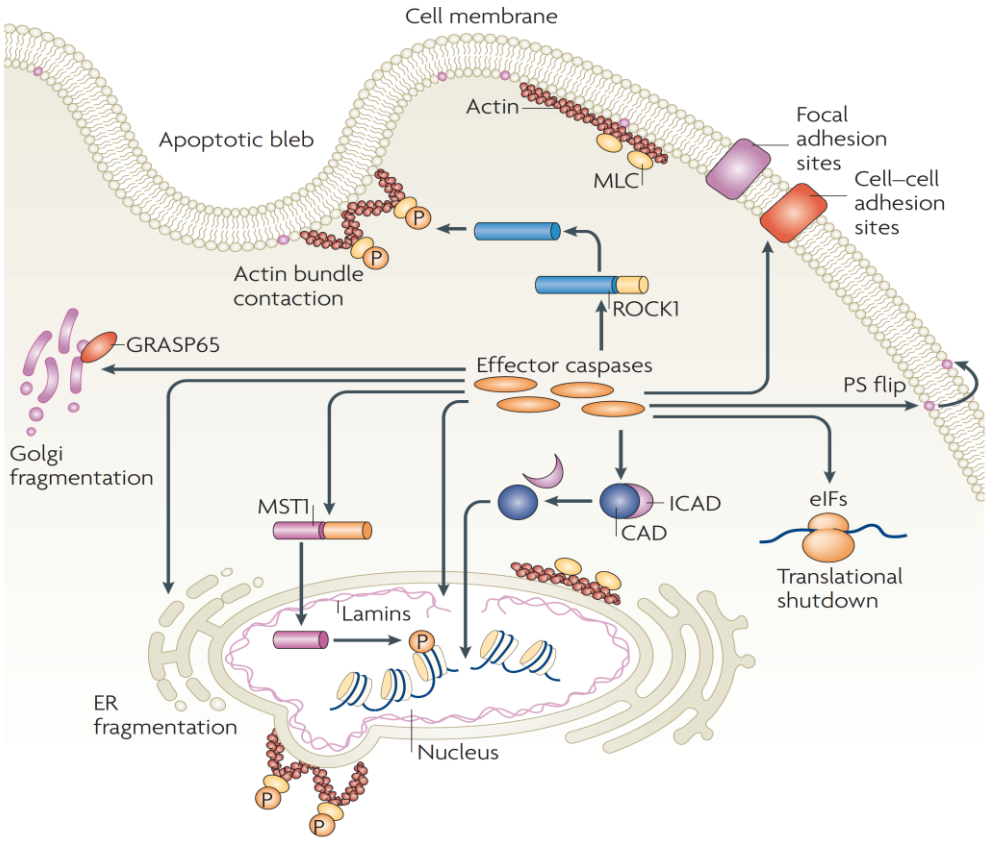


Figure 1.6 *Caspases as responsible agents for cellular organelles demolition (from [45]).* Caspases activity leads to the destruction of cellular organelles such as the Golgi complex (caspases provoke the cleavage of the Golgi-stacking protein GRASP65 and other Golgi proteins) and endoplasmic reticulum. Caspase-mediated cleavage of nuclear lamins weakens the nuclear lamina, allowing nuclear fragmentation, and nuclear envelope proteins are also proteolysed. Caspases are then responsible for the cleavage of the constituents of the cytoskeleton and subsequent dynamic membrane blebbing.

1.3.1. BCL-2 FAMILY OF PROTEINS

The apoptosis process can be divided into two main stages. An initialization stage when the process is triggered and the second step when the process effectively takes place. This second stage is regulated by the proteins belonging to the Bcl-2 family of proteins, which contain up to four regions of sequence homology (BH1 - BH4). The proteins belonging to this family can be divided in to three different subfamilies according to their structure and function – figure 1.7:

- The anti-apoptotic proteins that can contain the four BH domains, and includes Bcl-2, Bcl-x_L, Bcl-w, MCL1, Bcl2A1 and Bcl-b, for example;
- The pro-apoptotic membrane permeabilizing proteins that contain the homology domains 1 to 3, including for example BAX, BAK and BOK;
- The pro-apoptotic proteins that contain only the BH3 domain, that include for instance BIK, HRK, BIM, BAD, BID, PUMA, NOXA and BMF.

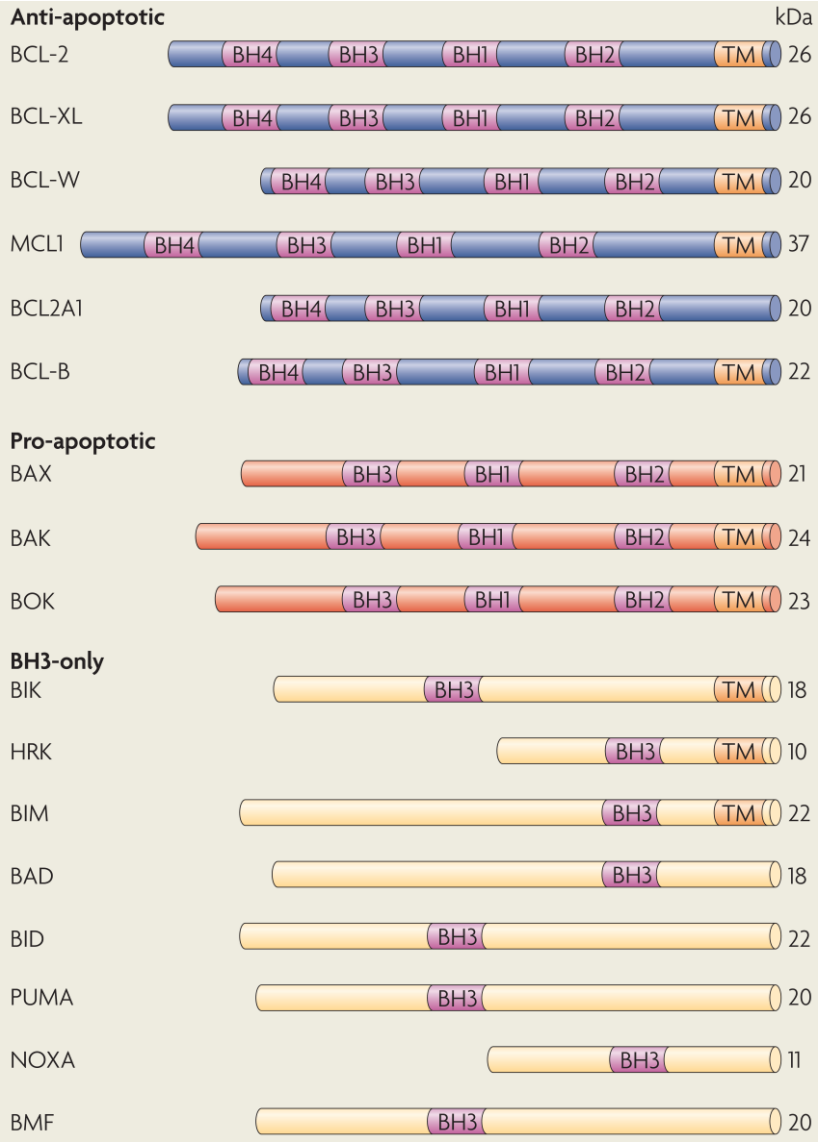


Figure 1.7 *Bcl-2 family of proteins.* The anti-apoptotic members of this family contain all four homology domains (1-4). The pro-apoptotic BAX-like subfamily lacks BH4 domain and promotes apoptosis by forming pores in mitochondrial outer membranes. The BH3-only subfamily is a structurally diverse group of proteins that only display homology within the small BH3 motif. A great number of the members of this family contain a transmembrane domain (TM) (from [45]).

Although with often very different primary sequence and function, all the members of the Bcl-2 family whose structure has been determined show a very similar fold, consisting of two central mainly hydrophobic helices surrounded by 6 or 7 amphipathic helices [47]. Another characteristic feature is a long unstructured loop connecting the first two α-helices.

Several important roles have been attributed to the Bcl-2 family of proteins. Among them is the important role that these proteins may have in cancer, namely some of the anti-survival proteins of this family that can act as tumor repressors as they promote cell death, therefore destroying cancer cells. On the other hand, all pro-survival Bcl-2-like proteins can be oncogenic as their main function is the cell death inhibition [48]. Since Bcl-x_L and Bcl-2 proteins act as anti-

apoptotic proteins, they are often responsible for the survival of cancer cells. Due to this, several Bcl-2 anti-sense nucleotides have been studied and some of them are already on clinical trials.

The expression levels of the members of this family of proteins are mainly regulated by cytokines, but other death-survival signals at different levels are also important in this regulation. To maintain homeostasis, a tight control of the ratio of pro-apoptotic and anti-apoptotic proteins is essential.

An interesting feature of some pro-apoptotic and anti-apoptotic members of this family is that they can heterodimerize, a process that is essential for BH3-only proteins; in addition they can titrate each other, a mechanism that is proposed to be controlled by their relative concentrations in the cell [49].

According to the results obtained by Adams *et al*, from mutagenesis studies, the BH1, BH2 and BH3 domains of these proteins influence homodimerization and heterodimerization. For instance, Bcl-x_L which is a very important anti-apoptotic protein, heterodimerizes due to the presence of an extensive surface that includes both BH1 and BH2 domains, with pro-apoptotic proteins, such as BAK protein. The BH3 domain of Bcl-x_L is responsible for the binding to anti-apoptotic proteins [48].

Regarding the BH4 domain, several studies [50-52] proposed that this domain is essential to the anti-apoptotic capacity of the anti-apoptotic proteins from Bcl-2 family of proteins. Removal of this domain leads to the incapacity of these proteins to induce cell apoptosis. Moreover, studies on isolated mitochondria showed that recombinant Bcl-x_L prevents both Ca^{2+}-induced mitochondrial membrane potential ($\Delta\psi$) loss and cytochrome c release performed by the voltage-dependent anion channel (VDAC). Recombinant ΔBH4 Bcl-x_L presents no anti-apoptotic activity. This is due to the fact that the BH4 domain of Bcl-x_L inhibits VDAC activity, probably by closing this channel [53].

1.3.2. BH3-ONLY PROTEINS

The subfamily of pro-apoptotic proteins can itself be divided into two sub-groups: the first one with pro-apoptotic proteins that contain the first 3 homology domains, BH1, BH2 and BH3, such as BAX, BAK and BOK, and the second group with BH3-only proteins that, as the name indicates, only contain the BH3 domain, such as BIK, BIM and BID. The pro-apoptotic proteins belonging to the first group undergo conformational changes that will lead to their oligomerization and subsequent insertion in the outer mitochondrial membrane. Genetic studies on *C. elegans* showed that BH3-only proteins act as death effectors. Being responsible for cell apoptosis, BH3-only proteins have an important role in tumorigenesis as they can act by provoking tumour cells death [54]. Besides being involved in cell apoptosis, Hetz and Glimcher showed that BH3-only proteins may also be involved in other cell functions, namely in cell

cycle regulation [55]. In a different study, Scatizzi *et al* showed that BIM and BID, two BH3-only proteins, protect cells against rheumatic arthritis [56, 57].

Interestingly, many proteins not belonging to the Bcl-2 family of proteins – figure 1.8 - have been identified as containing BH3-like domains and having death-inducing activity and/or the capacity to interact with anti-apoptotic Bcl-2 family of proteins. The existence of these proteins indicates that structurally different proteins, that contain the BH3-domain, can have the same function and mechanism as BH3-only proteins belonging to the Bcl-2 family of proteins [58].

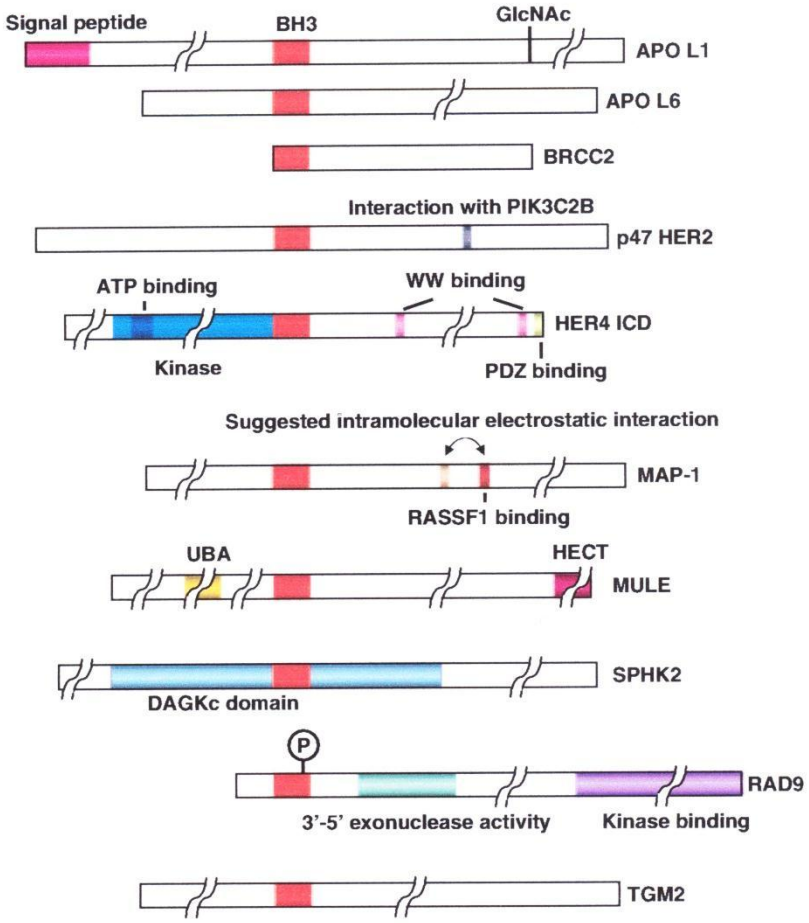


Figure 1.8 *Domain structures of some BH3-only-like proteins.* Example of BH3-like proteins and corresponding domain functions (adapted from [58]).

The BH3 motif in the Prosite [59] database (PS01259) consists of 15 amino acids, with a hydrophobic residue in the first position and conserved residues Leu, Gly and Asp at positions 5, 9 and 10, respectively. Surprisingly, these BH3-only domain proteins bind specifically to anti-apoptotic proteins that contain the four domains (BH1, BH2, BH3 and BH4), binding only to a limited number of anti-apoptotic proteins that contain only domains 1 to 3, which may suggest some role of the BH4-domain in binding of these proteins.

Transcriptional and post-transcriptional regulation is observed in order to maintain the steady-state level of BH3-only proteins. In post-transcriptional regulation, phosphorylation of BH3-

only proteins has been reported to influence apoptotic activity either by increasing or decreasing this activity. Some of the BH3-only proteins have been predicted as containing transmembrane domains, which indicates that these proteins are targeted to intracellular membranes. BIK, BIM, BNIP3, BAD, BID and NOXA, for instance, do not have any obvious domain for membrane interaction.

The way BH3-only proteins act is so far described by two distinct models, both involving BAX/BAK. This first model suggests the ‘direct binding and activation of BAX/BAK’ and the second the ‘neutralization of anti-apoptotic Bcl-2 family proteins and displacement of BAX/BAK’. The direct model postulates that BH3-only proteins directly activate Bax/Bak. The second model postulates that BAX/BAK activation occurs when these proteins are liberated from Bcl-2 proteins by BH3-only proteins.

1.4. BIOMOLECULAR CRYSTALLOGRAPHY

Proteins are essential components to all living organisms, virtually participating in every process within the cell. Enzymes, for instance, are very important for the catalysis of biochemical reactions, playing an essential role in metabolism. Proteins are also involved in functions like cell signaling, immune response and the cell cycle. To understand the biological processes at the molecular level, it is critical to study the protein's structure; the molecular structure provides the framework for the dynamics and function of proteins. One of the best examples of the importance of structural information is the elucidation of the DNA molecule structure; structural information on this double helical lead to the elucidation of the genetic code and the subsequent exponential growth in the genetics field. Cryo Electron Microscopy, Nuclear Magnetic Resonance and X-ray Crystallography are the main techniques used to determine protein structure and therefore understand protein function and dynamics.

These two last techniques are nowadays strongly used in drug design studies. X-ray Crystallography is a very powerful technique to investigate the interaction of small molecules with proteins. Once good diffracting crystals of a protein-ligand complex are obtained and the structure is solved, the specific interactions of a particular drug with a protein can be studied in detail. The difficulty often faced in obtaining good diffracting crystals together with the recent advances in NMR spectroscopy has resulted in this technique becoming more commonly used in protein structure determination and drug screening. NMR is also a very powerful tool for studying molecular recognition and the interactions of small ligands with biologically relevant macromolecules [60].

1.4.1. INTRODUCTION TO X-RAY CRYSTALLOGRAPHY

The basic single-crystal diffraction experiment is conceptually quite simple: a single crystal is placed into a focused X-ray beam, which will be scattered by the electrons in the molecules of the crystal, and the diffraction images are collected. Fourier methods are used to reconstruct the electron density that represents the atomic structure of the molecule in the crystal. With regard to X-ray Crystallography, the process from pure protein to the final macromolecular structure can be described in five main steps: crystallization, data collection, phasing, model building, and refinement and validation [61]. The range of things one can see is limited by the wavelength used. Therefore, things that are much smaller than the wavelength used cannot be seen. This is the reason for the use of X-ray radiation in protein structure determination - the wavelength of this radiation has the same magnitude as the interatomic bond lengths, allowing us to 'magnify' the internal organization of the studied protein.

Historically, the first biological diffraction pattern was acquired by Bernal and Crowfoot, in 1934, for pepsin [62]. However, the first protein structure (from myoglobin) was only

determined in 1958 by Kendrew *et al* [63]. In the last 50 years, huge developments have been made in this technique; 'in-house' X-ray sources are now routinely substituted by synchrotron X-ray sources – in contrast to laboratory sources, synchrotrons generate X-rays over a wide range of wavelengths essential for Multiple-wavelength Anomalous Diffraction (MAD) experiments. Moreover, the high brilliance of synchrotron X-rays allows the collection of data from very small crystals (10 μm and sometimes below). As ionizing radiation, X-rays are capable of separating high energy electrons from inner atomic levels; these free electrons can directly interact with other atoms, break bonds, or generate free radicals which can induce damage to the protein crystal. In order to prevent radiation damage to crystals and increase their lifetime, crystals that were initially mounted in sealed glass capillaries and exposed to the X-ray beams are nowadays mounted in loops and routinely cryocooled (which allow the solvent surrounding the crystal and in the solvent channels to glassify, avoiding ice formation and subsequent crystal damaging) and X-ray experiments performed at around 100 K, which delays crystal destruction due to radiation damage.

In recent years, another great improvement occurred with the introduction of crystallization robots. The first crystallization robots appeared in the 1980s although they were only capable of reproducing crystallization conditions. In 1990, Douglas Instruments, Ltd. (www.douglas.co.uk) commercialized the first crystallization robot that could miniaturize the drop volume. This was an important step in X-ray Crystallography as it meant that more crystallization conditions with less protein could be tested. To some extent, it represents the turning point when crystallization and X-ray structure determination became a High Throughput process.

It is important to emphasize that one of the main problems in using crystallization robots is that the crystals obtained are often quite small making it difficult to perform the X-ray experiment, which poses the need to scale-up the crystallization conditions which, in turn, can be a very difficult process.

1.4.2. The 'bottleneck' of Macromolecular X-ray Crystallography and data collection

Macromolecular X-ray Crystallography is based on the diffraction of X-ray radiation emanating from its interactions with a protein crystal. A single molecule if subjected to an X-ray beam would cause the scattering of this radiation. However, this scattering would be translated into a very weak signal. Therefore, in order to amplify this signal, the molecules must be arranged in a periodic crystal lattice, and the microscale molecular scattering contributions will amplify the signal resulting in sharp diffraction spots at the diffraction angles described by Bragg's law – equation 1.4 and figure 1.10 (section *1.4.3*).

The basic principle that needs to be fulfilled in protein crystal formation is to reduce the protein solubility. The protein solutions used in Crystallography are supersaturated and for that reason the molecules tend to collide against each other; if this occurs in orientations that allow favorable contacts, the local binding energy tends to overcome the loss of entropy due to the order increase. Thermodynamically, this represents a lowering in the local energy of the system, which leads to a favorable event. This can happen until the formation of a nucleus is observed. Once the nucleus reaches a critical size, the collision of other molecules or aggregates can result in the connection of these to the previously formed nuclei. This process can lead to protein crystal formation.

The nucleation process summarized previously can be graphically described if one adds kinetic information to the thermodynamic information. This representation is called 'crystallization diagram'- figure 1.9. The main goal of this diagram is to aid the understanding of the crystallization process (one has to keep in mind that in these diagrams only the solubility line can be experimentally determined, with the remaining information being not so well defined).

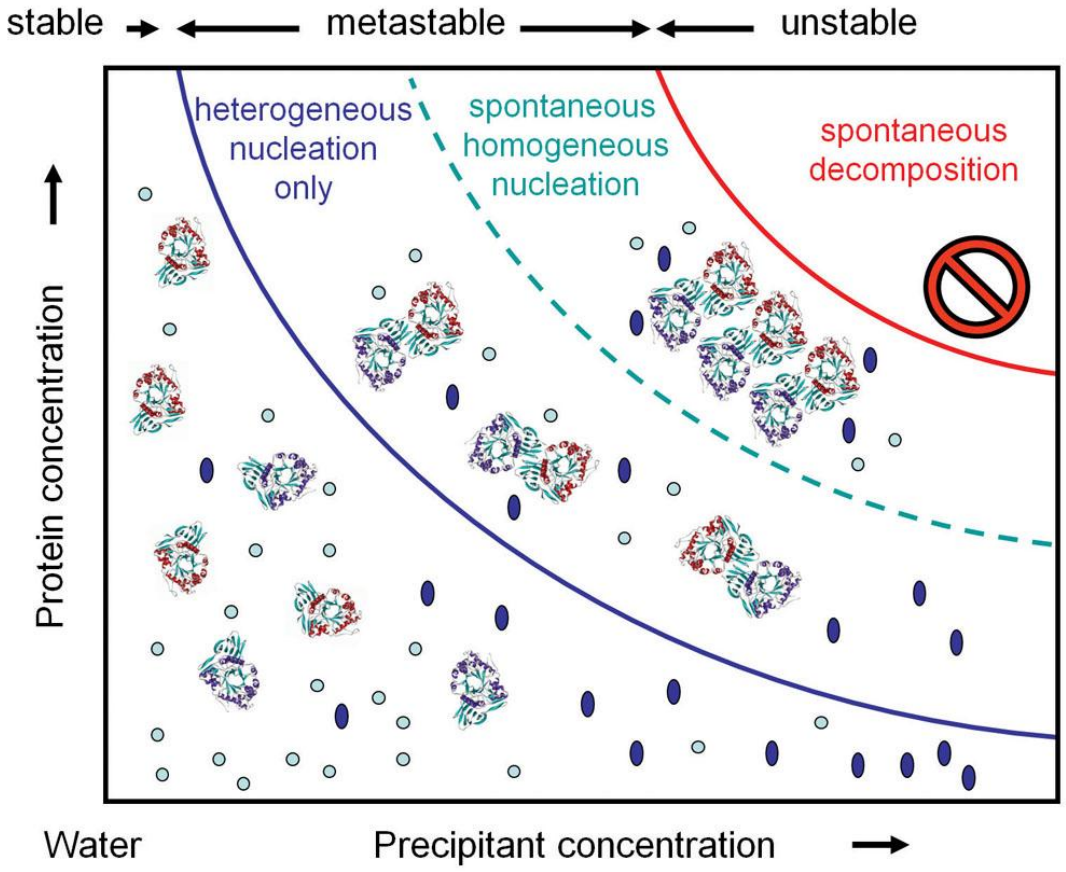


Figure 1.9 *Crystallization diagram (from [61]).* The light blue circles represent water molecules and the dark blue ovals represent precipitant molecules. As general rule, higher saturation will promote spontaneous formation of stable crystallization nuclei (homogeneous nucleation).

Once pure protein is obtained, getting a good diffracting crystal is the main goal in the process of determining the three-dimensional structure of a macromolecule. For that purpose, an

enormous number of conditions can be tried and can include different precipitant reagents, pH, temperature and crystallization methods. Once one has a starting condition for protein crystallization, it is often good idea to a) add or remove salt – due to the salting in - salting out effect, b) make the protein solubility change, c) add organic compounds, d) vary pH (the protein should be in weak buffers (approximately around 10 mM concentration) in order to be the crystallization cocktail driving the pH of the crystallization drop), e) vary temperature, f) change the dielectric constant of the solvent, etc, in order to optimize the obtained crystals. Among the known crystallization methods, the hanging-drop vapour diffusion methods are by far the most used in protein crystallization. In these methods, a protein/precipitant mixture is allowed to equilibrate over a reservoir containing the precipitant solution in larger amount. A siliconized glass slide is used to seal this compartment. On the hanging-drop vapour diffusion method, the protein/precipitant mixture is put in this cover slip. On the sitting-drop vapour diffusion method, the protein/precipitant mixture is put in a support inside the reservoir. The protein and precipitant concentrations on the drop will slowly increase by the water transfer to the more concentrated solution in the reservoir. This water transfer will stop once the protein solubility limit is reached. If our crystallization conditions are within the supersaturating zone, nucleation can occur and the formation of crystalline nuclei may lead to the formation of a protein crystal. Other methods used are the microbatch under oil, microdialysis and free-interface diffusion [64].

Some *a priori* information regarding the protein of interest can be of great importance to this process, namely, the secondary structure prediction once the protein sequence is obtained. Moreover, the net charge of the protein is a very important characteristic that can influence its crystallization. This parameter can be calculated according to the protein solution pH and number of charged amino acids. However, it is not possible to determine the local charge along the protein which makes it impossible to systematically predict the most probable conditions to promote protein crystallization. Bioinformatics tools like GlobProt [65], ONDR [66], DisEMBL [67], and others, can be used to predict disordered regions in the protein of interest to consequently try to remove or mutate some amino acid residues in these regions in order to induce stability of the protein and consequently facilitate its crystallization. Experimentally, the protein stability and conformational state can be studied before starting the crystallization experiments. A brand new approach is the use of thermofluor stability assays which allow the study of the effect of buffers, additives and cofactors on the protein stability [68]. Two other excellent techniques for the study of the protein oligomerization state are Native (non-denaturating) polyacrylamide gel electrophoresis (PAGE) and size exclusion chromatography (SEC). NMR spectroscopy can be used to infer the folding of the protein. ^{1}H, ^{15}N Heteronuclear

Single-Quantum Coherence (HSQC) experiments show if the protein is properly folded or partially/totally unstructured.

It is often very difficult to obtain good diffracting crystals even after a wide range of crystallization conditions have been tried. For this reason, it is often important, before exhaustively trying more and more crystallization cocktails, to re-think the protein of interest. Protein engineering can, as a consequence, be of great importance in macromolecular X-ray Crystallography.

In summary, due to the arbitrary nature inherent in the process of crystallizing a protein it is often designated as the "bottleneck" step in the dynamic process of solving a protein's structure by X-ray Crystallography.

Once a protein crystal is obtained, data collection is the next step in the process of protein structure determination. At this stage, some important decisions will have to be taken in order to optimize the process and increase the chances of solving the protein's structure. The first diffraction pattern will show the extent to which the crystal diffracts. Low resolution data (> 4 Å) can be useful for phasing but not enough to refine a good quality model. The other parameter which is very critical in data collection is the exposure time. If it is too short the signal to noise ratio will be low and will give low resolution data; long exposure time will lead to highly saturated spots. One should also estimate the crystal mosaicity (angular measure of the degree of long range order of the unit cell within a crystal) in order to determine the best rotating angle. Data are usually collected in snapshots taken during a small rotation of the crystal (φ angle). The larger the oscillation angle, the more reflections are collected on a single exposure. If the oscillation is too big , reflections may overlap; high mosaicity crystals require the use of fine slicing in data collection, *i.e.*, the rotating angle must be small, 0.1 - 0.3 Å, in order to have the complete profile of every spot. Lower mosaicity indicates better ordered crystals. Therefore, it is essential to set up a strategy for data collection that will be highly dependent on the purpose of data collection, for example, molecular replacement phasing or anomalous diffraction data.

Once the X-ray experiment is performed, raw data must be spatially integrated and scaled to take into account that the X-ray reflections were temporally separated during recording and the fluctuations that may have occurred due to the change in the X-ray beam intensity.

The commonly used indicator of data quality is the linear merging R-value – equation 1.1, which quantifies the overall quality of collected data but can also be calculated to specific resolution intervals.

$$R_{merge} = \frac{\sum_h \sum_{i=1}^{N} \left| I_{(h)i} - \overline{I_{(h)}} \right|}{\sum_h \sum_{i=1}^{N} I_{(h)i}}$$

Equation 1.1 *Linear merging R-value*. The summation takes into account all N redundant observations for a given reflection, h, and $\overline{I_{(h)}}$ is the averaged intensity of each reflection.

However, the R_{merge} does not take into account the data redundancy. The value that it asymptotically approaches is the redundancy-independent merging R-value, R_{rim} or R_{meas}, given by equation 1.2.

$$R_{rim} = \frac{\sum_h \left(\frac{N}{N-1}\right)^2 \sum_{i=1}^{N} \left| I_{(ih)} - \overline{I_{(h)}} \right|}{\sum_h \sum_{i=1}^{N} I_{(ih)}} = R_{meas}$$

Equation 1.2 *Redundancy-independent merging R-value.* N is the total redundant observations for a given reflection, h. $\overline{I_{(h)}}$ is the averaged intensity of each reflection.

Being an asymptote of Rmerge value, it is always higher than the R_{merge} value. The more data we merge, the more precise the intensities become; to account for this, the term $1(N\text{-}1)^{1/2}$, must be introduced in the linear merging R-value.

$$R_{pim} = \frac{\sum_h \left(\frac{1}{N-1}\right)^{1/2} \sum_{i=1}^{N} \left| I_{(ih)} - \overline{I_{(h)}} \right|}{\sum_h \sum_{i=1}^{N} I_{(ih)}}$$

Equation 1.3 *Precision-indicating merging R-value, R_{pim}.* For a given reflection, h, N redundant observations are considered. The average intensity of each reflection is given by $\overline{I_{(h)}}$.

The R_{merge} and R_{pim} are very important quality evaluation parameters and have to be taken into account during data collection and onwards.

1.4.3. THE 'PHASE PROBLEM'

Once a protein crystal is obtained, an X-ray diffraction experiment is performed. Once the X-rays hit the crystal they will be scattered in all possible directions by the atoms present in the protein crystal. According to Bragg's law (equation 1.4 and figure 1.10), the diffracted beam will consist only of the waves that result from a constructive interference with these atoms. The diffracted beam is recorded as diffraction spots, designated reflections, with each reflection having the contribution from all the atoms in the crystal at the specific diffraction angle.

$$n\lambda = 2\, d_{hkl} \sin\theta$$

Equation 1.4 *Bragg's Law.* In Bragg´s law, n is an integer, λ is the wavelength of the incident wave, d is the spacing between the planes in the atomic lattice, and θ the angle between the incident ray and the scattering planes.

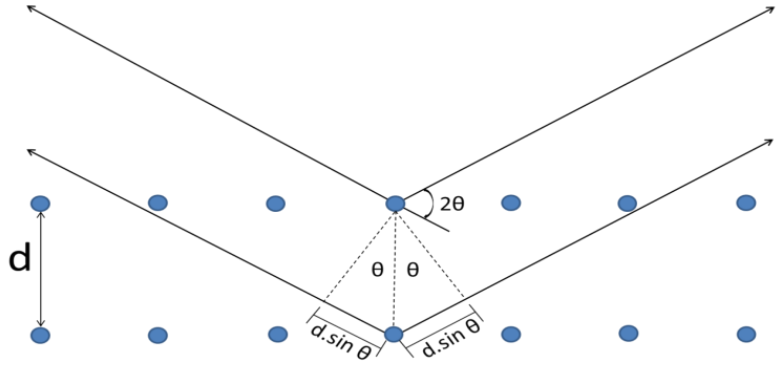


Figure 1.10 *Bragg's law graphical interpretation.* The Bragg's law can be graphically interpreted allowing the understanding of an X-ray experiment as the reflection on a set of imaginary planes in the crystal.

The total contribution from the total scattering of a crystal in a given direction of the reciprocal lattice, is proportional to the sum of all scattering elements in the unit cell. Thus, each and every atom *j* in the unit cell contributes a partial wave to every reflection, **h**. The complex structure factor, $\mathbf{F_h}$ can then be written as the summation of the contribution of partial waves of *j* atoms of scattering factor f_j at position $\mathbf{x}_j$.

$$\boldsymbol{F}_h = \sum_{j=1}^{atoms} f_{S,j}^{0} \,.\, exp\left(2\pi i \boldsymbol{h}\boldsymbol{x}_j\right)$$

Equation 1.5 *Complex structure factor.*

A structure factor can be represented as complex vectors in a complex plane – figure 1.11.

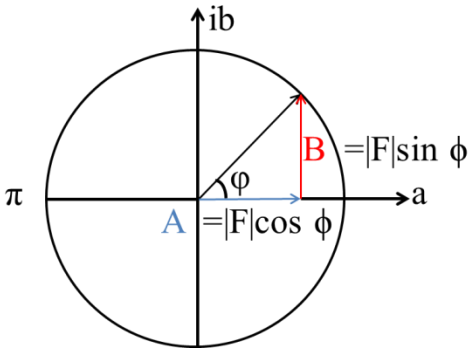


Figure 1.11 *Two-dimensional representation of a structure factor.* The vector length is equal to the amplitude of the structure factor and φ is the phase angle of the structure factor.

The term in the exponent contains information regarding the phase angle of partial waves from each atom, which depends on the direction of the scattering, **h**, and the positions of the atoms *j* in relation to the origin, given the fractional coordinate vector $\mathbf{x}_j$. The information regarding the phase angle is physically lost during data acquisition and this is the so-called 'phase problem'. The process of diffraction transforms information of the real space domain R into the reciprocal

space domain R*. Since this process results from scattering of X-ray photons by the electrons surrounding the atoms, mathematically reversing the diffraction process results in obtaining an electron density distribution, ρ(**r**), of the diffracting molecule through the mathematical formalism of Fourier transformation.

The electron density distribution can be defined according to equation 1.6.

$$\rho(x,y,z) = \frac{1}{V}\sum_{h=-\infty}^{+\infty}\sum_{k=-\infty}^{+\infty}\sum_{l=-\infty}^{+\infty} \boldsymbol{F}(hkl).exp[-2\pi i(hx+ky+lz)]$$

Equation 1.6 *Electron density distribution*. x, y, z are the fractional grid positions coordinates. 1/V is the normalized factor, with the unit cell dimensions defined in Å.

From experimental data, and after the Fourier transformation calculation, one knows that the scalar structure factor amplitude is proportional to the square root of its intensity, *I*(h). The measured intensities are affected by several factors. The geometry of the diffraction experiment equipment is one of these factors and is described by the Lorentz factor (L). A second factor interfering with the measured intensities is the polarization of the reflected light, which is taken into account by adding a correction factor, P. During data collection, radiation absorption, A, by the crystal is observed.

$$I_h = |F_h|^2\ LPA$$

Equation 1.7 *Measured intensity of a generic reflection, h.*

However, the information about the phase angle, φ(**h**) is still missing.

$$\sum_{h=-\infty}^{+\infty} F(\boldsymbol{h}).exp[-2\pi i(h.r) + i\varphi(\boldsymbol{h})] = \rho(\boldsymbol{r})$$

Equation 1.8 *Real space electron density.* Electron density definition for a general position vector, **r**.

In this thesis, two techniques were used to overcome the 'phase problem' and determine the hSOUL protein structure. Initially, independent phases were obtained by Single-wavelength Anomalous Dispersion, SAD, using a selenomethionine derivative of the protein. Due to the low resolution of the data, the protein structure determined was used as a model for molecular replacement, MR, against X-ray diffraction data previously acquired to confirm structure solution correctess.

1.4.3.1. SINGLE-WAVELENGTH ANOMALOUS DISPERSION, SAD

The most applicable method to overcome the phase problem with *de novo* phasing methods is to determine a marker atom substructure. The basic idea is to use a marker with an electron difference relative to an isomorphous reference structure. This difference leads to the existence of different atomic scattering factors, resulting in different structure factor amplitudes and

intensities. This difference is then used to calculate a smaller structure, substructure. This can be done, for instance, for a heavy atom soaked into the crystal or addition of an extra marker atom such as Se in selenomethionine derivative crystals. As a consequence, each complex structure factor will be the sum of the partial structure factor contributions of the marker atom, $\mathbf{F}_A$, and from the protein, $\mathbf{F}_{PA} = \mathbf{F}_P + \mathbf{F}_A$, being the basis for isomorphous replacement methods.

Any heavy atom that can be soaked and incorporated into the protein or even natively present sulfur atoms can be the source of anomalous signal. The anomalous scattering contribution from the heavy atom structure will cause the existence of anomalous Bijvoet differences between centrosymmetric related wedges. These methods are called anomalous diffraction and can be divided in to Single- or Multiple-wavelength Anomalous Dispersion, SAD or MAD, respectively.

The phasing equations can be graphically solved in the complex plane (Harker diagram), if one considers a structure factor **F**, with known magnitude, *F*, but unknown phase φ. This structure factor can lie on a circle of radius *F* in the complex plane. Figure 1.12 shows how the complex structure factors of the protein, derivative and heavy atom are related, in an acentric reflection, *hkl*. So one can have a circle with radius F_P and can then draw a circle with radius F_{PA} for each derivative structure factor amplitude. By the substructure solution we know the phase and magnitude of the heavy atom structure factor, $\mathbf{F}_A$ – figure 1.12.

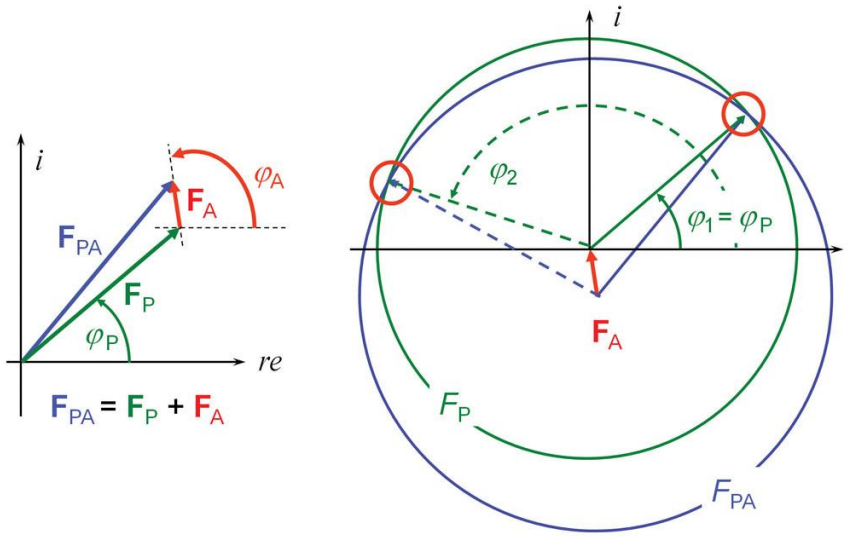


Figure 1.12 *Graphical solution of the phasing equations (from [61]).* The left panel shows the complex structure factors for the protein, derivative and heavy atom, given a generic reflection *hkl*. It is possible to determine the protein's phase angle by drawing a circle with radius F_{PA} and center with an offset of F_A from the origin and a circle with radius F_P. The interception of these two circles will give the two possible phase angles, φ_1 or φ_2. At this stage it is not possible to determine which of the two, φ_1 or φ_2, is the correct angle.

The ambiguity in the phase angle in Single Isomorphous Replacement (SIR) – figure 1.12, can be overcome by the addition of a second derivative (MIR – Multiple Isomorphous Replacement) at a different position from the first derivative. This will allow the drawing of a third circle with

different center and known radius F_{PA2} that will intercept the two previously drawn circles in only one of the two intercepting points. This unique angle is the initially unknown phase angle and phase problem is solved – figure 1.13.

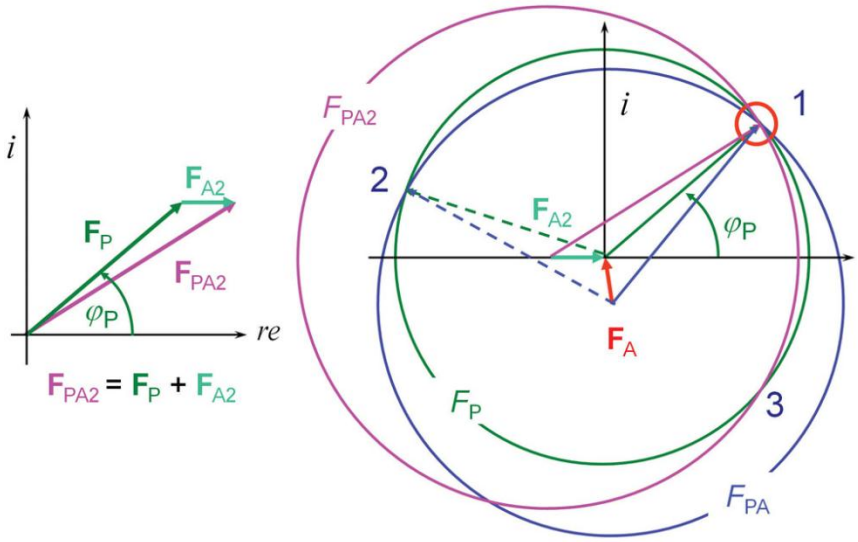


Figure 1.13 *The classical MIR case of breaking phase ambiguity*. Drawing the third circle (F_{PA2}) solves the phase ambiguity as the interception of the three circles (1) determines the previously unknow phase angle, φ_P.

SAD phasing can also be graphically solved – figure 1.14. From experimental data, the structure factor intensities of the Bijvoet pairs, F_{A+} and F_{A-} are known.

In order to comprehend SAD method, it is important to understand the phenomenon of X-ray absorption and associated anomalous dispersion which is becoming more and more important in protein crystallography. The phenomenon arises when X-rays are tuned to be near the resonance of an element. This leads to anomalous scattering and therefore anomalous diffraction near an X-ray absorption edge (for instance, regarding Se, this element presents an absorption edge at around 12 keV that can be used in SAD/MAD experiments with selenomethionine derivative proteins). On the basis of the atomic scattering function definition, one will have by the formal expansion of the atomic scattering function equation, two features important for X-ray protein crystallography: the two wavelength-dependent terms in the atomic scattering factor definition, $f'_{(\pi)}$ and $f''_{(\pi)}$, in addition to f_S^0.

Thus, to choose the optimum wavelength before collecting data for anomalous phasing, an X-ray absorption edge scan, generally a fluorescence scan in the form of an excitation spectrum, should be performed in order to get the highest anomalous dispersive signal. The imaginary anomalous scattering factor, f'' is proportional to the atomic absorption coefficient, μ, at that wavelength and the real part f', is indirectly obtained. For a SAD experiment, the peak of imaginary f'' should be the wavelength initially chosen to collect data since this is where the biggest anomalous difference between Bijvoet pairs is.

In the absence of anomalous differences, the centrosymmetry of the reciprocal lattice results in the conjugate pairs of structure factors, $\mathbf{F_h}$ and $\mathbf{F_{-h}}$ which have the same magnitude but opposite phase angle. For this reason the reflection intensities $I_{\mathbf{h}}$ and $I_{\mathbf{-h}}$ are equal – Friedel's law; the corresponding reflection pairs are called Friedel pairs. In the presence of an anomalous scattering atom, the Friedel's law is broken as f'' lags the phase by 90º. For this reason the centrosymmetry is broken, changing both magnitude and phase of the reflections of the Friedel or Bijvoet pair.

The vector diagram for a Bijvoet pair (reflection pair $\mathbf{F}_+$ and $\mathbf{F}_-$) is represented in figure 1.14 including the anomalous components $\mathbf{F''}_{A+}$ and $\mathbf{F''}_{A-}$. Instead of the average F_{PA} intensity observed in the case of SIR, graphically represented in figured 1.12, in the specific case of SAD for each reflection a pair of amplitudes F_{PA+} and F_{PA-} is available. In addition, we know the vector $\mathbf{F}_{A+}$ [61, 69].

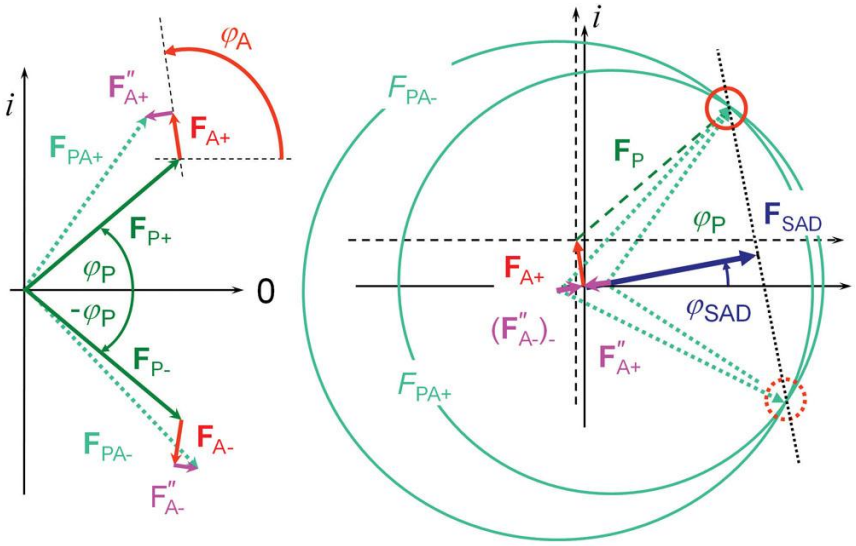


Figure 1.14 *Graphical SAD phasing equations solution(from [61]).* The magnitudes of the structure factors, F_{PA+} and F_{PA-} are known as well as the position of the anomalous scatterer and as a consequence $\mathbf{F}_{A+}$ (or $\mathbf{F}_{A-}$). Like in the SIR case, only one of the phase angles (φ_P) is correct.

Phase ambiguity is broken by collecting centrosymmetrically related reflections due to the presence of anomalous signal. From anomalous data one obtains anomalous differences for the derivative. Graphically four circles can be drawn with radius F_P, F_{PA}, F_{PA+} and F_{PA-}. The interception of the four gives the right phase angle solution. Similar procedure, i.e. the addition of anomalous signal can be performed with SIR and MIR methods: SIRAS and MIRAS.

A SAD structure can also be solved by using dispersive and anomalous signal together (MAD) with the help of density modification procedures or by direct methods.

1.4.3.2. MOLECULAR REPLACEMENT

Molecular replacement is another method to determine the initial phases. Although the use of the term is 'replacement', what is actually carried out in this method is to use a known structure model (usually with ~30 % or higher sequence identity to the protein of interest) to relocate the known structure on the unit cell or asymmetric unit until the solution with higher score is obtained. Generally speaking the idea is to place and score a real space probe given a reciprocal space data. For that, a 6-dimensional search is performed usually divided in a 3-dimensional rotational search followed by a 3-dimensional translation search (one of the reasons to do so is to optimize computational time).

The molecular replacement method can be performed using different protocols. Multi-dimensional search is a brute force method where the molecular search probe is put in every grid point of the asymmetric unit of the unknown cell, and a correlation score is attributed after varying the orientation of the probe. This process is computationally very slow. A second methodology is based on rotation-translation methods in which the search is divided into 2 stages. In the first one, the probe is properly oriented in the unit cell (rotation); once correctly oriented the probe is put in the right location (translation).

The rotation methods are based on the calculation and superposition of Patterson maps, and the orientation of a molecule is based on the match of intramolecular Patterson vectors, which can be calculated by different functions, for instance, real space rotation function, fast rotation function or direct rotation function. The same approach is used for the translation search and therefore location of our search probe in the unit cell. This is achieved by applying translational Patterson searches. In this case one has to determine the match of the intermolecular Patterson vectors, calculated by different fast translational functions.

The search probe is usually chosen according to its sequence identity to the protein of interest - usually it has to have ~30% or higher sequence identity. Even if this requisite is fulfilled, it is often difficult to use NMR models as search probes due to the variance in the coordinates of NMR ensembles. Different strategies can be taken into account to use the NMR ensemble: use only one out the twenty conformations determined or calculate an average structure. Using only one model has the disadvantage of not having the right weights to the atomic contributions of the scattering factors which reflect the precision of atomic positions (equivalent information is only embedded if one uses the NMR ensemble). If these approaches fail, one can also prepare a model by removing (or down-weighting) regions where large local structural variations are observed [70].

Another approach on molecular replacement methods is the use of Maximum likelihood functions for experimental phasing; these functions have been proven to be more realistic as they take into account the errors and incompleteness of the obtained models. For these reasons,

these functions are implemented in the program Phaser [71]. In a very basic description, these functions describe the probability of observing an experimental value given the model obtained.

1.4.4. MODEL BUILDING, REFINEMENT AND STRUCTURE VALIDATION

Once the phase problem is solved and experimental phases are obtained, model building and refinement are the next steps in protein structure determination. With accurate phases and high resolution data, amino acid side chains can be readily identified and the sequence fitted into the electron density maps. Low resolution data often requires the insertion of Cα atoms into the branching points of the polypeptide chain, allowing the construction of a polyalanine backbone model. Therefore, once a contiguous backbone fragment is achieved it is possible to fit the individual residues by a combination of real space refinement and real space geometry regularization tools.

The phase quality in experimental electron density maps is expressed as a figure of merit (f.o.m.), m, $0 \leq m \leq 1$, of the probability-averaged best (centroid) structure factors F_{BEST}.

Regarding molecular replacement phases, they are derived from a given model being often incomplete and incorrect in many regions as the electron density maps calculated will be biased due to the tendency to reconstruct the model density – model bias.

The model atoms parameters (coordinates x, y, z and individual B-factors, except in low resolution data) together with overall parameters such as scale factors and overall B-factors, bulk solvent corrections and anisotropy corrections are refined against the experimental data. The final goal is to minimize the defined target-function (residual between the observed experimental structure factors and the model structure factors amplitudes).

The overall parameter that quantifies the fit between the diffraction data and the resulting model is the R-value, which correlates the scaled structure factor amplitudes, F_{obs} and F_{calc}, as described in equation 1.9.

$$R = \frac{\sum_h |F_{obs} - F_{calc}|}{\sum_h F_{obs}}$$

Equation 1.9 *R-factor equation.* R factor for a given reflection, h. F_{obs} and F_{calc} are the observed and calculated structure factors, respectively.

This parameter is based on the restrained refinement evaluation and results from the fact that the structure factor of each reflection, h, is a nonlinear function of each and every atom in the unit cell – equation 1.10.

$$F_h = \sum_{j}^{atoms} f_j^B e^{(2\pi i h x_j)}$$

Equation 1.10 *Structure factor definition.* The structure factor, F_h, is therefore a summation of partial waves of *j* atoms with scattering factor f_j at position x_j.

Due to the parameterization used in any refinement protocol, it is important to perform a cross-validation to our model to make sure it makes sense. For that purpose, the cross validation method used consists of excluding a set of the experimental data (usually 5 – 10 % of the reflections) that will not be taken into account in any refinement cycles (if reflections are not independent, as it is the case of refinement against anomalous data, the Friedel mate, -h-k-l, of an excluded reflection, hkl, must also be excluded in the R_{work} calculation). Therefore, the agreement between fitted data and the model is computed separately for the "working data" - R_{work} (data used in the refinement procedure) and the "free data" - R_{free} (data that is initially excluded and not refined). The cross validation R-values are then the R_{work}, for the "working data, and R_{free} for the excluded reflections.

$$R_{free} = \frac{\sum_{h \in free} |F_{obs} - kF_{calc}|}{\sum_{h \in free} F_{obs}}$$

Equations 1.11 R_{free} *equation.* Before the first cycles of refinement a percentage of experimental data is excluded and used to calculate the R_{free}.

$$R_{work} = \frac{\sum_{h \notin free} |F_{obs} - kF_{calc}|}{\sum_{h \notin free} F_{obs}}$$

Equation 1.12 R_{work} *equation.* After every round of model building, completion and addition of parameters will make both R_{free} and R_{work} to convergence as the model becomes more complete and accurate.

The R_{free} value is, as a consequence, a measure for phase accuracy and therefore of the model quality in contrast with the R_{work} that can be reduced without a corresponding model improvement. For this reason, during model phase refinement, these two R-values must be analyzed together in order to correctly evaluate the refinement.

A very important issue regarding the refinement of low resolution data (which is the case of the results presented in this thesis), is that the independent initial phases should be used in the refinement procedure as these phases are independent and unbiased. In practical terms, this information can be imported as Hendrickson-Lattman coefficients or in the form of calculated phases and corresponding figure of merit.

More recently TLS (Translation/Libration/Screw) refinement has been introduced in macromolecular crystallography, especially after its implementation in software program REFMAC [72]. TLS describes more complex, anisotropic motions. In this refinement, the first step is to define TLS groups which can be different domains of a multi domain protein or

grouped according to identical B-factor values. For this purpose, TLSMD is a very powerful online tool to define the TLS groups to be used in refinement [73]. This TLS parameterization describes the displacement of an atom from one place to the other, which can be described as a combination of a rotational component and a translational component with the advantage that TLS parameterization contributes only with 20 parameters *per* group. At the final stage of refinement, the TLS tensors must be converted to the traditional anisotropic displacement parameter (ADP).

The evaluation of the refinement is mainly based on the residual R (R_{work}) and R_{free} values. If the two improve together, it may imply that there is room for some more refinement; if the two values tend to diverge it means something is going wrong with the refinement procedure. Figure 1.15 shows how the difference R_{free}-R varies with the resolution, according to the deposited PDB structures. Moreover, the values expected for these two R values depend on a wide range of parameters such as data quality, data resolution and the correctness of model building and refinement.

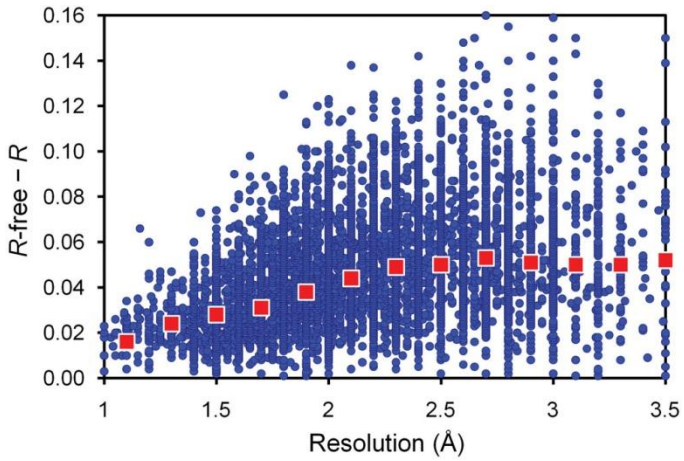


Figure 1.15 *Variation of the difference R_{free}-R (from [61]).* The mean value difference between R_{free} *and R is plotted in red full squares as a function of structure resolution (data extracted from the Protein Data Bank, PDB, http://www.pdb.org/pdb/home/home.do).*

In the process of validating a structure, one has to keep in mind that global refinement parameters like R and R_{free} values or coordinate root mean square (r.m.s.) do not take into account local errors that can exist in the structure. The local quality of the model becomes even more relevant when studying, for instance, the active site of an enzyme, the effects of mutations or the ligand binding. Therefore, to have a complete and correct analysis of the structure, a combination between the real space electron density correlation and the location of geometric outliers is necessary. The backbone conformation of the polypeptide chain is defined by its torsion angles; this can be evaluated by analysis of the Ramachandran plot, a 2D scatter plot of the φ- ψ backbone torsion angle pairs of each residue of the polypeptide chain.

The electron density map is the most important evidence given by crystallography, and the ultimate validation procedure is to analyze the fit of our model to the electron density map. The fastest way is to calculate the real space correlation coefficient (RSCC) or real space R-value (RSR) which correlates the model with the corresponding electron density map.

The RSCC is defined as a linear correlation coefficient between observed and calculated electron density as shown in equation 1.13.

$$RSCC = \frac{\sum_r \left(\rho(r)_{obs} - \overline{\rho(r)_{obs}} \cdot \left(\rho(r)_{calc} - \overline{\rho(r)_{calc}}\right)\right)}{\left(\sum_r \left(\rho(r)_{obs} - \overline{\rho(r)_{obs}}\right)^2 \cdot \sum_r \left(\rho(r)_{calc} - \overline{\rho(r)_{calc}}\right)^2\right)^{1/2}}$$

Equation 1.13 *Real space correlation coefficient, RSCC*. Correlation between the observed electron density map, $\rho(r)_{obs}$, and the calculated electron density map, $\rho(r)_{calc}$.

For low resolution structures, where side chains are mostly not visible it is often difficult to clearly assign the sequence in the beginning of the process. The presence of heavy atoms and dispersive atoms positions are often useful to lock the sequence properly on the electron density map.

1.5. PROTEIN NUCLEAR MAGNETIC RESONANCE

Nuclear Magnetic Resonance (NMR) is a very powerful tool in the field of structural biology and has been used in protein structure determination and in protein dynamics and protein-ligand/protein-protein interactions studies.

Historically, the first published NMR experiment on a biological molecule (studies of deoxyribonucleic acid solutions and gels in order to investigate their high viscosity, osmotic pressure and dielectric constant due to the possible presence of large hydration shells) was in 1954 [74], the first intact protein (ribonuclease) was studied in 1957 [75], and the the first protein structure determined in 1985 [76]. Nowadays, NMR spectroscopy of biological molecules has a wide range of applications in structural biology, protein function studies and drug design experiments.

1.5.1. BASIC PRINCIPLES OF NMR

This technique is based on the nuclear magnetic dipole moment that is a consequence of the spin angular momentum of the nucleus. Protons (and neutrons) have a spin angular moment that, when in a nucleus, pair between each other in an antiparallel fashion, with net spin of zero. However, all nuclei with an odd mass number (for example, ^{1}H, ^{13}C, ^{15}N) have spin angular momentum as they have an unpaired proton. Nuclei with an even mass number and an odd charge also have spin angular momentum, giving a nuclear spin quantum number, I. As an example one can think of ^{2}H nucleus, with I = 1. Therefore, the nucleus will have an associated magnetic moment (μ) that is dependent on the value of the spin quantum number, I. The spin angular moment of a nucleus can go from + I to – I, in integral steps. This value is known as the magnetic quantum number, m, with a total of (2I + 1) angular moment states. The magnetic moment of a nuclear spin is related to its spin angular moment and to I with proportionality constant, γ, the gyromagnetic ratio. This constant is characteristic for each nucleus and indicates the frequency with which a nucleus will precess in a fixed external magnetic field – equation 1.14 [77, 78].

$$\mu = \gamma I$$

Equation 1.14 *Nuclear spin magnetic moment.* This equation defines the magnetic moment of a nuclear spin, μ, which is related to the nuclear spin quantum number, I, and with a proportionality constant, γ, the gyromagnetic constant.

In table 1.1, some important active nuclei in NMR and some of their important characteristics are listed.

Table 1.1 *Properties of some NMR active nuclei (from [77]).* Some important NMR active nuclei in the study of biomolecules and polymers with the corresponding gyromagnetic constant, γ, nuclear spin quantum number, I, and natural abundance.

Nuclei	γ (10^6 rad × sec^{-1} × T^{-1})	I	Natural abundance (%)
^{1}H	267.513	½	99.980
^{2}H	41.065	1	0.016
^{13}C	67.262	½	1.108
^{15}N	-27.116	½	0.370
^{19}F	251.815	½	100.000
^{31}P	108.291	½	100.000

For high resolution protein NMR studies, the low natural abundance of the isotopes ^{13}C and ^{15}N is a problem. The general solution to this problem is to prepare biomolecules that are enriched with stable isotopes. Thus, in the case of nitrogen and carbon, samples are prepared by recombinant expression of the protein in *E. coli* cultured in M9 minimal medium containing usualy $^{15}NH_4Cl$ and ^{13}C uniformly labeled D-glucose as the sole nitrogen and carbon sources, respectively.

NMR Spectroscopy, like all spectroscopic techniques, is based on energy states and population distributions. The energy difference between energy states gives rise to the frequency to promote energy state transitions, whereas intensities of the spectral peaks are proportional to the population difference of the states. For a proton, the population ratio in the states is quantitatively described by the Boltzmann equation:

$$\frac{N_\beta}{N_\alpha} = e^{-\Delta E/kT} = e^{-(h/2\pi)\gamma B/kT} = \frac{1}{e^{-(h/2\pi)\gamma B/kT}}$$

Equation 1.15 *Boltzmann equation.* Equation describing the population distribution of two energy states, where N_α and N_β are the populations of the α and β states, respectively, T is the absolute temperature, *k* is the Boltzmann constant, h the Planck constant and B the magnetic field.

The energies of the states α and β arise from the interaction of a nuclear magnetic dipole moment with an intense external magnetic field.

At room temperature, the population of β state is slightly lower than that of α state. For example, the population ratio for protons at 800 MHz field strength is 0.99987. The consequence of this is that only a small fraction of the spins will contribute to the signal intensity due to the low energy difference and hence NMR spectroscopy is intrinsically a very insensitive technique. As seen by equations 1.15 and 1.17 stronger magnetic fields are necessary to obtain better sensitivity; higher magnetic fields will increase the population ratio between the ground state and the excited state and, consequently, the sensitivity. Recently, the use of cryoprobes has significantly increased NMR sensivity. The principle behind this technology is

the fact that the radio frequency electronics will generate a higher signal to noise ratio at lower temperatures. Therefore, by reducing the temperature of the NMR coil and preamplifier signal to noise ratio, signal enhancement can be achieved. Besides cryoprobes, the external magnetic field as described before and sample concentration influence NMR spectra sensivity. Higher magnetic fields and sample concentration will increase spectra sensitivity.

NMR takes advantage of the fact that when a nucleus with nonzero spin is placed under the influence of an external magnetic field, B_0, its angular moment oriention coincides with the field direction and precesses around it with a frequency, Larmor frequency or angular frequency – equation 1.17, which is dependent on the nucleus and the strength of the applied magnetic field. The actual magnetic field, B, felt in the nucleus is attenuated, or shielded, by the presence of electrons that surround the nucleus – equation 1.16.

$$B = (1 - \sigma)B_0$$

Equation 1.16 *Magnetic field at a given nucleus.* σ represents the degree of shielding and B_0 the strength of the applied magnetic field.

The Larmor equation is given by:

$$\omega_s = \gamma B$$

Equation 1.17 *Larmor equation.* ω_s is the resonance frequency of the shielded nucleus and is equal to γ, the gyromagnetic constant, multiplied by the strength of the magnetic field *at the nucleus*.

The irradiation of a sample with radiofrequency (RF) waves of the appropriate frequency (equation 1.17) will excite transitions from the ground to the excited state as a result of the interaction of the magnetic dipole with the oscillating magnetic field component of the electromagnetic radiation. This excitation field, B_1, must be orthogonal to the direction of the magnetic dipoles. The B_1 field can be applied to the sample either by scanning through multiple wavelengths (continuous wave NMR) or as short burst of high power RF that excites a broad range of transitions (pulsed NMR). The simplest pulsed NMR experiment consists of a short RF-pulse followed by detection of the signal. This experiment can be divided in three discrete time intervals: a) preparation, b) excitation and c) detection. The spins are initially at thermal equilibrium and subject to only the static B_0 field – preparation; during the excitation pulse the spins are subject to the static B_0 field and the oscillatory excitation field, B_1. In this period, the bulk magnetization moment of the sample is tipped from the z-axis to the x-y plane. After the pulse, the magnetization precesses about B_0 at a frequency ω_s inducing a current in the receiver coil. The excited spins precess under the static B_0 field, generating the free induction decay (FID). The NMR spectrum is obtained by Fourier transformation of the FID.

Another very important phenomena in NMR is the relaxation process, that, not only is essential because without it no NMR signal would be observed but also because it is this process that

imposes the molecular size limitations on NMR technique. In NMR, two relaxation parameters are defined:

- T_1, the spin-lattice, or longitudinal, relaxation time (or R_1 for spin-lattice relaxation rate, $R_1 \sim T_1^{-1}$);
- T_2, the spin-spin, or transverse, relaxation time (or R_2 for the spin-spin relaxation rate, $R_2 \sim T_2^{-1}$).

The parameter T_1 measures the efficiency with which the excited nuclear spins return to their ground state by exchanging energy with their surroundings. T_2 is a measurement of the efficiency with which spins exchange energy with each other. The more efficient this exchange, the shorter the relaxation time. In a single NMR experiment, one of two relaxation times can be studied to understand protein's dynamics, transverse (T_2) or longitudinal (T_1), in the timescale of milliseconds to seconds, respectively. In solution, the resonance line widths are inversely proportional to the T_2 relaxation time, which decreases with the increase in molecular size and tumbling time [79]. Therefore, these two relaxation times probe protein dynamics in the millisecond to second time scale.

Dipole-dipole interaction is probably the most important mechanism of relaxation pathway for protons in molecules containing contiguous protons and for carbons with directly attached protons. This is also the source of the Nuclear Overhauser Effect (NOE). Dipolar coupling occurs when the magnetic field of one nuclear dipole affects the magnetic field at another nucleus and depends on the distance between nuclei (further discussed in *1.5.2*) and takes place through-space.

Dipole-dipole relaxation is also dependent on the correlation time, τ_c. Small molecules tumble very fast and have short τ_c, usually in the order of picoseconds. Large molecules, such as proteins, usually move slowly and have long τ_c, usually in the order of the nanoseconds.

In summary, several dynamic processes can be studied by NMR; the different time scales of NMR observable phenomena are graphically represented in figure 1.16.

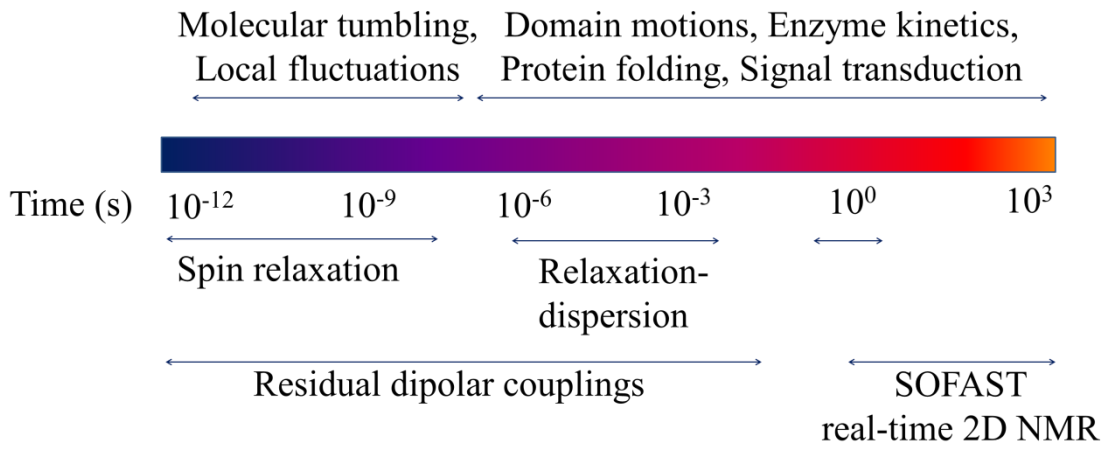


Figure 1.16 *Time scales of some important molecular dynamic processes and multidimensionl NMR methods available to study these processes (adapted from [80]).* Recent developments in NMR spectroscopy techniques made it a very important technique for the understanding of some of the most important dynamic processes in the cell such as, for instance, protein folding and enzyme kinetics.

1.5.2. PROTEIN NMR TECHNIQUES AND METHODOLOGIES

Due to the high natural abundance and large gyromagnetic ratio of the ^{1}H nucleus, protein NMR studies traditionally have utilized predominantly homonuclear ^{1}H spectroscopic techniques. Hundreds of 2D and 3D ^{1}H NMR experiments have been described in the literature; however many of these are not generally applicable. Multidimensional NMR experiments are useful to obtain scalar or dipolar correlations between different magnetic active nuclei. These experiments can be described in four periods: preparation, evolution, mixing and detection. The first period goes from the initial equilibrium state of the system until the first $\pi/2$ pulse is applied. Once this pulse is applied, the spins precess freely – evolution time. After that, all the magnetization transfer processes occur through dipolar or scalar coupling – mixing time. During this period magnetization is transferred out-and-back from the active nuclei, relaxation occurs and the NMR signal is measured – detection period. The multidimensional spectra are obtained by the accumulation of a series of one-dimensional experiments by incrementing the evolution time.

For small proteins (usually bellow 10 kDa) and peptides structure determination, two homonuclear bidimensional experiments, COSY (COrrelation SpectroscopY) and TOCSY (TOtal Correlation SpectroscopY), are used to give information between protons, due to the scalar coupling through covalent bonds. NOESY experiments (Nuclear Overhauser Effect SpectroscopY) are the most important multidimensional experiments in structure determination since they correlate protons through space, allowing the determination of the distances between close protons [81, 82]. These methodologies can be used to determine the structure of proteins up to 10 kDa. For larger molecules, heteronuclear experiments are necessary and require the isotopic labeling of the samples.

For larger than 10 - 12 kDa proteins, the previously described experiments are not suitable. The number of hydrogen atoms in proteins scales approximately linearly with molecular mass. In addition, the rotational correlation times of globular proteins, and therefore the linewidths of the NMR resonances, also increase linearly with molecular mass. The increased number and linewidth of the resonances in homonuclear ^{1}H NMR spectra result in extensive chemical shift overlap and degeneracy. For these reasons, conventional assignment procedures based on sequential NOE correlations become difficult or impossible. Heteronuclear NMR spectroscopy can overcome these problems for proteins of at least up to molecular masses of 25 - 30 kDa [83-85], provided that the proteins can be labeled with the NMR active isotopes ^{13}C and ^{15}N [86].

The magnetic fields fluctuate in time as the NMR spins are attached to molecules that tumble in solution. The magnitude of these fields scales with the gyromagnetic ratios of the nuclei that produce it. This is the basis for the use of deuterated proteins. Because the gyromagnetic ratio of deuteron is approximately 6.5 times less than a proton, the substitution of protons for deuterion atoms leads to a decrease in the relaxation rates of ^{13}C spins. For backbone assignment experiments and side chain triple resonance experiments, for instance, the slowed relaxation translates into increased sensitivity and resolution.

The high value of the proton gyromagnetic constant gives high sensitivity, but at the same time causes large dipole-dipole interactions that lead to rapid relaxation rates. The intrinsically small range of proton chemical shifts may cause severe resonance overlap, in particular in unfolded systems.

Due to the reasons mentioned above, labeling strategies have been recently developed to incorporate deuterium (^{2}H), in ^{15}N and ^{13}C-enriched proteins in order to eliminate the rapid decay of the NMR signal (spin relaxation) and thus increase the sensitivity and resolution of the spectra. The substitution of ^{1}H spin for ^{2}H spin reduces the rate of dipole-dipole relaxation of the observable proton; the dipolar coupling between a ^{1}H and a ^{2}H is much weaker than the dipolar coupling between two ^{1}H spins, because the gyromagnetic ratio of ^{2}H is 6.5 times smaller than that of ^{1}H giving NMR spectra with sharper and more intense lines [87, 88].

Two-dimensional and three-dimensional NMR experiments using double labeled samples take advantage of the large J-couplings between ^{15}N and ^{13}C nuclei and between these nuclei and their attached protons for efficient magnetization transfer. The scalar coupling J can be related with the dihedral angles, θ. Chemical shifts can be used to obtain restraints for the backbone dihedral angles ϕ and ψ.

In protein NMR, the HSQC (Heteronuclear Single Quantum Coherence) experiment is very important as it is a fast experiment to study the state of a protein in the sample. In a ^{15}N labeled protein only one N-H group is present for each amino acid residue of the protein sequence. Five exceptions to this are observed: asparagine, glutamine and histidine residues give two extra

signals due to the side chain amide groups and tryptophan will give one extra signal. Proline is the other exception as it does not contain an amide proton attached to nitrogen. For these reasons a ^{1}H, ^{15}N-HSQC spectrum is the 'fingerprint' of the protein.

The whole process of protein structure determination by NMR can be achieved in four major steps: data acquisition and processing, chemical shift assignment, restraints determination, and structure calculation. The first step after data acquisition is the Fourier transformation of the time-domain data into frequency-domain data and subsequent identification of NMR signals in this domain – chemical shift assignment. For structure determination, information must be acquired in the form of restraints on interatomic distances, torsion angles about chemical bonds and the relative directions of chemical bonds linking different atom pairs of the polypeptide chain [89].

A heteronuclear experiment is started by the generation of a transverse proton magnetization that is transferred to the heteronucleus and transferred back to the proton, which is then detected at the end of any experiment.

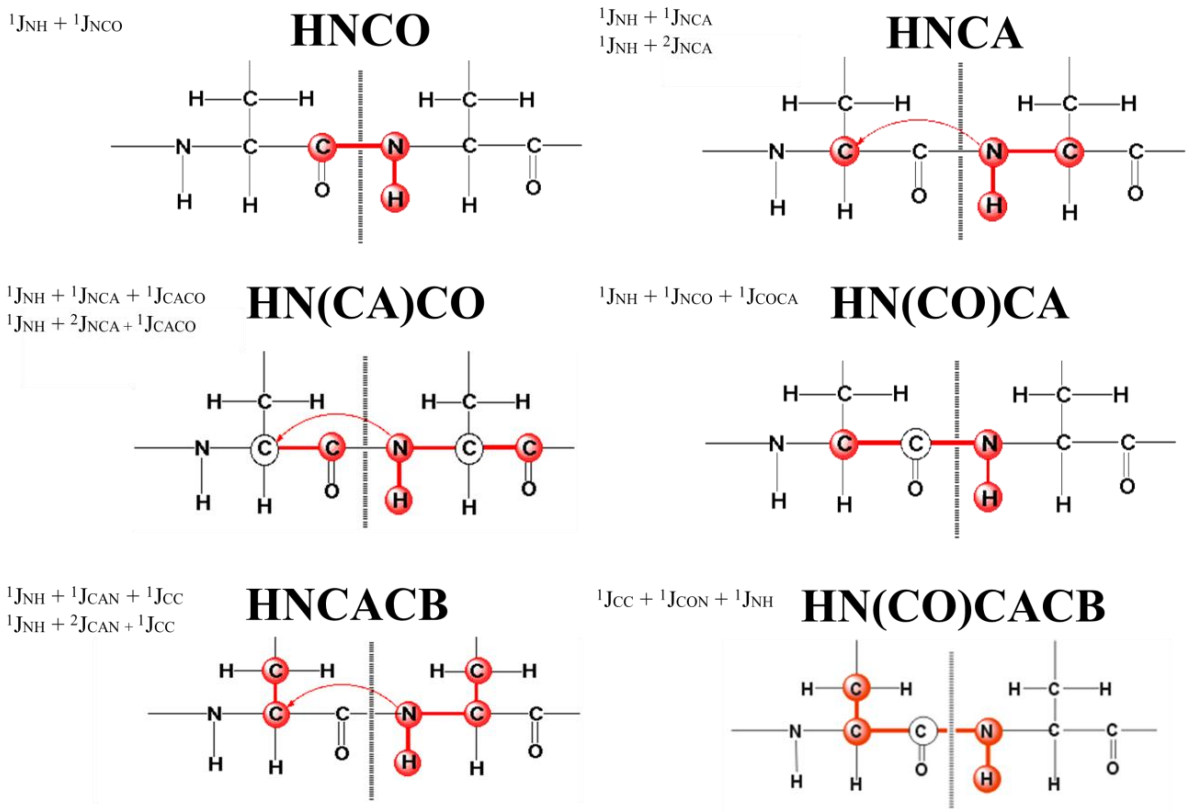


Figure 1.17 *Standard heteronuclear NMR experiments for protein backbone assignment (adapted from http://rmni.iqfr.csic.es/guide/eNMR/eNMR3Dprot/).* HNCO correlates ^{15}N-^{1}H pair of one residue with the carbonyl (^{13}CO) resonance of the preceding residue. The HNCA experiment correlates the ^{15}N and HN chemical shifts with the intra- and inter-residue ^{13}CA carbon shifts. The HN(CO)CA correlates the ^{15}N and HN chemical shifts with the inter-residue ^{13}CA carbon shifts. The HN(CA)CO correlates the inter- and intra-residue backbone connectivities between the amide ^{15}N-^{1}H pair and the carbonyl ^{13}CO resonance. The HNCACB spectrum correlates the ^{15}N-^{1}H pair with the intra- and inter-residue ^{13}CA and ^{13}CB carbon shifts. Finally, the HN(CO)CACB correlates the ^{15}N-^{1}H pair with the intra-residue ^{13}CA and ^{13}CB.

The experiments shown in figure 1.17 are the most commonly used in protein backbone assignment. Once all the NH groups are identified in the ^{1}H, ^{15}N- HSQC/TROSY-HSQC spectra, one has to correlate each NH group with the corresponding Cα and Cβ. The HNCACB correlates the NH group with the Cα and Cβ chemical shifts of its own residue (strongly) and of the preceding residue (weakly). For the identification of secondary structure elements knowing the amide and alpha proton chemical shifts is all the necessary information [90]

In 1997, Pervushin *et al* introduced the concept of Transverse Relaxation-Optimized SpectroscopY (TROSY), which generically improves the measurement of the dipolar couplings and the detection of scalar couplings across hydrogen bonds by reducing the T_2 relaxation time – figure 1.18. The NMR signal decays exponentially with a characteristic time constant – the transverse relaxation time, T_2. On the other hand, the line width of the NMR resonances is inversely proportional to T_2, which itself depends on size of the molecule under investigation –

the bigger the molecular weight of the molecule, the smaller the T_2 value which results in broader lines in the NMR spectra. This methodology can be demonstrated by using the example of the amide moiety in a polypeptide chain containing isotopic labeled nitrogen atoms (^{15}N); the ^{1}H nuclei couples (scalar coupling) to the ^{15}N nuclei, which translates in two lines in the NMR spectrum: protons attached to ^{15}N with spin up and protons attached to the ^{15}N nuclei with spin down relative to the external magnetic field. In the NMR spectra of a large protein, the two lines have different widths due to different interfering relaxation mechanism (dipole-dipole relaxation between the proton and nitrogen spins, and the chemical shift anisotropy (CSA) of the protons). Usually the two lines are joined together, which means they are decoupled. In the TROSY technique, the slower relaxation is chosen which leads to an improvement in the NMR spectrum [91, 92].

The optimal result can be achieved by the combination of TROSY techniques with deuterium labeling.

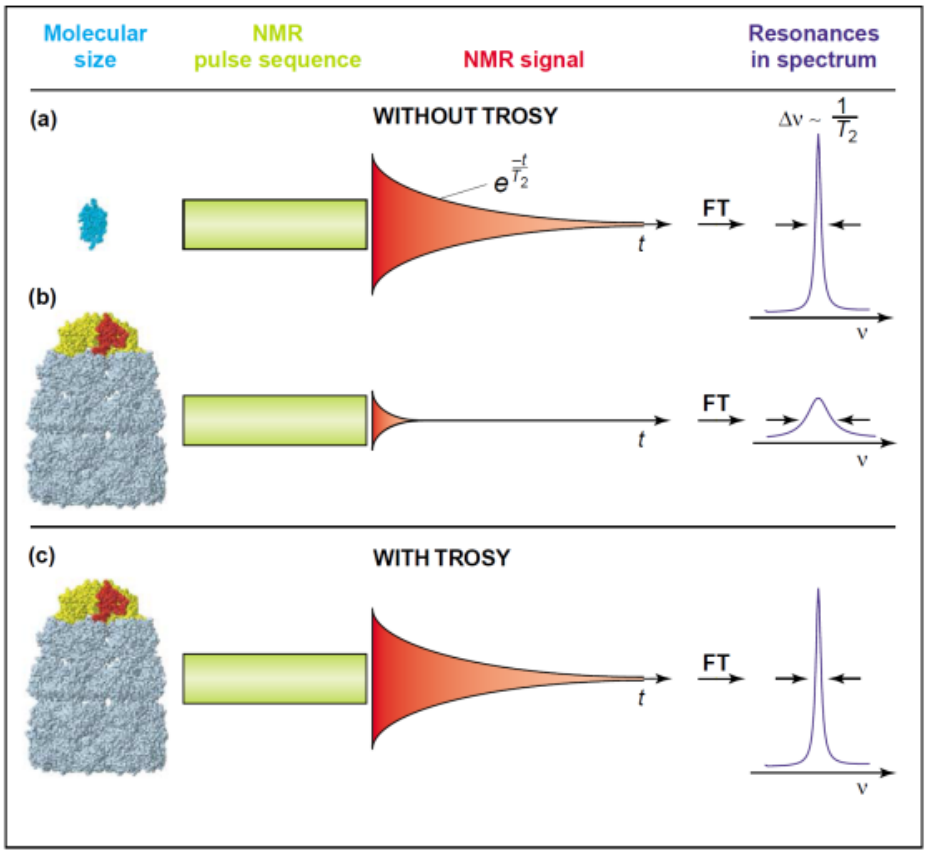


Figure 1.18 *TROSY effect on the transverse relaxation time,T_2, and line widths (adapted from [91]).* Schematic representation of the TROSY effect on the transverse relaxation time, T_2, and peak's line width. In a) the NMR signal from a small molecule relaxes slowly having a long transverse relaxation time (T_2) which gives raise to narrow line widths after Fourier transformation. In larger molecules (b), the T2 is smaller which results on weaker signals and broader lines. With the TROSY technique (c), an improvement in signals intensity and spectral sensitivity and resolution is observed.

The most important restraints necessary to determine a protein structure are obtained in result of the nuclear Overhauser effect (NOE) which defines the correlation between protons close in

space (within 5 - 6 Å distance) and is a consequence of dipole-dipole cross-relaxation in nuclear spin systems; intramolecular distance of protons that can be distant in the protein sequence but close in space can be estimated [79, 93, 94]. The cross-relaxation is a consequence from the fact that a given spin *I* relaxes spin *S*, and *vice-versa*. Considering a situation where *I* is initially in equilibrium but not *S*, the cross relaxation rate is given by,

$$\frac{dI}{dt} = -\sigma_{IS}(S - S^o)$$

where,

$$\sigma_{IS} = \frac{1}{10}K^2\tau_c\left[\frac{6}{1+(\omega_I+\omega_S)^2} - \frac{1}{1+(\omega_I-\omega_S)^2\,\tau_c^2}\right]$$

Equation 1.18 *Cross-relaxation rate.* $K = (\mu_0/4\pi)\gamma_I\gamma_S/r_{IS}^3$, γ_I and γ_S are the gyromagnetic ratios for nuclei *I* ans *S*, r_{IS} is the internuclear distance, ω_I and ω_S are the Larmor precession frequencies of nuclei *I* and *S*, and τ_c is the correlation time of the *IS* vector.

The relaxation rate between two spins, *I* and *S*, depends, as shown in equation 1.17, on the type of nuclei involved, on the distance between them and on the correlation time for the *IS* vector. Considering ^{1}H-^{1}H relaxation in a protein, under a 600 MHz magnetic field, ω_I and ω_S are close together making the first term in brackets (equation 1.17) negligible. This means that the magnetization vectors of *I* and *S* relax together in opposite directions. Therefore, cross-relaxation makes the magnetization spread around the molecule in a diffusive process – spin diffusion.

For all protons in a rigid protein, equation 1.17 can be simplified – equation 1.19.

$$\sigma_{IS} = cr_{IS}^{-6}$$

Equation 1.19 *Cross-relaxation rate between two nuclei, I and S.*

Thus, the cross-relaxation rate between two nuclei can be simply proportional to r^{-6}, and is measured as the intensity of NOESY cross peaks, allowing distances within the molecule to be measured. A reference calibration distance with a measured NOE intensity, A_{ref}, is needed – equation 1.20.

$$r_{IS} = r_{ref}\left[A_{IS}/A_{ref}\right]^{-1/6}$$

Equation 1.20 *Distance between two nuclei (I and S),* r_{IS}.

The correlation time of a molecule is related with its molecular weight. Small molecules have short correlation time, around 10^{-11} s or less, whereas proteins have correlation times of 5 ns or longer. Therefore, as the correlation time (molecular weight) increases, cross-relaxation gets faster and NOEs build up faster. As a consequence of equation 1.18, cross-relaxation rates for small molecules are positive and for proteins are negative. For medium size molecules, such as

small peptides, cross-relaxation rates can be positive or negative, which means very small (often not measurable) NOEs. This problem can be overcome by using ROE technique (rotating frame NOE), in which cross-relaxation occurs in the transverse plane instead of z (longitudinal plane).

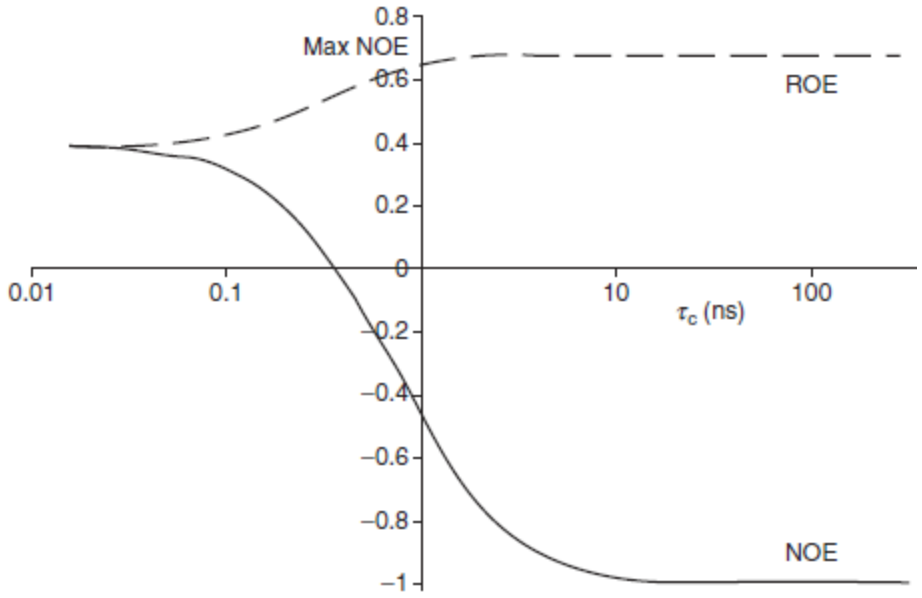


Figure 1.19 *Maximum NOE and ROE obtainable in NOESY (solid line) and ROESY experiment (dashed line).*

Macromolecular structure determination, since the introduction of triple resonance spectra using ^{2}H, ^{13}C and ^{15}N, has a standard procedure that can be divided in three stages: protein backbone assignment, side chain assignment, structure calculation (using distance restraint from NOE peaks).

To complement the information from NOE, residual dipolar couplings (RDC) can be determined, giving orientational restraints. The general form for the dipolar coupling Hamiltonian of two spins, I and S is described in equation 1.21.

$$H_D = \frac{(h/\pi)\gamma_I \gamma_S}{4\pi r_{IS}^3} \left[1 - 3\cos^3\theta\right] (3I_Z S_Z - I \times S)$$

Equation 1.21 *Dipolar coupling Hamiltonian*. Hamiltonian of two spins, I and S, dipolar coupling, where h is the Planck constant, γ is the gyromagnetic ratio, r is the inter-spin distance, θ is the angle between the inter-spin vector and the external magnetic field and I and S the spin operators.

The heteronuclear two-dimensional $\{^{1}H\}$-^{15}N nuclear Overhauser effect (hetNOE) is the most universally used NMR experiment to access protein dynamics on fast time scales (pico to nanosecond). With these experiments, flexible regions of the protein can be readily distinguished. Opposite sign NOE values can be used to identify unstructured parts of the protein. Values of HetNOE lower then 0.65, at 600 MHz, are indicative of a considerably flexibility on a picosecond timescale [95].

One important parameter in the hetNOE experiment is the relaxation delay. This cannot be too short as this will result in systematic errors in the hetNOE values; chemical exchange of amide protons with saturated water protons was shown to artificially increase hetNOE ratios [96, 97]. The HetNOE values are defined as the ratio of peak intensities with and without proton saturation and so, to measure the NOE, a pair of experiments is acquired.

$$NOE = \frac{I_{sat}}{I_{unsat}}$$

Equation 1.22 *{^{1}H}-^{15}N-NOE determination.* I_{sat} and I_{unsat} are the peak intensities with and without proton saturation, respectively.

The uncertainties of hetNOE values, *ΔNOE*, can be calculated using the well-established method [98].

$$\frac{\Delta NOE}{\mathrm{NOE}} = \sqrt{\left(\frac{\Delta \mathrm{I_{sat}}}{\mathrm{I_{sat}}}\right)^2 + \left(\frac{\Delta \mathrm{I_{unsat}}}{\mathrm{I_{unsat}}}\right)^2}$$

Equation 1.23 *{^{1}H}-^{15}N-NOE uncertainties determination.* I_{sat} and I_{unsat} are the peak intensities with and without proton saturation, respectively, and ΔI_{sat} and ΔI_{unsat} the corresponding uncertainties.

A flow chart summarizing same of the more important protein NMR experiments and their goal is shown in figure 1.20.

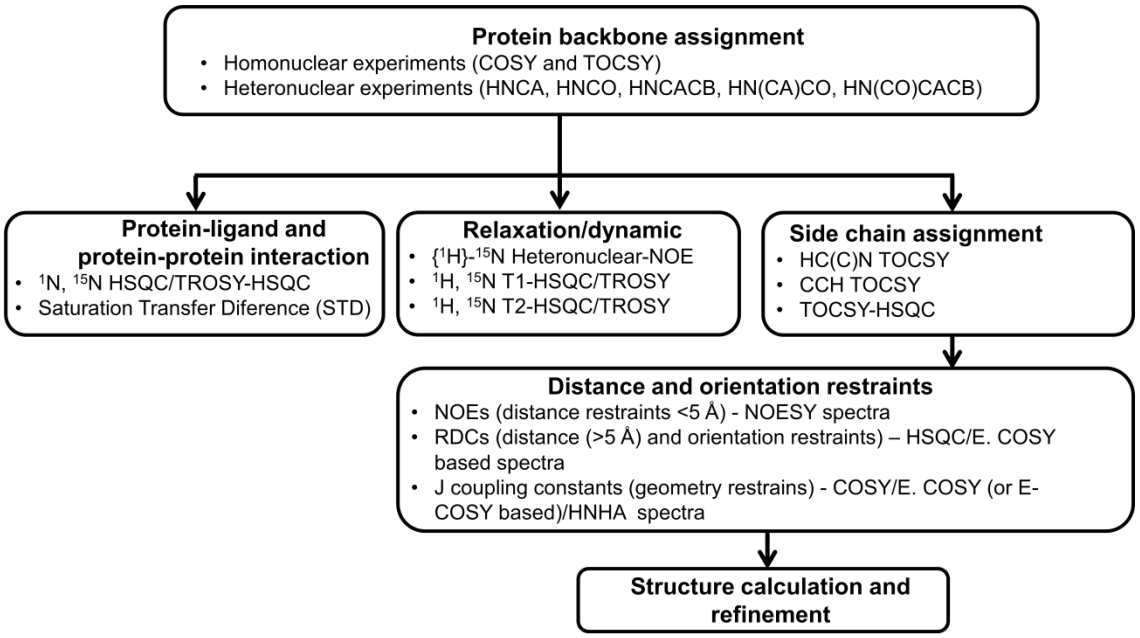


Figure 1.20 *Flow chart with some of the more important protein NMR experiments.* Depending on the protein size, homonuclear or heteronuclear experiments must be performed to do the protein backbone assignment. With this, 2D ^{1}H, ^{15}N HSQC/TROSY-HSQC spectra can be acquired upon ligand or protein addition to study protein-ligand and/or protein-protein interactions. Protein relaxation studies can be performed to determine protein relaxation, namely the hetNOE values and T_1 and T_2 time constants, for example. Protein structure determination is achieved using the distance and orientation restraints [94, 99].

Protein backbone assignment is the first step in NMR protein studies. Once this is done, several experiments can be performed depending on the goal to achieve. Protein-ligand interactions can be studied by STD NMR. ^{1}H, ^{15}N HSQC/TROSY spectra acquired upon ligand addition to a protein can be performed to understand if binding occurs and which residues are involved in the binding. Protein relaxation and dynamics can be studied by determining {^{1}H}-15N Heteronuclear-NOE values and T1 and T2 relaxation times. Finally, protein solution structure can also be determined by NMR. For that, several spectra need to be acquired to perform the backbone and side-chain protein assignment. Afterwards, distance, orientation and geometric restraints must be obtained in order to perform structure calculation – dipolar cross-relaxation (NOE) rate constants, scalar coupling constants, isotropic chemical shifts, and residual dipole-dipole coupling constants (RDCs) [99].

NMR spectroscopy has registered a great progress, namely in the field of protein NMR, in the last years: higher field magnets and cryoprobes were built and new acquisition pulse sequences designed. In addition, the development of recombinant protein overexpression with isotopic labeling allowed the study of larger molecules, extending the molecular weight limit to 100 kDa [100]. All these developments are expanding the applications and the limits of this technique.

1.6. COMBINING X-RAY CRYSTALLOGRAPHY AND NMR ON THE CHARACTERIZATION OF HSOUL/HBP HEME-BINDING FAMILY OF PROTEINS

X-ray Crystallography is a very powerful structural tool as it allows the determination of atomic coordinates and structures with high resolution can be obtained. In addition, the size limit inherent to NMR (approximately 50 kDA for a protein structure and approximately 100 kDA for local analysis) is not a limitation in Macromolecular Crystallography. On the other hand, NMR is much more powerful for studying the mobility and other dynamic processes such as the determination of pKa values, dissociation constants, etc.

The other main differences are related to sample preparation. In X-ray Crystallography it is mandatory to obtain good diffracting monocrystals obtained in very specific crystallization conditions. For NMR, the sample is in solution and as far as the sample is stable, a lot of variations on the solution conditions can be tested.

The number of deposited structures in the protein data bank - PDB (that include proteins, peptides, viruses, protein-nucleic acid complexes, nucleic acids and carbohydrates) has increased significantly, particularly since 2003. By analysis of the graph in figure 1.21, X-ray Crystallography is clearly the main technique used in the determination of the structure of biological molecules. So far, more than 67000 structures determined by X-ray Crystallography have been deposited in the PDB in contrast to approximately 9100 structures determined by Nuclear Magnetic Resonance.

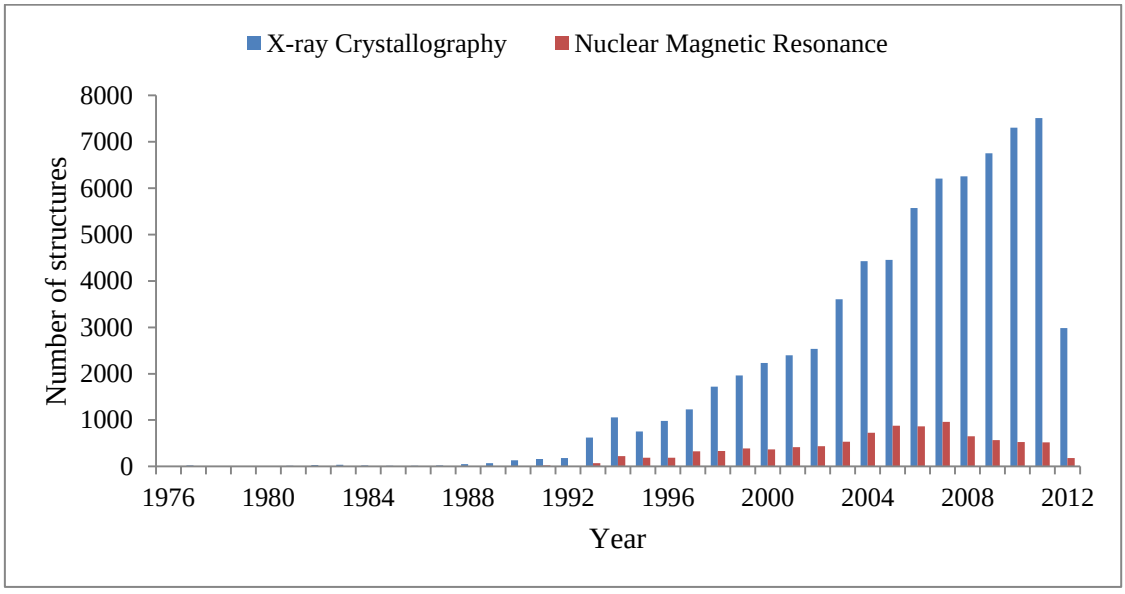


Figure 1.21 *Number of structures deposited in the Protein Data Bank (PDB, http://www.pdb.org/pdb/home/home.do).* The blue bars correspond to the number of deposited structures, solved by X-ray Crystallography. The red bars correspond to the number of solution structures, determined by Nuclear Magnetic Resonance, deposited in the PDB (data updated at May 2012).

Recently, Cryo-Electron Microscopy (cryo-EM) has been used in structure determination. This technique is used to determine the structure of macromolecules and biological aggregates with molecular resolution between 7 and 30 Å. The virtually unlimited size of the structure to be analyzed is the main advantage of this technique. In addition, smaller amounts of sample are required and the cryo methods allow the sample to be in their native aqueous medium, close to physiological conditions.

As mentioned in chapter *1.2.*, the SOUL/HBP family of heme-binding proteins has two members identified, heme-binding protein 1, p22HBP (approximately 22 kDa), and heme-binding protein 2, SOUL (approximately 23kDa). Until 2006 no structural information was available for any of the members of this family and until now no clear function, besides heme-binding, has been attributed to these proteins. Therefore, the two proteins are excellent candidates to be studied by Macromolecular Crystallography and NMR, in order to obtain more structural information, and to understand the dynamics of these proteins.
The solution structure of murine p22HBP was determined by NMR and chemical shift mapping was performed to identify the residues important in heme-binding [18]. Human p22HBP, due to the 86 % sequence identity to murine p22HBP, is expected to display a very similar three dimensional structure. For p22HBP, theoretical studies have been performed to confirm the residues participating in the heme-binding and determine the residues involved in this process in human p22HBP [19]. Molecular modeling studies determined the possible orientation of the heme molecule. The complete elucidation of the heme binding to murine and human p22HBP would be possible by determining the X-ray structure of the complexes murine p22HBP-hemin/ppIX and human p22HBP-hemin/ppIX.

The determination of SOUL protein structure would provide further structural characterization of this family of heme-binding proteins, and provide further understanding of SOUL protein function (a wide range of functions have been attributed to the protein including functions such as a heme transporter or being involved in necrotic cell death) [20, 31, 32].

Sato *et al* proposed that the heme binding for SOUL is through His42, the only histidine in hSOUL protein sequence [20]. The role of His42 can be probed by NMR. HSQC and HMQC pulse sequences can resolve the histidine cross-peaks of a protein as $^{15}N\delta_1$ and $^{15}N\varepsilon_2$ of histidine resonate at characteristic chemical shifts far from the backbone amides and other amino-acid side chains – figure 1.22. This methodology was initially used by Stockman *et al* for the histidine and tryptophan assignment of ^{15}N-labeled flavodoxin [101].
This methodology has also been applied in the study of human normal adult hemoglobin (Hb A). Hb A has been extensively studied, namely the elucidation of the relationship between its

structure and its physiologically important features, such as the cooperative oxygen binding and the allosteric interactions of oxygen and hydrogen ion-binding (Bohr Effect). However, the molecular basis of these mechanisms is not fully understood. For that, 2D NMR ^{1}H, ^{15}N-HMQC spectra of carbonmonoxy-hemoglobin A (HbCO A) were acquired to confirm the assignment of histidine residues in the heme pocket and in the protein surface [102]. These experiments were therefore used in this project to study the possible hSOUL heme binding *via* His42, by acquiring ^{1}H, ^{15}N-HSQC spectra of hSOUL upon hemin addition.

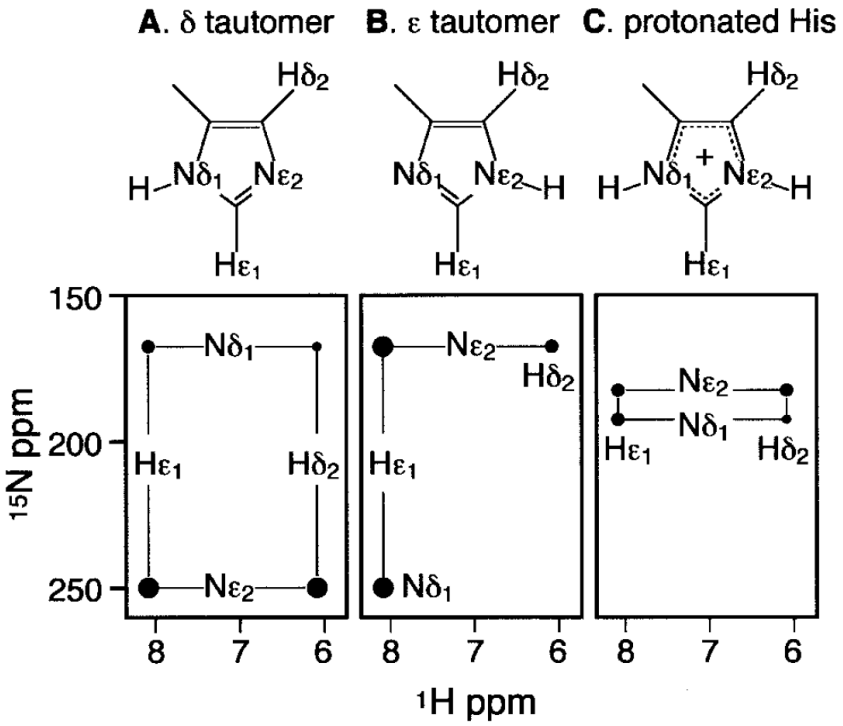


Figure 1.22 *Schematic diagram of theoretical expected ^{1}H, ^{15}N-HMQC spectrum of the imidazole-ring of the three possible protonation states of a histidyl residue (from [102]).*

In addition, ^{1}H, ^{15}N-HSQC/HSQC-TROSY spectra can be used to map the chemical shift changes upon heme addition and consequently to identify the residues involved in the process.

In summary, it should be emphasized that the work presented in this thesis confirms the importance of combining X-ray Crystallography and Nuclear Magnetic Resonance data to understand protein structure and dynamic and consequently give some insights into the protein function.

1.7. OBJECTIVES

The first goal of the work presented in this dissertation was to optimize the overexpression and purification of hSOUL protein in order to produce sufficient amounts of protein needed for the preparation of protein crystals for Macromolecular Crystallography experiments and samples for NMR. Additionally, it was necessary to optimize the overexpression of isotopic labeled protein, ^{13}C, ^{15}N and ^{2}H.

The determination of hSOUL protein structure, either by NMR or X-ray Crystallography, constituted the second goal of this work. Once the structure of the predicted heme-binding protein has been solved, it would be important to determine the protein-heme complex structure (by X-ray Crystallography) or at least identify the residues involved in the heme-binding (using Nuclear Magnetic Resonance). The dynamic of the heme-binding studied by NMR spectroscopy would be an important additional objective as it would provide significant insights for the elucidation of the protein mechanism and function, another goal of this thesis.

The solution structure of murine p22HBP has been determined by Dias *et al* and some theoretical calculations, NMR titrations and fluorescence quenching experiments have been performed in order to understand the mechanism of heme-binding to the protein [18, 19]. So far, no structural information is available for human p22HBP, although the 86% of sequence identity suggests a very similar protein folding. Therefore, another aim of this thesis was to determine the X-ray structure of the proteins, murine p22HBP and human p22HBP, bound to hemin/PPIX.

CHAPTER 2

HUMAN SOUL CLONING, OVEREXPRESSION AND PURIFICATION

CONTENTS

2.1. INTRODUCTION

Structural and dynamic studies by Nuclear Magnetic Resonance and X-ray Crystallography techniques require great amounts of pure protein. Even with a 600 MHz NMR spectrometer equipped with a cryoprobe, for a protein of around 25 kDa such as SOUL, protein concentration in the low mM range is required, typically > 0.5 mM. On the other hand, the 'bottleneck' for X-ray Crystallography is obtaining good diffraction crystals and it is often required to test thousands of different crystallization conditions. Besides this, solving the 'phase problem' in X-ray Crystallography is needed to determine the protein structure. Although different approaches can be followed, a very common technique used is the incorporation of seleno-methionines (substituting the methionine residues) in the protein to solve the 'phase problem' by MAD/SAD methods [103]. This can be achieved by cloning and overexpressing the protein of interest in highly studied and cheap systems, such as the bacteria *Escherichia coli (E. coli).* For these reasons, one of the main goals of this thesis was to clone human SOUL protein and optimize its overexpression in *E.coli* and subsequent purification.

As previously mentioned, bacterial expression systems are commonly used for recombinant protein cloning and overexpresion as they are cheap and well studied systems. The major drawbacks in these systems are the lack of secretation systems to the release of proteins to the growth media and incapacity to perform disulfide-bond formation and other posttranslational modifications. The introduction of DNA into host cells can be carried out by several vectors (plasmids, lambda phages, cosmids, etc.). The more commonly used vectors are plasmids which ensure easy cloning of the recombinant DNA into the host cell. These vectors must possess some specific features. All plasmids must have at least one DNA sequence that can act as an *origin of replication*, so they are able to multiply within the cell. In laboratory, antibiotic resistance is often used as a *selectable marker* to ensure that bacteria in a culture contain a particular plasmid. A *promoter* has to be present to induce protein expression. Gene cloning requires DNA molecules to be cut in a very precise and reproducible fashion. Each vector molecule must be cleaved at a single position, to open up the circle so that new DNA can be inserted. The DNA cleavage is performed in specific sites by restriction enzymes. The final step in the construction of a recombinant DNA molecule is the joining together of the vector molecule and the DNA to be cloned – ligation, a reaction catalyzed by DNA ligase. One of the most popular plasmids are the pET system plasmids, based on the T7 promoter. *E. coli* BL21 (DE3), for example, is a very common strain for protein overexpression and contains a T7 RNA polymerase gene, under the control of *lac*UV5 promoter. The protein overexpression can be induced, in this strain, by the addition of a lactose analog, isopropyl-β-D-thiogalactopyranoside (IPTG).

For purification purposes, the protein of interest is often cloned with a C- or/and N-teminal polyhistidine-tag (5 to 6 histidines). The purification is perfomed by Immobilized Metal Affinity Chromatography (IMAC), using an affinity media containing bound metal ions, either Ni^{2+} or Co^{2+}, to which the polyhistidine tag will bind with micromolar affinity. This approach was used for human SOUL, and human and murine p22HBP.

For hSOUL, the IMPACT™ (Intein Mediated Purification with an Affinity Chitin-binding Tag) was also used. This system utilizes the inducible self-cleavage activity of engineered protein splicing elements (termed inteins) to purify recombinant proteins by a single affinity column. In this work, pTYB12 plasmid was used. This plasmid adds AGH residues to the protein N-terminal. The protein purification is performed in one step. The intein contains a chitin binding domain which will interact with the affinity matrix containing chitin. The self cleavage of intein can be induced by the addition of thiols such as DTT, β-mercaptoethanol or cysteine.

2.2. hSOUL N-TERMINAL HISTIDINE TAG FUSION PROTEIN

2.2.1. MATERIALS AND METHODS

2.2.1.1. OVEREXPRESSION, PURIFICATION AND ISOTOPIC LABELING

hSOUL overexpression was achieved by growing the BL21(DE3) (Novagen) cells harboring the hSOUL/IOH3379-pDEST17-D18 plasmid (RZPD) in 2×YT media containing 25 $\mu g.ml^{-1}$ of zeocin (InvivoGen) and incubated at 310 K. hSOUL was cloned with an histidine tag (MSYY**HHHHHH**LESTSLYKKAGT) attached to the protein N-terminal. The culture was then inoculated in M9 minimal media (see appendix) and incubated at 310 K. Protein expression was induced at OD_{600nm} = 0.6 at a final concentration of 0.1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG), for 16 hours at 303 K. Different ITPG concentrations (0.1 mM, 0.5 mM and 1.0 mM) were initially tested. Three and five hours of induction time besides approximately 16 hours were also tested.

The harvested cells were resuspended in 50 mM phosphate buffer, pH 8.0 with 300 mM NaCl, ruptured by sonication (Hielscher - Ultraschall-Technologie) and centrifuged at 48384 × g (20000 rpm) for 1 hour.

The supernatant was loaded onto a Ni-NTA-agarose column (QIAGEN) previously equilibrated with the same buffer. The resin was washed in 2 steps with a buffer containing 50 mM phosphate at pH 8.0, 300 mM NaCl and 10 mM imidazole and with an identical buffer containing 20 mM instead of 10 mM imidazole. hSOUL was eluted in a discontinuous way with a buffer containing 50 mM phosphate at pH 8.0, 300 mM NaCl and 250 mM imidazole and with 50 mM phosphate at pH 8.0, 300 mM NaCl and 500 mM imidazole. Imidazole concentrations of 50 mM, 75 mM, and 175 mM between 20 mM and 250 mM were tested. The fractions containing hSOUL were concentrated and loaded (approximately 400 µl containing 10 mg of hSOUL) onto a Superdex 75 10/300 GL column (GE Healthcare pre-packed) coupled to an FPLC system (GE Healthcare) previously equilibrated with 50 mM phosphate at pH 8.0. The eluted fractions containing hSOUL were pooled together and concentrated in an Amicon concentrator equipped with an YM10 membrane. From hereby, this hSOUL clone is designated as hSOUL (histidine tag).

Isotopic labeling (^{13}C and ^{15}N) was achieved by using $^{15}NH_4Cl$ and U-^{13}C-glucose (CortecNet) as the sole nitrogen and carbon sources, respectively. To obtain a good isotopic labeling, the harvested cells, after overnight culture, were resuspended in M9 minimal medium containing $^{15}NH_4Cl$ (1 g *per* liter of minimal medium) and U-^{13}C-glucose (4 g *per* liter of minimal medium) and inoculated in the isotopic enriched M9 minimal medium. For triple isotopic labeling, the M9 minimal medium was prepared with 99.89 % ^{2}H atom D_2O (CortecNet).

2.2.2. RESULTS AND DISCUSSION

2.2.2.1. OVEREXPRESSION AND PURIFICATION

After protein overexpression optimization, yields of approximately 20 mg of soluble protein *per* liter of cell culture were achieved. After the two steps of purification, the immobilized metal affinity chromatography and gel filtration – figure 2.1, the protein is obtained with high purity for further studies.

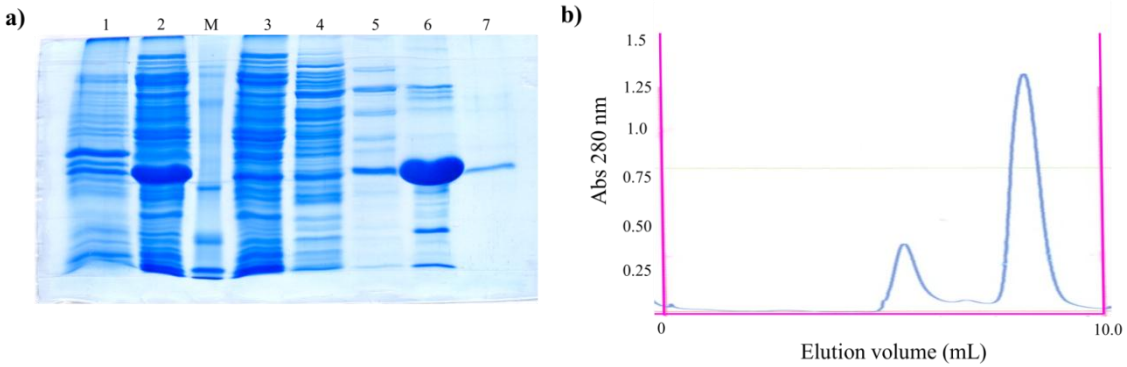


Figure 2.1 *Purification of overexpressed hSOUL (histidine tag)*. a) SDS-PAGE (15 % acrylamide) analysis of the different fractions obtained from the Ni-NTA Agarose column: 1 – insoluble fraction 2 – soluble fraction loaded on the Ni-NTA resin; M - Precision Plus Protein Unstained Standards - 10, 15, 20, 25, 37, 50, 75, 100, 150 and 250 kDa (Biorad); 3 – flow-through; 4 – resin wash with 10 mM imidazole; 5 - resin wash with 20 mM imidazole; 6 – hSOUL elution with 250 mM imidazole; 7 – hSOUL elution with 500 mM imidazole. b) Elution profile obtained from the gel filtration column (Superdex 75) loaded with hSOUL fractions (6+7) from Ni-NTA Agarose resin.

After gel filtration, hSOUL purity was confirmed by SDS-PAGE (15% acrylamide) gel analysis.

2.3. HSOUL C-TERMINAL HISTIDINE TAG FUSION PROTEIN

2.3.1. MATERIALS AND METHODS

2.3.1.1. CONSTRUCTION OF HSOUL PLASMID WITH C-TERMINAL HISTIDINE TAG, CLONING, OVEREXPRESSION AND PURIFICATION

Full-length hSOUL cDNA was isolated by polymerase chain reaction (PCR) amplification from the hSOUL/IOH3379-pDEST17-D18 plasmid (RZPD) [104] and engineered to contain a NdeI site overlapping the starting codon and a HindIII site immediately 5' of the amber codon. The amplified NdeI-HindIII hSOUL coding sequence was subsequently inserted into the NdeI-HindIII cleaved pET32a (+) plasmid with a C-terminal histidine tag fused to hSOUL in which the last two residues (NE) were replaced by a lysine and a leucine (KLAAALEHHHHHH). *E. coli* DH5α strain was used for cloning purposes and C-terminally His-tagged hSOUL was produced in the *E. coli* BL21 (DE3) host strain. These *Escherichia coli* BL21 cells were

cultured in 2×YT medium broth containing ampicillin (100 $\mu g.ml^{-1}$). Protein expression was induced at OD_{600nm} = 0.5-0.8 at a final concentration of 0.1 mM IPTG, for 16 hours at 303 K.
Cells were harvested and resuspended in 50 mM phosphate buffer pH 8.0, 300 mM NaCl. Cells were afterwards ruptured by sonication (Hielscher - Ultraschall-Technologie) and centrifuged at 48384 × g (20.000 rpm) for 1 hour.
The first step of purification procedure was similar to the one described for N-terminal histidine tag hSOUL fusion protein (section *2.2.1.1*) – before eluting with 50 mM phosphate at pH 8.0, 300 mM NaCl and 250 mM imidazole, the same buffer was used with 75 mM imidazole. Alternatively to the IMAC purification step, hSOUL purification was performed with Sephacryl S-200 HR (GE Healthcare) resin, equilibrated with 50 mM phosphate buffer pH 8.0.
Protein isotopic labeling was achieved using the protocol described in *section 2.2.1.1*.

2.3.1.2. NMR SAMPLE PREPARATION, DATA ACQUISITION AND PROCESSING

NMR spectra were acquired on a *Varian DirectDrive 600 MHz SB* equipped with a coldprobe (^{1}H, ^{15}N, ^{13}C, ^{2}H), at 293 K. SOFAST 2D ^{1}H, ^{15}N-HSQC and ^{1}H, ^{15}N-HSQC spectra on a 5 mm tube with the protein at approximately 0.3 mM, in 50 mM phosphate buffer, pH 8.0, 10 % D_2O were acquired. The spectra were processed using *NMRPipe [105]*.

2.3.2. RESULTS AND DISCUSSION

2.3.2.1. CLONING, OVEREXPRESSION, PURIFICATION AND ISOTOPIC LABELING

hSOUL was successfully cloned as an histidine tag fusion protein. The successful cloning results were confirmed by sequencing using T7 forward and reverse primers.
With a histidine tag attached to the protein, the first approach was to use the IMAC technique with a Ni-NTA agarose resin for purification. However, a SDS-PAGE gel (figure 2.2) showed that the protein did not bind to the resin, being eluted after the addition of buffer containing 10 mM imidazole.

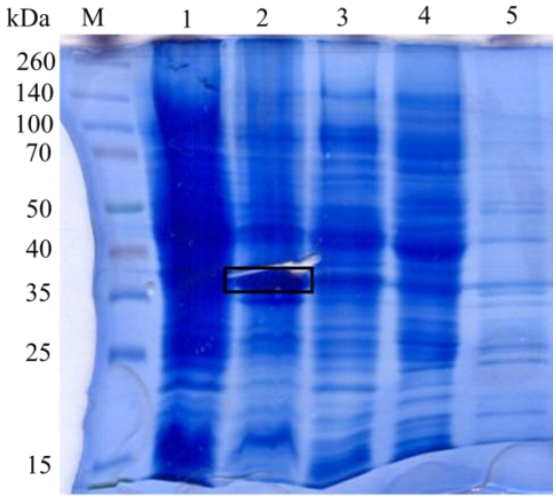


Figure 2.2 *Purification of overexpressed hSOUL (C-terminal histidine tag)*. SDS-PAGE (15 % acrylamide) analysis of the different fractions obtained from the Ni-NTA Agarose column: M - Spectra™ Multicolor Broad Range Protein Ladder (LadAid); 1 – flow-through; 2 – resin wash with 10 mM imidazole; 3 - resin wash with 20 mM imidazole; 4 – hSOUL elution with 75 mM imidazole; 5 – hSOUL elution with 250 mM imidazole. hSOUL protein band is identified in the black rectangle.

Analyzing lane 2 of the SDS-PAGE gel in figure 2.2, one can see an intense band at approximately 35 kDA, corresponding to hSOUL protein; the same band can be observed in lanes 3-5. This result indicates that the protein did not interact with the resin. One possible explanation for this occurrence may be the fact that the histidine tag is in the core of the protein, which means the histidine residues are not exposed to the immobilized Ni^{2+} ions, therefore no interaction occurs. Another possible explanation may be the unspecific cleavage of hSOUL protein near the histidine tag.

In order to overcome this problem, a gel filtration was performed using Sephacryl S-200 – figure 2.3.

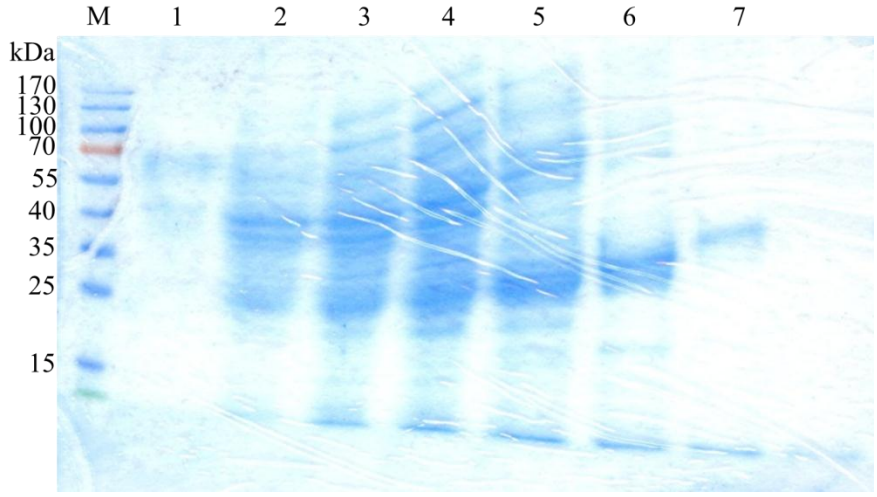


Figure 2.3 *Purification of overexpressed hSOUL.* SDS-PAGE (15 % acrylamide) analysis of the different fractions obtained from the Sephacryl S-200 resin equilibrated with 50 mM phosphate buffer pH 8.0: M – Molecular weight markers (Fermentas®); 1 - 8 - collected samples.

The fraction corresponding to lane 7 was concentrated to approximately 500 µl corresponding to a protein concentration of 0.3 mM. Deuterium oxide - D_2O (Sigma Aldrich) was added to the concentrated sample to a final concentration of 10% (v/v) for further NMR data acquisition. To analyse the protein folding a SOFAST 2D ^{1}H, ^{15}N-HSQC and a ^{1}H, ^{15}N-HSQC spectra (figures 2.4 and 2.5, respectively) were acquired.

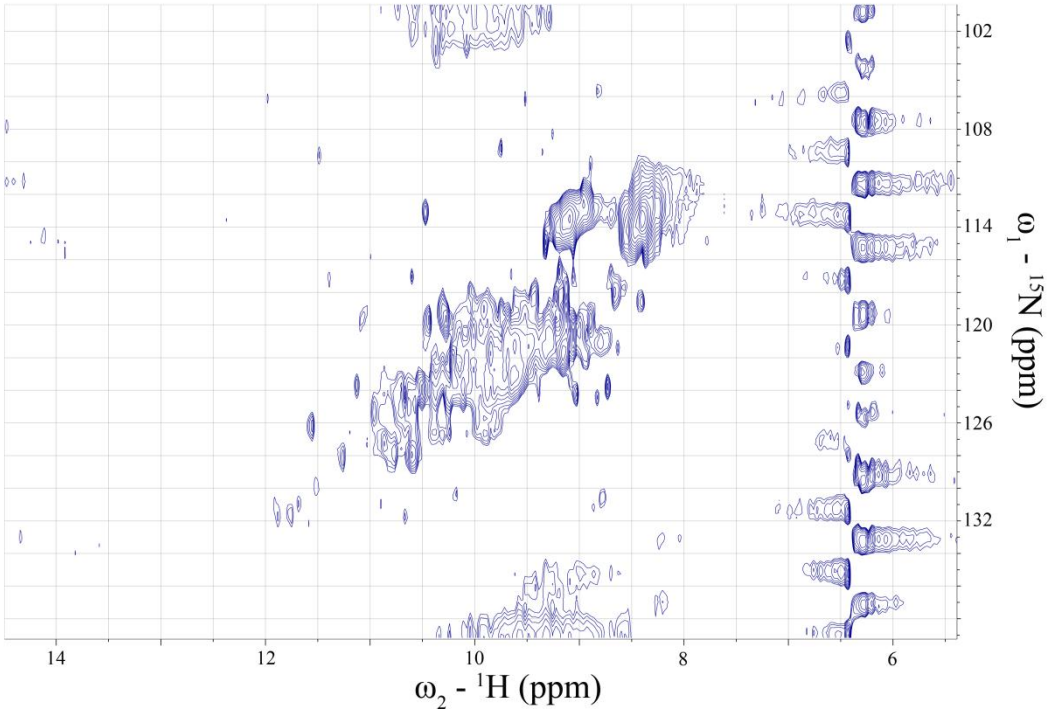


Figure 2.4 *SOFAST ^{1}H, ^{15}N- HSQC spectrum of C-terminal his tagged hSOUL.* 0.3 mM ^{15}N-labeled hSOUL sample spectrum acquired on a 600 MHZ NMR spectrometer with cryoprobe, at 293 K.

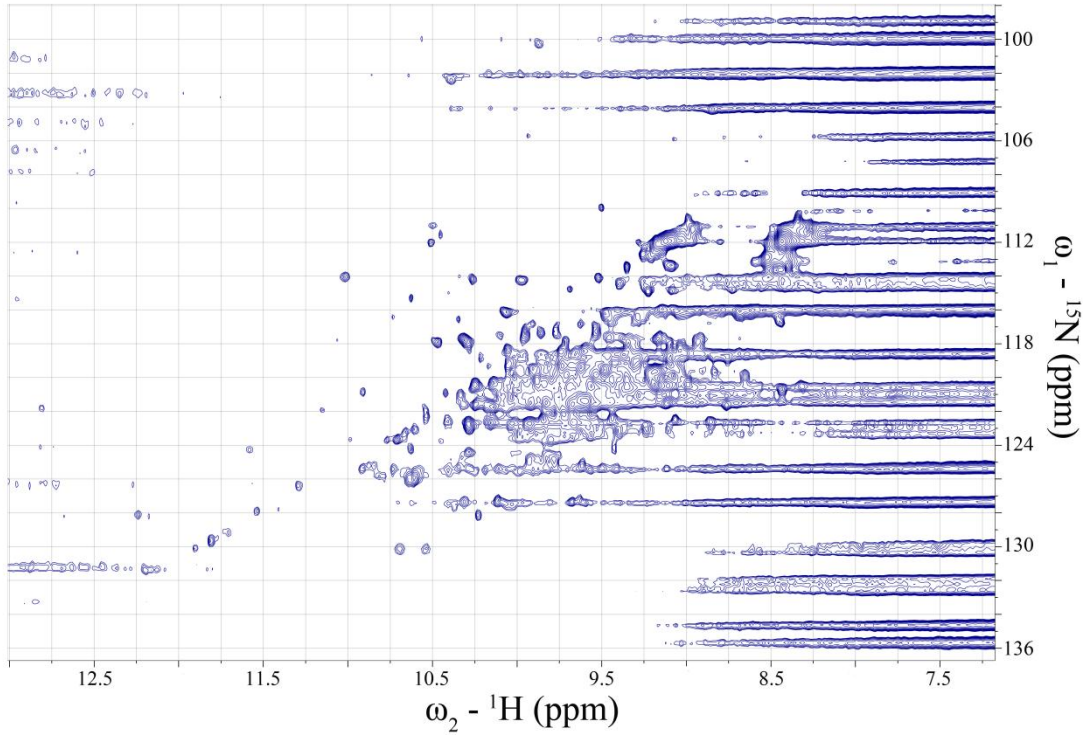


Figure 2.5 *^{1}H, ^{15}N- HSQC spectrum of C-terminal his tagged hSOUL.* 0.3 mM ^{15}N-labeled hSOUL sample spectrum acquired on a 600 MHZ NMR spectrometer with cryoprobe, at 293 K.

Spectra analysis shows that the protein is not properly folded, which prevented from continuing the studies with this hSOUL clone. Probably, due to the well structured C-terminal domain of human SOUL protein, the addition of extra residues may have disturbed the protein folding.

2.4. hSOUL-INTEIN FUSION PROTEIN

2.4.1. MATERIALS AND METHODS

2.4.1.1. CONSTRUCTION OF hSOUL PLASMID WITH INTEIN TAG, CLONING, OVEREXPRESSION, PURIFICATION AND ISOTOPIC LABELING

Full-length hSOUL cDNA was isolated by polymerase chain reaction (PCR) amplification from the hSOUL/IOH3379-pDEST17-D18 plasmid (RZPD) [104] and cloned into the pCR-BluntII-TOPO plasmid (Invitrogen). hSOUL cDNA was subsequently inserted into the pTYB12 plasmid of the Impact-CN kit (New England BioLabs$^{®}_{Inc}$) (Nde I and Eco RI restriction sites), to give the pTYB-SOUL construct with an N-terminal intein tag fused to hSOUL. The hSOUL protein cleaved from the fusion product gave wild type hSOUL with 3 additional amino acids at the N-terminus (AGH).

hSOUL protein was produced by first transforming the hSOUL-pTYB12 expression vector into competent *E. coli* DH5α cells and then into *E. coli* ER2566 competent host strain. These

Escherichia coli ER2566 cells were cultured in Terrific Broth (TB) medium broth containing ampicillin (100 μg/ml) at 310 k. On a first approach, 250 μL of the overnight culture were harvested and resuspended in 50 mL TB medium supplemented with ampicillin (100 μg/mL) and left for approximately 16 hours at 303 K, with shaking (150 rpm). Cells were harvested, ressuspended in TB medium, grown to mid-exponential phase (A_{600} = 0.5 - 0.8), harvested, resuspended in M9 minimal medium and left for 2 hours, at room temperature. Isopropyl β-D-thiogalactopyranoside (IPTG) was then added to a final concentration of 0.5 mM, and the cultures were incubated for further 16 hours at room temperature. hSOUL overexpression was optimized by harvesting the cells from the initial overnigh culture and ressupend them in 1 liter 2×YT medium broth (supplied with ampicillin 100 μg/mL). The cells were grown until mid-exponential phase was achieved. Cells were then harvested, ressuspended in M9 minimal medium, and left for 2h hours, at room temperature. Isopropyl β-D-thiogalactopyranoside (IPTG) was then added to a final concentration of 0.5 mM, and the cultures were incubated for further 16 hours at room temperature. The cells were harvested and resuspended in 50 mM phosphate buffer at pH 8.0 with 250 mM NaCl and 1 mM ethylenediaminetetraacetic acid (EDTA) and lysed by sonication (Hielscher - Ultraschall-Technologie). The cell extract was incubated for approximately 3 hours with 20 ml of chitin beads (New England BioLabs$^{®}{}_{Inc}$) at 277 K, with stirring. The resin was transferred into a column and washed with 10 volumes of 'washing buffer' containing 50 mM phosphate buffer pH 8.0, 250 mM NaCl and 1 mM EDTA. Two volumes of washing buffer with 50 mM dithiothreithol (DTT) were added to the resin and left for 16 hours at room temperature. The protein was then eluted with 10 volumes of washing buffer containing 50 mM DTT. The purity of the sample was confirmed by 15 % SDS-PAGE gel analysis (figure 2.4, *section 2.4.2.1*). From hereby, this hSOUL clone is designated as hSOUL (intein tag).

hSOUL (intein-tag) fusion protein isotopic labeling was achieved using $^{15}NH_4Cl$ and U-^{13}C-glucose (Cortecnet) as the sole nitrogen and carbon sources, respectively. To obtain a good isotopic labelling, the harvested cells, after overnight culture, were resuspended in M9 minimal medium containing 1 g $^{15}NH_4Cl$ *per* liter of minimal medium and 4 g of U-^{13}C-glucose *per* liter of minimal medium. The remaining procedure is described in section *2.4.1.1.*

2.4.2. RESULTS AND DISCUSSION

2.4.2.1. CLONING, OVEREXPRESSION AND PURIFICATION

The optimized overexpression protocol led to a protein yield of approximately 4 mg of protein *per* liter of cell culture. A high level of purity was obtained with only one purification step – figure 2.6.

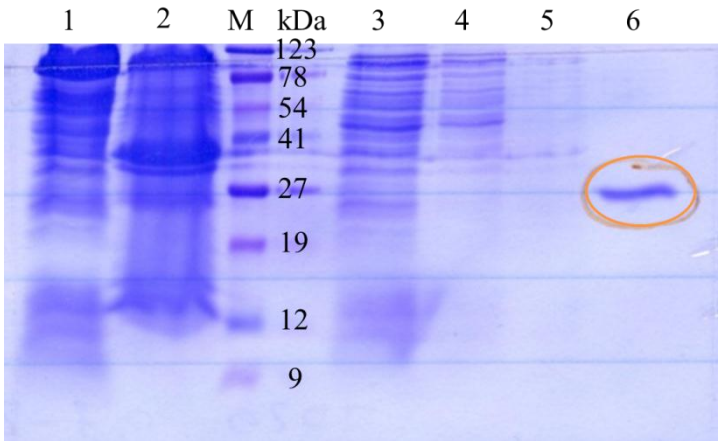


Figure 2.6 *Purification of overexpressed hSOUL.* SDS-PAGE (15 % acrylamide) analysis of the different fractions obtained from the chitin beads column: M – Protein Marker (NZYTech, genes enzymes, Ltd NZYTech); 1 – insoluble fraction; 2 –soluble fraction loaded on the chitin beads resin; 3 – flow-through; 4 – washing column step; 5 – after DTT addition; 6 – hSOUL elution. hSOUL protein band is identified by the orange circle.

By analysis of the SDS-PAGE gel on figure 2.6 it is clear that pure protein is obtained after this purification step.

The author of this thesis is especially thankful to Dr. Jean-Marc Moulis, IRTSV/LCBM, CEA-Grenoble who was responsible for the cloning of hSOUL described in sections *2.3.1.1* and *2.4.1.1*.

2.5. FINAL REMARKS

The human SOUL protein was first cloned and overexpressed as a fusion protein, with a long N-terminal 'histidine tag' (25 amino acid residues that include 6 consecutive histidine residues, essential for the immobilized metal affinity chromatography (IMAC) purification of the protein). The yield of overexpression was approximately 20 mg of protein *per* liter of culture and after two steps of purification (IMAC and gel filtration) the soluble protein was obtained with a high degree of purity. ^{1}H, ^{15}N-HSQC/TROSY-HSQC NMR spectra showed a folded and structured protein as shown in *chapter 4*, however the region around 8 ppm on proton dimension showed a big overlap of resonances. In addition, 3D NMR experiments, namely TROSY-HNCO, TROSY-HNCA and TROSY-HNCACB showed much fewer resonances than expected (see *chapter 4.3.2*.).
Additionally, several attempts were made in order to crystallize hSOUL protein construct which was achieved as described in *section 3.2.3*, although the crystals show poor diffraction power.

In order to overcome these problems three different approaches were considered although only two are hereby described. The first attempt was to clone hSOUL as a GST fusion protein, where hSOUL was overexpressed fused with the protein glutathione-S-transferase (GST). During the purification procedure a protease is added in order to cleave GST-hSOUL, and obtain hSOUL with no extra N-terminal amino acids. Once the protein was cloned, overexpression attempts were performed, however the protein was overexpressed as inclusion bodies, as this approach was therefore droped.

hSOUL was also cloned as a fusion protein with a C-terminal histidine tag and afterwards purified as described in *section 2.3.1.1*. Two drawbacks were however faced. First, the protein did not interact with the Ni-NTA resin, turning the purification process more difficult. Secondly, ^{1}H, ^{15}N-HSQC spectrum revealed a limited dispersion of proton chemical shifts, indicative of an unfolded protein.

The final approach was to clone hSOUL protein using the IMPACT Kit (New England Biolab$^{®}_{Inc.}$), which utilizes the inducible self-cleavage activity of engineered protein splicing elements (termed inteins) to purify recombinant proteins by a single affinity column. This system distinguishes itself from other protein fusion systems by its ability to separate a recombinant protein from the affinity tag without the use of a protease that requires a further purification step. In this case, intein tag cleavage was induced by DTT. High level of purity is achieved in a single purification step, since the intein fused with the protein of interest possesses a chitin beading domain (CBD) that will bind the resin.

^{1}H, ^{15}N HSQC spectrum showed a properly folded protein and allowed us to carry out hSOUL structure determination and heme-binding studies (further described in *chapters 3* and *4*, respectively).

CHAPTER 3

STRUCTURAL CHARACTERIZATION OF HUMAN SOUL BY X-RAY CRYSTALLOGRAPHY

CONTENTS

3.1. INTRODUCTION

Sato *et al* performed the first studies to understand how SOUL binds heme and concluded that the binding occurs *via* the only histidine present in the protein, constituting the axial heme ligand [20]. More recently, biochemical studies have associated the protein with the process of necrotic cell death, by permeabilization of the inner and outer membranes of the mitochondria. In addition, database searches revealed that SOUL possesses a BH3-like domain in its primary amino acid sequence. The lack of functional and structural information regarding hSOUL was the driving force for the studies presented in this dissertation. Determining the structure of hSOUL, especially as a complex hSOUL-heme, would be very important to not only reveal the possible heme-binding site, but also to understand the possible protein binding to anti-apoptotic members of the Bcl-2 family of proteins.

3.2. MATERIALS AND METHODS

3.2.1. SELENO-METHIONINE HSOUL DERIVATIVE

To produce seleno-L-methionine hSOUL, the *E. coli* methionine auxotroph B834 (DE3) was transformed with the pTYB-SOUL plasmid. The protocol for overexpression and purification was similar to that described in section *2.4.1.1.*, with the fundamental difference that, after the cells were grown to mid-exponential phase they were harvested and resuspended in M9 minimal medium supplemented with seleno-L-methionine (instead of methionine) and the other 19 naturally occurring amino acids.

3.2.2. INDUCTIVELY COUPLED PLASMA-ATOMIC EMISSION SPECTROMETRY

The pure seleno-methionine containing protein was analyzed by Inductively Coupled Plasma-Atomic Emission Spectrometry (ICP-AES) to confirm the presence of selenium.

3.2.3. SIZE EXCLUSION CHROMATOGRAPHY

The oligomerization state of hSOUL protein (in its *apo* form and bound to heme) was previously studied and was thought to be a dimer in the *apo* form, oligomerizing into a hexameric form upon heme binding. To test this, size exclusion chromatography was used to run hSOUL and hSOUL:hemin (molar ratio 1:1) samples in two different buffers – 100 mM phosphate buffer pH 8.0, 0.5 mM EDTA and 100 mM Tris-HCl buffer pH 8.0 and 0.5 mM EDTA. Four model proteins were used as references (MW = 78.5, 66.5, 16.9 and 13.7 kDa). hSOUL, murine p22HBP and hSOUL-hemin (incubated before injection in a molar ratio of 1:1 – hemin prepared according to Dias *et al* [18]). Experiments were performed on a Superdex 75-10/300 GL column (GE Healthcare pre-packed coupled to a FPLC system).

3.2.4. CRYSTALLIZATION AND DATA COLLECTION

A wide range of crystallization screens were initially used in order to obtain hSOUL (histidine tagged) crystals, namely, an in-house prepared sparse matrix screen of 80 conditions (table A.1, appendix), Crystal Screen 2 (Hampton Research), Emerald Wizard I (Emerald Biostructures), Emerald Wizard II (Emerald Biostructures), Crystallization Basic Kit for Membrane Proteins (Sigma) and JBScreen Classic 1-10 (Jena Bioscience). Experiments were performed using the hanging drop vapour diffusion method both at 277 K and 293 K, with drops consisting of 1µl of protein solution (10 mg/ml and 15 mg/ml in 10 mM Tris-HCl, pH 8.0), 1µl of reservoir solution and 700 µl of precipitant solution in the reservoir. The best crystallization conditions were obtained from the in-house prepared sparse matrix screen of 80 conditions (- 2 M ammonium sulphate, 0.1 M MES 6.5; - 0.2 M ammonium sulphate, 0.1 M cacodylate pH 6.5, 30% polyethylene glycol 8000 and - 2 M ammonium sulphate, 0.1 M Tris-HCl 8.5). For these

conditions, several ionic liquids were tried as additives (C_4MIMCl - 1-*n*-butyl-3-methylimidazolium chloride (Fluka), C_4MIM PF_6 - 1-butyl-3-methylimidazolium hexafluorophosphate (Solchemar), C_4MIM Otf - 1-butyl-3-methylimidazolium trifluoromethanesulfonate and cytosine bistriflimide. For that 0.5 µL of the ionic liquid were added to the drop (1 µL of protein, 1 µL of the reservoir solution and 0.5 µL of the ionic liquid) to give a final ionic liquid concentration of 0.2 and 0.4 M.

hSOUL crystallization trials were also attempted at the High Throughput Crystallization Laboratory, at the European Molecular Biology Laboratory (EMBL), Grenoble; Hampton Crystal Screen 1 to 6 were tested on hSOUL (15 mg/ml in 10 mM Tris-HCl, pH 8.0) on a 1:1 volume ratio drops at 277 K and 293 K using the sitting drop vapour diffusion method. All the conditions described above were also tried for hSOUL (15 mg/ml) previously incubated with hemin (prepared according to Dias *et al* [18]) in a 1:1 molar ratio solution. For *apo* hSOUL protein some promising crystallization conditions were observed and the manual scale up was tried. Protein crystals were obtained with 1.8 M Na/K phosphate buffer pH 5.6 as precipitant solution, at 277 K.

Once crystals of the *apo* protein hSOUL were obtained, they were stabilized with the corresponding harvesting buffer. Crystals were afterwards cryocooled using the harvesting buffer supplemented with 25 % glycerol (w/v), previously tested on the in-house X-ray generator (FR591 Enraf-Nonius) with Cu rotating anode and imaging plate (MAR-Research) as a detector, coupled to an Oxford cryo-system.

Several data sets were collected using single crystals grown under similar conditions in search for data of the highest diffraction quality. Data collection was performed on various beamlines, namely ID14-EH2, ID14-EH4, ID23-EH1 and ID29 at the European Synchrotron Radiation Facility – ESRF (Grenoble). All crystals were found to belong to the hexagonal point group 622 with a translational screw axis component, which could not be determined from the data alone, although examination of the 00l reflections suggested the existence of a 6_2 or 6_4 screw axis. After an exhaustive search for crystals of acceptable diffraction quality, the best data were collected on ID14-EH2 from a SOUL crystal (figure 3.4, section *3.3.2*) that diffracted to beyond 3.5 Å resolution. The unit-cell parameters were found to be $a = b = 144.7$ Å, $c = 60.2$ Å which, after calculation of the Matthews coefficient ($V_M = 3.6$ Å^3.Da^{-1}), indicated the presence of one molecule of hSOUL (25 530 Da) in the asymmetric unit and a solvent content of 66 % [106]. All data were integrated with the program MOSFLM [107] and scaled with SCALA from the CCP4 suite of programs [108]. A summary of the data collection statistics is shown in table 3.2, section *3.3.3*.

Crystal optimization attempts were made by varying the buffer pH, precipitant concentration and temperature based on previous hits. Also, Additive Screen 1 (Hampton Research) and

Additive Screen 2 (Hampton Research) were tested at 277 K and 293 K. All crystals obtained were poor diffracting and no useful data were collected.

In order to solve hSOUL protein structure using experimental phasing protocols some approaches were tried in order to incorporate an anomalous scatterer in the protein. For that purpose, hSOUL was incubated for approximately 1 hour with sodium phosphomolybdate hydrate ($Mo_{12}Na_3O_{40}P.H_2O$) (Aldrich) and sodium phosphotunsgate hydrate ($Na_3O_{40}PW_{12}$) (Aldrich) in a 1:1 molar ratio solution and crystallization trials were attempted with the best crystallization conditions obtained for the hSOUL protein (- 2 M ammonium sulphate, 0.1 M MES 6.5; - 0.2 M ammonium sulphate, 0.1 M cacodylate pH 6.5, 30% polyethylene glycol 8000; - 2 M ammonium sulphate, 0.1 M Tris-HCl 8.5 and - 1.8M Na/K phosphate buffer pH 5.6) at 277 K and 293 K.

All the experiments described so far on this section (*3.2.4.*) were performed with hSOUL (histidine tag).

The crystallization trials of the hSOUL (intein tag) seleno-methionine derivative were performed using previously described conditions at 277 K and 293 K as crystallization conditions. The drops consisted of 1 µl of protein (15 mg/ml, 10 mM Tris-HCl 8.0, 10 mM DTT) and 1 µl of the reservoir solution. The best diffracting crystals were obtained with 1.8 M Na/K phosphate buffer pH 5.6 and were stabilized with the corresponding harvesting buffer – 2.2 M Na/K phosphate buffer pH 5.6 and afterwards cryocooled, using the harvesting buffer supplemented with 25 % glycerol (w/v). Preliminary analysis of the diffraction data from crystals obtained at both temperatures showed that the crystals obtained at 293 K were twinned.
A Single-wavelength Anomalous Dispersion (SAD) experiment was conducted on beamline ID23-EH1 at the ESRF, using an ADSC Quantum-4 CCD detector. After an X-ray absorption scan, data were collected at 0.9793 Å wavelength to maximize the anomalous signal. After indexing an initial diffraction image using MOSFLM [107], the program STRATEGY [109] was used to determine the optimal range to collect complete data using a minimal oscillation sweep. A total of 120 images with 1º oscillation for 0.7 seconds (with a detector-to-crystal distance of 537 mm) were collected to 2.7 Å resolution. The data were processed with MOSFLM, and scaled with SCALA [110], part of the CCP4 suite of programs [108]. The statistics are shown in table 3.2, section *3.3.3.*
The hSOUL seleno-methionine derivative crystal (obtained with 1.8M Na/K phosphate buffer pH 5.6 at 277 K) belonged to the space group $P6_222$ with cell constants $a = b = 146.4$ Å and $c = 133.0$ Å. Assuming two molecules in the asymmetric unit, the calculated Matthews coefficient (V_M) is 4.4 $Å^3\ Da^{-1}$ and the solvent content is 72 % [106]. The possibility of having three copies

of hSOUL in the asymmetric unit was also considered, as indicated by a $V_M = 3.0$ Å^3 Da^{-1} and a more probable solvent content of 58%. However, a self-rotation function calculation with program MOLREP [111], indicated clear peaks at chi = 180º, confirming the presence of a homodimer in the asymmetric unit – figure 3.1.

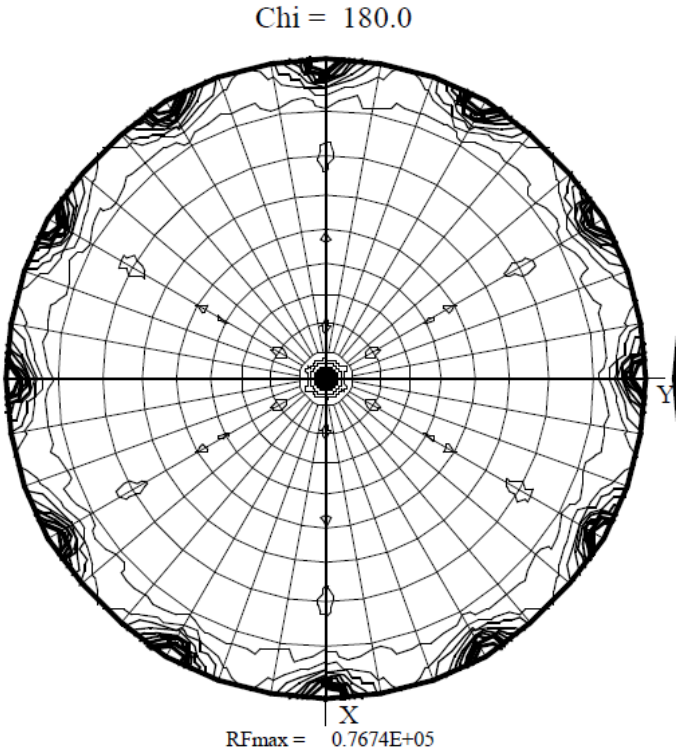


Figure 3.1 *Self-rotation function.*

3.2.5. STRUCTURE SOLUTION, MODEL BUILDING AND REFINEMENT

Before the availability of a seleno-methionine derivative, attempts to solve the 3D structure of hSOUL were based on molecular replacement (MR) methods. The available solution structures (PDB accession codes: 2GOV and 2HVA), obtained by NMR spectroscopy, were used as search models in the BALBES molecular replacement system from the CCP4 suite of programs [108, 112]. These models show around 27 % sequence similarity to hSOUL, and the CHAINSAW [113] module of BALBES produced a search model from these, modifying the template average structures on a residue-by-residue basis. The search was performed for all choices of possible space groups, and a molecular replacement solution was found for space group $P6_422$, with 1 molecule of SOUL in the asymmetric unit, using the program MOLREP, implemented in CCP4 suite of programs [108] with a MR-score of 2.14, a C-score of 0.9087 and a figure-of-merit of 0.44. This solution was further confirmed using the program PHASER [114] (using the lowest energy NMR structure from each search model) that found a clear solution for space group $P6_422$, with a z-score of 9.63 and a refined LL-gain of 90.45, against a z-score of 5.46 and a LL-gain of 20.37 for space group $p6_22$.

For some of the tested crystals, the presence of systematically weak reflections suggested the existence of a pseudo-translation vector. This was confirmed by the presence of a strong peak (with around 20 % of the origin peak height) in the calculated native Patterson map (figure 3.2). In these cases, the cell constants were doubled along the *c* axis, suggesting the presence of 2 molecules of SOUL in the crystals' asymmetric units. PHASER was not able to produce a molecular replacement solution from these data sets. Attempts to find a solution also involved the program MOLREP, using the known pseudo-translation vector, although a clear MR solution was not found. These data were eventually discarded in the further attempts to solve hSOUL's crystal structure.

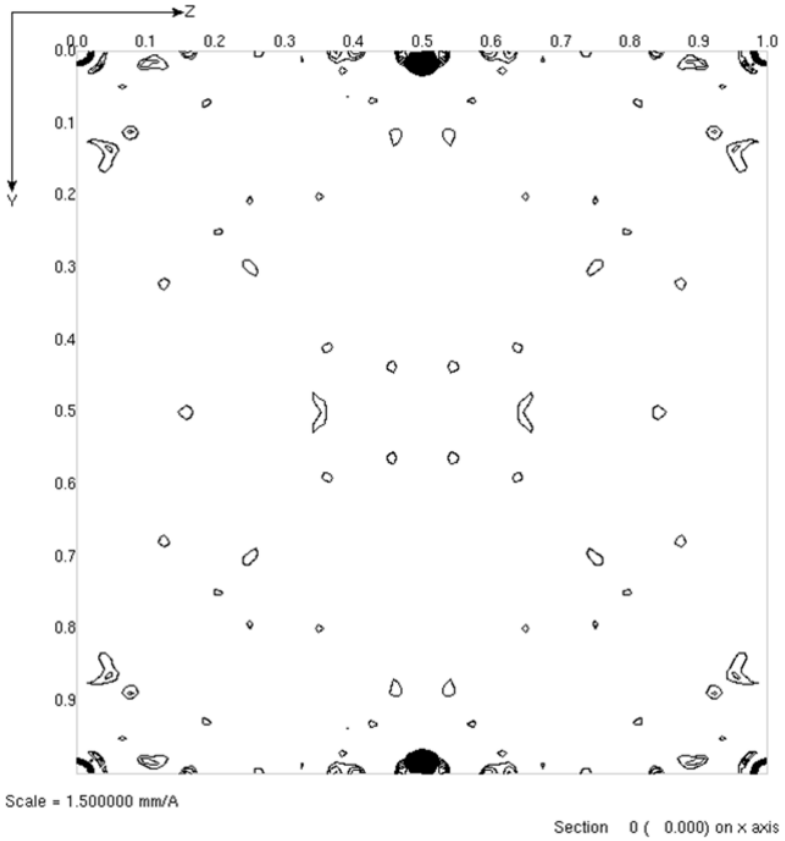


Figure 3.2 *Native Patterson map.* Patterson map of hSOUL where pseudo-translational symmetry was detected due to the strong off-origin peak.

Independent phases were obtained by crystallizing the hSOUL seleno-methionine derivative. An X-ray absorption spectrum of the crystal with a clear white line was collected near the selenium *K* absorption edge by measuring the fluorescent signal perpendicular to the beam during an energy scan performed at beamline ID23-EH1, ESRF.

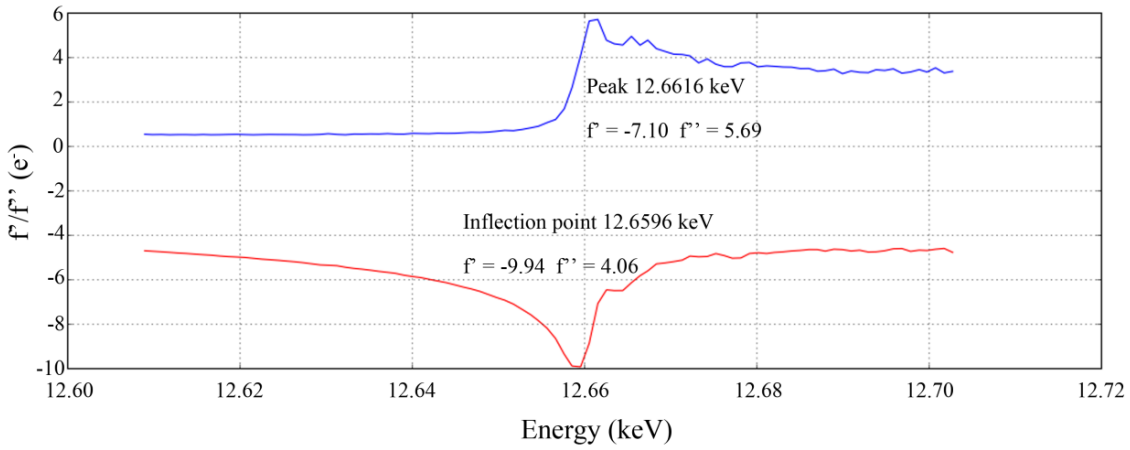


Figure 3.3 *Selenium K-edge fluorescence scan.* Anomalous and dispersive Se scattering factors across the *K* edge derived from fluorescence scan at beamline ID23-EH1, ESRF.

Eight out of the ten expected selenium sites in SeMet-SOUL derivative were detected by visual inspection of the anomalous Patterson maps, and were confirmed by SOLVE's (Automated crystallographic structure solution for MIR, SAD, and MAD software) automated Patterson search [115]. The positions were refined, and phases were calculated using AutoSol [116] implemented in PHENIX [117]. The list of selenium refined positions and occupancies is presented in table 3.1, section *3.3.3.* The electron density map at 3.50 Å calculated after density modification by solvent flattening in space group $P6_222$, displayed extensive well defined regions revealing continuous stretches of main chain density, with unambiguous density for carbonyl oxygen atoms and side chains. A model comprising 2 chains, A and B, with 184 and 180 amino acids respectively, was built from the initial map with program RESOLVE. Program COOT [118] was used for visual inspection of calculated $2mF_o$-DF_c and F_o-F_c electron density maps and model adjustments. Real-space refinement of selected zones of model was performed taking into account geometry terms (refinement weight matrix was set to 20.0) and Ramachandran restraints. Refinement was carried out using Refmac5 [119] and a weighting term of 0.01, relating reflection data and geometry restraints, and bulk solvent scaling. The refinement was performed with TLS [120] parameterization using Hendrickson-Lattman coefficients from SOLVE/RESOLVE [121]. In all rounds of refinement, Non-Crystallographic Symmetry (NCS) between chain A and B was taken into account, choosing medium main-chain and loose side-chain restraints. The statistics for structure refinement are shown in table 3.2, section *3.3.3.*

After building a complete main chain, the NCS assembly was chosen based on the results given by the PISA server (Protein interfaces, surfaces and assemblies service PISA at European Bioinformatics Institute (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html) [122]. The analysis of the protein interfaces did not indicate a stable quaternary structure between chains A and B, with a surface area of approximately 17500 $Å^2$ and a buried area of 960 $Å^2$. The calculated ΔG

(int) and ΔG (diss) were -9.6 kcal.mol^{-1} and -1.1 kcal.mol^{-1}, respectively. Although this NCS assembly may not be stable in solution, it was chosen as the asymmetric unit for model building and refinement. The model quality was monitored using the validation tools implemented in COOT [118] and the web server Molprobity [123]. Validation procedures reported 8 Ramachandran outliers (chain A: S181, V183; chain B: A31, P35, S97, G98, P182, V183), corresponding to 1.9 % of the total number of residues. Analysing hSOUL structure, these residues are found in more flexible regions of the protein.

The seleno-methionine derivative model of hSOUL (Se-SAD-hSOUL) was also refined against diffraction data collected at the ESRF, ID14-EH2, from the hSOUL (histidine tag) crystal grown in different conditions (previously described) – table 3.2 [104]. A clear molecular replacement solution was found by using the program PHASER [71], in space group $P6_422$, with a Z score of 29.16 and a refined LL gain of 721.83. After density modification using DM, refinement was performed using REFMAC [119], and a weighting term of 0.005, relating reflection data and geometry restraints, and bulk solvent scaling. A final round of refinement including TLS [120] parameterization using Hendrickson-Lattman coefficients was performed. Statistics for data processing and refinement are shown in table 3.2.

3.3 RESULTS AND DISCUSSION

3.3.1. ICP-AES ANALYSIS

The pure Se-Methionine containing protein was analyzed by Inductively Coupled Plasma - Atomic Emission Spectrometry (ICP-AES) to confirm the presence of selenium. The concentration of selenium was found to be 11.28 mg/ml that corresponds to the expected 5 selenium atoms per molecule of protein.

3.3.2. CRYSTALLIZATION AND DATA COLLECTION

The best hSOUL (histidine tag) crystals (0.2 × 0.2 × 0.2 mm^3, figure 3.4) were obtained within 4 days in drops consisting of 1µl of protein solution (15 mg/ml in 10 mM Tris-HCl, pH 8.0) and 1µl of reservoir solution (2 M ammonium sulphate, 0.1 M MES pH 6.5), at 293 K with 700 µl of precipitant solution in the reservoir. hSOUL crystals were also obtained with 0.2 M ammonium sulphate, 0.1 M cacodylate pH 6.5 and 30% polyethylene glycol 8000. Crystals could also be grown in the presence of 2 M ammonium sulphate buffered with 0.1 M Tris-HCl pH 8.5, although, in these conditions, but crystals were multiple.

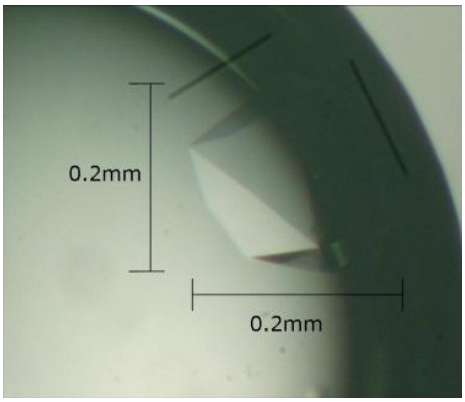


Figure 3.4 *hSOUL protein crystal*. hSOUL (histidine tag) protein diffracting crystal belonging to the space group $P6_422$ and cell unit $a = b = 144.7$ Å, $c = 60.2$ Å, grown in 2M ammonium sulphate, 0.1M MES 6.5.

Several data sets were collected using single crystals grown under similar conditions, in the search for data with the highest diffraction quality.

All the crystallization attempts to crystallize hSOUL in the presence of ionic liquids resulted in no crystals or crystals that did not diffract any better. In addition, the few attempts to crystallize hSOUL with sodium phosphomolybdate hydrate and sodium phosphotunsgate hydrate were unsuccessful.

In order to obtain independent phases, the Single Wavelength Anomalous Dispersion (SAD) method was used on the Se-Met containing protein. Crystals of hSOUL seleno-methionine derivative were grown (within 2 days) using the hanging-drop vapor-diffusion method with an equal volume (1 μl) of protein (15 mg/ml) and reservoir solution (1.8 M Na/K phosphate buffer pH 5.6) at 277 K.
Crystals (figure 3.5) were harvested in a solution containing 2.2 M Na/K phosphate, buffered to pH 5.6. Glycerol (25 % v/v) was added to this mother liquor as a cryoprotectant and crystals were flash-frozen in this solution.

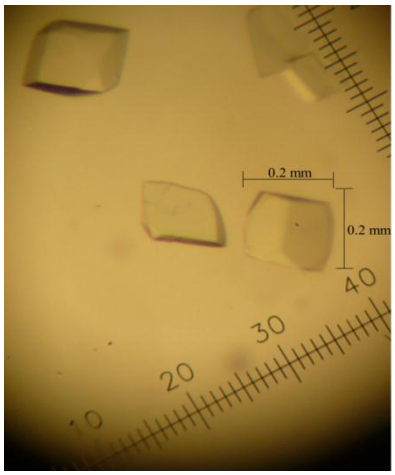


Figure 3.5 *Se-Met protein crystal.* Protein diffracting crystal belonging to the space group $P6_222$ and cell unit $a = b = 146.4$ Å, $c = 133.0$ Å, grown in 1.8M Na/K phosphate buffer pH 5.6.

A single-wavelength SAD experiment was conducted on beamline ID23-EH1 at the European Synchrotron Radiation Facility (ESRF) in Grenoble, France, using an ADSC Quantum-4 CCD detector which allowed to determine hSOUL protein 3D structure. Being a tunable beam line, the initially planned approach was to perform a MAD experiment. However, due to radiation damage it was only possible to collect images at the Se peak wavelength (0.9793 Å).

3.3.3. Crystal structure of hSOUL

From the initial hSOUL (histidine tag) crystals a preliminary model of the protein was determined by molecular replacement. This model was very incomplete; segments G24-Y43, W48-M56, S60-N70, Y72-G75, N77-M85, A87-V93, G98-S101, S103-F128, E130-E134, T136-Q154, T157-E164, Y175-V183 and L186-V192 could be traced but no refinement was performed.

This preliminary model of hSOUL resulted from data collected to 3.5 Å resolution, but was severily affected by model bias that prevented structure completion. Independent phases were clearly needed and afterwards obtained by preparing and crystallizing a seleno-methionine derivative of hSOUL (intein tag). The Se-Met derivative crystallized in space group $P6_222$ and gave a final structure to 3.50 Å resolution. The selenium refined positions and occupancies are indicated in table 3.1. Although comparable to the native X-ray diffraction data (the selenium derivative did not improve the resolution) the estimation of independent phases was crucial, as the electron density map was of higher quality and more informative. RESOLVE was used for automated model building and refinement of the polypeptide chain in the electron density map.

Table 3.1 *Se atoms coordinates, occupancies, figure-of-merit (FOM), f' and f" values.*

Atom	Coordinates			Occupancy
	x	y	z	
Se1	0.4826	0.6058	0.0393	0.89
Se2	0.3329	0.3601	0.0117	0.95
Se3	0.3116	0.9039	0.1338	1.00
Se4	0.1979	0.7784	0.0791	0.92
Se5	0.2172	0.6600	0.0989	0.59
Se6	0.4746	0.4701	0.0576	0.86
Se7	0.1031	0.5801	0.0798	0.69
Se8	0.1982	0.4686	0.9266	0.41
FOM acentric/FOM centric			0.320/0.099	
f'			- 7.10	
f"			5.73	

In figure 3.6, the ribbon structure of hSOUL (monomer A) is represented with special emphasis to the four selenium atoms in chain A. For structure determination, the eight selenium atoms were used.

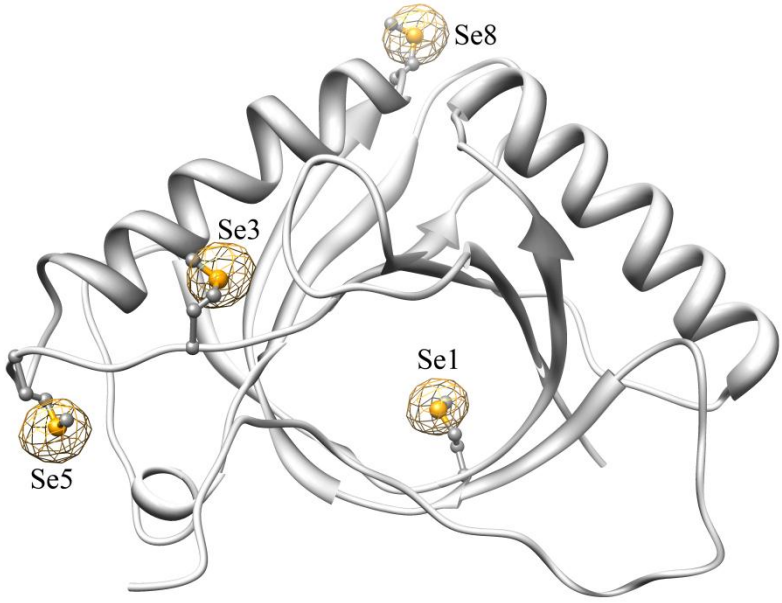


Figure 3.6 *Ribbon representation of human SOUL structure (chain A) superimposed on the anomalous difference map.* Four Se-methionine residues are represented as ball-and-stick models, with selenium atoms shown in yellow. Superposed on the structure is the anomalous difference Fourier map, confirming the selenium positions and corresponding seleno-methionine residues. The anomalous difference Fourier map, calculated from the anomalous contribution of selenium atoms at wavelength of 0.9793 Å, is shown in yellow and contoured at 2 σ. Picture was produced with program CHIMERA [124].

However, since the previously reported diffraction data from $P6_422$ native crystals had better data collection statistics (Rmerge 15.2 (75.0) % vs 23.9 (97.8) %, cf table 3.2) we decided to use the Se-SAD model for MR and refinement using the previously obtained data. The new structure presented here is designated by MR-hSOUL (from native crystals).

The N- and C- terminal in the MR-hSOUL model lack 18 (plus the 25 residues belonging to the His-tag fused to the protein) and 8 amino acids residues, respectively. In addition, residues 30-35, 98-99 and 181-183 are missing in the model. MR-hSOUL model was refined to R_{work} of 27.1 % (R_{free} = 30.6 %). Full refinement statistics are presented in table 3.2.

Table 3.2 *Data collection and refinement statistics.* Data collection and structure refinement statistics (values in parentheses are for the lowest/highest resolution shells).

Data collection		
Crystals	Se-SAD-hSOUL hSOUL (intein tag)	MR-hSOUL hSOUL (histidine tag)
Beamline	ESRF, ID23-EH1 (SAD)	ESRF, ID14-EH2 (MR)
Cell parameters *a* (Å)=*b* (Å) *c* (Å)	 146.4 133.0	 144.7 60.2
Space Group	$P6_222$ (2 mol/a.u.; 72 % solvent)	$P6_422$ (1 mol/a.u.; 66 % solvent)
Wavelength, Å	0.9793	0.931
Resolution of data (outer shell), Å	50.00 - 3.50 (3.69 - 3.50)	47.35 – 3.50 (3.69 – 3.50)
R_{pim} (outer shell), % [a]	6.5 (27.6)	2.7 (13.0)
R_{merge} (outer shell), % [b]	23.9 (97.8)	15.2 (75.0)
Mean I/σ(I) (outer shell)	11.0 (4.1)	24.6 (6.2)
Total number of observations (outer shell)	150665 (22286)	154889 (22813)
Number of unique observations (outer shell)	11094 (1575)	5025 (709)
Completeness (outer shell), %	100.0 (100.0)	100.0 (100.0)
Anomalous Completeness (outer shell), %	100.0 (100.0)	-
Redundancy (outer shell)	13.6 (14.1)	30.8 (32.2)
Anomalous Redundancy (outer shell)	7.3 (7.4)	-
FOM for 8 Se-sites (before / after solvent flattening) [c]	0.32/0.70	-/-
Structure refinement		
No. of protein atoms	2727	1264
Resolution used in refinement, Å	50.00 – 3.50	47.35 – 3.50
No. of reflections	10509	4541
No. of R_{free} reflections	1082	484
R_{work} / R_{free} (%) [d]	23.4/ 27.0	27.1 / 30.6
rms bond lengths (Å)	0.007	0.005
rms bond angles (degrees)	1.031	0.846
rms deviation chiral volume (Å^3)	0.068	0.051
Avg B factors (Å^2) Molecule A main-chain atoms Molecule A side-chain atoms Molecule B main-chain atoms Molecule B side-chain atoms	 45.7 44.9 50.6 49.5	 46.6 46.7 - -

[a], $R_{p.i.m.} = \left(\sum_{hkl}\sqrt{\frac{n}{n-1}}\sum_{j=1}^{n}\left|I_{hkl,j} - \langle I_{hkl}\rangle\right|\right) \Big/ \left(\sum_{hkl}\sum_{j} I_{hkl,j}\right)$ where $\langle I_{hkl}\rangle$ is the average of symmetry-related observations of a unique reflection.

[b], $R_{sym} = \left(\sum_{hkl}\sum_{j}\left|I_{hkl,j} - \langle I_{hkl}\rangle\right|\right) \Big/ \left(\sum_{hkl}\sum_{j} I_{hkl,j}\right)$ where $\langle I_{hkl}\rangle$ is the average of symmetry-related observations of a unique reflection.

[C], Figure-of-merit as computed by SOLVE/RESOLVE.

[d], $R_{work} = \left(\sum_{hkl}\left|F_{hkl}^{obs} - F_{hkl}^{calc}\right|\right) \Big/ \left(\sum_{hkl} F_{hkl}^{obs}\right) \times 100$ where F^{calc} and F^{obs} are the calculated and observed structure factor amplitudes, respectively. R_{free} is calculated for a randomly chosen 10% of the reflections.

The 3D coordinates of both hSOUL structures, Se-SAD-hSOUL and MR-hSOUL, have been deposited in the PDB and can be accessed under the codes 4ayz and 4b0y, respectively. Both structures are essentially identical within the 3.5 Å resolution limits with an rmsd of 0.5 Å for 164 Cα atoms. The structural analysis and comparison with the previously solved NMR structures of murine p22HBP that follows is carried out using the Se-SAD-hSOUL model.

Although the main-chain atoms are fully defined in the density, several solvent-exposed side-chains exhibit high B-factors and poorly defined density, when compared to the core residues. Despite the significantly high B-factors (table 3.2), the structure is well ordered in the electron density maps and 89% of the total amino-acid residues could be built. The crystal packing reveals extended zones of disordered solvent, while the asymmetric unit (two monomers A and B related by a two-fold axis) exhibits a total area of 12180 $Å^2$ of solvent accessibility.
As found for the murine p22HBP NMR structure, the hSOUL crystal structure is built-up of a predominantly hydrophobic central core flanked by two α-helices (Trp58 – Gln74 and Ala148 – Asp165). The central core consists of a distorted β-barrel, composed of an eight-stranded antiparallel β-sheet (Glu39 - Tyr43, Ala46 – Ser55, Val89 – Glu94, Ser103 – Tyr110, Val127 – Arg132, Met135 – Phe142, Tyr174 – Gly178 and Asn190 – Ile195) - figure 3.7. An additional β-sheet (Trp25 – Lys 26) and α-helix (Ser113 – Glu116) are observed.

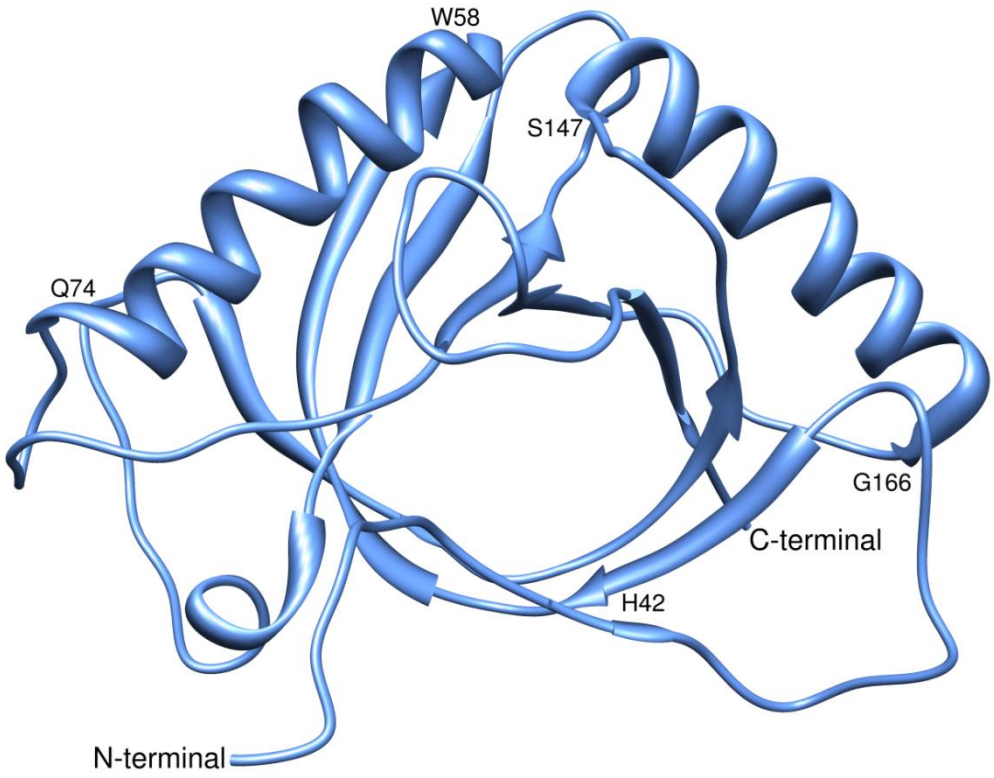


Figure 3.7 *X-ray structure of Se-SAD-hSOUL*. The central core of the protein consists of an eight-stranded antiparallel β-sheet surrounded by two α-helices.

A pseudo two-fold symmetry axis is clearly visible along the hSOUL monomer, which is suggested to result from a gene duplication phenomenon. This defines the two sub-domains in hSOUL, in a similar and even clearer way to murine p22HBP: β_2-β_3-α_1-β_4-β_5 (comprising residues 39 to 110) equivalent to sub-domain β_6-β_7-α_3-β_8-β_9 (comprising residues 127 to 195). In the case of hSOUL, superposition of these sub-domains generates an rmsd of 2.9 Å for 70 alpha-carbon atoms, although their sequence identity is only 5.3 % (for murine p22HBP the rmsd is 3.2 Å for 72 α-carbons with 8% identity) – figure 3.8.

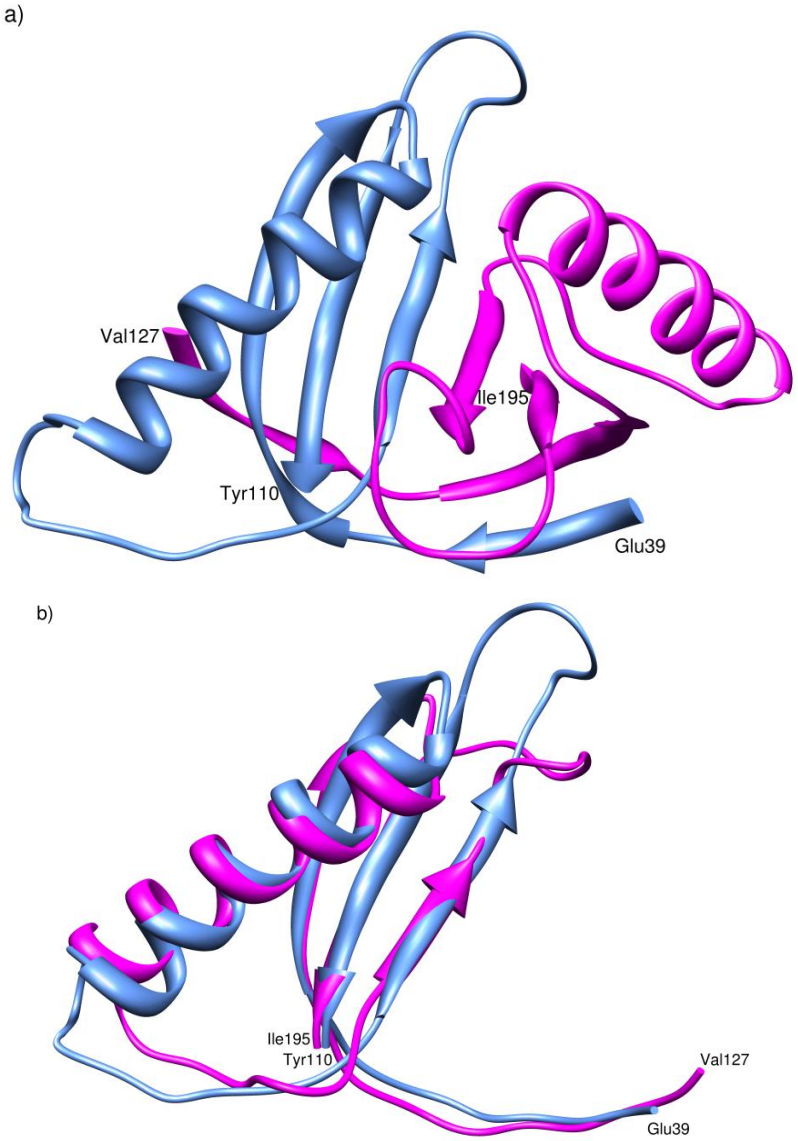


Figure 3.8 *Ribbon representations of hSOUL sub-domains.* a) hSOUL representation with the 2 sub domains identified: Glu39-Tyr110 (blue) and Val127-Ile195 (magenta); b) Superposition of the 2 sub-domains with β-β-α-β-β motif, in result of gene duplication.

The previously used His-tagged recombinant murine SOUL was purified as a dimer in the absence of heme, which upon heme-binding displayed an hexameric structure [20]. Although there is no biochemical evidence that hSOUL may form a functional dimer, PISA [122] analysis

of possible interfaces among neighboring molecules, has revealed that the largest interface area (approximately 830 $Å^2$) occurs between one of the monomers (chain A) and a 2-fold symmetry related molecule of the same chain along the *c* edge of the unit cell, with 24 interfacing residues. There are 14 inter residue hydrogen bonds between the two symmetry-related molecules. Although this may suggest a possible functional dimer, there is nevertheless no strong structural evidence of dimerization in hSOUL. NMR spectra collected before and after hemin addition also did not show any sign of protein oligomerisation in phosphate buffer, at pH 8.0; the resonances are sharp and remain so in all spectra – figures 4.7 and 4.8, section *4.3.3*. The monomeric structure of hSOUL in solution was confirmed by running samples before and after incubation with hemin on a gel filtration column with 100 mM phosphate buffer pH 8.0.

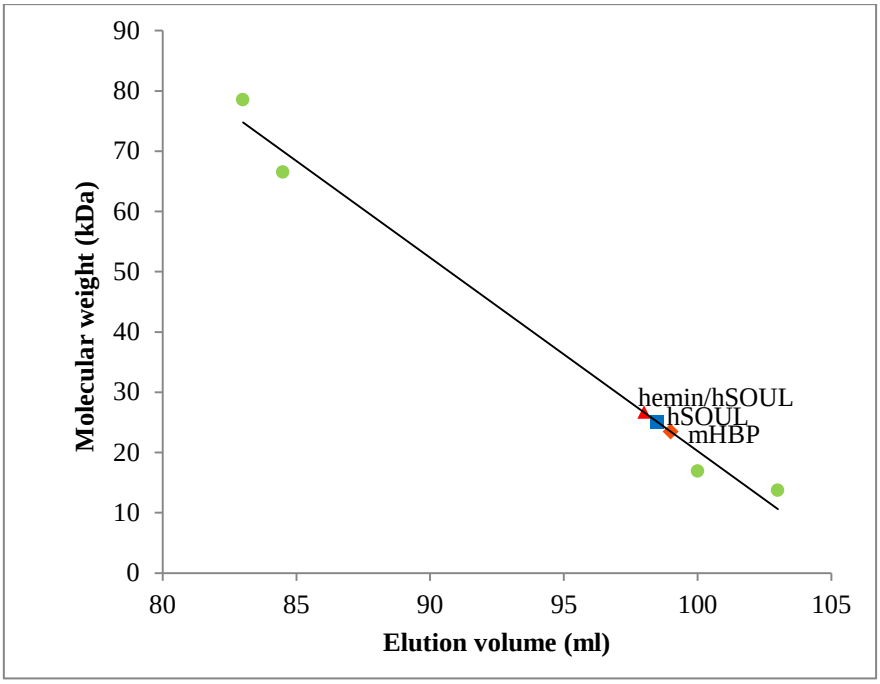


Figure 3.9 *Molecular weight of various proteins (green circles; MW = 78.5, 66.5, 16.9 and 13.7 kDa) as a function of the elution volume of gel filtration in order to determine the oligomerization state of apo-hSOUL (blue square) and hemin/hSOUL (red triangle).* In addition, *apo* murine p22HBP (orange diamond) was used as a control protein. hSOUL (25.1 kDa), hemin/hSOUL (26.7 kDa) and murine p22HBP (23.4 kDa) molecular weights were estimated according to the elution volume on the gel filtration, showing that the three proteins are eluted as monomers. Experiments were performed in 100 mM phosphate buffer, pH=8, on a Superdex 75-10/300 GL column (GE Healthcare, pre-packed coupled to a FPLC system).

When Tris/HCl buffer was used as elution buffer, both *apo* SOUL and SOUL-hemin were eluted as tetramers. Previous studies performed in similar conditions indicate that the protein is a dimer in the *apo* form, becoming hexameric upon heme binding [20]. All the studies presented in this thesis indicate the protein to be a monomer in both *apo* and heme-binding form, which is also true for the other member of SOUL/HBP family of proteins, p22HBP. These results may indicate a possible interaction between Tris ions (at this concentration) and hSOUL protein, promoting the self-association of the protein.

3.3.4. Structural similarity of hSOUL to murine p22HBP

Sequence alignment of hSOUL and murine p22HBP (figure 1.3), performed with ClustalW [23], indicates a 27% identity between the two proteins, and that His42 is not conserved in murine p22HBP, where the corresponding amino acid residue is an alanine. Structural alignment using the DALI server [125] indicates an rms deviation of 2.8 Å (for 184 Cα atoms) between hSOUL and the average structure of murine p22HBP determined by NMR (PDB accession code 2GOV). 3D superposition of the two structures shows that, near His42 (the putative heme binding site, vide infra), the most significant structural difference between the two proteins occurs in a loop that comprises 8 residues from Pro28 to Pro35 (figure 3.10). Apart from this loop and loop Tyr179-Arg188, the crystal structure of hSOUL shows relatively good conservation with the global structure described for murine p22HBP.

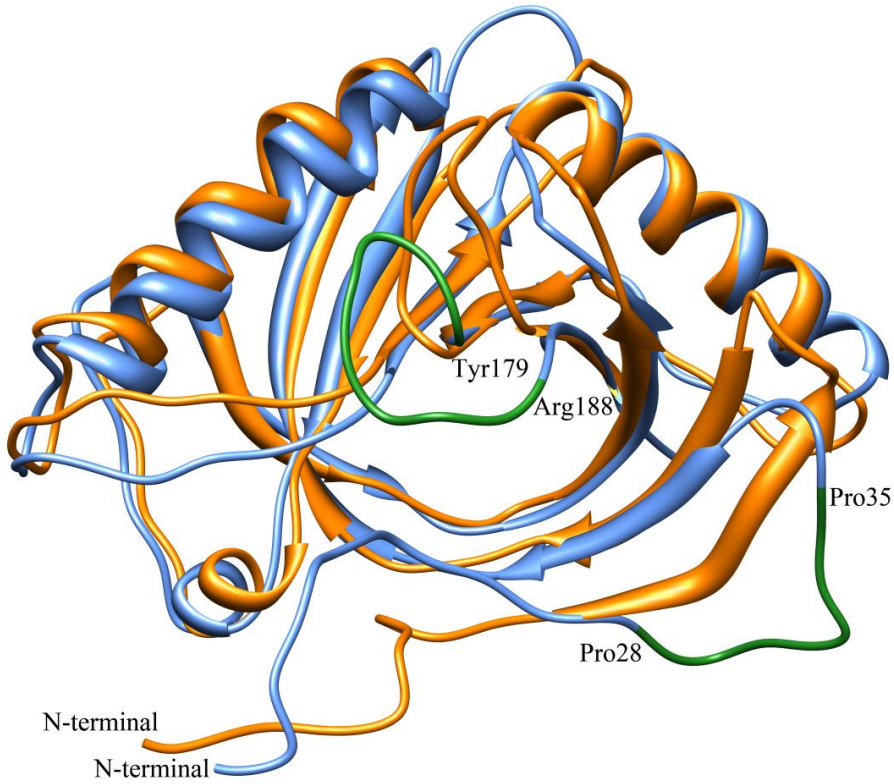


Figure 3.10 *Overlay of hSOUL X-ray structure and murine p22HBP solution structure.* In orange, solution structure of murine p22HBP and in blue hSOUL structure (monomer A). The two loops that show more significant differences regarding mHBP structure are highlighted in forest green.

Chemical shift mapping has revealed that murine p22HBP binds one heme molecule per subunit with no specific axial ligand coordination of the Fe(III) heme. The tetrapyrrole binding site was defined by a hydrophobic cleft with residues from helix α_1 and the β8-β9 loop [18]. In the X-ray structure of hSOUL, this helix extends from residue Trp58 to Gln74 and shows a good structural conservation in murine p22HBP. This hydrophobic patch could clearly be identified

in the electrostatic surface calculated for the murine p22HBP structure – figure 3.11 (same orientation as hSOUL, figure 3.7, and left panel, figure 3.12) [18].

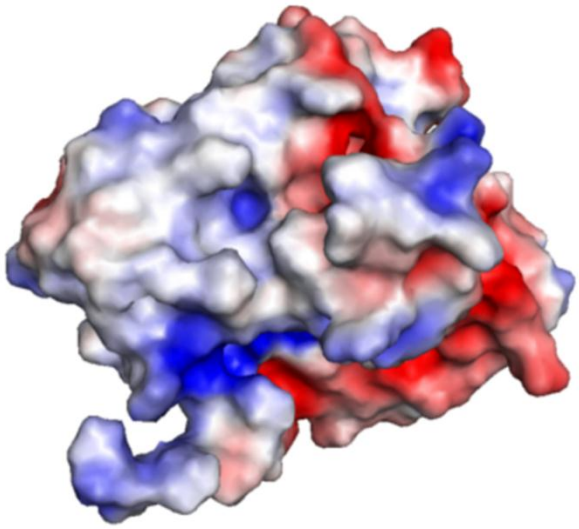


Figure 3.11 *Electrostatic surface potential for murine p22HBP (calculated using APBS [126] at pH 8.0).*

As shown in figure 3.12 no such patch exists in the hSOUL surface. Also Met59 and Met63 residues in murine p22HBP (found close to the bound tetrapyrrole and both experiencing large chemical shifts deviations when titrated with PPIX) are replaced by Phe66 and Asn70 in hSOUL when comparing the two proteins sequences.

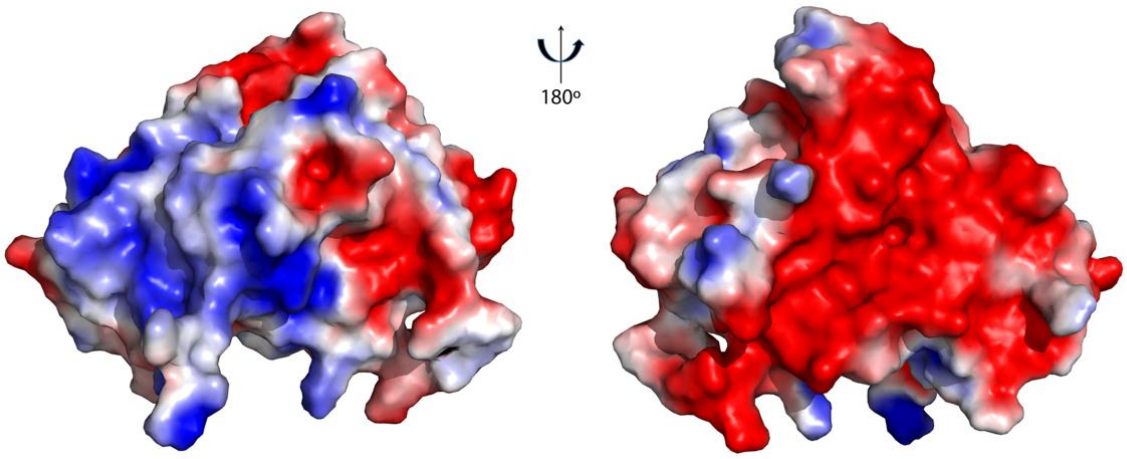


Figure 3.12 *Electrostatic surface potential (calculated using APBS [126] at pH 8.0) for the hSOUL monomer structure, viewed in 2 perpendicular orientations.* A significantly more negative surface is visible on the right side representation, which is rotated 180º with respect to the orientation in figure 3.7, which is highly solvent exposed when the crystal packing is considered.

A search for proteins with structural similarity using the VAST server [127] found 10 structures, all from bacterial sources apart from mHBP. The VAST structural alignment used all the β-sheet elements in hSOUL and the α_3 helix (containing part of the sequence defining the BH3

domain). Five of the structures were generated by structural genomics studies and they have no associated literature. The remaining 3 structures (SbmC protein from *E. coli* [128], the C-terminal domain of the *E. coli* transcription factor Rob [129], and the C-terminal multidrug-binding domain of transcription activator BmrR from *B. subtilis* [41, 42]), excluding 2GOV and 2HVA, had already been detected and discussed when the murine p22HBP structure was solved by NMR [18].

3.3.5. THE BH3 DOMAIN IN HSOUL

According to the work of Szigeti *et al*, SOUL has been proposed to be a novel member of the BH3-only domain-containing protein family which under oxidative stress conditions, can facilitate mitochondrial membrane permeabilization and necrotic cell death. The hSOUL BH3 domain, was identified by sequence comparisons as the Leu158-Asp172 fragment (figure 3.13), and, when 9 amino acids of this domain were deleted, the cellular sensitivity to H_2O_2 was no longer potentiated by SOUL [32].

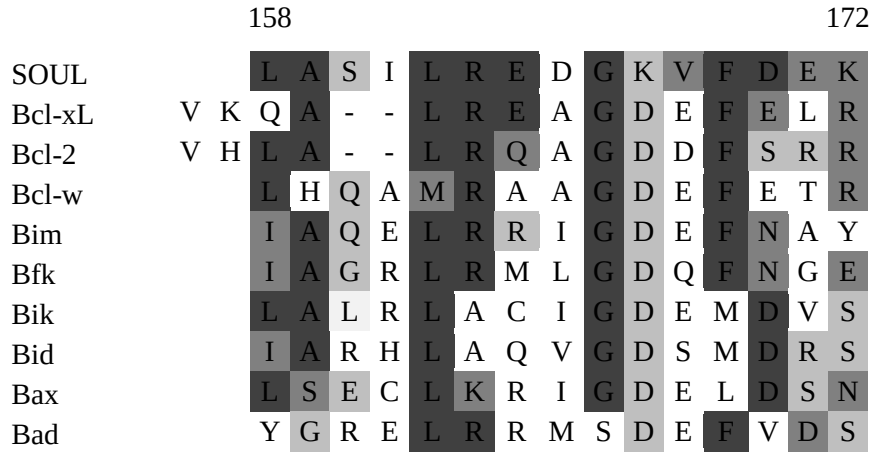


Figure 3.13 *Comparison of the BH3 domain of hSOUL protein with members of the Bcl-2 family of proteins.* Black-shaded amino acids are identical, grey-shaded amino acids are conserved substitutions, and light gray-shaded amino acids are semiconserved substitutions (adapted from [32]).

The BH3 motif in the Prosite [59] database (PS01259) has 15 amino acids with an hydrophobic residue at position 1 and conserved residues Leu, Gly and Asp at positions 5, 9 and 10, respectively. Position 15, not so well conserved, has Arg, Asn or Ser as possible residues. The sequence of SOUL displays Lys in both positions 10 and 15 of this consensus sequence having also an hydrophobic amino acid in position 1 – leucine, leucine in position 5 and glycine in position 9. All BH3 domain containing proteins with structures in the PDB (ca. 70 in the Prosite entry, not including models) are made up of almost exclusively α-helices. In the hSOUL crystal structure described here, the Leu158-Lys172 BH3 domain corresponds to a well-defined region comprising part of the α2 helix and part of the connecting loop to β7 (figure 3.14). In both

chains A and B this loop is solvent exposed. So, approximately half of the BH3 domain in hSOUL has no clear secondary structure.

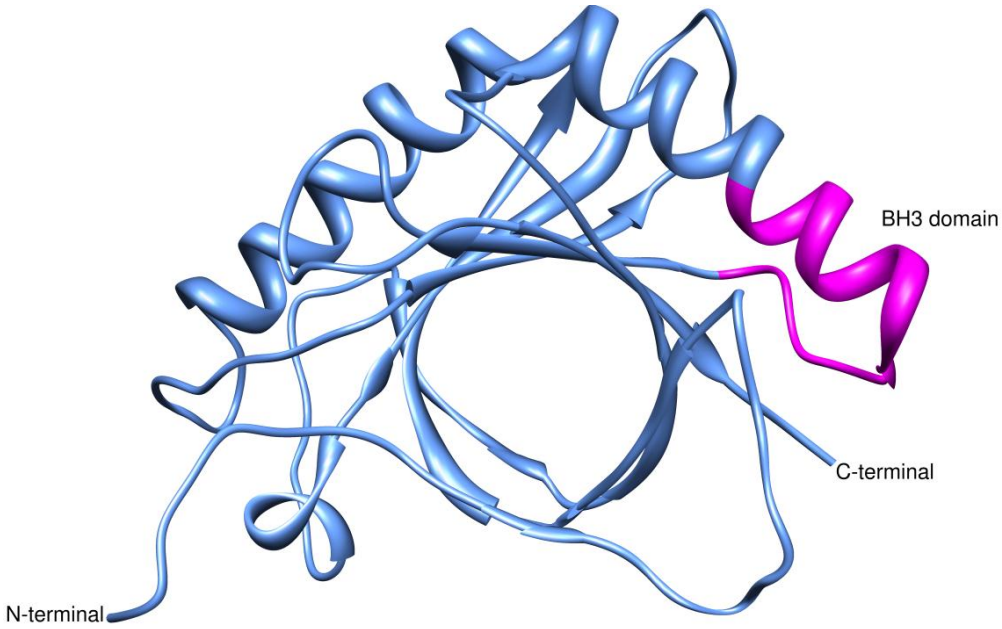


Figure 3.14 *The BH3 domain on hSOUL.* In magenta the BH3 domain consisting in part of helix α2 and following loop.

There are also a number of protein structures in the PDB involving short BH3 domains bound to, in the majority of cases, Bcl-2, Bcl-x_L and Mcl-1. In figure 3.15 is shown, as an example, Bax BH3 domain peptide bound to Bcl-2.

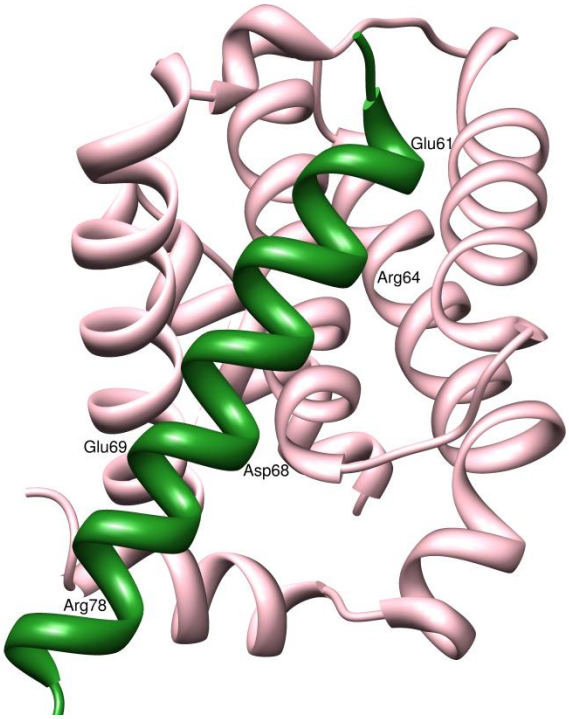


Figure 3.15 *Example of a BH3 domain bound to pro-survival proteins of the Bcl-2 family of proteins.* a) Bax BH3 peptide (chain C, forest green) bound to Bcl-2 (chain A, pink) through residues Glu61, Arg64, Asp68, Glu69 and Arg78 of the BH3 peptide [130].

Bcl-2, Bcl-x_L and Mcl-1 are pro-survival proteins of the Bcl-2 family that contain all (1 to 4) BH domains. In Bcl-2 regulated apoptosis, the proteins Bak and Bax are required to associate and induce mitochondrial outer membrane permeabilization. The BH3-only proteins, that include Bim, Puma, Noxa, Bik, Bmf, Bad, Hrk and Bid, act upstream from Bax/Bak oligomerisation by activating it or by inactivating the pro-survival proteins [131]. The structures in the PDB containing short BH3 domains (varying from 16 to 35 amino acids in length) bound to Bcl-2 or Mcl-1 are all 100 % α-helical. However, BH3 domain folding may occur upon binding since isolated short peptides are mainly unstructured in solution [132]. If the sequence alignments having identified the putative BH3 domain of SOUL are to be believed, it appears that the BH3 domain in hSOUL is the first to fold as an α + turn structure. With these structural features, it is very unlikely that hSOUL can function as a BH3-only protein, i.e. can bind to Bcl-2 and similar proteins, since the interaction site of the latter is always an elongated hydrophobic groove [133, 134]. The fold of the putative BH3-domain does not appear to fill the geometric requirements to bind to these grooves. Furthermore, considering the compact core of the hSOUL structure, it is also very unlikely that significant parts of the molecule exhibit conformational dynamics allowing it to transform the α + turn arrangement of residues 158-172 into a fully extended α-helix upon binding to a putative protein partner. As there appears to be a strong correlation between the presence of the BH3 domain in hSOUL and enhanced susceptibility to oxidant-triggered death, and the fact that overexpression of Bcl-2 and Bcl-x_L counteracts this effect, the actual mechanism of hSOUL action on mitochondria needs to be clarified. It should be noted that the folding of the BH3 domain-deleted SOUL has not been compared to that of full length SOUL [32].

CHAPTER 4

HEME-BINDING INTERACTIONS STUDIES ON HUMAN SOUL

CONTENTS

4.1. INTRODUCTION

In vitro studies by Sato *et al* using recombinant N-terminally hexa-His-tagged SOUL from mouse, suggested His42 as the axial ligand of hemin iron, since upon mutation to Ala, the spectroscopic (UV-visible and Raman) evidence for heme binding was lost. In addition, gel filtration of hSOUL indicates that the protein exists as a dimer and becomes hexameric upon heme binding [20].
In the monomeric crystal structure of hSOUL (shown in *chapter 3*), the side-chain of His42 residue, although solvent-exposed, is very well defined in the electron density map (figure 4.8, *section 4.3.3.*).

NMR spectroscopy is very powerful to study protein-ligand interactions. Chemical shift mapping can be used to follow protein changes upon ligand addition. hSOUL interaction with PPIX and hemin was studied by NMR. Several ^{15}N labeled hSOUL samples were prepared with increasing amounts of PPIX and hemin. 2D ^{1}H, ^{15}N-HSQC and ^{1}H, ^{15}N-TROSY-HSQC spectra were collected for each sample and chemical shift changes could be monitored upon the progressive addition of either PPIX or hemin. For the analysis of these data, it is necessary to previously carry out the backbone assignment of the protein in order to know which residues may be involved in ligand binding to the protein. For large proteins, such as human SOUL, this is a difficult task as described in *chapter 1.5.2*.

Tryptophan, tyrosine and phenylalanine are three aromatic amino acid residues that contribute to the intrinsic fluorescence of a protein. Due to stronger fluorescence and higher quantum yield of tryptophan residues, the excitation of these residues is responsible for the majority of the fluorescence emission from a protein. In addition, intrinsic fluorescence quenching can be observed due to the binding of small molecules to the protein in the vicinity of the fluorophore. This property can be used to measure the binding affinity by which a ligand binds to a protein. Possible hSOUL interaction with either PPIX or hemin was therefore studied by Fluorescence Quenching.

Heme has a characteristic absorptium spectrum, presenting a Soret band resulting from the π-π^* transition of the porphyrin ring [135]. This feature can be used to study possible binding of heme to peptides/proteins. The binding of peptides or/and proteins to heme may distort the porphyrin ring and increase the energy gap of the π-π^* transition. Heme binding to hSOUL was also studied by performing a UV/visible titration in which several hSOUL additions to hemin would cause a change in the Soret band change if interaction occurred.

4.2. MATERIAL AND METHODS

4.2.1. SAMPLE PREPARATION AND NMR DATA ACQUISITION AND PROCESSING

NMR spectra were recorded on a 600 MHz Bruker avance III spectrometer (Bruker, Wissembourg, France) equipped with a 5 mm inverse detection triple resonance z-gradient probe, at 293 K and 298 K. 3mm Thin Wall, 7" Lgth (Wilmad LabGlass) tubes were used in all NMR experiments.

hSOUL (histidine tag and intein tag) experiments for backbone assignment were performed with double (^{13}C and ^{15}N) (triple labeled protein (^{2}H, ^{13}C and ^{15}N) at 0.7 – 1.0 mM concentration, in 50 mM phosphate buffer pH 8.0, 10 % D_2O (CortecNet) – table 4.1.

Table 4.1 *hSOUL NMR experiments.* hSOUL NMR spectra parameters, including FID size, number of scans, spectral width and corresponding pulse program, acquired for backbone assignment.

Experiment	FID size	Scans	Spectral width	Pulse program
^{1}H, ^{15}N-HSQC	1024 × 512	8	9615 Hz (^{1}H) 2311 Hz (^{15}N)	hsqcfpf33gpphwg [136-139]
^{1}H,^{15}N-TROSY-HSQC	1024 × 256	8	9615 Hz (^{1}H) 2311 Hz (^{15}N)	trosyf3gpphsi19.2 [140-145]
3D TROSY-HNCO (trHNCO)	2048 × 40 × 128	16	10776 Hz (^{1}H) 3320 Hz (^{13}C) 2189 Hz (^{15}N)	trhncogp2h3d [146]
TROSY-HNCA (trHNCA)	2048 × 40 × 128	16	9615 Hz (^{1}H) 4829 Hz (^{13}C) 2311 Hz (^{15}N)	trhncagp3d2 [147, 148]
TROSY-HN(CO)CA (trHN(CO)CA)	2048 × 40 × 128	16	9615 Hz (^{1}H) 4829 Hz (^{13}C) 2222 Hz (^{15}N)	trhncocagp2h3d [149, 150]
TROSY-HNCACO (trHNCACO)	2048 × 40 × 128	32	10775 Hz (^{1}H) 3409 Hz (^{13}C) 2189 Hz (^{15}N)	trhncacogp3d [149]
TROSY-HNCACB (trHNCACB)	2048 × 40 × 128	32	9615 Hz (^{1}H) 11364 Hz (^{13}C) 2222 Hz (^{15}N)	trhncacbgp2h3d [147, 148]
TROSY-HN(CO)CACB (trHN(CO)CACB)	2048 × 40 × 128	32	9615 Hz (^{1}H) 11364 Hz (^{13}C) 2222 Hz (^{15}N)	trhncocacbgp2h3d [149, 150]

2D ^{1}H, 15N-HSQC, 2D ^{1}H, ^{15}N-TROSY-HSQC, 3D TROSY-HNCO, TROSY-HNCA, TROSY-HN(CO)CA, TROSY-HNCACO, TROSY-HNCACB and TROSY-HN(CO)CACB (with the same parameters as described in table 4.1) were initially acquired for double labeled (^{13}C and ^{15}N) hSOUL samples.

2D ^{1}H, ^{15}N-TROSY-HSQC and 3D TROSY-HNCO, TROSY-HNCA, TROSY-HN(CO)CA, TROSY-HNCACO, TROSY-HNCACB spectra with the same conditions as before were also acquired for triple labeled (^{2}H, ^{13}C and ^{15}N) hSOUL (intein tag) sample at 298 K.

Heteronuclear {^{1}H}-^{15}N-NOE were determined from the ratio of two experiments with and without saturation [151, 152] The heteronuclear experiments (trnoef3gpsi [153]) described above were collected with 2048 × 256 points, 32 scans and 10 s of relaxation delay. The central frequency for proton was set on the solvent signal (water) and for nitrogen was set on the center of the amide region. The spectral widths were 9615 Hz for ^{1}H and 2311 Hz for ^{15}N, at 293 K.

In order to understand hSOUL-heme interaction, **2D ^{1}H, ^{15}N-HSQC** (1024 × 256 points, 32 scans, spectral widths were 9615 Hz for ^{1}H and 2311 Hz for ^{15}N; pulse program: hsqcfpf3gpphwg [140, 153, 154]) and **2D ^{1}H, ^{15}N-TROSY-HSQC** (1024 × 256 points, 16 scans, spectral widths were 9615 Hz for ^{1}H and 2311 Hz for ^{15}N; pulse program: trosyf3gpphsi19.2 [140-145]) experiments were acquired on ^{15}N-labelled hSOUL samples in the presence of hemin or protoporphyrin IX, at 293 K. Porphyrins were added to final molar ratios of hSOUL:hemin/PPIX of 0.5, 1:1, 1:2 and 1:5.

To study the possible interaction of heme with His42, ^{1}H, ^{15}N-HSQC experiments (pulse program: hsqcetfpf3gpsi2 [147, 155-157]) were acquired to detect NHδ-His residues resonances. Spectra were collected with 1024 × 256 data points, 32 scans, with the proton carrier frequency set on the water resonance and the nitrogen carrier at 175 ppm. Spectral widths of 24 ppm for ^{1}H and 200 ppm for 15N were used, at 293 K.

The spectra were processed and analyzed using TopSpin® 2.1, NMRPipe [105], CARA 1.8.4.2 [158] and SPARKY software [159].

4.2.2. TETRAPYRROLE PREPARATION

Tetrapyrrole solutions (hemin and protoporphyrin IX – Fluka) used in all the experiments described in this thesis (including hemin solution used in co-crystallization trials for hSOUL, murine 22HBP and human p22HBP) were prepared as described here and according to Dias *et al* [18]. Porphyrins (approximately 1 mg) are initially dissolved in ammonium hydroxide 25% (Sigma-Aldrich) and followed by dilution in water. After addition of a surfactant, Tween 80 1.5 % (v/v), pH is adjusted to 8.0 with NaH_2PO_4 (Merck).

4.2.3. INTRINSIC TRYPTOPHAN FLUORESCENCE QUENCHING

All Fluorescence quenching measurements were carried out on a HORIBA Jobin Yvon FluoroMax-3 Spectrofluorometer using 1 cm path length 3 ml cuvettes at room temperature. Titrations were carried out using 2 ml of protein solution at a concentration of 100 nM, in 50 mM phosphate buffer at pH 8.0. This solution was prepared by dilution of a stock solution at 0.5 mM where the concentration was estimated using UV spectroscopy (ε_{280} of 33920 $M^{-1}.cm^{-1}$ was used). The tetrapyrrole stock solutions (mM) used for titrations were prepared according to Dias *et al* [18] and subsequently diluted to the required concentration (µM). Concentrations were estimated by measuring the absorbance at 280 nm and using molar absortivities calculated via serial dilutions (ε_{400} of 32257 $M^{-1}.cm^{-1}$ was used for hemin and 97071 $M^{-1}.cm^{-1}$ for PPIX).
Dissociation constants (K_d) were obtained by nonlinear fitting of the titration curves carried out using a model that accounts for ligand depletion or for non-specific binding [18, 160].

$$y = I_0 - (I_0 - I_{int})\frac{\sqrt{K_D - 3.[Protein].x - (K_D - [Protein].x)^2}}{2.[Protein]}$$

Equation 4.1 *K_d determination equation.* The protein emission maxima (y) are plotted as a function of porphyrin concentration (x). I_0 and I_{int} are the intensities at zero and saturating porphyrin concentrations, and [Protein] the concentration of the protein of interest.

The FQ titrations of hSOUL were run at the same time as murine p22HBP FQ titrations and performed according to Dias *et al* [18]. The results from murine p22HBP fitting curves were used as internal control/standard.

4.2.4. HSOUL/HEMIN UV-VISIBLE TITRATION

Hemin-hSOUL titrations were followed by UV/visible absorption spectroscopy. Measurements were performed on an Ultrospec 2100 pro spectrometer. Hemin was diluted with 50mM phosphate buffer pH 8.0 to give a final concentration of 12 µM, which was titrated with hSOUL (final concentration of 4.6 µM, 9.2 µM, 13.6 µM and 22 µM).

4.3. RESULTS AND DISCUSSION

4.3.1. ISOTOPIC LABELING

1D ^{1}H NMR spectrum was used to confirm protein deuteration. The decrease in the intensity of the peaks in the aliphatic region of the spectrum reflects the effectiveness of the deuteration – figure 4.1.

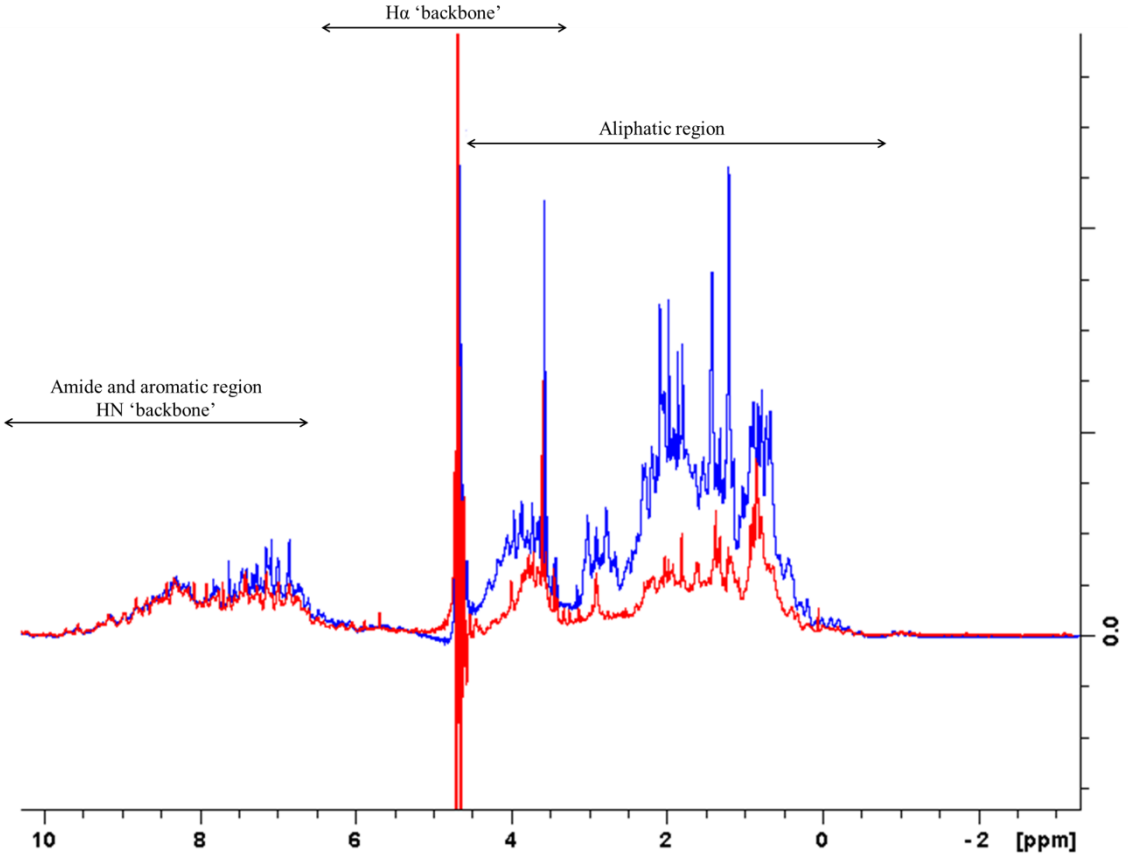


Figure 4.1 *1D ^{1}H NMR spectra of hSOUL.* Double labeled (^{13}C, ^{15}N) sample spectrum - blue and triple labeled (^{2}H, ^{13}C, ^{15}N) sample spectrum - red, acquired on a 600 MHz with cryoprobe, at 293 K, both samples on 50 mM phosphate buffer pH 8.0, 10% D_2O.

Protein overexpression in 99 % D_2O solution but with protonated carbon and nitrogen sources should give average enrichment levels of ~60-80 %.

Analysing the spectra in figure 4.1, it is possible to conclude that the protein is partially deuterated; deuteration led to a decrease of the nonexchangeable aliphatic and aromatic protons, a clear demonstration of fractional deuterium labeling.

4.3.2. PROTEIN BACKBONE ASSIGNMENT AND HETERO-NOE ANALYSIS

^{1}H, ^{15}N-TROSY-HSQC spectrum of hSOUL (histidine tag) showed good chemical shift dispersion however a number of intense overlap of peaks was observed in the region of approximately 8 ppm in the ^{1}H dimension. In order to investigate these intense peaks, ^{1}H, ^{15}H-HSQC spectra for 15N transverse relaxation rate (R_2) determination were acquired. Figure 4.2 shows the superposition of ^{1}H, ^{15}N-HSQC spectrum and ^{1}H, ^{15}N-HSQC spectrum with a relaxation period of 0.016 s of a 1.0 mM ^{15}N labeled hSOUL sample.

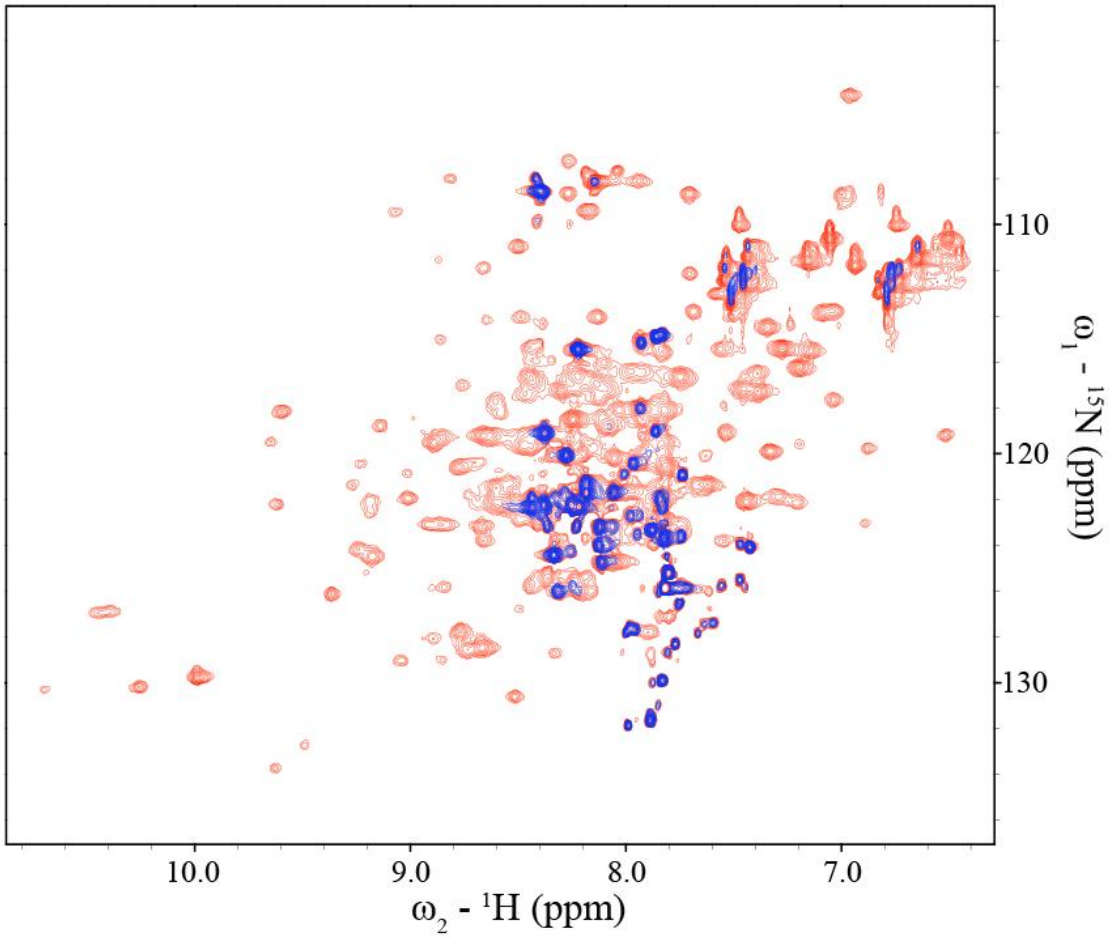


Figure 4.2 *^{15}N labeled hSOUL ^{1}H, ^{15}N-HSQC spectra*. Overlay of ^{1}H, ^{15}N-HSQC spectrum (red) with ^{1}H, ^{15}N-HSQC spectra with relaxation period of 0.016 s (blue) for a 1.0 mM ^{15}N labeled hSOUL sample in 50 mM phosphate buffer pH 8.0, 10 % D_2O, at 293 K.

Analysing the spectra on figure 4.2 it is evident the existence of a protein region displaying a very characteristic dynamic.

Chemical shift assignments for hSOUL (histidine tag) were obtained by standard methods. HN , CO, C_α and C_β resonances were manually assigned using 2D ^{1}H, ^{15}N-TROSY-HSQC and 3D TROSY-HNCO, TROSY-HNCA, TROSY-HNCACB, TROSY-HN(CO)CACB, with hSOUL double (^{13}C and ^{15}N) and triple labeled samples (^{2}H, ^{13}C and ^{15}N). Resonance peaks were

initially picked in 2D ^{1}H, ^{15}N-TROSY-HSQC spectrum and subsequenctly loaded onto trHNCO spectrum, allowing the identification of the spin systems. This peaklist was afterwards loaded onto trHNCA, trHNCACB, trHN(CO)CACB. In these spectra the alpha and beta carbon resonances were identified for each spin system. The sequential assignment was performed by comparing the Cα and Cβ chemical shifts for each spin system with the expected value (table A.2 in Appendix section) for each amino acid in the sequence.

However, even with triple isotopic labeled samples, only 25 % of the protein backbone for hSOUL (histidine tag) could be assigned – G10-A11, A16-A19, G24-A27, K47-T51, I62-G65, Q74-E78, I129-M135, V137-G144, and D165-G178 were the main segments assigned. The strong overlay of chemical shifts in the region of 8 ppm in the ^{1}H direction may be the result of a very different dynamic behavior for the N-terminal region of the protein contributing to poor NMR spectra.

3D experiments showed much less resonances than expected and this was one of the main reasons to clone hSOUL protein in a different cloning system.

Therefore, because of the results previously discussed, hSOUL backbone assignment was performed with hSOUL (intein tag) protein construct. ^{1}H, ^{15}N-TROSY-HSQC spectrum of hSOUL (intein tag) showed good chemical shift dispersion indicative of a properly folded protein - figure 4.3.

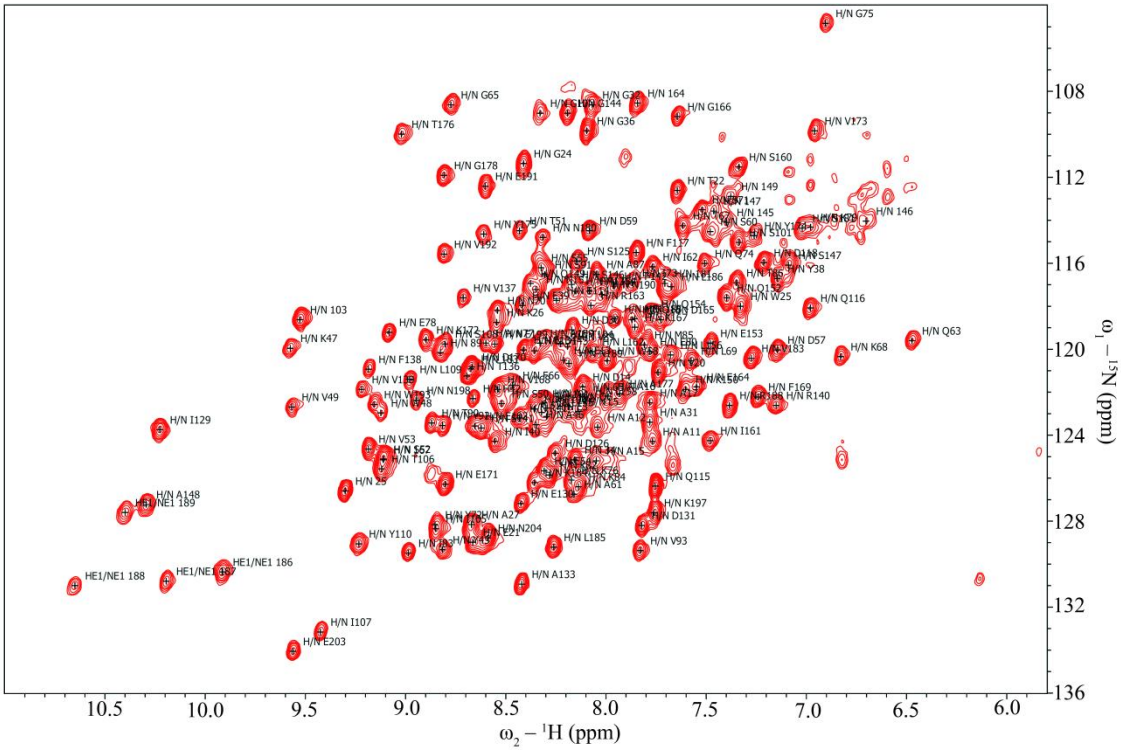


Figure 4.3 *^{1}H, ^{15}N-TROSY-HSQC spectrum of hSOUL.* Resonance assignments are indicated.

The same procedure as before was used for hSOUL (intein tag) backbone assignment allowing the assignment of approximately 77 % of the protein backbone – figure 4.4. Analysing the PDB, only 204 solution structures by NMR were so far determined for proteins with 200 or more amino acids. This number emphasizes the difficulty inherent to the process of backbone assignment and structure determination of proteins with the molecular mass of hSOUL or higher.

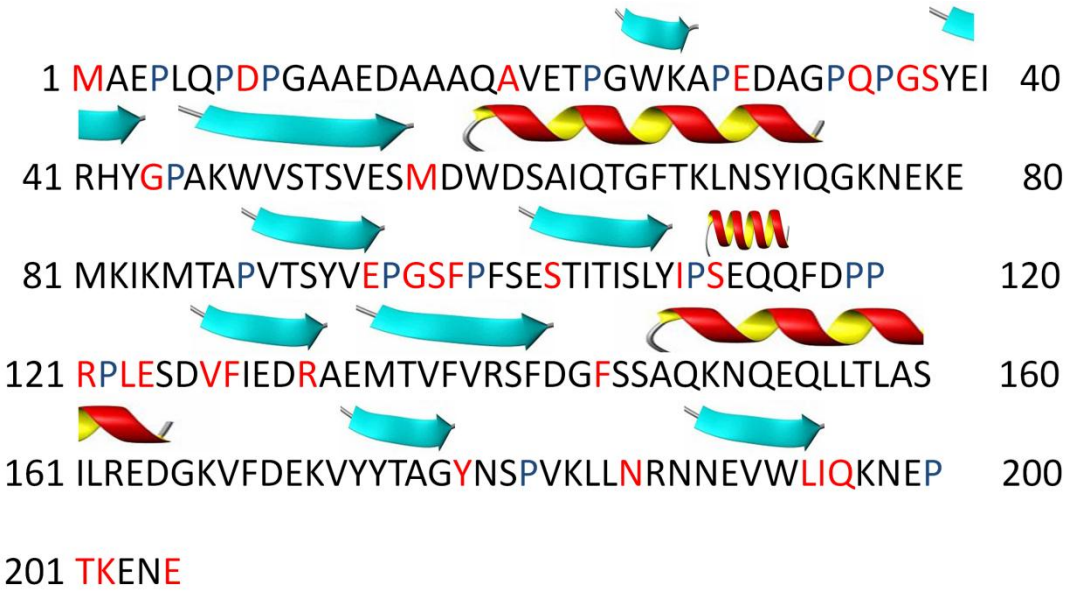


Figure 4.4 *hSOUL protein backbone assignment.* Residues in red were not assigned. Prolines, that could not be assigned with the NMR spectra acquired, are represented in blue. Secondary structure elements observed in the crystal structure of hSOUL - PDB code 4ayz (α-helices in red and yellow, β-sheets in light blue) are shown with ribbon representation, above the corresponding amino acids. Right and left panel numbering represent the aminoacid position from the first and last residue in each row.

Resonance peaks were initially picked in the 2D ^{1}H, ^{15}N–TROSY-HSQC. Afterwards, TROSY-HN(CO)CA, TROSY-HNCA, TROSY-HN(CO)CACB and TROSY-HNCACB spectra were used to identify the α and the β carbon resonances for each spin system and to determine sequential connectivities between different spin systems. Figure 4.5 shows the sequential assignment from Tyr38 to Tyr43, as an example of the backbone assignment.

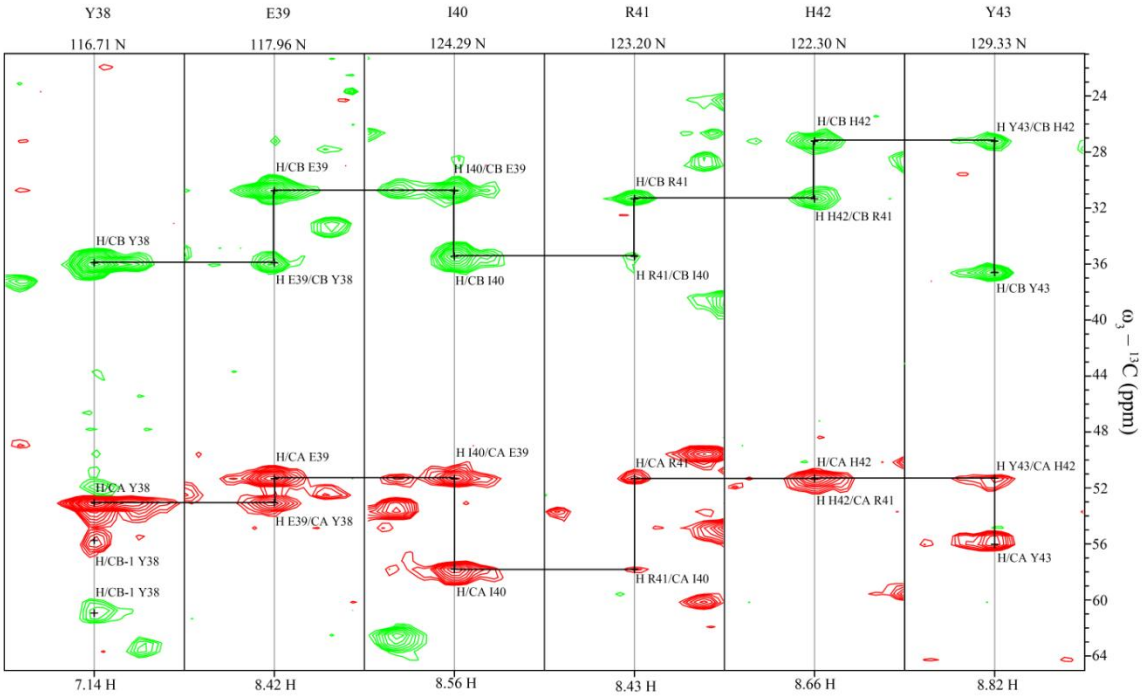


Figure 4.5 *Region of the trHNCACB ^{2}H, ^{13}C, ^{15}N-hSOUL spectrum.* Sequential assignments of the resonances from residue Tyr38 to Tyr43 using the 3D trHNCACB spectrum.

The N, H, C, Co Cα and Cβ chemical shifts of the assigned amino acids are shown in table A.3 (appendix).

As previously mentioned, the assignment of the spectral resonances for hSOUL was carried out by comparison of the Cα and Cβ chemical shifts for each spin system with the expected chemical shift for each amino acid in the sequence (using the mean Cα and Cβ chemical shifts, typical for each amino acid type) [78]. Some specific characteristics for these chemical shifts were used to initially identify particular amino acids. Glycines have typical Cα chemical shift value (45.33 ppm) and presents no Cβ. Alanines have a very characteristic Cβ – 18.95 ppm. Some of the serine and threonine amino acids could also be identified, but not distinguished, *via* their characteristic Cβ values (63.77 ppm and 69.94 ppm, respectively).

The backbone assignment was confirmed with the output results from the MANI PINE server v2.0 [161]. For the PINE server, several input data were used, namely the protein sequence and the peak lists from the 2D ^{1}H, ^{15}N–TROSY-HSQC, TROSY-HNCO, TROSY-HN(CO)CA, TROSY-HNCA, TROSY-HN(CO)CACB and TROSY-HNCACB spectra. The results given by the PINE server are in agreement with the manual backbone assignment.

Although the very dynamic behavior of the C- and N- terminus of hSOUL, similar to that observed in murine p22HBP, and reinforced by the absence of electron density in the electron

density maps of hSOUL in these regions, it was possible to assign the majority of the amino acids at the N-terminal and a great number of amino acids at the C-terminal – figure 4.4. Three short segments in the amino acid sequence can be pointed out as the main sites where assignment could not be carried out; residues E94-P99 (EPGSFP), consisting of a region with no secondary structure, between two β sheets (β4 and β5) and with two prolines that cannot be assigned with the spectra acquired. The second region comprises residues P119 to E124 (PPRPLE) between a short α-helix (α2) and a β-sheet (β6). A third region is longer and comprises the C-terminal residues from L194 to E205 in which seven residues could not be assigned (L194, I195, Q196, P200, T201, K202, E205) that can be explained, as previously mentioned, by the higher mobility of this region.

Secondary structure of proteins can be determined by the Chemical shift index (CSI) procedure that is based on chemical shift differences with respect to some predefined 'random coil' values. It can be applied from the measured Hα, Cα, Cβ and CO chemical shifts for each residue in a protein. hSOUL secondary structure was predicted using the TALOS+ server [162], giving the protein sequence and the NH, Cα, Cβ and CO chemical shifts as input – table 4.2 and figure 4.6.

Table 4.2 *hSOUL secondary structure from X-ray structure and predicted from NMR data (NH, Cα, Cβ and CO chemical shifts).*

Secondary structure	hSOUL crystal structure	Prediction from NMR data
β1	W25 - K26	W25
β2	E39 – Y43	E39 – H42
β3	A46 – S55	W48 – S50, S52 – V53
β4	V89 – E94	V89 – V93
β5	S103 – Y110	I105 – Y110
β6	V127 – R132	V127 – E130
β7	M135 – F142	T136 – S141
β8	Y174 – G178	Y175 – G178
β9	N190 – I195	E191 – I195
α1	W58 – Q74	D59 – I73
α2	S113 – Q116	E114 – Q116
α3	A148 – D165	A148 – E164

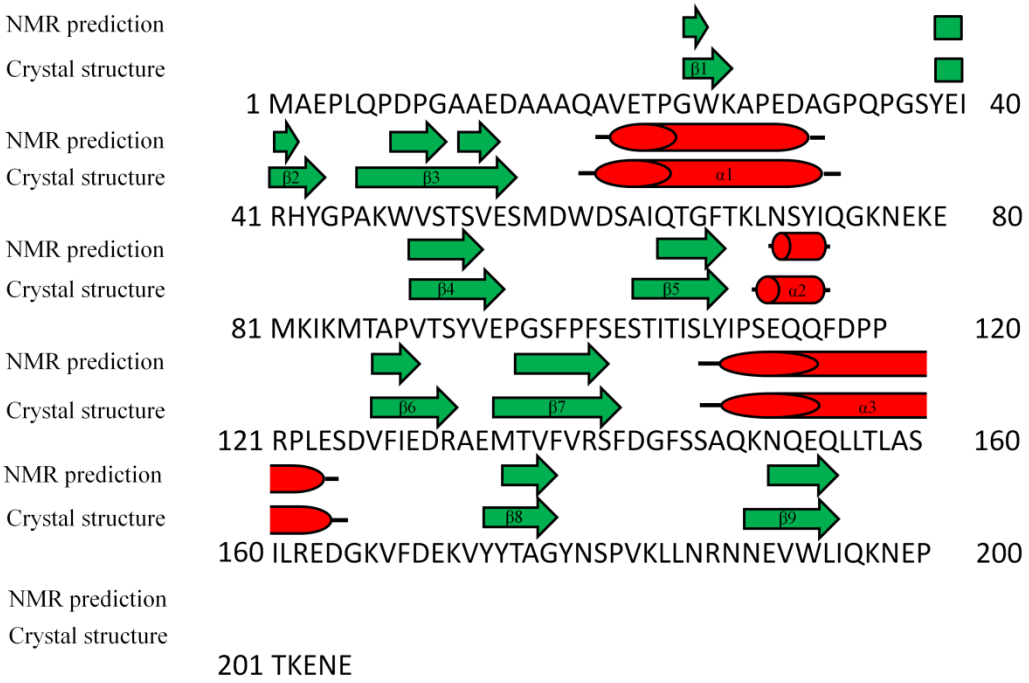


Figure 4.6 *hSOUL protein secondary structure schematic representation.* hSOUL protein sequence with secondary structure from X-ray crystal structure - PDB code 4ayz (Crystal structure) and from TALOS+ server (NMR prediction). β-sheet (green) and α-helix (red) are cartoon represented.

The results from TALOS+ are, in general, in agreement with the crystal structure of hSOUL. Residues W25, E39, V89 and G178, although contained in secondary structure elements, were classified as ambiguous. The remaining residues were unambiguous classified has belonging to the secondary structure segments listed in table 4.2.

After sequential identification, dynamic information can be obtained from $\{^{1}H\}$-^{15}N-NOE values. Figure 4.7 shows these values as a function of hSOUL protein sequence residues. These experiments allow the study of the local dynamic, i.e., the flexible residues in the protein.

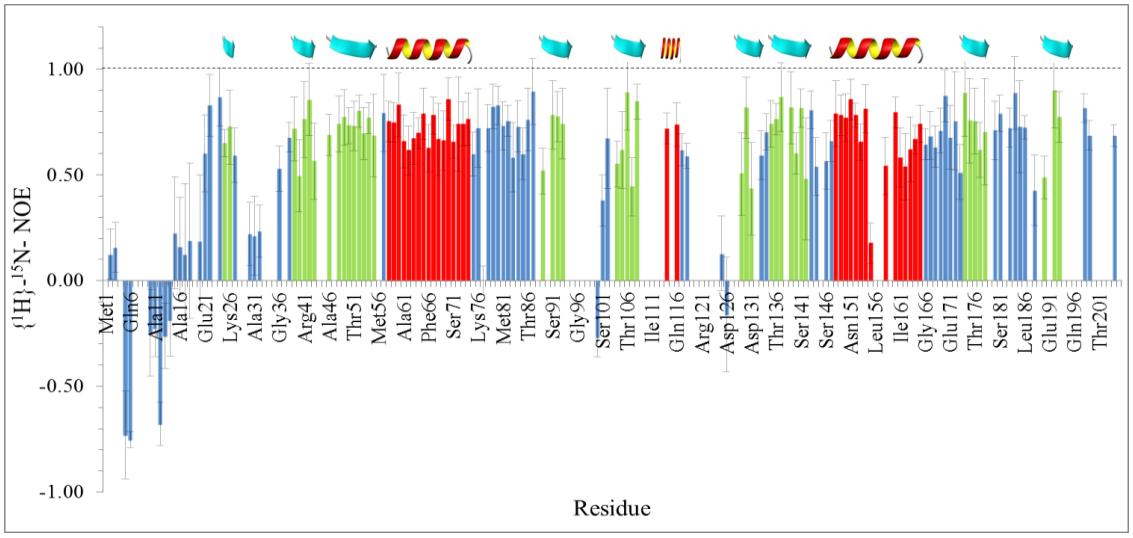


Figure 4.7 *{^{1}H}-^{15}N-NOE values plotted as a function of hSOUL protein sequence.* Red bars correspond to amino acids in α-helices, green bars correspond to amino acids belonging to β-sheets and blue bars correspond to amino acids in regions displaying no secondary structure. The NOE uncertaintanties are represented by the error bars in the graphic. Besides the residues that could not be assigned (Met1, Asp8, Ala19, Glu29, Gln34, Gly36, Ser37, Gly44, Met56, Glu94, Gly96-Phe98, Ser103, Ile111, Ser113, Arg121, Leu123, Glu124, Val127-Phe128, Arg132, Phe145, Tyr179, Asn187, Leu194-Gln196, Thr201-Lys202, Glu205 and prolines) the hetero-NOE values are not shown for residues Lys47, Asn77, Thr90, Leu109, Lys110, Gln115, Phe138, Leu156, Ala 159, Asn189, Glu191, Lys 197 and Glu 203. The dashed line ({^{1}H }-^{15}N- NOE = 1) represents the theoretical maximum value for {^{1}H }-^{15}N- NOE.

{^{1}H}-^{15}N-NOE relaxation data is highly sensitive to motions of the polypeptide backbone on a pico to nanosecond time scale. The {^{1}H}-^{15}N-NOE values for hSOUL remain fairly constant throughout the amino acid sequence with the exception of some regions that show much lower NOE values. These lower NOE values, indicative of more flexible regions, are almost exclusively located in the N-terminal region in agreement with the X-ray Crystallographic data presented and discussed in *chapter 3*. The residues from Glu21 to Lys26 show comparatively higher NOE values; this was also an expected result, since electron density maps are well defined in this protein segment. On the other hand, the regions Asp30-Gly32 and Ser125-Asp126 show comparatively low NOE values; these segments are in regions with no secondary structure, therefore they have higher mobility. The high NOE uncertainties values reflect the low signal to noise ratio.

Comparing the hSOUL {^{1}H}-^{15}N-NOE values with murine p22HBP {^{1}H}-^{15}N-NOE values determined by Dias *et al* [18], the hSOUL protein contains regions where the protein backbone is flexible (hetero NOE values below 0.65), whereas murine p22HBP displays constant and high NOE values throughout the protein sequence, indicative of a less flexible structure. Interestingly, two regions show significant differences. In hSOUL, in the region from Pro28 to Ser37, the majority of the residues could not be assigned (Pro28, Glu29, Pro33, Gln34, Pro35, Gly36 and Ser37) and the residues assigned show very low hetero NOE values which indicate a region with high backbone flexibility. In murine p22HBP, this region is much less flexible

according to the hetero NOE values. This is in agreement with the structure of both proteins (figure 3.10, *chapter 3*). Analysing the overlay between the two protein structures, this region in hSOUL is represented in forest green and shows significant differences to murine p22HBP: in murine p22HBP, this region contains part of β-sheet and a contiguous loop different from hSOUL where this region displays no secondary structure element. A second region that shows significant differences in the hetero NOE values comprises residues Tyr172 to Arg181 in murine p22HBP and Tyr179-Arg188 in hSOUL; for murine p22HBP, the only hetero value determined was for Tyr179 while for hSOUL, 8 out 11 hetero NOE values could be determined. Analysing figure 3.10, *chapter 3*, this is another region showing significant structural differences between the two proteins. These results agree with the conclusions drawn from molecular modelling analysis performed by Micaelo *et al* [19], that indicate this loop in murine p22HBP as a flexible loop, involved in heme-binding to the protein.

To complement the relaxation information it would be important to determine *R*1 and *R*2 parameters; *R*1 values provide information about motional properties with a frequency of approximately $10^8 – 10^{12}$ s^{-1}, whereas *R*2 values, in addition to depending on motions occurring at these frequencies, are also sensitive to dynamics on the micro-millisecond time scale. Hence, by measuring both *R*1 and *R*2, dynamic information over a large motional regime could be obtained [163-165]. The internal dynamic behaviour of a protein can be characterized by the order parameter, S^2. The order parameter gives the amplitude of motion, i.e., how far the atoms move from an average position. This parameter has been proven to be related to conformational entropy and can be used to estimate changes in conformational entropy [164, 166].

4.3.3. THE PUTATIVE HSOUL HEME-BINDING SITE

Although the hSOUL structure was solved at low resolution (3.5 Å), the electron density of the histidine 42 side chain is clear and solvent exposed – figure 4.8, allowing possible heme binding.

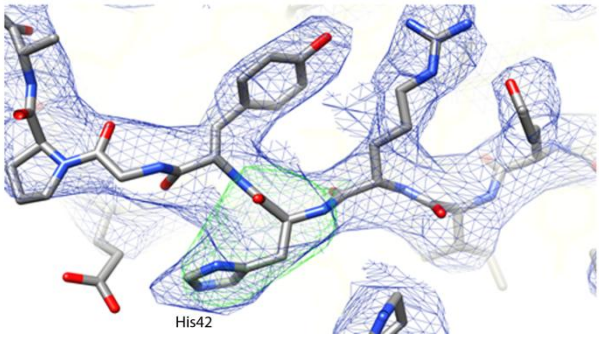


Figure 4.8 *Closer view of the side chain of His42*. The simulated annealing omit map (calculated with program phenix.refine from the PHENIX package and contoured at 1 σ) is shown in green superimposed with the 2mFo-DFc difference Fourier map (shown in blue), contoured at 2 σ.

As mentioned in *section 3.2.4*, several attempts were performed in order to co-crystallize the hSOUL protein with hemin and PPIX. In some drops prepared with protein and hemin some crystals were grown. These crystals were tested and did not show any anomalous signal in the iron fluorescence edge nor were good diffracting data obtained. Similar conclusions were reported by Ambrosi *et al* [34].

In previous studies, His42 was identified as the axial heme ligand of hSOUL protein [20]. Experiments to follow the backbone NH and the Nδ of this residue in the protein hSOUL (hSOUL construct has His42 and an additional single His in the small 3 amino acid N-terminal tag left after intein cleavage) were carried out using ^{15}N-enriched hSOUL samples and PPIX/hemin:^{15}N-hSOUL 5:1, 2:1, 1:1, 0.5 samples. HSQC spectra centered on the histidine side chain Nδ-proton region were acquired.

The signals in these spectra would be expected to shift significantly upon binding to heme iron. Paramagnetic relaxation via contact shift effects (Fe^{3+}, S=5/2) should broaden the His42 side chain Nδ resonance beyond detection. Figure 4.9a shows this region, where two patterns for the ε tautomer can be clearly seen [102]. The more intense signals were assigned to His42, the putative binding site, and the less intense to the extra His present in the N-terminal tag. Heme addition causes no significant alterations for these resonances, which implies that Nδ-His42 is not involved in iron binding. For comparison, HSQC spectra acquired in the same region with a 5:1 and 1:1 PPIX: ^{15}N-hSOUL mixture (figure 4.9b) gave similar patterns as compared to the labeled recombinant protein alone.

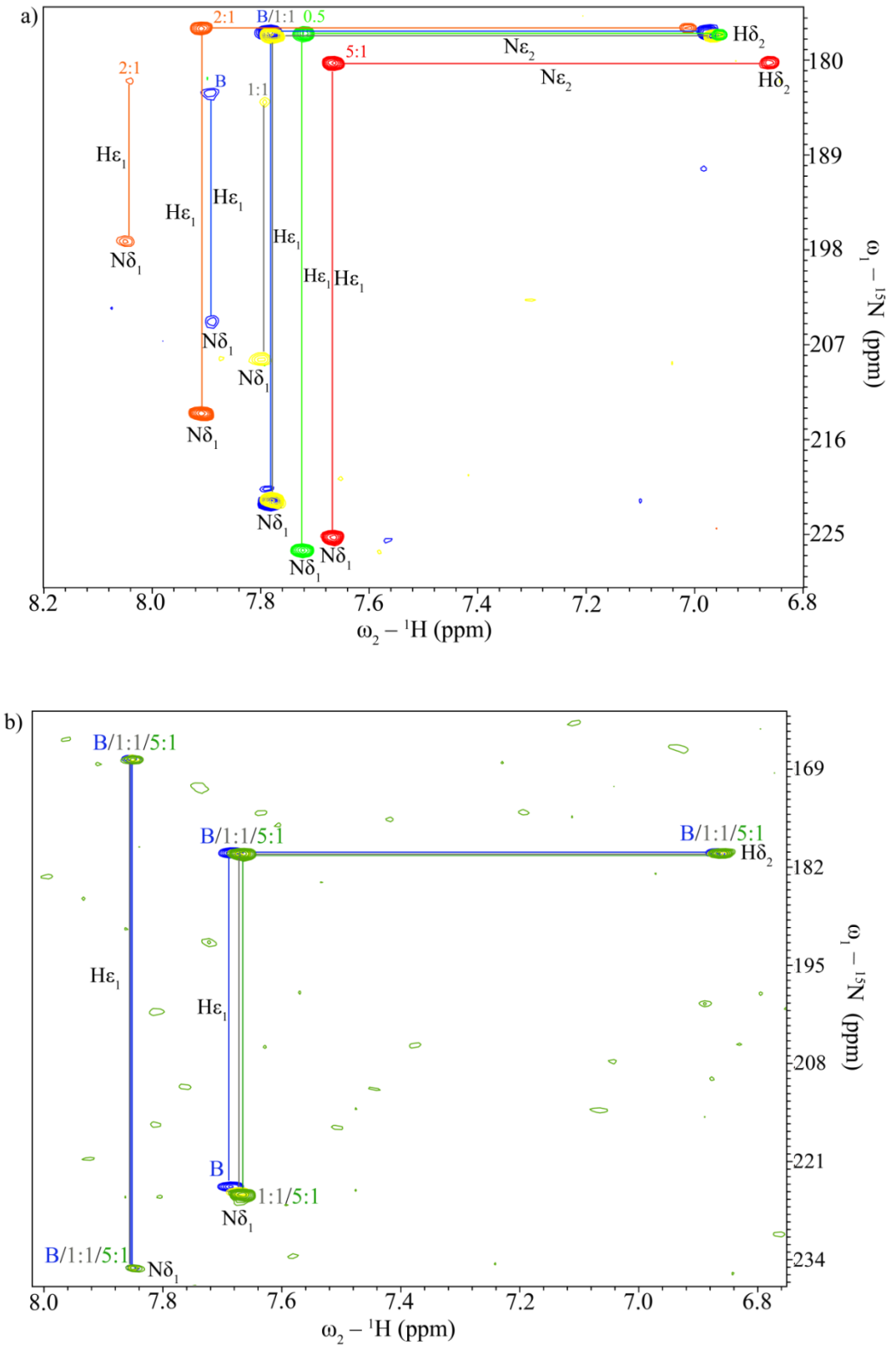


Figure 4.9 ^{1}H, ^{15}N *HSQC spectra, centered on the histidine side chain Nδ proton region.* a) hemin-^{15}N-hSOUL at molar ratio of 0.5 (green), 1:1 (yellow), 2:1 (orange), 5:1 (red) and ^{15}N-hSOUL alone (blue). b) PPIX:^{15}N-hSOUL at molar ratio of 1:1 (yellow) , 5:1 (green) and ^{15}N-hSOUL alone (blue).

It can also be observed, on both figures 4.9a and 4.9b, that the $H\delta_2$ cross-peaks from the N-terminal histidine residue (remaining after DTT cleavage) do not appear and the remaining cross-peaks are shifted to the left (higher ppm values in proton dimension) in comparison with cross-peaks from His42. Interestingly, hemin: hSOUL samples with molar ratios of 0.5 and 5:1 do not present any cross-peaks from the N-terminal histidine which can be a result of same unspecific interaction or paramagnetic effect due to iron heme, leading to a drastic chemical shift change of these peaks.

The tetrapyrrole binding site in murine p22HBP was defined by a hydrophobic cleft with residues from helix α1 and the β8-β9 loop [18]. In the crystal structure of hSOUL (figure 3.7, *chapter 3*), this helix extends from residue Trp58 to Gln74 and shows good structural conservation in murine p22HBP. The hydrophobic patch, clearly identified in the electrostatic surface calculated for the murine p22HBP [18], does not exist in hSOUL (figure 3.12). Also Met59 and Met63 residues in murine p22HBP (found close to the bound tetrapyrrole and both experiencing large chemical shifts deviations when titrated with PPIX) are replaced by Phe66 and Asn70 in hSOUL when comparing both proteins primary sequences.
Nevertheless, and to corroborate the previously drawn conclusions, NMR chemical shift mapping was used to follow heme binding to hSOUL. The same samples used for the above experiments were used to record ^{1}H, ^{15}N-HSQC and TROSY-HSQC spectra. The results indicate that, in contrast to murine p22HBP where loss of specific NH resonances close to the binding site was observed [18], no large chemical shifts are seen for hSOUL (figures 4.10 and 4.11). Careful inspection of the TROSY-HSQC and HSQC spectra upon hemin addition showed no significant changes (figure 4.10). Similar results are observed upon PPIX addition (figure 4.11).

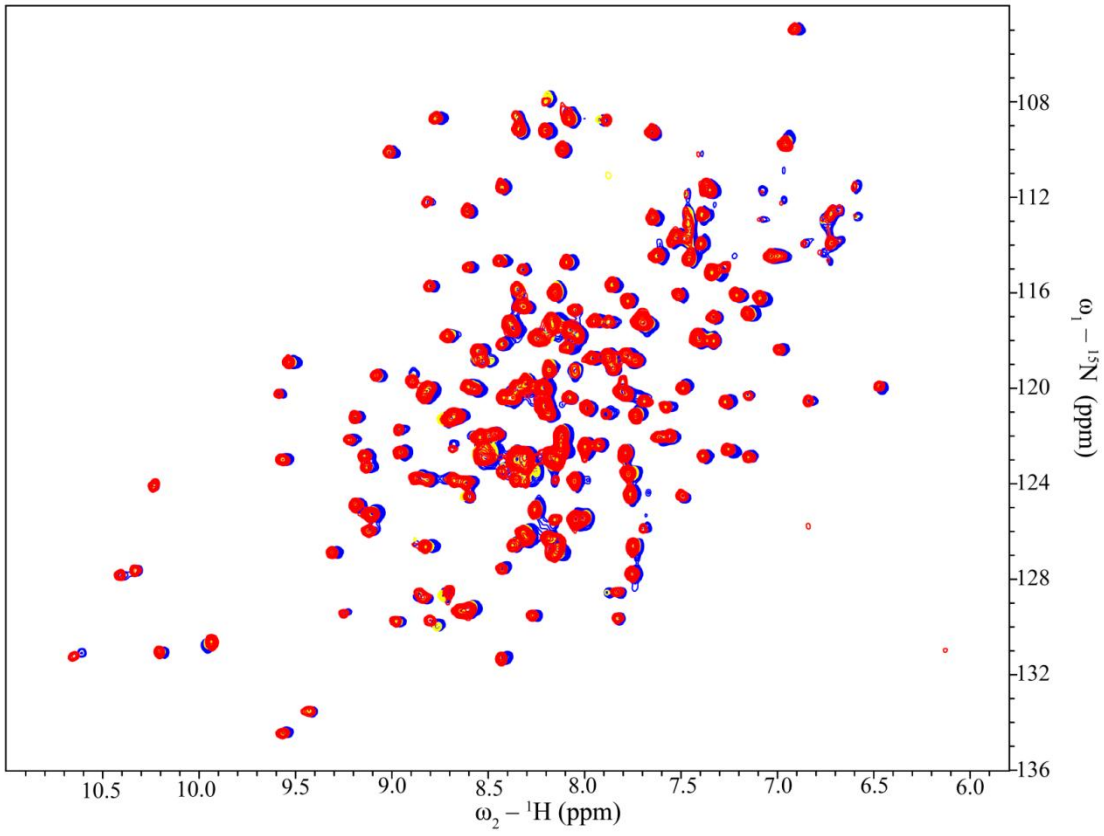


Figure 4.10 *^{1}H, ^{15}N-TROSY-HSQC spectra of hemin hSOUL.* ^{15}N-hSOUL:hemin at molar ratio of 5:1 (red), 1:1 (yellow), and ^{15}N-hSOUL alone (blue).

^{1}H, ^{15}N-TROSY-HSQC spectra of hemin:hSOUL at molar ratio of 0.5 and 2:1 are shown in appendix, figures A.1 and A2.

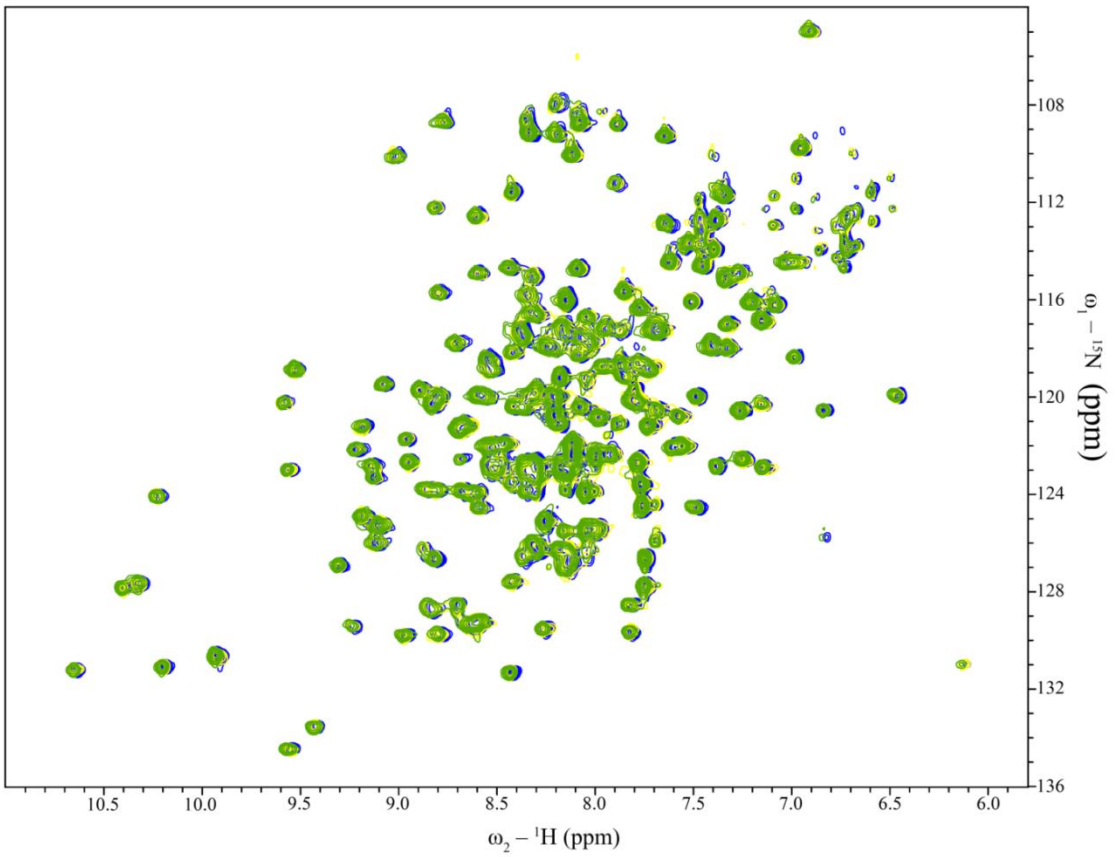


Figure 4.11 *^{1}H, ^{15}N-TROSY-HSQC spectra of PPIX: hSOUL*. PPIX: ^{15}N-hSOUL at molar ratio of 5:1 (green), 1:1 (yellow), and ^{15}N-hSOUL alone (blue).

In summary, there is no evidence from the NMR experiments for heme interaction with hSOUL.

Previous studies by Sato *et al* reported a dissociation constant value, K_d, for hSOUL-heme in the nanomolar range obtained by stopped flow experiments [20]. The dissociation constant was obtained by determining the association constant rate, k_{on}, of Fe(II)-CO heme with mSOUL by consecutive additions of *apo* mSOUL to Fe(II)-CO heme and the dissociation constant rate, k_{off}, of Fe(III)-hemin. The dissociation constant is therefore calculated from the k_{on} and k_{off} which may not reflect the correct values. The same methodology was used to determine the K_d for murine p22HBP. Both constants are indicated in table 4.3.

In order to compare results for hSOUL with results from murine p22HBP, fluorescence quenching experiments were carried out to determine the dissociation constant between hSOUL and some specific tetrapyrroles (hemin, PPIX, CPI (data not known) and CPIII (data not shown)). hSOUL contains 4 tryptophan residues, 3 near the N-terminus and 1 at the C-terminus. Assuming that hSOUL binds to heme with nanomolar K_d, a low concentration of protein (0.1 µM) was used and titrated with µM concentrations of tetrapyrrole in a similar manner to a previous study of murine p22HBP heme binding [18]. A non-linear fitting of the resulting

binding curves was carried out using equation 4.1, to compensate for ligand depletion. The results are shown in table 4.3.

In the literature, different dissociation constants, K_d, for p22HBP and SOUL proteins with porphyrins are reported [1, 13, 18, 20]. Table 4.3 includes published data and experimental results from this thesis.

Table 4.3 *Dissociations constants, K_d (and error associated with the measuerements, ΔK_d) for the complexes hSOUL:hemin/PPIX, human p22HBP:hemin/PPIX, murine p22HBP:hemin/PPIX, cHBP1:hemin/PPIX and cHBP2:hemin/PPIX*

Protein	**PPIX**		**Hemin**		**Technique**
	K_d (M)	ΔK_d	K_d (M)	ΔK_d	
hSOUL	181.8×10^{-9} 88.8×10^{-9}	57.0×10^{-9} 25.2×10^{-9}	226.2×10^{-9} 107.7×10^{-9}	55.0×10^{-9} 33.1×10^{-9}	Fluorescence Quenching
mSOUL (Sato *et al* [20])	-	-	4.8×10^{-9}	-	Stopped flow
Human p22HBP	6.35×10^{-9} 4.70×10^{-9}	1.32×10^{-9} 3.43×10^{-9}	20.4×10^{-9} 25.5×10^{-9}	4.60×10^{-9} 1.25×10^{-9}	Fluorescence Quenching
Human p22HBP (Blackmon *et al* [1])	12100×10^{-9}		8800×10^{-9}		Fluorescence Quenching
Murine p22HBP	2.6×10^{-9} 9.7×10^{-9}	2.9×10^{-9} 1.1×10^{-9}	11.1×10^{-9} 8.8×10^{-9}	5.9×10^{-9} 5.6×10^{-9}	Fluorescence Quenching
Murine p22HBP (Dias *et al* [18])	0.5×10^{-9}	-	3.0×10^{-9}	-	Fluorescene Quenching
Murine p22HBP (Taketani *et al* [13])	-	-	26×10^{-9}	1.8×10^{-9}	Liquid Scintillation Counting
Murine p22HBP (Blackmon *et al* [1])	11500×10^{-9}	-	900×10^{-9}	-	Fluorescence Quenching
Murine p22HBP (Sato *et al* [20])	-	-	0.021×10^{-9}	-	Stopped flow
cHBP1 (Takahashi *et al* [44])	440×10^{-9}	-	380×10^{-9}	-	Fluorescence Quenching
cHBP2 (Takahashi *et al* [44])	160×10^{-9}	-	700×10^{-9}	-	Fluorescence Quenching

Comparing the K_d values for the human variants of both SOUL and p22HBP proteins, hSOUL presents higher (20 to 30 times regarding PPIX and 50 to 100 times more regarding hemin) dissociation constants for both hemin and PPIX then human p22HBP. These results show that p22HBP binds stronger to PPIX and hemin, however the K_d for hSOUL indicates that heme-binding is occuring. This result contradicts the remaining data described in this dissertation which lead to draw the conclusion that the fluorescence decay upon porphyrin addition is probably due to a non-specific interaction. Another explanation for the high K_d values of hSOUL towards hemin/PPIX may be the possible packing of the porphyrins. If this occurs during the fluorescence measurements, the fluorescence intensity values observed will be influenced by this aggregation. Therefore, the calculated K_d values may not corresponding to the real value as the fluorescence decay is influenced by porphyrin molecules interacting with each others.

The results for murine and human p22HBP determined and presented in this dissertation (table 4.3) are comparable to those initially reported Taketani *et al* [13] where afiinities in the low nanomolar range were observed for hemin and PPIX, however a different methodology was used. The results for murine and human p22HBP should therefore be compared to the results from Blackmon *et al* [1]. These results are significantly different from those determined and presented in this thesis and from the values determined by Sato *et al* [20] and Taketani *et al* [13]. The lack of information on Blackmon *et al* [1] regarding porphyrin concentration and the probable incorrect protein concentration stated (10 - 40 M) do not allow any comparisons to be made.

Based on a previous work from Goodfellow *et al* [167], a UV-visible titration was performed where hSOUL was added to a hemin solution in order to follow any possible changes in the Soret band of hemin at 392 nm due to protein binding.

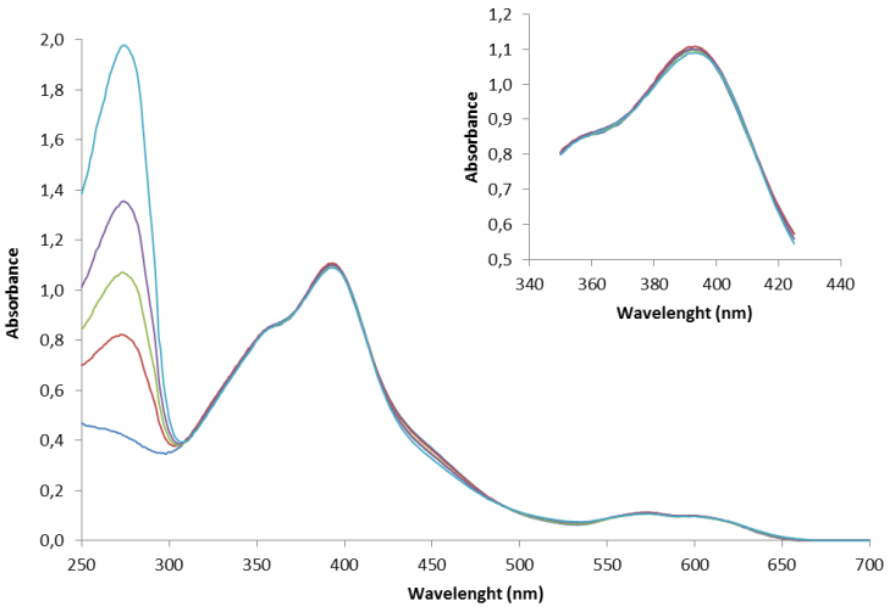


Figure 4.12 *UV-visible spectra of the hSOUL-hemin titration.* The addition of hSOUL, that can be seen by the increasing absorbance at 280 nm is not accompannied by an increase at 392 nm, which indicates the inexistence of the interaction of the hSOUL with hemin. hemin:hSOUL molar ratios were 2.6 (red), 1.3 (green), 0.8 (purple) and 0.5 (light blue).

Analysing the superimposed titration curves on figure 4.12, no change is observed in the Soret band confirming the NMR results and reinforcing the conclusion of the non-existing or non specific heme-binding is occurring in hSOUL.

CHAPTER 5

HEME-BINDING INTERACTIONS STUDIES ON P22HBP

CONTENTS

5.1 INTRODUCTION

Murine p22HBP was the first protein from the SOUL/HBP heme-binding family of proteins to be structurally characterized. The NMR solution structure of the *apo* protein was determined and revelead that it consists of a nine stranded β-barrel core surrounded by 2 α-helices [18].NMR chemical shift mapping was performed to identify the residues in this protein involved in the heme binding and, recently, molecular modeling and docking studies with several porphyrins were performed, confirming the results obtained by NMR [18, 19].

No information is yet available regarding human p22HBP; this protein presents 86 % sequence identity to murine p22HBP and this information together with murine p22HBP solution structure was used to create a protein model that was further used for identical molecular modeling and docking studies to murine p22HBP with the results presented in *section 1.2.2* [19]. Therefore, the other goal of this work was to determine the structures of both murine and human p22HBP bound to hemin/PPIX in order to confirm the theoretical calculation results and elucidate the heme-binding to p22HBP.

5.2. MATERIAL AND METHODS

5.2.1. OVEREXPRESSION AND PURIFICATION OF HUMAN AND MURINE P22HBP

The gene of human p22HBP was cloned in pET28a expression vector (NOVAGEN), in Nco I and Xho I sites. BL21 (DE3) cells were transformed with the plasmid for further protein overexpression (process performed by NzyTech, genes & enzymes, Ltd.). For that, a colony was inoculated in 20 mL Luria Broth (LB) medium supplement with kanamycin (50 µg/mL) and incubated at 310 k for 12 – 16 hours. The overnight culture was then inoculated in 150 mL of LB medium and incubated until O.D. = 0.5 – 0.8. Cells are therefore harvested and ressuspended in M9 minimal medium supplemented with kanamycin 50 µg/mL. After 2 hours, human p22HBP overexpression is induced by the addition of IPTG 0.1 mM.

The purification protocol is similar to the described in section *2.2.1.1;* after adding 50 mM phosphate buffer pH 8.0, 300 mM NaCl, 20 mM imidazole, four additional steps on the imidazole gradient were performed using the same buffer but with increasing imidazole concentrations (50, 75, 175 and 500 mM, respectively). The fractions containing human p22HBP were concentrated and loaded (approximately 400 µl containing 10 mg of human p22HBP) onto a Superdex 75 10/300 GL column (GE Healthcare pre-packed) coupled to an FPLC system (GE Healthcare) previously equilibrated with 50 mM phosphate at pH 8.0. The eluted fractions containing hSOUL were pooled together and concentrated in an Amicon concentrator equipped with a YM10 membrane. Human p22HBP protein was obtained with a high level of purity.

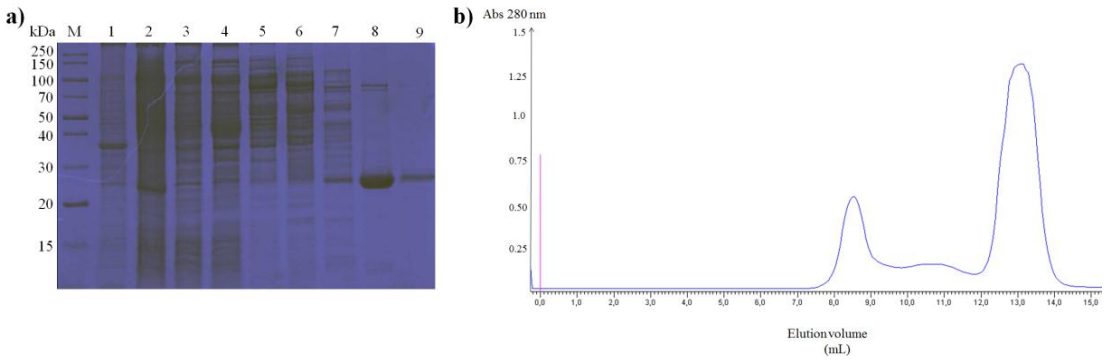


Figure 5.1 *Purification of overexpressed human p22HBP*. a) SDS-PAGE (15 % acrylamide) analysis of the different fractions obtained from the Ni-NTA Agarose column: M – Low molecular weight standards (Bio-Rad Laboratories); 1 – insoluble fraction; 2 - soluble fraction loaded on the Ni-NTA resin; 3- flow through; 4 - flow through; 5 – resin wash with 10 mM imidazole; 6 - resin wash with 20 mM imidazole; 7 - resin wash with 50 mM imidazole; 8 – human p22HBP elution with 75 mM imidazole; 9 – human p22HBP elution with 175 mM imidazole; b) Elution profile obtained from the gel filtration column (Superdex 75) loaded with human p22HBP fractions (8+9) from Ni-NTA Agarose resin.

Human p22HBP overexpression and purification optimization were performed by Leonildo Delgado, during his Master's thesis.

Murine HBP overexpression and purification were performed according to Dias *et al* [18].
Murine p22 HBP (residues 7 - 190) overexpression was achieved by growing the BL21 (DE3) cells harboring the pNJ2 plasmid in LB media supplemented with 50 µg/mL and incubated at 310 K. The culture was then inoculated into MOPS minimal media (phosphate limiting) – see appendix [168] and incubated for 16 hours at 303 K after which the culture was harvested. The cell extract was resuspended in 50 mM phosphate buffer pH 8.0 with 300 mM NaCl, ruptured by sonication and centrifuged at 48384 × g (20000 rpm) for 1 hour. The protein purification was performed with a Ni-NTA Agarose resin (Qiagen). The supernatant was loaded onto the Ni-NTA column equilibred with 50 mM phosphate buffer pH 8.0 with 300 mM NaCl. The resin was washed in two steps by adding the same buffer with 10 mM and 20 mM imidazole. Murine p22HBP was eluted in a discontinuous way with 50 mM phosphate buffer pH 8.0 with 300 mM containing 250 mM imidazole. The fractions containing murine p22HBP were pulled together and loaded to a Superdex 75 10/300 GL column (GE Healthcare pre-packed) coupled to an FPLC system (GE Healthcare) previously equilibrated with 50 mM phosphate at pH 8.0.

5.2.2. MURINE AND HUMAN P22HBP CRYSTALLIZATION

5.2.2.1. MURINE P22HBP

For the crystallization of *apo* murine p22HBP several crystallization screens were used, namely, an in-house prepared sparse matrix screen of 80 conditions, Crystal Screen 2 (Hampton Research), Emerald Wizard I (Emerald Biostructures), Emerald Wizard II (Emerald Biostructures), Crystallization Basic Kit for Membrane Proteins (Sigma) and Natrix crystallization screening (Hampton Research). Experiments were performed using the hanging drop vapour diffusion method both at 277 K and 293 K, with droplets consisting of 1µl of protein solution (12 mg/ml and 25 mg/ml in 10 mM Tris-HCl, pH 8.0), 1µl of reservoir solution and 700 µl of precipitant solution in the *reservoir*. Crystallization trials were also attempted at the High Throughput Crystallization Laboratory, at the EMBL Grenoble; Hampton Crystal Screen 1 to 6 were tested on *apo* murine p22HBP (15 mg/ml in 10 mM Tris-HCl, pH 8.0) on a 1:1 droplet at 277 K and 293 K using the sitting drop vapour diffusion method. Crystallization of murine p22HBP bound to hemin was also tried. For this purpose an in-house prepared sparse matrix screen of 80 conditions, JBScreen Classic 1-10 (Jena Bioscience), Natrix crystallization screening (Hampton Research), Crystallization Basic Kit for Membrane Proteins (Sigma), Emerald Wizard I (Emerald Biostructures) and Emerald Wizard II (Emerald Biostructures) screens were used. These experiments were performed using the hanging drop vapour diffusion method at 293 K. The conditions described above were tried for murine p22HBP (17 mg/ml) previously incubated for approximately 1 hour with hemin (prepared according to Dias *et al.* [18]) in a 1:1 molar ratio solution. Crystallization trials of murine p22HBP bound to hemin were also attempted at the High Throughput Crystallization Laboratory, at the EMBL Grenoble; Hampton Crystal Screen 1 to 6 were tested on murine p22HBP (15 mg/ml in 10 mM Tris-HCl, pH 8.0) pre-incubated with hemin on a molar ration of 1:1, on a 1:1 droplet at 277 K and 293 K using the sitting drop vapour diffusion method.
In total, approximately 1200 conditions for *apo* murine p22HBP and approximately 1100 for murine p22HBP pre-incubated with hemin were tested.

5.2.2.2. HUMAN P22HBP

Initial crystallization trials on *apo* human p22HBP were performed in the Oryx8 (Douglas Instruments Ltd.). Drops of 0.30 µL of protein (15 mg/mL in 10 mM Tris-HCl buffer pH 8.0) + 0.15 µL of the reservoir solution were set up using the sitting drop vapour diffusion method, at 277 K. Several crystallization screenings were used: in-house prepared sparse matrix screen of 80 conditions, Crystallization Basic Kit for Membrane Proteins (Sigma), JBScreen Classic 1-10

and JCSG+ (Jena Bioscience), PEG/Ion 4 k, PEG/Ion 8 k (Hampton Research) and SaltRX (HamptonResearch) and Morpheus protein crystallization screen (Molecular Dimensions)

The in-house prepared sparse matrix screen of 80 conditions and JBScreen Classic 1, 6, 7 and 10 were also used on initial crystallization trials, at 293 K, for *apo* human p22HBP (15 mg/mL in 10 mM Tris-HCl buffer pH 8.0). The droplets were performed by Oryx8 (Douglas Instruments Ltd.) and consisted in 0.30 µL of protein and 0.15 µL of the reservoir solution, using the sitting drop vapour diffusion method. Approximately 750 conditions were tested.

For the crystallization of human p22HBP bond to hemin, human p22HBP previously incubated for approximately 1 hour with hemin in a 1:1 molar ratio was used (porphyrin stock solutions were prepared according to Dias *et al* [18]). After incubation, dialysis was performed to exchange the solution buffer to 10 mM Tris-HCl pH 8.0 and the protein was further concentrated to the desired concentration (approximately 15 mg/mL). Crystallization trials were performed with the Oryx8 (Douglas Instruments Ltd.), using the sitting drop vapour at 277 K; droplets consisted in 0.30 µL of protein and 0.15 µL of the reservoir solution. Several crystallization screenings were used: in-house prepared sparse matrix screen of 80 conditions, Crystallization Basic Kit for Membrane Proteins (Sigma), JBScreen Classic 1-10 and JCSG+ (Jena Bioscience), PEG/Ion 4 k, PEG/Ion 8 k (Hampton Research) and SaltRX (HamptonResearch) and Morpheus protein crystallization screen (Molecular Dimensions) screens were used.

For the more promising crystallization conditions (obtained in the crystallization robot) for the protein pre-incubated with hemin: a) 12 % PEG 3350, 0.2 M amonium sulphate, 0.1 M acetate buffer 4.5; b) 12 % PEG 3350, 0.2 M magnesium chloride, 0.1 M Tris/HCl 8.5; c) 25 % PEG 400, 0.05 M magnesium acetate, 0.05 M magnesium acetate, optimization attempts were performed by varying the precipitant concetrations (PEG 3350 – from 8 % to 20 % and PEG 400 – from 20 % to 30 %) and buffer pH (0.1 M acetate buffer 4.5 – from pH 4.29 to pH 6.01 and 0.1 M Tris-HCl 8.5 – from pH 7.5 to pH 9.5).

5.3. RESULTS AND DISCUSSION

5.3.1. MURINE P22HBP

Despite the several crystallization conditions tried no protein crystals were yet obtained for murine p22HBP. In some crystallization conditions (a) 0.2 M calcium chloride, 0.1 M acetate buffer 4.5, 30 % 2-methyl-2,4-pentanediol and b) 0.2 M calcium chloride, 0.1 acetate buffer 4.5, 20 % isopropanol), for both murine p22HBP and murine p22HBP pre-incubated with hemin at 293 K, salt crystals were observed. Since these crystals could be protein crystals, they were tested at the ESRF, ID14-EH4, but found to be salt crystals. For murine p22HBP:hemin complex, salt crystals could also be observed in: a) 0.2 M ammonium acetate, 0.15 M magnesium acetate, 0.05 M HEPES buffer 7.0, 5 % PEG 3350; b) 0.1 M ammonium acetate, 0.02 M magnesium chloride, 0.05 M HEPES 7.0, 5 % PEG 8000; c) 16 % PEG 3350, 0.1 M Tris-HCl 8.5, 0.1 M magnesium chloride; d) 20 % PEG 3350, 0.1 M Tris-Hcl 8.5, 0.2 M calcium chloride and e) 30 % PEG 3350, 0.1 M acetate buffer 4.5, 0.1 M magnesium chloride.
For *apo* murine p22HBP sea urchins were observed, at 293 K, in 0.1 M NaCl, 0.1 M citrate buffer 5.5, 30 % PEG 400 but could not yet be reproduced or optimized.

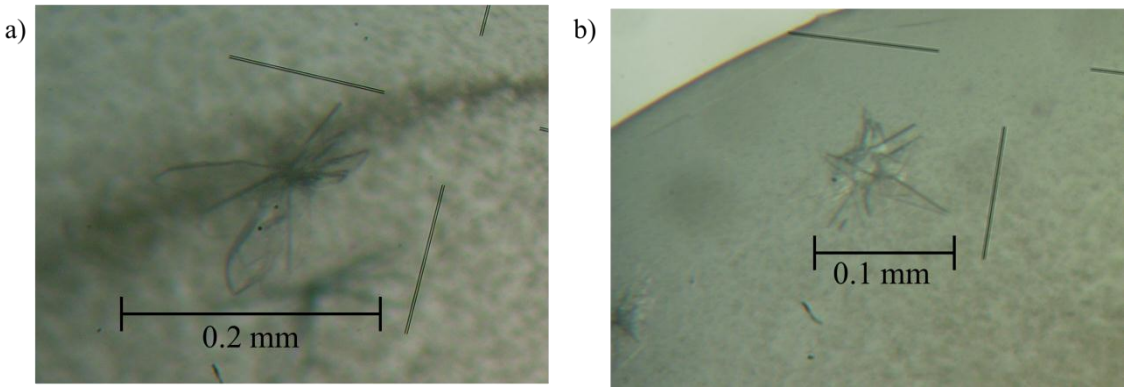


Figure 5.2 *Salt crystals.* Crystals obtained in a) 0.2 M calcium chloride, 0.1 M acetate buffer 4.5, 30 % 2-methyl-2,4-pentanediol and b) 0.2 M calcium chloride, 0.1 acetate buffer 4.5, 20 % isopropanol, in crystallization trials with murine p22HBP (*apo* form), at 293 K.

Several approaches can be performed in order to overcome the difficulties in crystallazing murine p22HBP; regarding the protein, several experiments can be performed, namely by thermofluor or NMR, in order to study its stability in different conditions; if the protein is not completely stable in the buffer used, the crystallization process tends to be very difficult. On the other hand, regarding crystallization trials, more crystallization screens, higher protein concentrations together with different drop volumes and different crystallization methods can be tried

5.3.2. HUMAN P22HBP

Some initial crystallization conditions were obtained for human p22HBP:hemin complex. In four promising conditions (a) 12 % PEG 3350, 0.2 M amonium sulphate, 0.1 M acetate buffer 4.5; b) 12 % PEG 3350, 0.2 M magnesium chloride, 0.1 M Tris/HCl 8.5; c) 25 % PEG 400, 0.05 M magnesium acetate, 0.05 M magnesium acetate and d) 0.2 M magnesium chloride, 25 % PEG 2000 MME), very thin needles (‘a’ and ‘c’), 2D plates (‘d’) and a crystal (‘b’) – figure 5.3, could be observed. In order to optimize these conditions, some crystallization trials were done varying the precipitant concentration and the buffer pH. No improvements were yet achieved for conditions ‘a’ and ‘c’. The crystal obtained in 12 % PEG 3350, 0.2 M magnesium chloride, 0.1 M Tris-HCl 8.5 (‘b’) was tested at the ESRF, ID14-EH4, and shown to be a salt crystal.

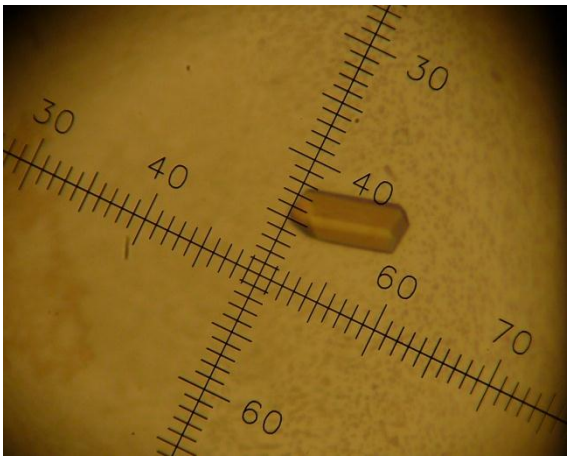


Figure 5.3 *Salt crystal.* Crystal obtained in 12 % PEG 3350, 0.2 M magnesium chloride, 0.1 M Tris-HCl 8.5 in a drop with 0.30 µl of protein:hemin and 0.15 µl of the precipitant solution.

From human p22HBP:hemin complex crystallization trials, salt crystals could also be observed in 2 M magnesium chloride, 0.1 M Tris-HCl pH 8.5.

Sea urchins were observed in several conditions, at 277 K, for human p22HBP:hemin complex: a) 1 M ammonium sulphate, 0.1 M acetate buffer 4.5; b) 20 % PEG 3350, 0.2 M sodium nitrate; c) 15 % PEG 400, 0.1 M calcium chloride, 0.1 M acetate buffer 4.5; d) 25 % PEG 400, 0.1 M acetate buffer 4.5, 0.1 M magnesium chloride; e) 30 % PEG 1000, 0.1 M Tris-HCl 8.5; f) 1 M magnesium sulphate, 0.1 M acetate buffer 4.5 and g) 1 M magnesium sulphate, 0.1 M Citrate buffer 5.6.

Interestingly, five out of the nine conditions where needles or sea urchins were observed, contain acetate buffer pH 4.5, which can mean that, at this pH, the human p22HBP:hemin complex tends to stabilize and crystallize.

In summary, for human p22HBP:hemin complex, some initial crystallization conditions where needles and sea urchins could be observed, were determined. Optimization experiments need to

be performed, starting from these preliminary results, in order to obtain good diffracting crystals of the protein-heme complex.

CHAPTER 6

CONCLUSIONS AND FUTURE PERSPECTIVES

CONTENTS

6.1 CONCLUSIONS

The solution structure of murine p22HBP was solved by NMR in 2006 [18], constituting the first structure of a member of the SOUL/HBP heme-binding family of proteins to be determined. Chemical shift mapping by NMR was performed and residues involved in the heme-binding were identified. Analyzing the electrostatic surface of the protein, together with the intrinsic characteristics of the amino acid residues, it was concluded that the heme-binding region is a hydrophobic pocket where the heme is thought to be buried. These conclusions were further confirmed by molecular modeling and docking studies [19].
The other member of this family of proteins is the SOUL protein, whose structural and dynamic characterization constituted the main subject of this dissertation. Little information was available on this protein at the beginning of the work. The protein was classified as a cytosolic heme-binding protein and histidine 42 was proposed to be heme axial ligand.

Due to the large amount of protein necessary for X-ray Crystallography and NMR studies, the first objective of the work reported on this thesis was to optimize the overexpression and purification of human SOUL protein. Two overexpression systems were used – N-terminal histidine tag fusion protein and hSOUL-Intein fusion protein, that made it possible to optimize the conditions for protein production and good yields could be achieved (20 mg and 4 mg *per* liter of cell culture, respectively).

Once the best conditions for hSOUL protein overexpression were obtained, the determination of the three-dimensional structure of the human SOUL protein (either by NMR or X-ray methods) as well as the elucidation of the protein heme binding mechanism were the main goals.

The X-ray structure of hSOUL was solved to 3.5 Å resolution consisting of an 8 stranded β-barrel core surrounded by two α-helices and a disorded N-terminal. The superposition of human SOUL structure with the solution structure of murine p22HBP shows that the two proteins from the SOUL/HBP family of proteins have a very similar overall fold. Analysis of the electrostatic surfaces of the two proteins show that the hydrophobic cleft present in murine p22HBP protein involved in the heme binding is not observed in hSOUL. Therefore, different experiments were designed in order to obtain the necessary information to understand the possible heme-binding to the protein. hSOUL backbone assignment was carried out and approximately 77 % of the amino acid residues could be assigned. This allowed us to further study heme binding by hSOUL using NMR, which constitutes this an important achievement. Analysing the PDB, the number of

structures determined by NMR for proteins with more than 200 amino acids is approximatley 200 out of approximately 9500, which results from the difficulty in acquiring good NMR spectra of large proteins. Despite the low resolution of the crystal structure, the electron density around histidine 42 was clear, showing that the His42 side chain is solvent exposed, therefore accessible for heme binding. To confirm the proposed binding, several experiments were performed. NMR data combined with UV-visible spectroscopy data from heme-hSOUL titrations showed no binding or a non-specific binding of heme to hSOUL protein.

Sequence alignment among hSOUL and some proteins from the Bcl-2 family of proteins (Bcl-x_L, Bcl-2, Bim, Bax, etc.) identified the presence of a BH3-domain in hSOUL sequence (from Leu158 to Lys172). In all the structures determined so far of BH3-domain containing proteins, this domain is always an α-helix. Careful examination of the three-dimensional structure of hSOUL protein shows that the BH3-domain is composed by part of helix α3 and a loop. Further experiments need to be performed to study the possible interaction between hSOUL and members of the Bcl-2 family of proteins. To explore this new function attributed to hSOUL protein some experiments can be designed, namely the titration of hSOUL protein with anti-apoptotic members of the Bcl-2 family of proteins, such as Bcl-2 and Bcl-x_L in order to follow the possible interaction by chemical shift perturbation. Moreover, if interaction between the proteins is observed, determining the X-ray structure of the complex would be the next step in this study.

In order to continue the p22HBP protein (murine and human) characterization, especially the heme binding mechanism, crystallization of murine and human p22HBP with PPIX and/or hemin and the concomitant structure determination of the complexes were attempted. Same promising crystallization conditions have been determined regarding human p22HBP:hemin complex and optimization experiments are being performed. Thus, more experiments need to be done to obtain suitable crystallization conditions for both proteins.

The results obtained and described in this dissertation show the importance of combining different approaches in the elucidation of biological relevant problems. For the aims of this dissertation, the determination of the three-dimensional structure of hSOUL protein was achieved by X-ray Crystallography. To study the heme-protein interaction both X-ray Crystallography and Nuclear Magnetic Resonance techniques were used. Solving the hSOUL protein structure enabled the location of the histidine position. NMR experiments brought significant new insights on the hSOUL heme-binding.

During the process of submitting to publication the results described in this thesis, a manuscript was published by Ambrosi *et al* describing the 3D structure of human SOUL to 2.85 Å resolution, and the complex of the anti-apoptotic protein Bcl-x_L with hSOUL BH3 domain peptide. The overall hSOUL structure is very similar to the described in this thesis with the exception of the N-terminal region. The rmsd between the chains A of hSOUL (4ayz and 3r8k) is 0.84 Å for 165 Cα atoms. In the same publication, hemin-hSOUL studies using X-ray Crystallography, NMR and UV-visible spectroscopy are described, leading to the conclusion that no interaction occurs between hemin and hSOUL. The binding of the BH3 domain peptide of hSOUL to the anti-apoptotic Bcl-x_L was also studied. When the BH3 domain peptide was added to Bcl-x_L significant chemical shift changes could be observed on the ^{1}H, ^{15}N HSQC spectra. These results were confirmed by Surface Plasmon Resonance. Similar experiments were performed by SPR using the intact hSOUL protein; in this case the results show no interaction between hSOUL and Bcl-x_L [34].

6.2 FUTURE PERSPECTIVES

The work presented in this thesis represents an important step torwards the characterization of hSOUL protein: the three-dimensional structure of the protein was determined and the possible heme-protein interaction was elucidated.

A big question mark however remains about SOUL protein biological function. None of the data so far obtained brought significant insights to the understanding of this issue.

The discovery of a BH3-like domain in hSOUL protein leads to the need to investigate its possible pro-apoptotic activity, similar to the one shown by BH3-only proteins belonging to the Bcl-2 family of proteins, briefly described in *section 1.3.2* of this dissertation. For that purpose, NMR titrations of hSOUL, followed by ^{1}H, ^{15}N-HSQC/TROSY-HSQC spectra, upon Bcl-2 anti-apoptotic proteins addition (namely, Bcl-2 and/or Bcl-x_L) and subsequent analysis of the chemical shift changes should identify possible interactions of these proteins and the residues involved in these interactions. If these interactions are observed, the next step would be solving the structure of protein-protein complex.

Crystallization trials must be performed in order to crystallize and determine the crystal structure of murine and human p22HBP-heme complexes.

REFERENCES

1. Blackmon BJ, Dailey TA, Xiao LC & Dailey HA (2002) Characterization of a human and mouse tetrapyrrole-binding protein. *Archives of Biochemistry and Biophysics* **407**, 196-201.
2. Suits MD, Jaffer N & Jia Z (2006) Structure of the Escherichia coli O157:H7 heme oxygenase ChuS in complex with heme and enzymatic inactivation by mutation of the heme coordinating residue His-193. *J Biol Chem* **281**, 36776-36782.
3. Ajioka RS, Phillips JD & Kushner JP (2006) Biosynthesis of heme in mammals. *Bba-Mol Cell Res* **1763**, 723-736.
4. Latunde-Dada GO, Simpson RJ & McKie AT (2006) Recent advances in mammalian haem transport. *Trends Biochem Sci* **31**, 182-188.
5. Donovan A, Brownlie A, Zhou Y, Shepard J, Pratt SJ, Moynihan J, Paw BH, Drejer A, Barut B, Zapata A, et al. (2000) Positional cloning of zebrafish ferroportin1 identifies a conserved vertebrate iron exporter. *Nature* **403**, 776-781.
6. McKie AT, Marciani P, Rolfs A, Brennan K, Wehr K, Barrow D, Miret S, Bomford A, Peters TJ, Farzaneh F, et al. (2000) A novel duodenal iron-regulated transporter, IREG1, implicated in the basolateral transfer of iron to the circulation. *Molecular Cell* **5**, 299-309.
7. Abboud S & Haile DJ (2000) A novel mammalian iron-regulated protein involved in intracellular iron metabolism. *J Biol Chem* **275**, 19906-19912.
8. Shibahara S, Yoshida T & Kikuchi G (1979) Mechanism of Increase of Heme Oxygenase Activity Induced by Hemin in Cultured Pig Alveolar Macrophages. *Archives of Biochemistry and Biophysics* **197**, 607-617.
9. Takeda A, Onodera H, Sugimoto A, Itoyama Y, Kogure K & Shibahara S (1994) Increased expression of heme oxygenase mRNA in rat brain following transient forebrain ischemia. *Brain Res* **666**, 120-124.
10. Taketani S (2005) Aquisition, mobilization and utilization of cellular iron and heme: endless findings and growing evidence of tight regulation. *Tohoku J Exp Med* **205**, 297-318.
11. Wang J & Pantopoulos K (2011) Regulation of cellular iron metabolism. *Biochemical Journal* **434**, 365-381.
12. Immenschuh S, Baumgart-Vogt E, Tan M, Iwahara S, Ramadori G & Fahimi HD (2003) Differential cellular and subcellular localization of heme-binding protein 23/peroxiredoxin I and heme oxygenase-1 in rat liver. *J Histochem Cytochem* **51**, 1621-1631.
13. Taketani S, Adachi Y, Kohno H, Ikehara S, Tokunaga R & Ishii T (1998) Molecular characterization of a newly identified heme-binding protein induced during differentiation of urine erythroleukemia cells. *J Biol Chem* **273**, 31388-31394.
14. Babusiak M, Man P, Sutak R, Petrak J & Vyoral D (2005) Identification of heme binding protein complexes in murine erythroleukemic cells: Study by a novel two-dimensional native separation - liquid chromatography and electrophoresis. *Proteomics* **5**, 340-350.
15. Zhou S, Zong Y, Ney PA, Nair G, Stewart CF & Sorrentino BP (2005) Increased expression of the Abcg2 transporter during erythroid maturation plays a role in decreasing cellular protoporphyrin IX levels. *Blood* **105**, 2571-2576.
16. Scortegagna M, Morris MA, Oktay Y, Bennett M & Garcia JA (2003) The HIF family member EPAS1/HIF-2alpha is required for normal hematopoiesis in mice. *Blood* **102**, 1634-1640.
17. Zylka MJ & Reppert SM (1999) Discovery of a putative heme-binding protein family (SOUL/HBP) by two-tissue suppression subtractive hybridization and database searches. *Brain Res Mol Brain Res* **74**, 175-181.

18. Dias JS, Macedo AL, Ferreira GC, Peterson FC, Volkman BF & Goodfellow BJ (2006) The first structure from the SOUL/HBP family of heme-binding proteins, murine P22HBP. *J Biol Chem* **281**, 31553-31561.
19. Micaelo NM, Macedo AL, Goodfellow BJ & Felix V (2010) Tetrapyrrole binding affinity of the murine and human p22HBP heme-binding proteins. *J Mol Graph Model* **29**, 396-405.
20. Sato E, Sagami I, Uchida T, Sato A, Kitagawa T, Igarashi J & Shimizu T (2004) SOUL in mouse eyes is a new hexameric heme-binding protein with characteristic optical absorption, resonance Raman spectral, and heme-binding properties. *Biochemistry* **43**, 14189-14198.
21. Lathrop JT & Timko MP (1993) Regulation by heme of mitochondrial protein transport through a conserved amino acid motif. *Science* **259**, 522-525.
22. Shin WS, Yamashita H & Hirose M (1994) Multiple effects of haemin binding on protease susceptibility of bovine serum albumin and a novel isolation procedure for its large fragment. *Biochemical Journal* **304 (Pt 1)**, 81-86.
23. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, et al. (2007) Clustal W and clustal X version 2.0. *Bioinformatics* **23**, 2947-2948.
24. Wagner V, Ullmann K, Mollwo A, Kaminski M, Mittag M & Kreimer G (2008) The phosphoproteome of a Chlamydomonas reinhardtii eyespot fraction includes key proteins of the light signaling pathway. *Plant Physiol* **146**, 772-788.
25. Rolland N, Atteia A, Decottignies P, Garin J, Hippler M, Kreimer G, Lemaire SD, Mittag M & Wagner V (2009) Chlamydomonas proteomics. *Curr Opin Microbiol* **12**, 285-291.
26. Bohn H & Winckler W (1991) Isolation and Characterization of 5 New Soluble Placental Tissue Proteins (Pp22, Pp23, Pp24, Pp25, Pp26). *Arch Gynecol Obstet* **248**, 111-115.
27. Gianazza E, Wait R, Begum S, Eberini I, Campagnoli M, Labo S & Galliano M (2007) Mapping the 5-50-kDa fraction of human amniotic fluid proteins by 2-DE and ESI-MS. *Proteomics Clin Appl* **1**, 167-175.
28. Hermes-Lima M (1995) How do Ca2+ and 5-aminolevulinic acid-derived oxyradicals promote injury to isolated mitochondria? *Free Radic Biol Med* **19**, 381-390.
29. Kroemer G & Reed JC (2000) Mitochondrial control of cell death. *Nat Med* **6**, 513-519.
30. Green DR & Reed JC (1998) Mitochondria and apoptosis. *Science* **281**, 1309-1312.
31. Szigeti A, Bellyei S, Gasz B, Boronkai A, Hocsak E, Minik O, Bognar Z, Varbiro G, Sumegi B & Gallyas F, Jr. (2006) Induction of necrotic cell death and mitochondrial permeabilization by heme binding protein 2/SOUL. *FEBS Lett* **580**, 6447-6454.
32. Szigeti A, Hocsak E, Rapolti E, Racz B, Boronkai A, Pozsgai E, Debreceni B, Bognar Z, Bellyei S, Sumegi B, et al. (2010) Facilitation of mitochondrial outer and inner membrane permeabilization and cell death in oxidative stress by a novel Bcl-2 homology 3 domain protein. *J Biol Chem* **285**, 2140-2151.
33. Maxwell J, Marinou I, Gowers I, Hinks A, Donn R, Worthington J & Wilson AG (2008) HEBP2 and IFNGR1, but not TNFAIP3 are differentially expressed in rheumatoid arthritis compared to healthy controls. *Rheumatology* **47**, Ii60-Ii60.
34. Ambrosi E, Capaldi S, Bovi M, Saccomani G, Perduca M & Monaco HL (2011) Structural changes in the BH3 domain of SOUL protein upon interaction with the anti-apoptotic protein Bcl-xL. *Biochem J* **438**, 291-301.
35. Taketani S, Adachi Y, Kohno H, Ikehara S, Tokunaga R & Ishii T (1998) Molecular characterization of a newly identified heme-binding protein induced during differentiation of urine erythroleukemia cells. *Journal of Biological Chemistry* **273**, 31388-31394.

36. Gell DA, Westman BJ, Gorman D, Liew CK, Liew JJ, Weiss MJ & Mackay JP (2006) A novel haem-binding interface in the 22 kDa haem-binding protein p22HBP. *Journal of Molecular Biology* **362**, 287-297.
37. Babusiak M, Man P, Sutak R, Petrak J & Vyoral D (2005) Identification of heme binding protein complexes in murine erythroleukemic cells: study by a novel two-dimensional native separation -- liquid chromatography and electrophoresis. *Proteomics* **5**, 340-350.
38. Lee HY, Lee SY, Shin EH, Kim SD, Kim JM, Lee MS, Ryu SH & Bae YS (2007) F2L, a peptide derived from heme-binding protein, inhibits formyl peptide receptor-mediated signaling. *Biochem Biophys Res Commun* **359**, 985-990.
39. Romanowski MJ, Gibney SA & Burley SK (2002) Crystal structure of the Escherichia coli SbmC protein that protects cells from the DNA replication inhibitor microcin B17. *Proteins* **47**, 403-407.
40. Kwon HJ, Bennik MHJ, Demple B & Ellenberger T (2000) Crystal structure of the Escherichia coli Rob transcription factor in complex with DNA. *Nat Struct Biol* **7**, 424-430.
41. Zheleznova EE, Markham PN, Neyfakh AA & Brennan RG (1999) Structural basis of multidrug recognition by BmrR, a transcription activator of a multidrug transporter. *Cell* **96**, 353-362.
42. Newberry KJ & Brennan RG (2004) The structural mechanism for transcription activation by MerR family member multidrug transporter activation, N terminus. *J Biol Chem* **279**, 20356-20362.
43. Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen MY, Pieper U & Sali A (2006) Comparative protein structure modeling using Modeller. *Curr Protoc Bioinformatics* **Chapter 5**, Unit 5 6.
44. Takahashi S, Ogawa T, Inoue K & Masuda T (2008) Characterization of cytosolic tetrapyrrole-binding proteins in Arabidopsis thaliana. *Photochem Photobiol Sci* **7**, 1216-1224.
45. Taylor RC, Cullen SP & Martin SJ (2008) Apoptosis: controlled demolition at the cellular level. *Nat Rev Mol Cell Biol* **9**, 231-241.
46. Edlich F, Banerjee S, Suzuki M, Cleland MM, Arnoult D, Wang CX, Neutzner A, Tjandra N & Youle RJ (2011) Bcl-x(L) Retrotranslocates Bax from the Mitochondria into the Cytosol. *Cell* **145**, 104-116.
47. Petros AM, Olejniczak ET & Fesik SW (2004) Structural biology of the Bcl-2 family of proteins. *Bba-Mol Cell Res* **1644**, 83-94.
48. Adams JM & Cory S (1998) The Bcl-2 protein family: Arbiters of cell survival. *Science* **281**, 1322-1326.
49. Vaux DJT, Helenius A & Mellman I (1988) Spike Nucleocapsid Interaction in Semliki Forest Virus Reconstructed Using Network Antibodies. *Nature* **336**, 36-42.
50. Hunter JJ & Parslow TG (1996) A peptide sequence from Bax that converts Bcl-2 into an activator of apoptosis. *J Biol Chem* **271**, 8521-8524.
51. Huang DC, Adams JM & Cory S (1998) The conserved N-terminal BH4 domain of Bcl-2 homologues is essential for inhibition of apoptosis and interaction with CED-4. *EMBO J* **17**, 1029-1039.
52. Borner C, Martinou I, Mattmann C, Irmler M, Schaerer E, Martinou JC & Tschopp J (1994) The protein bcl-2 alpha does not require membrane attachment, but two conserved domains to suppress apoptosis. *J Cell Biol* **126**, 1059-1068.
53. Shimizu S, Konishi A, Kodama T & Tsujimoto Y (2000) BH4 domain of antiapoptotic Bcl-2 family members closes voltage-dependent anion channel and inhibits apoptotic mitochondrial changes and cell death. *Proc Natl Acad Sci U S A* **97**, 3100-3105.

54. Conradt B & Horvitz HR (1998) The C. elegans protein EGL-1 is required for programmed cell death and interacts with the Bcl-2-like protein CED-9. *Cell* **93**, 519-529.
55. Hetz C & Glimcher L (2008) The daily job of night killers: alternative roles of the BCL-2 family in organelle physiology. *Trends in Cell Biology* **18**, 38-44.
56. Scatizzi JC, Bickel E, Hutcheson J, Haines GK & Perlman H (2006) Bim deficiency leads to exacerbation and prolongation of joint inflammation in experimental arthritis. *Arthritis Rheum* **54**, 3182-3193.
57. Scatizzi JC, Hutcheson J, Bickel E, Haines GK & Perlman H (2007) Pro-apoptotic Bid is required for the resolution of the effector phase of inflammatory arthritis. *Arthritis Res Ther* **9**, -.
58. Lomonosova E & Chinnadurai G (2008) BH3-only proteins in apoptosis and beyond: an overview. *Oncogene* **27 Suppl 1**, S2-19.
59. Hulo N, Bairoch A, Bulliard V, Cerutti L, De Castro E, Langendijk-Genevaux PS, Pagni M & Sigrist CJA (2006) The PROSITE database. *Nucleic Acids Res* **34**, D227-230.
60. Lepre CA, Moore JM & Peng JW (2004) Theory and applications of NMR-based screening in pharmaceutical research. *Chem Rev* **104**, 3641-3676.
61. Rupp B (2010) *Biomolecular crystallography : principles, practice, and application to structural biology*. Garland Science, New York.
62. Bernal JD & Crowfoot D (1934) X-ray photographs of crystalline pepsin. *Nature* **133**, 794-795.
63. Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff H & Phillips DC (1958) A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* **181**, 662-666.
64. Papoyan VV, Pervushin VN & Smirichinski VI (1998) Conformal quantum cosmology: Integrable models and Friedmann observables. *Phys Atom Nucl+* **61**, 1908-1913.
65. Linding R, Russell RB, Neduva V & Gibson TJ (2003) GlobPlot: Exploring protein sequences for globularity and disorder. *Nucleic Acids Res* **31**, 3701-3708.
66. Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, Oldfield CJ, Campen AM, Ratliff CM, Hipps KW, et al. (2001) Intrinsically disordered protein. *J Mol Graph Model* **19**, 26-59.
67. Linding R, Jensen LJ, Diella F, Bork P, Gibson TJ & Russell RB (2003) Protein disorder prediction: implications for structural proteomics. *Structure* **11**, 1453-1459.
68. Nettleship JE, Brown J, Groves MR & Geerlof A (2008) Methods for protein characterization by mass spectrometry, thermal shift (ThermoFluor) assay, and multiangle or static light scattering. *Methods Mol Biol* **426**, 299-318.
69. Wang BC (1985) Resolution of phase ambiguity in macromolecular crystallography. *Methods Enzymol* **115**, 90-112.
70. Chen YW (2001) Solution solution: using NMR models for molecular replacement. *Acta Crystallogr D* **57**, 1457-1461.
71. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC & Read RJ (2007) Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674.
72. Winn MD, Isupov MN & Murshudov GN (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr D* **57**, 122-133.
73. Painter J & Merritt EA (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr D Biol Crystallogr* **62**, 439-450.
74. Jacobson B, Anderson WA & Arnold JT (1954) A Proton Magnetic Resonance Study of the Hydration of Deoxyribonucleic Acid. *Nature* **173**, 772-773.

75. Saunders M, Wishnia A & Kirkwood JG (1957) The Nuclear Magnetic Resonance Spectrum of Ribonuclease. *Journal of the American Chemical Society* **79**, 3289-3290.
76. Williamson MP, Havel TF & Wuthrich K (1985) Solution conformation of proteinase inhibitor IIA from bull seminal plasma by 1H nuclear magnetic resonance and distance geometry. *J Mol Biol* **182**, 295-315.
77. Rule GS & Hitchens TK (2006) *Fundamentals of protein NMR spectroscopy*. Springer, Dordrecht.
78. Cavanagh J (2007) *Protein NMR spectroscopy : principles and practice*. 2nd edn. Academic Press, Amsterdam ; Boston.
79. Grzesiek S & Sass HJ (2009) From biomolecular structure to functional understanding: new NMR developments narrow the gap. *Curr Opin Struc Biol* **19**, 585-595.
80. Felli IC & Brutscher B (2009) Recent advances in solution NMR: fast methods and heteronuclear direct detection. *Chemphyschem* **10**, 1356-1368.
81. Wuthrich K (1990) Protein structure determination in solution by NMR spectroscopy. *J Biol Chem* **265**, 22059-22062.
82. Braun W, Bosch C, Brown LR, Go N & Wuthrich K (1981) Combined use of proton-proton Overhauser enhancements and a distance geometry algorithm for determination of polypeptide conformations. Application to micelle-bound glucagon. *Biochim Biophys Acta* **667**, 377-396.
83. Fesik SW & Zuiderweg ER (1990) Heteronuclear three-dimensional NMR spectroscopy of isotopically labelled biological macromolecules. *Q Rev Biophys* **23**, 97-131.
84. Clore GM & Gronenborn AM (1991) Structures of larger proteins in solution: three- and four-dimensional heteronuclear NMR spectroscopy. *Science* **252**, 1390-1399.
85. Grzesiek S & Bax A (1993) Amino acid type determination in the sequential assignment procedure of uniformly 13C/15N-enriched proteins. *J Biomol NMR* **3**, 185-204.
86. McIntosh LP & Dahlquist FW (1990) Biosynthetic incorporation of 15N and 13C for assignment and interpretation of nuclear magnetic resonance spectra of proteins. *Q Rev Biophys* **23**, 1-38.
87. Griswold IJ & Dahlquist FW (2002) Bigger is better: megadalton protein NMR in solution. *Nat Struct Biol* **9**, 567-568.
88. Gardner KH & Kay LE (1998) The use of 2H, 13C, 15N multidimensional NMR to study the structure and dynamics of proteins. *Annu Rev Biophys Biomol Struct* **27**, 357-406.
89. Billeter M, Wagner G & Wuthrich K (2008) Solution NMR structure determination of proteins revisited. *J Biomol NMR* **42**, 155-158.
90. Wishart DS, Sykes BD & Richards FM (1991) Relationship between Nuclear-Magnetic-Resonance Chemical-Shift and Protein Secondary Structure. *Journal of Molecular Biology* **222**, 311-333.
91. Fernandez C & Wider G (2003) TROSY in NMR studies of the structure and function of large biological macromolecules. *Curr Opin Struct Biol* **13**, 570-580.
92. Pervushin K, Riek R, Wider G & Wuthrich K (1997) Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci U S A* **94**, 12366-12371.
93. Kanelis V, Forman-Kay JD & Kay LE (2001) Multidimensional NMR methods for protein structure determination. *IUBMB Life* **52**, 291-302.
94. Teng Q & SpringerLink (Online service) (2005) Structural Biology: Practical NMR Applications. In. Springer Science+Business Media, Inc., Boston, MA.

95. Tjandra N, Kuboniwa H, Ren H & Bax A (1995) Rotational dynamics of calcium-free calmodulin studied by 15N-NMR relaxation measurements. *Eur J Biochem* **230**, 1014-1024.
96. Grzesiek S & Bax A (1993) The Importance of Not Saturating H2O in Protein Nmr - Application to Sensitivity Enhancement and Noe Measurements. *Journal of the American Chemical Society* **115**, 12593-12594.
97. Skelton NJ, Palmer AG, Akke M, Kordel J, Rance M & Chazin WJ (1993) Practical Aspects of 2-Dimensional Proton-Detected N-15 Spin Relaxation Measurements. *J Magn Reson Ser B* **102**, 253-264.
98. Farrow NA, Muhandiram R, Singer AU, Pascal SM, Kay CM, Gish G, Shoelson SE, Pawson T, Forman-Kay JD & Kay LE (1994) Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. *Biochemistry* **33**, 5984-6003.
99. Cavanagh J (2007) Protein NMR spectroscopy principles and practice. In, pp. xxv, 885 p. Academic Press, Burlington, Mass. ; London.
100. Marley J, Lu M & Bracken C (2001) A method for efficient isotopic labeling of recombinant proteins. *J Biomol NMR* **20**, 71-75.
101. Stockman BJ, Reily MD, Westler WM, Ulrich EL & Markley JL (1989) Concerted Two-Dimensional Nmr Approaches to H-1, C-13, and N-15 Resonance Assignments in Proteins. *Biochemistry* **28**, 230-236.
102. Simplaceanu V, Lukin JA, Fang TY, Zou M, Ho NT & Ho C (2000) Chain-selective isotopic labeling for NMR studies of large multimeric proteins: application to hemoglobin. *Biophys J* **79**, 1146-1154.
103. Hendrickson WA, Horton JR & LeMaster DM (1990) Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J* **9**, 1665-1672.
104. Freire F, Romao MJ, Macedo AL, Aveiro SS, Goodfellow BJ & Carvalho AL (2009) Preliminary structural characterization of human SOUL, a haem-binding protein. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **65**, 723-726.
105. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J & Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**, 277-293.
106. Matthews BW (1968) Solvent content of protein crystals. *J Mol Biol* **33**, 491-497.
107. Leslie AGW (1992) Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 and ESF-EACBM Newsletters on Protein Crystallography* **26**.
108. (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50**, 760-763.
109. Ravelli RBG, Sweet RM, Skinner JM, Duisenberg AJM & Kroon J (1997) STRATEGY: a program to optimize the starting spindle angle and scan range for X-ray data collection. *J Appl Crystallogr* **30**, 551-554.
110. Evans P (2006) Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* **62**, 72-82.
111. Vagin A & Teplyakov A (1997) MOLREP: an automated program for molecular replacement. *J Appl Cryst* **30**, 1022-1025.
112. Long F, Vagin AA, Young P & Murshudov GN (2008) BALBES: a molecular-replacement pipeline. *Acta Crystallogr D Biol Crystallogr* **64**, 125-132.
113. Stein N (2008) CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *J Appl Crystallogr* **41**, 641-643.
114. Storoni LC, McCoy AJ & Read RJ (2004) Likelihood-enhanced fast rotation functions. *Acta Crystallogr D Biol Crystallogr* **60**, 432-438.

115. Terwilliger TC & Berendzen J (1999) Automated MAD and MIR structure solution. *Acta Cryst* **D55**, 849-861.
116. Terwilliger TC, Adams PD, Read RJ, McCoy AJ, Moriarty NW, Grosse-Kunstleve RW, Afonine PV, Zwart PH & Hung LW (2009) Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallogr D Biol Crystallogr* **65**, 582-601.
117. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221.
118. Emsley P, Lohkamp B, Scott WG & Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501.
119. Vagin AA, Steiner RA, Lebedev AA, Potterton L, McNicholas S, Long F & Murshudov GN (2004) REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr D Biol Crystallogr* **60**, 2184-2195.
120. Winn MD, Isupov MN & Murshudov GN (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Cryst D* **57**, 122-133.
121. Terwilliger T (2004) SOLVE and RESOLVE: automated structure solution, density modification, and model building. *Journal of Synchrotron Radiation* **11**, 49-52.
122. Krissinel E & Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**, 774-797.
123. Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS & Richardson DC (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21.
124. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC & Ferrin TE (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612.
125. Hasegawa H & Holm L (2009) Advances and pitfalls of protein structural alignment. *Curr Opin Struct Biol* **19**, 341-348.
126. Baker NA, Sept D, Joseph S, Holst MJ & McCammon JA (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* **98**, 10037-10041.
127. Gibrat JF, Madej T & Bryant SH (1996) Surprising similarities in structure comparison. *Curr Opin Struct Biol* **6**, 377-385.
128. Romanowski MJ, Gibney SA & Burley SK (2002) Crystal structure of the Escherichia coli SbmC protein that protects cells from the DNA replication inhibitor microcin B17. *Proteins* **47**, 403-407.
129. Kwon HJ, Bennik MH, Demple B & Ellenberger T (2000) Crystal structure of the Escherichia coli Rob transcription factor in complex with DNA. *Nat Struct Biol* **7**, 424-430.
130. Ku B, Liang C, Jung JU & Oh BH (2011) Evidence that inhibition of BAX activation by BCL-2 involves its tight and preferential interaction with the BH3 domain of BAX. *Cell Res* **21**, 627-641.
131. Giam M, Huang DCS & Bouillet P (2009) BH3-only proteins and their roles in programmed cell death. *Oncogene* **27 Suppl 1**, S128-136.
132. Hinds MG, Smits C, Fredericks-Short R, Risk JM, Bailey M, Huang DCS & Day CL (2007) Bim, Bad and Bmf: intrinsically unstructured BH3-only proteins that undergo a localized conformational change upon binding to prosurvival Bcl-2 targets. *Cell Death Differ* **14**, 128-136.

133. Stewart ML, Fire E, Keating AE & Walensky LD (2010) The MCL-1 BH3 helix is an exclusive MCL-1 inhibitor and apoptosis sensitizer. *Nat Chem Biol* **6**, 595-601.
134. Gavathiotis E, Reyna DE, Davis ML, Bird GH & Walensky LD (2010) BH3-triggered structural reorganization drives the activation of proapoptotic BAX. *Mol Cell* **40**, 481-492.
135. Zhang L & Guarente L (1995) Heme binds to a short sequence that serves a regulatory function in diverse proteins. *EMBO J* **14**, 313-320.
136. Bodenhausen G & Ruben DJ (1980) Natural Abundance N-15 Nmr by Enhanced Heteronuclear Spectroscopy. *Chem Phys Lett* **69**, 185-189.
137. Piotto M, Saudek V & Sklenar V (1992) Gradient-Tailored Excitation for Single-Quantum Nmr-Spectroscopy of Aqueous-Solutions. *Journal of Biomolecular Nmr* **2**, 661-665.
138. Sklenar V, Piotto M, Leppik R & Saudek V (1993) Gradient-Tailored Water Suppression for H-1-N-15 Hsqc Experiments Optimized to Retain Full Sensitivity. *J Magn Reson Ser A* **102**, 241-245.
139. Mori S, Abeygunawardana C, Johnson MO & Vanzijl PCM (1995) Improved Sensitivity of Hsqc Spectra of Exchanging Protons at Short Interscan Delays Using a New Fast Hsqc (Fhsqc) Detection Scheme That Avoids Water Saturation. *J Magn Reson Ser B* **108**, 94-98.
140. Czisch M & Boelens R (1998) Sensitivity enhancement in the TROSY experiment. *Journal of Magnetic Resonance* **134**, 158-160.
141. Pervushin KV, Wider G & Wuthrich K (1998) Single transition-to-single transition polarization transfer (ST2-PT) in [N-15,H-1]-TROSY. *Journal of Biomolecular Nmr* **12**, 345-348.
142. Meissner A, Schulte-Herbruggen T, Briand J & Sorensen OW (1998) Double spin-state-selective coherence transfer. Application for two-dimensional selection of multiplet components with long transverse relaxation times. *Mol Phys* **95**, 1137-1142.
143. Weigelt J (1998) Single scan, sensitivity- and gradient-enhanced TROSY for multidimensional NMR experiments. *Journal of the American Chemical Society* **120**, 10778-10779.
144. Rance M, Loria JP & Palmer AG (1999) Sensitivity improvement of transverse relaxation-optimized spectroscopy. *Journal of Magnetic Resonance* **136**, 92-101.
145. Zhu G, Kong XM & Sze KH (1999) Gradient and sensitivity enhancement of 2D TROSY with water flip-back, 3D NOESY-TROSY and TOCSY-TROSY experiments. *Journal of Biomolecular Nmr* **13**, 77-81.
146. Gruber O & Bosch HS & Gunter S & Herrmann A & Kallenbach A & Kaufmann M & Krieger K & Lackner K & V M & Neu R, et al. (1999) Overview of ASDEX Upgrade results. *Nucl Fusion* **39**, 1321-1336.
147. Schweinzer J, Sandmann W, Haas G, Neuhauser J, Murmann H, Salzmann H, ASDEX-Upgrade-Team & NBI-Team (1999) Comparison of scrape-off layer behaviour between DIV-I and DIV-II operations on ASDEX-Upgrade. *J Nucl Mater* **266**, 934-939.
148. Waller M & Salzmann U (1999) Holocene vegetation changes in the Sahelian zone of NE Nigeria: The detection of anthropogenic activity. *Palaeoeco A* **26**, 85-102.
149. Salzmann M, Wider G, Pervushin K, Senn H & Wuthrich K (1999) TROSY-type triple-resonance experiments for sequential NMR assignments of large proteins. *Journal of the American Chemical Society* **121**, 844-848.
150. Eletsky A, Kienhofer A & Pervushin K (2001) TROSY NMR with partially deuterated proteins. *Journal of Biomolecular Nmr* **20**, 177-180.

151. Mulder FA, van Tilborg PJ, Kaptein R & Boelens R (1999) Microsecond time scale dynamics in the RXR DNA-binding domain from a combination of spin-echo and off-resonance rotating frame relaxation measurements. *J Biomol NMR* **13**, 275-288.
152. Farrow NA, Zhang O, Szabo A, Torchia DA & Kay LE (1995) Spectral density function mapping using 15N relaxation data exclusively. *J Biomol NMR* **6**, 153-162.
153. Latchman HA, Salzmann C, Gillet D & Bouzekri H (1999) Information technology enhanced learning in distance and conventional education. *Ieee T Educ* **42**, 247-254.
154. Schirra HJ, Renner C, Czisch M, Huber-Wunderlich M, Holak TA & Glockshuber R (1998) Structure of reduced DsbA from Escherichia coli in solution. *Biochemistry* **37**, 6263-6276.
155. Salzmann J, Limb GA, Charteris DG, Gregor Z & Luthert P (1999) Selective increase in TIMP-2 expression by retinal membranes of proliferative diabetic retinopathy. *Invest Ophth Vis Sci* **40**, S465-S465.
156. Dal Ponte D, Berman SS, Williams SK, Kleinert LB, Salzmann DL & Patula VB (1999) The healing characteristics of low-profile talent endoluminal grafts. *J Endovasc Surg* **6**, 84-84.
157. Ellmer K, Salzmann C, Bruns J & Klaer J (1999) Formation of polycrystalline CuInS2-films by plasma sulphurization of Cu-In bilayers in a magnetron discharge in an Ar/H2S-atmosphere. *Solid State Phenom* **67-8**, 379-384.
158. Keller R (2004) The Computer Aided Resonance Assignment Tutorial. In *The Swiss Federal Institute of Technology*. The Swiss Federal Institute of Technology.
159. Goddard TD & Kneller DG SPARKY 3. In.
160. Swillens S (1995) Interpretation of binding curves obtained with high receptor concentrations: practical aid for computer analysis. *Mol Pharmacol* **47**, 1197-1203.
161. Bahrami A, Assadi AH, Markley JL & Eghbalnia HR (2009) Probabilistic Interaction Network of Evidence Algorithm and its Application to Complete Labeling of Peak Lists from Protein NMR Spectroscopy. *Plos Comput Biol* **5**.
162. Shen Y, Delaglio F, Cornilescu G & Bax A (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR* **44**, 213-223.
163. Chi YH, Kumar TKS, Chiu IM & Yu C (2000) N-15 NMR relaxation studies of free and ligand-bound human acidic fibroblast growth factor. *Journal of Biological Chemistry* **275**, 39444-39450.
164. Jarymowycz VA & Stone MJ (2006) Fast time scale dynamics of protein backbones: NMR relaxation methods, applications, and functional consequences. *Chem Rev* **106**, 1624-1671.
165. Tjandra N, Feller SE, Pastor RW & Bax A (1995) Rotational diffusion anisotropy of human ubiquitin from N-15 NMR relaxation. *Journal of the American Chemical Society* **117**, 12562-12566.
166. Stone MJ (2001) NMR relaxation studies of the role of conformational entropy in protein stability and ligand binding. *Acc Chem Res* **34**, 379-388.
167. Goodfellow BJ, Dias JS, Ferreira GC, Henklein P, Wray V & Macedo AL (2001) The solution structure and heme binding of the presequence of murine 5-aminolevulinate synthase. *FEBS Lett* **505**, 325-331.
168. Garboczi DN, Hullihen JH & Pedersen PL (1988) Mitochondrial ATP synthase. Overexpression in Escherichia coli of a rat liver beta subunit peptide and its interaction with adenine nucleotides. *J Biol Chem* **263**, 15694-15698.

APPENDIX

M9 minimal medium (1 L)

Add 100 ml of 10 × M9 salts to 880 ml of deionized H_2O.

M9 Salts (10×) 1L

128g $Na_2HPO_4.7H_2O$

30g KH_2PO_4

10g NH_4Cl

5g NaCl

10 × M9 salts are prepared in deionized H_2O.

MOPS media (1 L)

745 mL deionized H_2O

200 mL M solution

2 mL O solution

0.1 mL P solution

1 mL S solution

20 mL glucose 20 % (w/v)

M Solution (1 L)

42 g MOPS

4 g Tricine

14.6 g NaCl

8 g KOH

2.55 g NH_4Cl

Adjust pH to 7.3 – 7.4

Add deionized H_2O to a final volume of 1 L.

T solution (100 mL)

8 mL HCl concentrated

18.4 mg $CaCl_2$

64 mg H_3BO_3

40 mg $MnCl_2.4H_2O$

18 mg $CoCl_2.6H_2O$

4 mg $CuCl_2.2H_2O$

340 mg $ZnCl_2$

605 mg $Na_2MoO_4.2H_2O$

Add deionized H_2O to a final volume of 100 mL.

O solution (50 mL)

Dissolve 0.1 g $FeCl_2.4H_2O$ in 10 mL concentrated HCl

Add 10 mL H_2O

Add 1 mL of T solution

Add 2.68 g $MgCl_2.6H_2O$

Add deionized H_2O to a final volume of 50 mL, filter sterilize and store at room temperature.

P solution

KH_2PO_4 1.0 M

S solution

K_2SO_4 0.276 M

Appendix

Table A1 *In-house sparse matrix screen (80!).*

Condition	
1	0.2 M Calcium chloride, 0.1 M Acetate buffer 4.5, 30 % 2-Methyl-2,4-pentanediol
2	1M K/Na tartrate, 0.1 M MES 6.5
3	0.4 M Ammonium phosphate
4	0.1 M Tris/HCl 8.5, 3M Ammonium sulphate
5	0.2 M Trisodium citrate, 0.1M HEPES 7.5, 30 % 2-Methyl-2,4-pentanediol
6	0.2 M Magnesium chloride, 0.1 M Acetate buffer 4.5, 30 % PEG 3350
7	1.2 M Sodium citrate, 0.1 M HEPES buffer 7.5
8	0.2M Trisodium citrate, 2 M Ammonium sulphate
9	0.2 M Ammonium acetate, 0.1 M Citrate buffer 5.5, 30 % PEG 400
10	0.1 M Acetate buffer 4.5, 1.5 M Ammonium phosphate
11	0.2 M Ammonium sulphate, 0.1 M HEPES buffer 7.5, 1.5 M Potassium phosphate/1.5M sodium phosphate
12	0.2 M Trisodium citrate, 0.1 M Tris/HCl 8.5, 20 % PEG 400
13	0.2 M Calcium chloride, 0.1 M HEPES buffer 7.5, 25 % PEG 3350
14	0.1 M Magnesium chloride, 0.1 M MES buffer 6.5, 30 % PEG 8000
15	0.2 M Lithium sulphate, 0.1M Citrate buffer 5.5, 30 % PEG 3350
16	1 M Lithium sulphate, 0.1 M Acetate buffer 4.5
17	0.2 M Ammonium phosphate, 0.1 M Tris/HCl 7.5, 30 % 2-Methyl-2,4-pentanediol
18	0.2 M Ammonium acetate, 0.1 M Tris/HCl 7.5, 1.5 M Potassium phosphate/1.5M sodium phosphate
19	0.1M Ammonium sulphate, 0.1 M Citrate buffer 5.5, 30 % PEG 8000
20	0.1 M MES buffer 6.5, 30% 2-Methyl-2,4-pentanediol
21	0.2 M Magnesium chloride, 0.1 M HEPES buffer 7.5, 30 % PEG 3350
22	0.2 M Sodium acetate, 0.1 M Tris/HCl 8.5, 30 % PEG 3350
23	0.1 M Tris/HCl 7.5, 1M K/Na tartrate
24	0.2 M Calcium chloride, 0.1 M Tris/HCl 8.5
25	0.5 M Ammonium acetate, 0.1 M Citrate buffer 5.5, 30 % 2-Methyl-2,4-pentanediol
26	2 M Sodium acetate. 0.1 M MES buffer 6.5
27	0.2 M K/Na tartrate, 0.1 M MES 6.5, 30 % PEG 8000
28	1 M K/Na Tartrate, 0.1 M HEPES 7.5
29	0.2 M Ammonium sulphate, 0.1 M Acetate buffer 4.5, 30 % PEG 400
30	0.1 M Ammonium sulphate, 0.1 M HEPES buffer 7.5, 20 % PEG 3350
31	2 M Ammonium sulphate, 0.1 M MES buffer 6.5
32	0.2 M Sodium chloride, 0.1 M MES 6.5, 30 % Ethanol

33	0.2 M Magnesium chloride, 0.1 M HEPES buffer 7.5, 30 % Ethanol
34	0.2 M Ammonium acetate, 0.1 M Tris/HCl 8.5, 30 % Ethanol
35	0.2 M Calcium chloride, 0.1 M Acetate buffer 4.5, 30 % Ethanol
36	0.2 M Sodium acetate, 0.1 M HEPES buffer 7.5, 30 % Ethanol
37	0.2 M Magnesium chloride, 0.1 M HEPES 7.5, 30 % Isopropanol
38	0.1 M Cacodilate buffer 6.5, 30 % 2-Methyl-2,4-pentanediol
39	0.1 M Acetate buffer 4.5, 2 M Sodium formate
40	0.2 M Trisodium citrate, 0.1 M Cacodylate buffer 6.5, 40 % isopropanol
41	0.1 M HEPES buffer 7.5, 20 % PEG 400, 10 % Isopropanol
42	0.1 M HEPES 7.5, 1M Lithium sulphate
43	0.2 M Lithium sulphate, 0.1 M Tris/HCl 8.5, 30 % PEG 3350
44	0.2 M Ammonium sulphate, 0.1 M Cacodylate buffer 6.5, 30 % PEG 6000
45	0.1 Acetate buffer 4.5, 1.5 M Sodium acetate
46	0.1 M Trisodium citrate, 1M Ammonium phosphate
47	4 M Sodium formate
48	0.1 M HEPES buffer 7.5, 1.2 M Trisodium citrate
49	0.4 M K/Na tartrate
50	0.2 M Magnesium chloride, 0.1 M Tris/HCl 8.5, 30 % PEG 3350
51	0.1 M Cacodylate buffer 6.5, 1.4 M Sodium acetate
52	0.2 M Ammonium acetate, 0.1 M Citrate buffer 5.5, 30 % PEG 3350
53	0.2 M Ammonium acetate, 0.1 M Acetate buffer 4.5, 30 % PEG 3350
54	0.2 M Calcium chloride, 0.1 M HEPES buffer 7.5, 28 % PEG 400
55	0.2 M Ammonium sulphate, 0.1 M Cacodylate buffer 6.5, 30 % PEG 8000
56	0.2 M Magnesium acetate, 0.1 M Cacodylate buffer 6.5, 20 % PEG 8000
57	0.2 M Ammonium acetate, 0.1 M Tris/HCl 8.5, 30 % Isopropanol
58	0.2 M Ammonium sulphate, 0.1 M Acetate buffer 4.5, 25 % PEG 3350
59	0.2 M Magnesium acetate, 0.1 M Cacodylate buffer 6.5, 20 % 2-Methyl-2,4-pentanediol
60	0.2 M Calcium chloride, 0.1 Acetate buffer 4.5, 20 % Isopropanol
61	0.1 M Imidazole buffer 7.0, 20 % Isopropanol
62	0.2 M TRisodium citrate, 0.1 Cacodylate 6.5, 20 % Isopropanol
63	0.2 M Sodium acetate, 0.1 M Cacodylate 6.5, 30 % PEG 8000
64	0.2 M Ammonium sulphate, 30 % PEG 8000
65	0.2 M Ammonium sulphate. 30 % PEG 3350
66	0.1 M HEPES buffer 7.5, 1.6 M K/Na phosphate
67	0.1 M Tris/HCl 8.5, 8 % PEG 8000

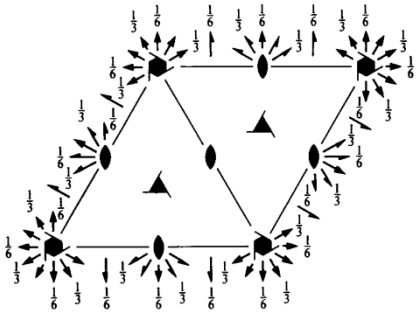
68	0.1 M Acetate buffer 4.5, 8 % PEG 3350
69	0.1 M HEPES buffer 7.5, 2% PEG 400, 2M Ammonium phosphate
70	0.1 M Citrate buffer 5.5, 20 % Isopropanol, 20 % PEG 3350
71	0.05 M Potassium phosphate, 20 % PEG 8 K
72	30 % PEG 8 K
73	0.2 M Magnesium formate
74	0.2 M Zinc acetate, 0.1 M Cacodylate buffer 6.5, 18 % PEG 8000
75	0.2 M Calcium acetate, 0.1 M Cacodylate 6.5, 18 % PEG 8000
76	0.1 M Acetate buffer 4.5, 2 M Ammonium sulphate
77	0.1 M Tris/HCl 8.5, 2 M Ammonium sulphate
78	1 M Lithium sulphate, 2 % PEG 8000
79	0.5 M Lithium sulphate, 15 % PEG 8000
80	0.2 M Ammonium acetate, 0.1 M Citrate buffer 5.5, 20 % Isopropanol, 20 % PEG 3350

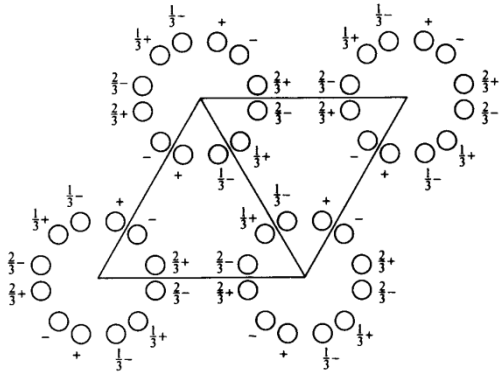
Space groups $p6_222$ and $p6_422$ (From the International Tables for Crystallography -Volume.A Space-Group Symmetry).

International Tables for Crystallography (2006). Vol. A, Space group 180, pp. 572–573.

$P6_222$ D_6^4 622 Hexagonal

No. 180 $P6_222$ Patterson symmetry $P6/mmm$





Origin at 222 at $6_2(2,1,1)\,(1,2,1)$

Asymmetric unit $0 \le x \le 1;\quad 0 \le y \le 1;\quad 0 \le z \le \frac{1}{6};\quad y \le x$

Vertices $0,0,0\quad 1,0,0\quad 1,1,0$

$0,0,\frac{1}{6}\quad 1,0,\frac{1}{6}\quad 1,1,\frac{1}{6}$

Symmetry operations

(1) 1	(2) $3^+(0,0,\frac{2}{3})$ $0,0,z$	(3) $3^-(0,0,\frac{1}{3})$ $0,0,z$
(4) 2 $0,0,z$	(5) $6^-(0,0,\frac{2}{3})$ $0,0,z$	(6) $6^+(0,0,\frac{1}{3})$ $0,0,z$
(7) 2 $x,x,\frac{1}{3}$	(8) 2 $x,0,0$	(9) 2 $0,y,\frac{1}{6}$
(10) 2 $x,\bar{x},\frac{1}{3}$	(11) 2 $x,2x,0$	(12) 2 $2x,x,\frac{1}{6}$

Generators selected (1); $t(1,0,0)$; $t(0,1,0)$; $t(0,0,1)$; (2); (4); (7)

Positions

Multiplicity, Wyckoff letter, Site symmetry			Coordinates						Reflection conditions
									General:
12	k	1	(1) x,y,z (4) $\bar{x},\bar{y},z$ (7) $y,x,\bar{z}+\frac{2}{3}$ (10) $\bar{y},\bar{x},\bar{z}+\frac{2}{3}$	(2) $\bar{y},x-y,z+\frac{2}{3}$ (5) $y,\bar{x}+y,z+\frac{2}{3}$ (8) $x-y,\bar{y},\bar{z}$ (11) $\bar{x}+y,y,\bar{z}$	(3) $\bar{x}+y,\bar{x},z+\frac{1}{3}$ (6) $x-y,x,z+\frac{1}{3}$ (9) $\bar{x},\bar{x}+y,\bar{z}+\frac{1}{3}$ (12) $x,x-y,\bar{z}+\frac{1}{3}$				$000l$: $l=3n$
									Special: as above, plus
6	j	. . 2	$x,2x,\frac{1}{2}$	$2\bar{x},\bar{x},\frac{1}{6}$	$x,\bar{x},\frac{5}{6}$	$\bar{x},2\bar{x},\frac{1}{2}$	$2x,x,\frac{1}{6}$	$\bar{x},x,\frac{5}{6}$	no extra conditions
6	i	. . 2	$x,2x,0$	$2\bar{x},\bar{x},\frac{2}{3}$	$x,\bar{x},\frac{1}{3}$	$\bar{x},2\bar{x},0$	$2x,x,\frac{2}{3}$	$\bar{x},x,\frac{1}{3}$	no extra conditions
6	h	. 2 .	$x,0,\frac{1}{2}$	$0,x,\frac{1}{6}$	$\bar{x},\bar{x},\frac{5}{6}$	$\bar{x},0,\frac{1}{2}$	$0,\bar{x},\frac{1}{6}$	$x,x,\frac{5}{6}$	no extra conditions
6	g	. 2 .	$x,0,0$	$0,x,\frac{2}{3}$	$\bar{x},\bar{x},\frac{1}{3}$	$\bar{x},0,0$	$0,\bar{x},\frac{2}{3}$	$x,x,\frac{1}{3}$	no extra conditions
6	f	2 . .	$\frac{1}{2},0,z$	$0,\frac{1}{2},z+\frac{2}{3}$	$\frac{1}{2},\frac{1}{2},z+\frac{1}{3}$	$0,\frac{1}{2},\bar{z}+\frac{2}{3}$	$\frac{1}{2},0,\bar{z}$	$\frac{1}{2},\frac{1}{2},\bar{z}+\frac{1}{3}$	$hkil$: $h=2n+1$ or $k=2n+1$ or $l=3n$
6	e	2 . .	$0,0,z$	$0,0,z+\frac{2}{3}$	$0,0,z+\frac{1}{3}$	$0,0,\bar{z}+\frac{2}{3}$	$0,0,\bar{z}$	$0,0,\bar{z}+\frac{1}{3}$	$hkil$: $l=3n$
3	d	2 2 2	$\frac{1}{2},0,\frac{1}{2}$	$0,\frac{1}{2},\frac{1}{6}$	$\frac{1}{2},\frac{1}{2},\frac{5}{6}$				$hkil$: $h=2n+1$ or $k=2n+1$ or $l=3n$
3	c	2 2 2	$\frac{1}{2},0,0$	$0,\frac{1}{2},\frac{2}{3}$	$\frac{1}{2},\frac{1}{2},\frac{1}{3}$				$hkil$: $h=2n+1$ or $k=2n+1$ or $l=3n$
3	b	2 2 2	$0,0,\frac{1}{2}$	$0,0,\frac{1}{6}$	$0,0,\frac{5}{6}$				$hkil$: $l=3n$
3	a	2 2 2	$0,0,0$	$0,0,\frac{2}{3}$	$0,0,\frac{1}{3}$				$hkil$: $l=3n$

Symmetry of special projections

Along [001] $p6mm$
$\mathbf{a}'=\mathbf{a}$ $\mathbf{b}'=\mathbf{b}$
Origin at $0,0,z$

Along [100] $p2mm$
$\mathbf{a}'=\frac{1}{2}(\mathbf{a}+2\mathbf{b})$ $\mathbf{b}'=\mathbf{c}$
Origin at $x,0,0$

Along [210] $p2mm$
$\mathbf{a}'=\frac{1}{2}\mathbf{b}$ $\mathbf{b}'=\mathbf{c}$
Origin at $x,\frac{1}{2}x,\frac{1}{6}$

Maximal non-isomorphic subgroups

I [2] $P6_211$ ($P6_2$, 171) 1; 2; 3; 4; 5; 6
[2] $P3_221$ (154) 1; 2; 3; 7; 8; 9
[2] $P3_212$ (153) 1; 2; 3; 10; 11; 12
[3] $P222$ ($C222$, 21) 1; 4; 7; 10
[3] $P222$ ($C222$, 21) 1; 4; 8; 11
[3] $P222$ ($C222$, 21) 1; 4; 9; 12

IIa none

IIb [2] $P6_122$ ($\mathbf{c}'=2\mathbf{c}$) (178)

Maximal isomorphic subgroups of lowest index

IIc [2] $P6_422$ ($\mathbf{c}'=2\mathbf{c}$) (181); [3] $H6_222$ ($\mathbf{a}'=3\mathbf{a},\mathbf{b}'=3\mathbf{b}$) ($P6_222$, 180); [7] $P6_222$ ($\mathbf{c}'=7\mathbf{c}$) (180)

Minimal non-isomorphic supergroups

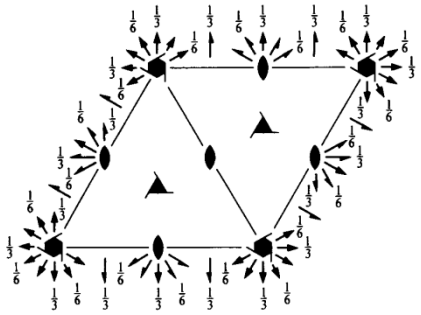
I none

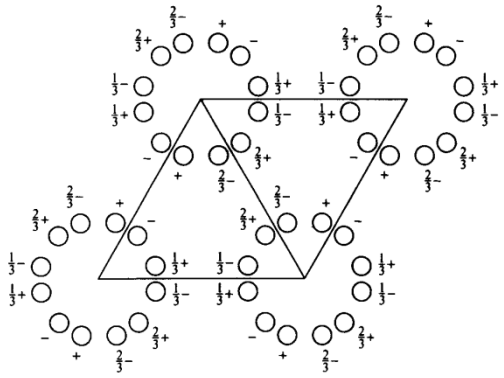
II [3] $P622$ ($\mathbf{c}'=\frac{1}{3}\mathbf{c}$) (177)

International Tables for Crystallography (2006). Vol. A, Space group 181, pp. 574–575.

$P6_422$ D_6^5 622 Hexagonal

No. 181 $P6_422$ Patterson symmetry $P6/mmm$





Origin at 2 2 2 at $6_4(2,1,1)\,(1,2,1)$

Asymmetric unit $0 \le x \le 1;\quad 0 \le y \le 1;\quad 0 \le z \le \frac{1}{6};\quad y \le x$

Vertices $0,0,0\quad 1,0,0\quad 1,1,0$

$0,0,\frac{1}{6}\quad 1,0,\frac{1}{6}\quad 1,1,\frac{1}{6}$

Symmetry operations

(1) 1	(2) $3^+(0,0,\frac{1}{3})$ $0,0,z$	(3) $3^-(0,0,\frac{2}{3})$ $0,0,z$
(4) 2 $0,0,z$	(5) $6^-(0,0,\frac{1}{3})$ $0,0,z$	(6) $6^+(0,0,\frac{2}{3})$ $0,0,z$
(7) 2 $x,x,\frac{1}{6}$	(8) 2 $x,0,0$	(9) 2 $0,y,\frac{1}{3}$
(10) 2 $x,\bar{x},\frac{1}{6}$	(11) 2 $x,2x,0$	(12) 2 $2x,x,\frac{1}{3}$

Generators selected (1); $t(1,0,0)$; $t(0,1,0)$; $t(0,0,1)$; (2); (4); (7)

Positions

Multiplicity, Wyckoff letter, Site symmetry			Coordinates						Reflection conditions
									General:
12	k	1	(1) x,y,z (4) $\bar{x},\bar{y},z$ (7) $y,x,\bar{z}+\frac{1}{3}$ (10) $\bar{y},\bar{x},\bar{z}+\frac{1}{3}$	(2) $\bar{y},x-y,z+\frac{1}{3}$ (5) $y,\bar{x}+y,z+\frac{1}{3}$ (8) $x-y,\bar{y},\bar{z}$ (11) $\bar{x}+y,y,\bar{z}$	(3) $\bar{x}+y,\bar{x},z+\frac{2}{3}$ (6) $x-y,x,z+\frac{2}{3}$ (9) $\bar{x},\bar{x}+y,\bar{z}+\frac{2}{3}$ (12) $x,x-y,\bar{z}+\frac{2}{3}$				$000l$: $l=3n$
									Special: as above, plus
6	j	. . 2	$x,2x,\frac{1}{2}$	$2\bar{x},\bar{x},\frac{5}{6}$	$x,\bar{x},\frac{1}{6}$	$\bar{x},2\bar{x},\frac{1}{2}$	$2x,x,\frac{5}{6}$	$\bar{x},x,\frac{1}{6}$	no extra conditions
6	i	. . 2	$x,2x,0$	$2\bar{x},\bar{x},\frac{1}{3}$	$x,\bar{x},\frac{2}{3}$	$\bar{x},2\bar{x},0$	$2x,x,\frac{1}{3}$	$\bar{x},x,\frac{2}{3}$	no extra conditions
6	h	. 2 .	$x,0,\frac{1}{2}$	$0,x,\frac{5}{6}$	$\bar{x},\bar{x},\frac{1}{6}$	$\bar{x},0,\frac{1}{2}$	$0,\bar{x},\frac{5}{6}$	$x,x,\frac{1}{6}$	no extra conditions
6	g	. 2 .	$x,0,0$	$0,x,\frac{1}{3}$	$\bar{x},\bar{x},\frac{2}{3}$	$\bar{x},0,0$	$0,\bar{x},\frac{1}{3}$	$x,x,\frac{2}{3}$	no extra conditions
6	f	2 . .	$\frac{1}{2},0,z$	$0,\frac{1}{2},z+\frac{1}{3}$	$\frac{1}{2},\frac{1}{2},z+\frac{2}{3}$	$0,\frac{1}{2},\bar{z}+\frac{1}{3}$	$\frac{1}{2},0,\bar{z}$	$\frac{1}{2},\frac{1}{2},\bar{z}+\frac{2}{3}$	$hkil$: $h=2n+1$ or $k=2n+1$ or $l=3n$
6	e	2 . .	$0,0,z$	$0,0,z+\frac{1}{3}$	$0,0,z+\frac{2}{3}$	$0,0,\bar{z}+\frac{1}{3}$	$0,0,\bar{z}$	$0,0,\bar{z}+\frac{2}{3}$	$hkil$: $l=3n$
3	d	2 2 2	$\frac{1}{2},0,\frac{1}{2}$	$0,\frac{1}{2},\frac{5}{6}$	$\frac{1}{2},\frac{1}{2},\frac{1}{6}$				$hkil$: $h=2n+1$ or $k=2n+1$ or $l=3n$
3	c	2 2 2	$\frac{1}{2},0,0$	$0,\frac{1}{2},\frac{1}{3}$	$\frac{1}{2},\frac{1}{2},\frac{2}{3}$				$hkil$: $h=2n+1$ or $k=2n+1$ or $l=3n$
3	b	2 2 2	$0,0,\frac{1}{2}$	$0,0,\frac{5}{6}$	$0,0,\frac{1}{6}$				$hkil$: $l=3n$
3	a	2 2 2	$0,0,0$	$0,0,\frac{1}{3}$	$0,0,\frac{2}{3}$				$hkil$: $l=3n$

Symmetry of special projections

Along [001] $p6mm$
$\mathbf{a}'=\mathbf{a}$ $\mathbf{b}'=\mathbf{b}$
Origin at $0,0,z$

Along [100] $p2mm$
$\mathbf{a}'=\frac{1}{2}(\mathbf{a}+2\mathbf{b})$ $\mathbf{b}'=\mathbf{c}$
Origin at $x,0,0$

Along [210] $p2mm$
$\mathbf{a}'=\frac{1}{2}\mathbf{b}$ $\mathbf{b}'=\mathbf{c}$
Origin at $x,\frac{1}{2}x,\frac{1}{3}$

Maximal non-isomorphic subgroups

I [2] $P6_411$ ($P6_4$, 172) 1; 2; 3; 4; 5; 6
[2] $P3_121$ (152) 1; 2; 3; 7; 8; 9
[2] $P3_112$ (151) 1; 2; 3; 10; 11; 12
[3] $P222$ ($C222$, 21) 1; 4; 7; 10
[3] $P222$ ($C222$, 21) 1; 4; 8; 11
[3] $P222$ ($C222$, 21) 1; 4; 9; 12

IIa none

IIb [2] $P6_522$ ($\mathbf{c}'=2\mathbf{c}$) (179)

Maximal isomorphic subgroups of lowest index

IIc [2] $P6_222$ ($\mathbf{c}'=2\mathbf{c}$) (180); [3] $H6_422$ ($\mathbf{a}'=3\mathbf{a},\mathbf{b}'=3\mathbf{b}$) ($P6_422$, 181); [7] $P6_422$ ($\mathbf{c}'=7\mathbf{c}$) (181)

Minimal non-isomorphic supergroups

I none

II [3] $P622$ ($\mathbf{c}'=\frac{1}{3}\mathbf{c}$) (177)

Table A2 *Amino acids chemical shifts (from Biological Magnetic Resonance Data Bank, BMRB).*

Aminoacid	C (ppm)	Cα (ppm)	Cβ (ppm)	Hα (ppm)	Hβ (ppm)	N (ppm)
Ala	177.75	53.16	18.95	4.26	1.36	123.18
Arg	176.46	56.81	30.63	4.29	1.8 1.77	120.79
Asn	175.34	53.53	38.64	4.67	2.81 2.77	119.00
Asp	176.45	54.65	40.82	4.60	2.72 2.68	120.74
Cys	174.81	57.98	33.25	4.68	2.95 2.90	120.18
Gln	176.36	56.56	29.14	4.27	2.05 2.02	119.92
Glu	176.96	57.36	29.97	4.25	2.03 2.01	120.72
Gly	173.96	45.33	-	3.97 3.90	-	109.70
His	175.26	56.47	30.19	4.62	3.11 3.05	119.54
Ile	175.85	61.57	38.58	4.19	1.79	121.51
Leu	177.01	55.65	42.24	4.31	1.62 1.54	121.81
Lys	176.71	56.96	32.74	4.27	1.79 1.76	121.04
Met	176.21	56.14	32.99	4.40	2.04 2.00	120.06
Phe	175.54	58.15	39.88	4.61	3.00 2.95	120.60
Pro	176.72	63.31	31.81	4.39	2.07 2.02	132.10
Ser	174.58	58.71	63.77	4.49	3.88 3.85	116.27
Thr	174.58	62.18	69.64	4.47	4.17	115.49
Trp	176.10	57.65	29.98	4.69	3.19 3.14	121.78
Tyr	175.45	58.10	39.26	4.63	2.91 2.86	120.66
Val	175.65	62.43	32.68	4.18	1.98	121.14

Table A3 *hSOUL NMR chemical shifts*. Chemical shifts for hSOUL assigned nuclei of triple labeled sample (^{2}H, ^{13}C, ^{15}N) and based on trHNCA, trHNCACO, trHNCOCA, trHNCACB, trHNCOCACB spectra acquired on a 600 MHz magnetic field spectrometer, at 293 K.

Residue	δ ^{2}HN (ppm)	δ ^{15}N (ppm)	δ ^{13}CO (ppm)	δ ^{13}Cα (ppm)	δ ^{13}Cβ (ppm)
Met1					
Ala2	8.31	122.50	172.40	50.69	24.85
Glu3	8.30	123.13	172.00	49.06	37.83
Pro4					
Leu5	8.14	122.89	175.04	51.92	38.39
Gln6	8.18	122.66	171.44	50.16	25.38
Pro7					
Asp8					
Pro9					
Gly10	8.33	109.02	171.89	41.92	
Ala11	7.77	124.29	175.54	49.58	15.76
Ala12	8.04	123.63	175.88	49.59	
Glu13	8.21	120.52	174.38	53.69	26.61
Asp14	8.12	121.75	174.33	51.31	37.80
Ala15	8.05	125.19	176.05	50.18	15.38
Ala16	7.99	122.18	175.92		
Ala17	7.78	122.49	173.20	49.58	
Gln18	7.87	118.63		52.54	
Ala19					
Val20	7.74	121.09	172.80	58.98	29.58
Glu21	8.66	128.98	172.05	53.69	27.82
Thr22	7.64	112.61	170.24	54.26	67.80
Pro23					
Gly24	8.41	111.36	170.35	41.95	
Trp25	7.33	118.00	172.02	51.91	27.29
Lys26	8.55	118.75	173.42	51.27	32.48
Ala27	8.67	128.16	173.65	47.22	14.77
Pro28					
Glu29					
Asp30	8.17	119.10	173.43	50.75	37.21
Ala31	7.78	123.39	175.79	49.52	15.99
Gly32	8.07	108.63	175.72	41.34	
Pro33					

Gln34					
Pro35					
Gly36	8.10	109.84	172.02	41.31	
Ser37					
Tyr38	7.14	116.71	170.24	53.06	35.91
Glu39	8.42	117.96	172.45	51.34	30.76
Ile40	8.56	124.29	172.64	57.80	35.45
Arg41	8.43	123.20	171.35	51.33	31.34
His42	8.66	122.30	173.65	51.31	27.21
Tyr43	8.82	129.33	172.40	56.02	36.60
Gly44					
Pro45					
Ala46	8.35	123.51	173.45	49.58	19.88
Lys47	9.58	119.96	174.30	53.11	30.73
Trp48	9.12	122.96	171.70	52.86	28.20
Val49	9.57	122.69	171.34	56.66	29.81
Ser50	8.52	122.53	169.01	53.80	65.83
Thr51	8.43	114.49	169.01	58.38	66.82
Ser52	9.11	125.11	171.13	55.39	61.93
Val53	9.19	124.64	171.80	57.50	31.94
Glu54	8.31	125.66	173.80	51.92	26.64
Ser55	8.32	116.21		54.94	60.75
Met56					
Asp57	7.14	120.03	175.55	52.47	39.57
Trp58	7.99	120.54	173.67	53.54	26.60
Asp59	8.08	114.49	176.61	55.39	37.44
Ser60	7.48	114.53	174.34	57.60	59.57
Ala61	8.16	126.75	177.10	51.93	15.43
Ile62	7.77	116.17	174.58	58.97	32.78
Gln63	6.47	119.60	176.59	55.45	24.85
Thr64	8.08	117.30	175.20	63.08	65.60
Gly65	8.77	108.63	172.42	43.68	
Phe66	8.47	121.67	174.23	59.57	35.45
Thr67	7.62	114.25	174.21	63.95	65.52
Lys68	6.83	120.35	177.35	56.06	28.99
Leu69	7.58	120.54	176.44	54.28	37.21

Asn70	8.54	118.20	175.01	52.46	34.90
Ser71	7.52	113.51	170.91	58.97	
Tyr72	8.85	128.17	172.72	51.30	
Ile73	7.87	116.91	174.48	60.75	
Gln74	7.51	116.01	174.03	51.93	24.27
Gly75	6.90	104.84	172.71	41.39	
Lys76	8.17	126.09	172.54	51.47	25.68
Asn77	8.60	119.71	173.77	49.51	40.11
Glu78	9.08	119.21	173.83	56.43	27.21
Lys79	6.98	114.32	172.42	51.76	29.58
Glu80	7.78	120.11	173.31	53.70	24.28
Met81	7.96	118.59	173.16	52.02	32.93
Lys82	8.28	125.85	173.40	52.44	27.83
Ile83	8.99	129.49	172.05	56.63	36.63
Lys84	8.14	126.40	172.64	54.27	29.55
Met85	7.81	119.80	173.60	53.09	
Thr86	7.35	116.91	168.59	56.66	68.12
Ala87	8.05	116.52	170.75	46.04	18.97
Pro88					
Val89	8.19	119.88	175.72	58.97	28.97
Thr90	8.87	123.43	171.88	57.21	69.55
Ser91	8.31	116.53	168.30	53.68	64.61
Tyr92	8.82	123.55	172.40	51.93	37.81
Val93	7.83	129.37	171.35	58.98	28.99
Glu94					
Pro95					
Gly96					
Ser97					
Phe98					
Pro99					
Phe100	8.36	122.79	173.65	53.66	
Ser101	7.34	115.03	172.09	55.00	61.32
Glu102	8.65	123.57	174.16	53.72	26.60
Ser103					
Thr104	8.23	119.76	173.60	59.56	66.67
Ile105	8.85	128.35	170.95	57.28	37.06

Thr106	9.12	125.57	171.49	58.86	66.82
Ile107	9.42	133.16	172.85	57.23	36.06
Ser108	8.80	119.75	170.37	55.08	64.04
Leu109	8.98	121.40	173.09	49.98	42.09
Tyr110	9.23	129.06	173.31	53.77	33.77
Ile111					
Pro112					
Ser113					
Glu114	8.25	117.73	173.42	50.15	36.06
Gln115	7.75	126.37	178.71	55.17	27.60
Gln116	6.98	118.08	174.38	56.04	24.86
Phe117	7.85	115.50	173.80	56.62	35.30
Asp118	7.21	115.97	168.66	48.97	37.28
Pro119					
Pro120					
Arg121					
Pro122					
Leu123					
Glu124					
Ser125	8.14	115.89	172.29	58.99	66.67
Asp126	8.25	124.84		52.99	
Val127					
Phe128					
Ile129	10.23	123.74	174.03	58.96	37.22
Glu130	8.42	127.18	173.29	54.28	26.60
Asp131	7.82	128.20	172.56	52.08	30.65
Arg132					
Ala133	8.42	130.93	175.54	48.40	16.07
Glu134	8.41	120.03	172.67	53.06	26.45
Met135	8.15	117.30	170.59	52.51	30.76
Thr136	8.69	121.25	170.86	58.41	66.62
Val137	8.71	117.61	170.41	54.39	31.42
Phe138	9.19	120.93	173.11	51.94	38.40
Val139	9.22	121.87	173.58	58.97	31.41
Arg140	7.15	122.61	175.19	59.29	33.70
Ser141	8.62	123.67	171.51	53.69	62.49

Phe142	7.96	117.06	170.48	52.47	37.13
Asp143	8.36	120.07	174.47	50.73	38.97
Gly144	8.19	109.02	171.98	41.32	
Phe145					
Ser146	8.17	116.99	170.55	55.45	61.35
Ser147	7.09	116.09	171.48	53.51	63.12
Ala148	10.29	127.22	177.97	52.50	15.46
Gln149	8.38	116.94	176.06	55.91	24.46
Lys150	7.60	121.87	175.08	54.78	27.82
Asn151	8.35	117.22	175.01	52.25	33.24
Gln152	7.40	117.61	175.59	63.66	24.29
Glu153	7.48	119.72	177.55	56.07	26.62
Gln154	7.77	118.35	175.77	54.61	24.27
Leu155	8.32	122.94	176.88	54.87	38.81
Leu156	7.68	120.27	178.35	54.97	37.81
Thr157					
Leu158	8.10	122.42	174.43	54.81	37.28
Ala159	8.33	119.72	176.77	51.99	14.84
Ser160	7.34	111.52	174.81	58.83	59.72
Ile161	7.48	124.25	175.90	61.33	34.87
Leu162	8.07	120.11	177.62	54.42	37.51
Arg163	8.08	117.96	178.89	56.64	26.03
Glu164	7.55	121.75	175.70	56.07	25.53
Asp165	7.73	118.63	174.25	51.35	38.20
Gly166	7.64	109.17	172.71	43.17	
Lys167	7.86	118.98	173.11	50.14	28.81
Val168	8.54	121.87	172.61	59.57	
Phe169	7.24	122.22	171.70	50.77	39.42
Asp170	8.67	120.83	174.83	51.31	38.73
Glu171	8.80	126.28	175.16	54.27	26.04
Lys172	8.90	119.56	173.96	54.86	30.73
Val173	6.96	109.84	173.09	56.11	31.94
Tyr174	7.26	114.72	170.06	54.26	36.64
Tyr175	8.61	114.64	174.79	53.13	40.16
Thr176	9.02	109.99	170.08	57.20	67.22
Ala177	7.93	122.06	172.20	48.39	17.81

Gly178	8.81	111.91	170.95	42.65	
Tyr179					
Asn180	8.32	114.80	171.66	49.57	
Ser181	7.02	114.37	171.11	52.77	61.40
Pro182					
Val183	7.27	120.42	172.15	56.65	30.72
Lys184	8.36	126.21	173.00	53.80	27.13
Leu185	8.26	129.21	174.16	53.41	40.75
Leu186	7.67	117.09	173.59	50.14	44.00
Asn187					
Arg188	7.38	122.61	171.88	51.33	29.58
Asn189	8.19	120.66	173.31	49.56	35.15
Asn190	8.02	117.46	170.46	50.76	36.63
Glu191	8.60	112.42	176.28	51.33	31.94
Val192	8.81	115.58	171.83	57.76	31.33
Trp193	9.16	122.57	174.29	49.56	30.16
Leu194					
Ile195					
Gln196					
Lys197	7.76	127.61	170.81	58.41	29.57
Asn198	8.95	122.38	169.99	51.92	38.41
Glu199	8.56	119.76	172.89	55.37	31.71
Pro200					
Thr201					
Lys202					
Glu203	9.56	134.06	172.04	50.75	28.97
Asn204	8.59	128.78	173.56	50.56	36.65
Glu205					

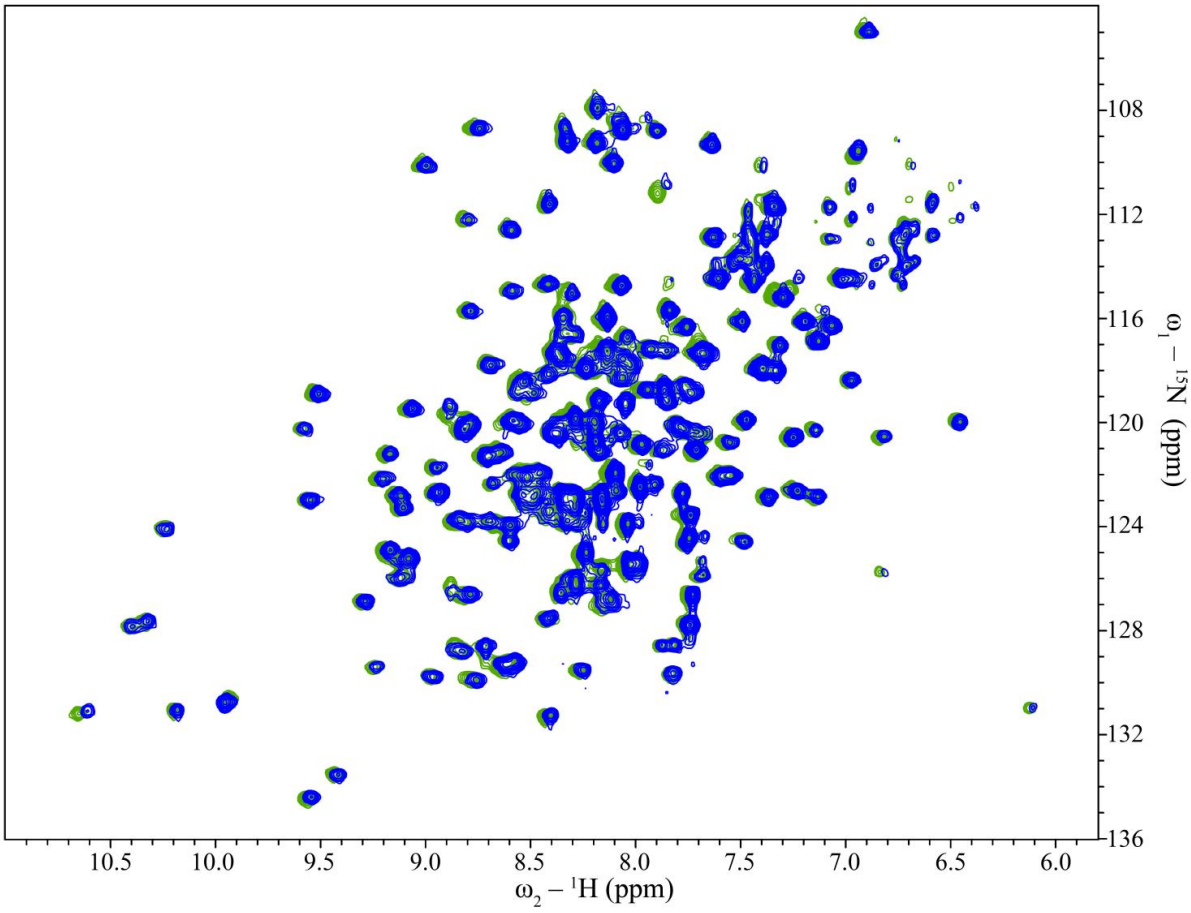


Figure A1 *^{1}H, ^{15}N-TROSY-HSQC spectra of hemin hSOUL.* ^{15}N-hSOUL:hemin at molar ratio of 0.5 (green) and ^{15}N-hSOUL alone (blue).

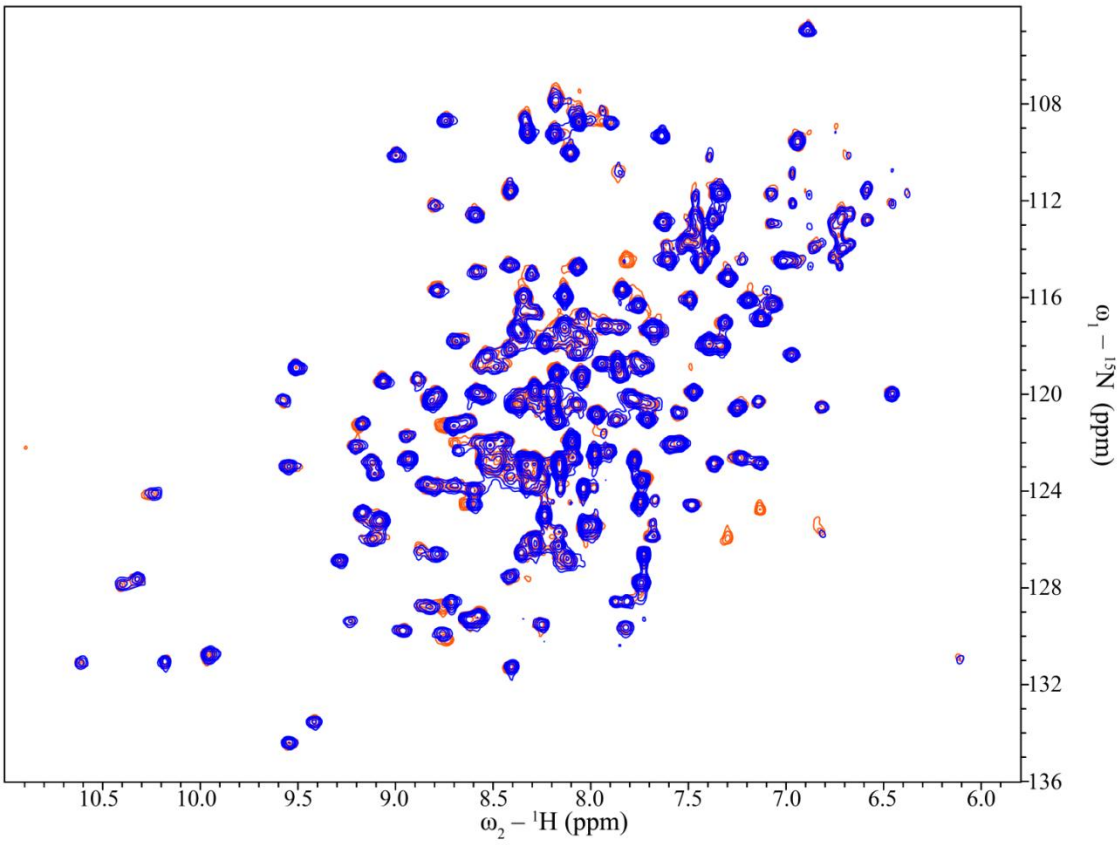


Figure A2 *^{1}H, ^{15}N-TROSY-HSQC spectra of hemin hSOUL*. ^{15}N-hSOUL:hemin at molar ratio of :1:2 (orange) and ^{15}N-hSOUL alone (blue).