UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS

DEPARTAMENTO DE BIOLOGIA ANIMAL

Computational and validation approaches in proteomics discovery of disease biomarkers

André Abrantes da Costa

Mestrado em Biologia Humana e Ambiente

Dissertação orientada por:
Doutora Deborah Penque, PhD, Instituto Nacional de Saúde Dr Ricardo Jorge
Professora Doutora Deodália Dias, PhD, Universidade de Lisboa

2019

The references of this project meet the criteria of the Journal of Proteomics – Elsevier

ACKNOWLEDGMENTS

A year ... a year of hard work that made me grow and learn a lot.
A long trip with immense difficulties, but I learned from these flaws and how to get around the situation so that they can be overcome.
It was a year that is marked by the positive, not only for the work and the area in which I worked but also for all the experiences and people I met throughout this stage.

I want to thank Professor Deodália Dias for all the support provided over the two years of masters. And for all the trust, affection and sympathy.

I want to give a special thank you to Doctor Deborah Penque who allowed me to do the dissertation in the Laboratory of Proteomics of the National Institute of Health Doctor Ricardo Jorge, and who always guided me in order to make me think of the best solution to my problems.

I also thank Doctor João Lavinha for guidance and knowledge sharing related to Sickle cell disease. And advice on important parts to incorporate in this dissertation.

There is also a thankfulness of love to everyone present at the lab, Fátima, Cristina, Sofia and Inês, who accompanied me and helped me along this journey. The laughs and the good times that made this experience very good. From silly in pictures with orange peel in the mouth to moments of reflection and teachings that made me overcome some difficulties.

I want to thank my parents, Fernando and Isabel, for all the love they have always given me. For all the trust you placed in me. And for all their efforts to study abroad (the rent is expensive in Lisbon haha) and make my dreams come true. Without you it was not possible, thanks!

My brothers, Stefan and Rodrigo, who have always been present, I want to thank you for trusting me too. And for putting up with me whenever I go home.

I want to thank my best friend, João, who has always been present throughout this journey and who with huge coffees always made me vent and improve my mood when I was less well. Several night outings to be able to relax and enjoy the good that life can give us not to get lost in work and stress.

Thank my best friend Ana Maria, who besides being my colleague since graduation, going through the same masters, endured all my boring conversations about what was happening and advised me always and did not let me go down when something went less well.

To my master's colleagues, Marta and Rita, who spent a lot of lunch hours drinking our coffee between work and stress. They have often helped to relax and unwind. Eaten ice cream and laughs.

And last but not least, to all my friends who have always listened to me and taken me into the bad life of the night, to take away all the pressure I had and helped to pass this year with more tranquillity and love.

A **THANK YOU** TO ALL!

RESUMO

O estudo em larga escala de proteínas, a proteómica, está a mudar amplamente a nossa compreensão das funções dos genes na era pós-genómica. Depois da revolução na genómica pelos métodos de sequenciação de ADN (ácido desoxirribonucleico), a proteómica tem vindo a aumentar o nosso conhecimento sobre a variabilidade, localização, função e vias metabólicas das proteínas na célula, tecido ou organismo [1].

O desenvolvimento de biomarcadores, como uma componente relevante na tomada de decisões, tanto nos processos clínicos quanto no desenvolvimento de novos medicamentos, é uma área emergente onde a proteómica tem vindo a ganhar relevo. As tecnologias proteómicas permitem uma analise comparativa, qualitativa e quantitativa de milhares de proteínas de células/tecidos de doentes versus indivíduos controles, assim como, de doentes antes e depois de um determinado tratamento. As proteínas identificadas diferencialmente abundantes e/ou modificadas pós-traducionalmente na condição de doença ou em resposta a terapia, são possíveis candidatas a biomacardores destas condições [2]. No entanto, a interpretação da enorme quantidade de dados proteómicos, gerados principalmente por experiências baseadas em espectrometria de massa (MS), requer suporte computacional para processamento e análise dos dados de forma efetiva e robusta.

A Proteómica computacional estuda os métodos computacionais, algoritmos, bases de dados e metodologias utilizadas para processar, gerir, analisar e interpretar os dados produzidos em experiências proteómicas na identificação de potenciais biomarcadores.

O Laboratório de Proteómica do INSA (Instituto Nacional de Saúde Doutor Ricardo Jorge), através da busca de biomarcadores proteómicos para compreensão de doenças tais como a Anemia das Células Falciformes (ACF), tem produzido grandes dados de MS que necessitam de análise computacional, sendo este o principal propósito deste projeto (ver abaixo objetivo deste estudo).

A ACF, também denominada por drepanocitose ou anemia drepanocítica, é um distúrbio monogénico autossómico recessivo, clinicamente heterogéneo, caracterizado por episódios recorrentes de hemólise grave, vaso-oclusão e infecção. Vários modificadores genéticos e ambientais foram sugeridos para modular o início e o curso da ACF. Especificamente, os componentes vasculares da patologia (por exemplo, acidente vascular cerebral) foram submetidos a pesquisas intensivas e o uso de metodologias proteómicas promete oferecer novas percepções moleculares sobre a fisiopatologia da ACF. A mudança do estado estacionário para a crise ainda é em grande parte imprevisível. A fim de descobrir biomarcadores putativos para essa exacerbação, o laboratório do INSA analisou por proteómica de *shotgun* MS, amostras de plasma e glóbulos vermelhos (GV), de um grupo de pacientes com ACF em estado estacionário e em crise (episódio de vaso-oclusão).

Objetivo do estudo:

Este estudo teve como objetivo analisar os dados de *shotgun* MS gerados para a ACF através de plataformas de proteómica computacional de código aberto, nomeadamente o PatternLab [3] e o MaxQuant [4] , no sentido de identificar proteínas como possíveis candidatos a biomarcadores da ACF e em particular da ACF associada a vaso-oclusão.

O PatternLab *for Proteomics* é um ambiente computacional integrado para análise de proteómica *shotgun*, formatando bancos de dados de sequências, que realizam a correspondência de espectro peptídico, filtrando estatisticamente e organizando dados por proteómica diferencial, exibindo resultados em formato de gráficos, realizando estudos orientados por similaridade com dados de sequenciação de novo, ajudando à compreensão do significado biológico dos dados à luz da Ontologia Genética (*Gene Ontology*).

O MaxQuant é um conjunto de algoritmos, que inclui a detecção de picos e a pontuação de péptidos, realiza a calibração em massa e pesquisas em bancos de dados para identificação de proteínas, quantifica proteínas identificadas e fornece estatísticas resumidas.

Para validar os achados proteómicos, algumas das proteínas identificadas diferencialmente na patologia por essas plataformas computacionais, foram selecionadas para validação (verificação) através de abordagens como *Western blot*. O Western blot é uma técnica bioquímica imunológica na qual uma mistura de proteínas é separada por gel 1DSDS-PAGE (Sodium Dodecyl Sulfate - PolyAcrylamide Gel Electrophoresis), transferida para uma membrana, posteriormente incubada com anticorpo específico contra a proteína de interesse. A reação, geralmente visualizada por quimioluminescência, pode ser quantificada por densitometria **[5]**.

De todas as 111 proteínas diferencialmente expressas identificadas (74 da fração citoplasmática e 37 da fração membranar) associadas ao evento de crise na ACF, a peroxiredoxina-2, a catalase, a Hsp70 e Hsp 90 foram validadas por *Western blot*. Destas 4 proteínas, apenas a peroxiredoxina-2 apresentou significância estatística. Das proteínas diferencialmente expressas que hipoteticamente podem estar associadas à ACF, um promissor biomarcador de crise, nomeadamente, a *Voltage-dependent anion-selective channel protein 1* (VDAC1) foi encontrada diminuida. Os resultados sugerem que episódios de vaso-oclusão em doentes com ACF podem estar associados à diminuição da VDAC1 nos glóbulos vermelhos.

Os resultados deste projeto após apresentação e discussão podem contribuir para um melhor entendimento das vias moleculares associadas à ACF, bem como para a identificação de modulações protéicas específicas, como possíveis candidatos a biomarcadores para essas patologias.

Palavras-chave: Proteómica; Bioinformática; Anemia das Células Falciformes; Biomarcadores; *Western blot*

ABSTRACT

The Laboratory of Proteomics at INSA (Instituto Nacional de Saúde Doutor Ricardo Jorge) by searching biomarkers for Sickle-cell disease (SCD) has been produced considerable Mass Spectrometry (MS) data, needing computational analysis to process, manage, analyze and interpret the data to reveal relevant biomarkers.

SCD is a clinically heterogeneous autosomal recessive monogenic disorder characterized by recurrent episodes of severe haemolysis, vaso-occlusion and infection. Several genetic and environmental modifiers have been suggested to modulate the onset and course of SCD. The vascular components of the pathology have been thus subjected to intensive research and the usage of proteomics methodologies promises to offer novel unbiased molecular insights into the pathophysiology of SCD.

The objective of this project is to analyze by different bioinformatics tools the MS raw data that have been generated by INSA's Lab in order to investigate biological/molecular mechanisms responsible for protein changes that might be related with the development of SCD. The most pertinent proteins identified by these computational approaches associated with those pathologies will be selected for further validation as candidate biomarkers by using western blot methods.

From all the identified 111 differentially expressed proteins (74 cytoplasmatic fraction and 37 membrane fraction) associated with SCD crisis event, Peroxiredoxin-2, Catalase, 70-kDa Heat shock protein and Heat shock protein 90 were validated with Western blot. From these 4 proteins only Peroxiredoxin-2 showed statically significant. Of the differentially expressed proteins that hypothetically may be associated with SCD, a promise candidate biomarker of crisis namely the Voltage-dependent anion-selective channel protein 1 (VDAC1) was found decreased. Our results suggest that vaso-occlusion episodes in SCD patients may be associated with decreased VDAC1 in their RBCs.

This study indicated that SCD patients at crisis-state are under oxidative stress and the proteins such as PRDX2 and VDAC1 are promising candidates biomarkers for SCD crisis-state.

In summary, the main objective of this project is to contribute to a better understanding of the molecular mechanisms associated with these pathologies, as well as to discover new diagnostic, prognostic or monitoring biomarkers for these diseases, leading to the development of new methods that would increase the quality of life of these patients.

Keywords: Proteomics; Bioinformatics; Sickle Cell Disease; Biomarkers; Western blot

PUBLICATIONS

André Costa*, Sofia Neves, Fátima Vaz, Peter James, Paula Faustino, João Lavinha, Deborah Penque. "*Red Blood Cell proteome modulation associated with vaso occlusion in Sickle Cell Disease*" (*manuscript in preparation*).

André Costa*, Sofia Neves, Fátima Vaz, Inês L. Martins, Peter James, João Lavinha, Deborah Penque. Sickle-cell disease investigated by computational proteomics approaches. VI International Caparica Conference on Analytical Proteomics and *Jornadas da ToxOmics*. 08th – 11th July 2019 and 28th October 2019. Caparica, Portugal and Lisbon, Portugal. Poster Presentation (Annex D)

INDEX

LIST OF FIGURES

TABLE LIST

ABBREVIATIONS

MS	Mass Spectrometry
DNA	Deoxyribonucleic Acid
INSA	*Instituto Nacional de Saúde Doutor Ricardo Jorge*
SCD	Sickle Cell Disease
HbS	Sickle Haemoglobin
VOC	Vaso-Occlusive Crisis
Hb	Haemoglobin
ROS	Reactive Oxygen Species
NO	Nitric Oxide
LC	Liquid Chromatography
PSM	Peptide−Spectrum Match
FDR	False-Discovery Rate
TDA	Target−Decoy Approach
XIC	eXtracted-Ion Chromatograms
PEP	Posterior Error Probability
PTM	Post-Translational Modification
LFQ	label-free quantification
CAT	Catalase
PRDX2	Peroxiredoxin-2
RBC	Red Blood Cells
Hsp70	70-kDa Heat shock protein
Hsp	Heat shock protein
Hsp90	Heat shock protein 90
NIAF	Normalized Ion Abundance Factor
DAVID	Database for Annotation, Visualization and Integrated Discovery
GO	Gene Ontology
SDS-PAGE	Sodium Dodecyl Dulfate - PolyAcrylamide Gel Electrophoresis
PVDF	PolyVinyliDene Fluoride
LDS	Lithium Dodecyl Sulfate
PBS-T	Phosphate Buffered Saline - Tween 20
PBS	Phosphate-Buffered Saline
ECL	Enhanced ChemiLuminescence
HS	Heat Shock
mRNA	messenger Ribonucleic Acid
VDAC1	Voltage-dependent anion-selective channel protein 1
ATP	Adenosine TriPhosphate

STATE OF THE ART

1. What is Proteomics?

Since the late 1970s, Proteomics existed after the formation of protein databases initiated by researchers using the recently developed bi-dimensional gel electrophoresis technique. It is related to a set of technologies that aim the separation and identification of proteins in more complex biological samples. Most cellular processes are controlled by proteins, which occur in great diversity, acting as antibodies, cell receptors, hormones, enzymes and structural components. The study of this set of expressed proteins led to the development of new technologies [6]. Proteome is a *portmanteau* (a linguistic blend of words, in which parts of multiple words or their sounds are made into a new word) of protein and genome and was given by Marc Wilkins in 1994 while he was a PhD student at Macquarie University, Australia [7].

The entire set of proteins that are produced or modified by an organism or system at a given condition and moment, is the proteome. Proteomics empowered the identification of an always increasing number of proteins. This changes over time and distinct requirements (or stresses) that an organism or cell suffers [8].

Proteomics is a large-scale discovery strategy, that refers to the set of technologies applied to explore the proteome, permitting evaluation of hundreds to thousands of proteins at certain time and under certain conditions. The advantages of proteomics are to obtain information regarding changes in protein abundance, protein proteoforms, protein–protein interactions in a cell, tissue, or organism that may be very important to understand the true molecular phenotype of a disease, which may also reveal disease biomarkers and disease targets for the development of new drugs [9].

Proteomic studies initially aimed to identify on a large scale all proteins present in a cell or tissue. Nowadays, it consists on a simultaneous analysis of complex mixtures of proteins such as those from cell lysis and tissue extracts to detect quantitative and qualitative differences in protein expression [6].

Its goals have been expanded for the investigation of a few practical parts of proteins, for example, post-translational changes, protein-protein interactions, presence of isoforms, activities and structures. The field of activity of this science reaches out to the disclosure of new drugs, treatments, diagnostics, microbiology, biochemistry. Proteomic research makes possible the identification and characterization of biomarkers, that is, endogenous or exogenous molecules specific to a certain pathological state. The ability to identify these molecules is extremely useful in the early diagnosis of diseases and in the follow-up of treatment evolution [10].

So, is proteomics really necessary?

Recently, in the field of proteomics an important progress can be seen, with the improvement of alternative methods to separate and identify proteins, such as chip-based technologies, the protein arrays, a direct analysis of protein complexes by mass spectrometry (MS) [6].

Researchers are recognizing that just having a complete sequence of genomes is not sufficient to clarify biological function, duo to an increase of the amount of DNA (Deoxyribonucleic acid) sequences in databases. For its survival a cell is normally dependent to a large number of metabolic and regulatory pathways. Genes and the proteome of a cell do not have a strict linear relationship. Since it focuses on the gene products, active agents in cells, proteomics is complementary to genomics. This is the reason why proteomics has a direct contribute to foster a better understanding of disease processes, develop new biomarkers for diagnosis and early detection of disease, and accelerate development of new therapy approaches [1].

Advances in genome sequencing have great potential to promote understanding of diseases, but by themselves are not sufficient for a complete description of the disease phenotypes and their

interaction/dynamic modulation with the environment. Proteome information provides a broad picture of patient phenotype rather than indicating unchanging probabilistic risks that can be inferred from sequence analysis of genomic DNA. MS-based proteomics has emerged as a large-scale platform to improve medical disease assessment by focusing on the discovery of new protein biomarkers. However, inherent in protein biochemistry and sample complexity, proteomic methods face a major challenge in identifying relevant candidate biomarkers [9].

The laboratory of proteomics at INSA (*Instituto Nacional de Saúde Doutor Ricardo Jorge*) has been applied proteomics to discover new biomarkers for diagnosis, prognosis and monitoring of diseases such as Obstructive Sleep Apnea (OSA) and Sickle Cell Disease (SCD) [11 - 13].

The development and application of innovative and dedicated proteomic strategies obviously have a great influence on the discovery of biomarkers for SCD. One of them depends on the power of data integration systems biology has become a key approach for identifying new disease mediators, potential diagnostic and therapeutic targets for many diseases [9].

2. Sickle cell disease

SCD is s multisystem disorder, caused by a single gene mutation. SCD is a consequence of a point mutation in the haemoglobin β-globin chain, causing a change in the sixth position of hydrophilic amino acid glutamic acid by hydrophobic valine amino acid. It is in the short arm of chromosome 11 that we find the β-globin gene [14]. Episodes of acute illness and continuous organ damage are common (nearly every organ in the body can be affected). It's the most common monogenic disorder [15 - 17]. Sickle cell anemia is the most common form and accounts for 70% of cases of SCD in patients of African ethnicity [18].

Figure 1.1 illustrates abnormal and sickle red blood cells trapped at the branch point of a vein. The image also shows a cross section of a sickle cell with long strands of polymerized sickle cell haemoglobin (HbS), stretching and distorting the shape of the cell to look like a crescent, and a normal red blood cell.

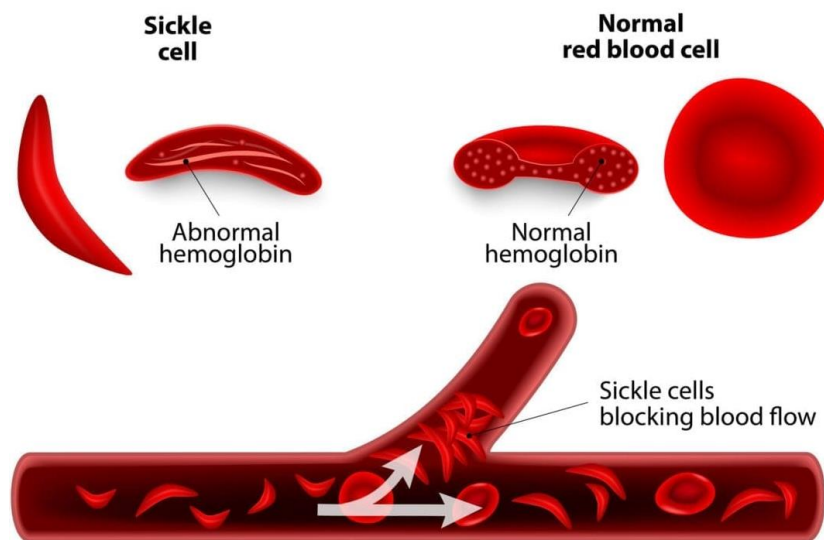**Figure 1.1** Illustration of a blood flow with abnormal and sickle red blood cells, with the cross section of red blood cells. (Source: [19]).

Over the past seven decades scientists have characterized three major pathobiological processes leading to SCD (HbS polymerization, vaso-occlusion, and hemolysis-mediated endothelial dysfunction). Recently, a fourth process has emerged, sterile inflammation [20].

2.1. HbS polymerization

This mutation present in SCD patients will allow the polymerization of HbS molecules into polymers under hypoxia. Due to deoxygenation and the conformational alignment which is associated with haemoglobin in the R (relaxed) state and T (tense) state, the polymerization of HbS molecules begins. HbS molecules aggregate rapidly when a nucleus forms, thus leading to fiber growth. The subsequent branching of the fibers was shaped as a double nucleation mechanism. This increase in HbS fibers will promote distortion of Red Blood Cell (RBC) and thus impair the change in cell shape, which will contribute to the potential initiation and propagation of vaso-occlusion, which is a feature of SCD. These recurrent and quite unpredictable vaso-occlusive crisis (VOC) episodes lead to stroke, frequent painful crisis, and other serious complications including acute chest syndrome, splenic sequestration, and aplastic crisis [21].

2.2. Vaso-occlusion

Vaso-occlusion leading to ischemia is the main pathophysiology responsible for acute systemic painful VOC and the demand for emergency medical care by SCD patients. Blood rheology is imposed by hematocrit, plasma viscosity and erythrocyte deformability. Decreased blood flow through capillaries and postcapillary venules of tissues with high oxygen demand is a result of increased plasma viscosity, which occurs as a result of chronic hemolysis and reduced sickle cell erythrocyte deformability due to polymerization and dehydration of Haemoglobin (Hb).

Patients with SCD show elevated levels of neutrophils, monocytes and platelets at baseline, and high levels of circulating aggregates of neutrophils-platelets and monocytes-platelets in human blood of SCD correlate with disease severity.

In addition, thrombocytopenia is a major predictor of VOC progression in SCD patients with potentially lethal lung injury also known as acute chest syndrome (ACS), suggesting a role for platelet sequestration at vessel occlusion sites.

VOC is often initiated by an inflammatory or environmental stimulus, including infection, hypoxia, dehydration, acidosis, or other unidentified factors [20].

2.3. Hemolysis-mediated endothelial dysfunction

Episodic and sustained vaso-occlusion is the result of HbS-containing erythrocytes with Hb intracellular polymer being less deformable and trapped in the microcirculation. Pulmonary hypertension left diastolic heart disease, and renal dysfunction (proteinuria, albuminuria, and chronic renal dysfunction) are manifested because patients with higher hemolysis rates have lower steady-state Hb levels and are more likely to develop vascular injury and organ dysfunction as they get older.

Cell-free Hb also promotes the formation of Reactive oxygen species (ROS), critically altering the vascular redox balance of steady-state Nitric oxide (NO) production to ROS production (decreased NO–ROS balance). NO is essential for vasodilation and regulates platelet function, inflammation, cell smooth muscle proliferation, and oxidative stress, and the elimination of NO by cell-free plasma haemoglobin impairs endothelial function and promotes proliferative vasculopathy of the pulmonary and systemic vasculature [20].

2.4. Sterile inflammation

Vaso-occlusion contributes to ischemia-reperfusion injury, which, together with the release of erythrocyte damage-associated molecular pattern molecules (eDAMPs), promotes the progression of sterile inflammation. Vaso-occlusion also contributes to the progression of sterile inflammation in SCD. Repeated episodes of vaso-occlusion and reperfusion contribute to ischemia-reperfusion injury,

promoting transient hypoxia, ROS generation, microvascular dysfunction, activation of innate and adaptive immune responses, and cell death [20].

2.5. Distribution

The disease prevalence is relatively high throughout large areas in Sub-Saharan Africa, the Mediterranean basin, the Middle East, and India as a result of a remarkable protection that the sickle cell trait provides against severe malaria [17], as shown in Figure 1.2.
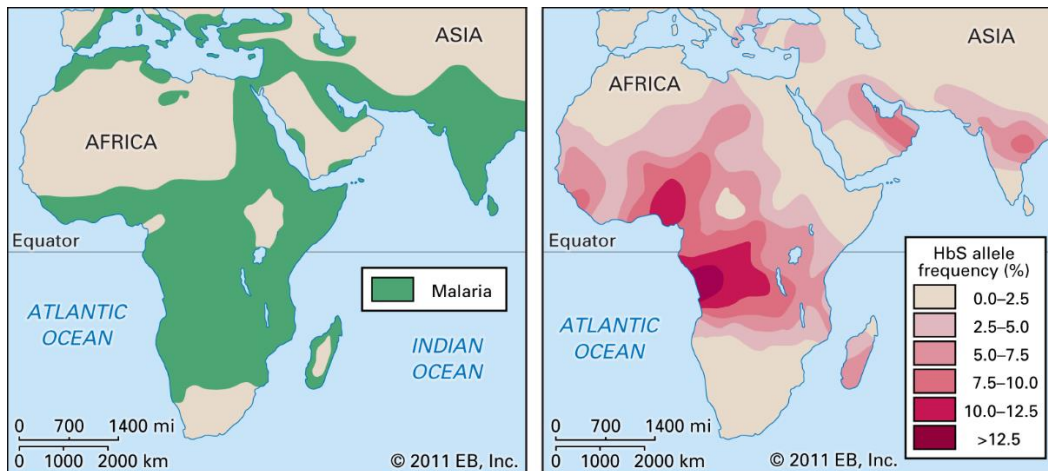

Figure 1.2 Malaria and Sickle cell disease distribution (Source: [22]).

In areas of Africa, Southern Asia and the Mediterranean, the distribution of malaria and sickle cell anemia overlap. It was explained by the fact that heterozygous people are resistant to malaria that there is a persistence of the HbS gene, which causes sickle cell anemia. When a person's sickle cell red blood cells adhere to the blood vessel walls, they become deoxygenated, take the shape of the sickle and are destroyed, and the red blood cell invader malaria is destroyed with them.

There are no existing data on antimalarial drugs currently recommended in patients with SCD, but WHO (World Health Organization) has issued a recommendation to patients in these endemic areas for antimalarial prophylaxis to patients with this type of condition [23].

2.6. Symptomatology

It is described by chronic hemolytic anemia and intermittent vaso-occlusive events. These events result in tissue ischemia that lead to acute and chronic pain, and also in organ damage affecting any organ system such as, the bones, spleen, brain, lungs, liver, joints and kidneys. Swelling of the hands or feet and/or pain (dactylitis) is mostly the first manifestation of SCD. When it occurs in children, a "splenic sequestration" occurs in which the blood cells present in the spleen become engorged. The major cause of mortality in SCD is an acute chest syndrome [16].

2.7. Treatments

However, no drugs have been developed that target the pathophysiology of this disease, and this is due to a basic clinical management, even though there is some evidence lending support to the use of blood transfusion and hydroxycarbamide in some situations [15].

The beginning of the development of programs in areas with a high prevalence of SCD (India and some African) are really important. These programs include premarital, antenatal, and neonatal screening that have also been implemented in some high-income countries, inclusive some parts of the

Middle East and the United States. Specially in some rural areas per sub-Saharan Africa and

India, the development of inexpensive and confidential point-of-care diagnostic tests with a high sensibility and specificity could immensely facilitate screening for SCD in these least developed countries. Notwithstanding, genotypic identification is nearly insignificant, if the diagnosis isn't followed by preventive interposition and treatment with cheap oral agent to avoid complications of acute disease [16].

To get more widely relevant as new techniques for development are stem-cell transplantation and gene therapy, using induced pluripotent stem cells offering the most promise [15]. Concerning sickle-cell research, clinical trials and natural history studies have provide progressive growth in our acquaintance concerning this monogenic event that results in a systemic disorder of enormous complexity. Even though some problems have been successfully tackled, and gene therapy ensure promise, many questions stay [16].

3. Computational Proteomics

The available MS based proteomic technologies can generate large amounts of data such as the mass spectrum itself, which contains continuous and discrete information simultaneously. MS measures the mass-to-charge ratio (m/z) of ionized molecules such as proteins/peptides and produces spectra containing m/z values versus intensity count. A typical shotgun MS strategy to identify proteins (bottom up proteomics), consists first of digesting proteins present in a sample usually by trypsin (which predominantly cleaves at the carboxyl side of lysine "K" and arginine "R" except when either is followed by proline). The tryptic peptides generated from protein digestion are injected into a liquid chromatography system (for peptide separation) online coupled with a mass spectrometer, which employs two stages of mass analysis (tandem MS, MS/MS). The first stage is MS, which corresponds to mass scans of peptides eluted at different time points from the Liquid chromatography (LC) system. In the second stage, some peptides, usually the most abundant, are automatically selected for fragmentation generating MS/MS spectra. Mass differences between peaks in the MS/MS spectrum can be directly correlated to amino acid residues and a short peptide sequence can be obtained. Protein and peptide identification (known by 'protein inference' identification) is achieved by computer-based matching of the experimental tandem mass spectra with the theoretical tandem mass spectra generated from in silico digestion of protein sequence databases. Initially applied for full identification and characterization of proteomes, the shotgun approach was soon extended to quantitative and comparative studies in biomarker discovery using different strategy, which includes label-free quantitative approaches. Label-free quantification can be based on precursor signal intensity or on spectral counting. The computational framework of label free approach includes detecting peptides, matching the corresponding peptides across multiple LC-MS data, quantifying and selecting discriminatory peptides [24, 25].

As in several scientific disciplines, the ultimate goal of Computational Proteomics is to infer knowledge models (for example, ascertain a hypothesis or identify proteins involved in a disease) from the inspection of biological samples [24].

Two software used in this study will be described below in a simplistic way.

3.1. PatternLab for Proteomics

PatternLab for Proteomics is an integrated computational environment for analyzing shotgun proteomic by formatting sequence databases, performing peptide spectrum matching, statistically filtering and organizing shotgun proteomic data, extracting quantitative information from label-free and chemically labeled data, performing statistics for differential proteomics, displaying results in a variety of graphical formats, performing similarity-driven studies with de novo sequencing data, analyzing

time-course experiments, and helping with the understanding of the biological significance of data in the light of the Gene Ontology. Databases of protein sequences are necessary in order that theoretical mass spectra generated from those databases can be compared with experimental spectra [3].

Following this, a target-decoy database needs to be generated before searching with PatternLab's integrated version of Comet.

This type of analysis is based on the fact that incorrect "decoy" sequences added to the search space will correspond to incorrect search results that could be considered correct. It will not only evaluate how many incorrect results are in a final data set but also to use decoy hits to guide the outline of filtering criteria that sensitively partition a data set into correct and incorrect identifications. This strategy is a simplistic way and powerful way to bring false positive estimations. Also, it can be applied to nearly any MS/MS workflow [26].

Proteins are then inferred by matching those identified peptide sequences to the sequences in the generated database. Proteins can be moreover grouped according to a maximum parsimony criterion, as peptides can match more than one protein [3]. Parsimony analysis is applied to derive the least protein list enough to account for the observed peptide identifications. These tools aid decrease sequence redundancy problems [27].

The best-scoring candidates are indicated as a peptide−spectrum match (PSM) with the score of the candidate. Given thousands of spectra in a data set, an general quality measure for the PSMs is checked by statistical means, often by manage the false-discovery rate (FDR), which is estimated using the previously mentioned target−decoy approach (TDA) [28].

PatternLab additionally permits quantification by extracted-ion chromatograms (XIC), which are frequently used in single-shot analyses and are gotten by plotting the intensity of a given m/z value, plus or minus a given tolerance, over a given range of time. The region or area underneath this curve, or integral, would then be able to be utilized as a surrogate for a peptide's relative abundance in the mixture and in that capacity give a premise to correlation against the XIC of the same peptide in various mixtures [3].

3.2. MaxQuant

The MaxQuant suite is a set of algorithms, which includes peak detection and scoring of peptides, it performs mass calibration and database searches for protein identification, it quantifies identified proteins, and provides summary statistics [4].

MaxQuant can recognize more than one peptide from each MS/MS spectrum, as it performs a 'second peptide' search specifically looking for signals resulting from co-fragmentation of additional precursors.

MaxQuant employs the TDA. Estimate and control the extent of false-positive identifications. Within the TDA MaxQuant uses the notion of posterior error probability (PEP) to unite multiple peptide properties, such as length, charge and number of modifications.

While the PEP provides statistical proof for individual PSMs, further workflows control global FDRs at several tiers, including PSM, protein group and Post-translational Modification (PTM) site levels. Frequently used techniques embrace label-free quantification (LFQ).

The MaxLFQ (MaxQuant LFQ) workflow is a fundamental part of MaxQuant that empowers accurate proteome-wide quantification without labeling, even for samples with peptide or protein pre-fractionation before mass spectrometric analysis.

Here, the set of all MS/MS spectra that were not identified in the Andromeda search are submitted to another search level against the already identified MS/MS spectra as database. For that reason, an especially designed match score has been built up that, based on the precursor mass difference between identified and unidentified MS/MS spectra, attempts to situate a speculative modification on each residue. The algorithm then checks whether the unidentified MS/MS spectrum can be explained as

a modified form of an already-identified spectrum. In this way, the modifications are not limited to a few predetermined masses yet rather a modification of any mass or composition can, in principle, be detected in an unbiased way [29].

4. Biomarkers

'Biomarker', a *portmanteau* of 'biological marker', relate to a broad subcategory of medical signs, in other words, objective indications of medical state observed from outside the patient, which can be measured accurately and reproducibly [30].

From pathogenesis of a disease to its clinical manifestations, there are a lot of links in the chain of events than can lead to this. A biomarker can be used at any part of this chain, at the molecular, cellular, or organ levels. Also, in order to try to manipulate the disease (symptomatically or therapeutically), a therapy might be developed to attack any of these links. Any measurement short of the actual outcome could be regarded as a surrogate endpoint biomarker. However, although all surrogate endpoints are biomarkers, not all biomarkers are useful surrogate endpoints [31].

It is significant to identify the intention that the biomarker serves in the drug development and evaluation process, when the measurement of a biomarker is considered in the evaluation of a response to therapeutic intervention [32].

A biomarker is a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes or pharmacologic responses to a therapeutic intervention [33].

Biomarkers, as in the biomedical research venture, play a critical role in improving the drug development process. To improve our understanding of normal and healthy physiology and to increase our arsenal of treatments for all diseases, it is necessary to understand well the relationship between measurable biological processes and clinical outcome. The normal physiology of a biological process, the pathophysiology of this process in the disease state and the effects of an intervention in these processes, are of great interest, since by understanding these processes, biomarkers could serve only as substitutes for the clinically relevant [30].

Biomarkers have clear potential benefits. All because information can be obtained faster and cheaper. But all this event chair of a disease process that will link the pathogenesis to the result is very fragile and the more we understand the nature of the whole path the disease takes and the whole pharmacology of a drug that affects the best biomarkers, we can in turn develop in the diagnosis, staging and monitoring of the disease in question and its response to therapy [31].

Use of appropriate biomarkers that measure biological disease parameters and therapeutic responses in humans allows a better evaluation of disease intervention strategies. High levels of scientific scrutiny and rigor, and a greater understanding of the binding of biomarkers to clinical parameters, allow the realization of the potential benefits that surrogate endpoints can bring to accelerate the development of safe and effective therapies [32].

For a certain molecule to be considered a disease biomarker, it must have:

- Sensitivity and specificity, which allows the diagnosis of the pathology in question;
- Correlation with disease severity, thus enabling prognosis;
- Response to a given treatment, predicting its effectiveness;
- Correlation with pathological mechanisms that, once directed to therapies, can change their expression levels, which translates into modifications aiming at the physiological recovery of this mechanism [34, 35].

OBJECTIVES

In this work the shotgun MS data that have been generated for SCD by INSA's lab will be studied by open-source computational proteomics platforms. The aim is by using these bioinformatics platforms to identify differentially abundant proteins related to VOC event in SCD.

The results of this study can contribute to a better understanding of this disease as well as to the identification of putative monitoring and prognostic biomarkers involved in an exacerbation VOC event.

This will contribute to a development of better health condition to patients with SCD.

MATERIALS & METHODS

1. Raw MS Data files / samples

The Raw MS data files available for computational study corresponded to the MS analysis (by ESI-LQT Orbitrap-XL, Thermo, Waltham mass spectrometer in collaboration with Lund University, Prof Peter James) of the following samples [36]:

Cytoplasmatic and membrane fractions of RBCs haemoglobin depleted obtained from children patients with SCD (n=6) at steady-state and vaso-occlusion exacerbation episode (crisis) and children control subjects (n=4) (table 3.1). Table 3.2 shows the demographic and haematological data of SCD patients.

Table 3.1 Experimental design.

Red Blood Cells Samples		Cytoplasmatic fraction	Membrane fraction
SCD (n=6)	Steady-state	6	6
	Crisis	6	6

Table 3.2 SCD Patients's demographic and haematological data.

				Subject	Crisis Events[a]	Average (T1-T0)	Range	Age at onset
Age (years)	9.6			10012	5	6.4	4-8	6.4
Gender (F/M)	3/3			10016	4	10	6-15	0.7
Ethnicity	African			10021	14	5.8	2-13	0.8
T1-T0 (months)	1-15			10022	11	6.5	1-12	8.8
Average T1-T0 (months)	6			10026	22	6	3-9	0.6
Blood analysis	Steady-state	Crisis		10033	3	4.7	2-8	1.7
Haemoglobin (g/dL)	7	7.5		Total	59	6.6[b]	1-15	3.2[c]

[a] Without two outliers (52 and 81 months)
[b] Average of averages
[c] Average age of onset (age of first symptoms)

Blood analysis	Steady-state	Crisis
Haemoglobin (g/dL)	7	7.5
Erithrocytes (×10^12 /L)	2.7	3.1
MCV (fL)	85	80
MCH (pg)	28.2	25.7
Haematocrit (%)	23	24
RDW (%)	22.5	22.5
Platelet (%)	415	376
PDW (%)	41.3	51.8
Leukocytes	11.94	16.19
Neutrophil	6.35	11.12
LDH	1401	1649
Total Bilirubin	2.43	3.2

T1 - T0: Intercrisis timelapse range
MCV: Mean cell volume
MCH: Mean Corpuscular Haemoglobin
RDW: Red cell Distribution Width
PDW: Platelet Distribuition Width
LDH: Lactate Dehydrogenase

2. Bioinformatics and Statistical Analysis

The generated MS raw data files were analysed by the "PatternLab for Proteomic" (version 4.1.0.17) [3] and "MaxQuant" (version 1.6.0.1) [4] bioinformatic platform for protein identification and label-free quantification XIC. The Homo sapiens proteome database used for this study was acquired from UniProt (The Universal Protein Resource), obtained in 2018. The statistical analysis was performed trough Microsoft® Office Excel (Microsoft, Redmond), IBM SPSS Statistics 25 (IBM, Chicago) and the integrated software tool from MaxQuant, denominated Perseus [29]; using a non-parametric approach.

2.1. Protein Identification

The two search machines "Comet" and "Andromeda" included in PatternLab for Proteomic and in MaxQuant, were used for protein identification.

For both algorithms, the same parameters had to be followed in order to have the best possible comparison of the obtained results. The following table 3.3 resumes the most important criteria used.

Table 3.3 Parameters pre-selected in both computational platforms

Post-translational Modification (PTM)	carbamidomethylation of cysteine (fixed) and several variable PTM[1]
Enzyme	trypsin
Enzyme Specificity	semi-specific
Search Mass Range (Da)	500 to 5500
Protein Score	1.8
High Resolution MS1	yes
Minimum Number of Peptides	2
Only Unique Peptides	yes
Protein False Discovery Rate	1%
Missed Cleavages	2

1 - Described in annex B.1

Golden rules for increasing confidence in protein identification were applied in the MS raw data analysis as follows:

- one or more unique peptide was considered to infer protein identification;
- identified proteins should be present at least 80% in 1 of the 2 studied sample groups;
- all potential contaminant proteins were excluded from analysis.

2.2. Computational Protein Quantification

After protein identification, the protein quantification was performed based on Normalized Ion Abundance Factor (NIAF) and Label free quantification (LFQ) Intensity, using PatternLab and MaxQuant, respectively.

From all the identified proteins, those present in at least 1 of the 2 studied groups were subjected to statistical analysis referring to their variance using the Wilcoxon-T non-parametric statistical test ($\alpha = 0.05$).

The list of proteins identified as differentially expressed by both Pattern-Lab and MaxQuant was analysed by Venn Diagram **[37]** in order to show the intersection between both computational platforms.

3. DAVID Pathway Analysis

The list of differentially expressed proteins identified by both bioinformatic platforms, were combined and analysed by the Database for Annotation, Visualization and Integrated Discovery (DAVID) **[38]** to integrate the identified protein into signaling pathways with biological meaning. The "Functional Annotation Tool" and the "Gene Ontology (GO) Biological Process Term" were selected for the analysis and data filtered by a p-value ≤ 0.05. Protein sequence and functional information was obtained by the Universal Protein Resource (UniProt) database (https://www.uniprot.org/)

4. Validation by Western Blot

A group of proteins identified as differentially expressed in SCD RBC cytoplasmatic fraction associated with vaso-occlusion crisis was selected for biochemical validation by Western blot method. The proteins peroxiredoxin-2, catalase, Hsp90 and Hsp70 were selected for validation based on the availability of antibodies in the laboratory. Validation test on membrane fractions were not performed due to inexistence of these samples.

Samples of cytoplasmatic fraction, previously prepared and kept at -80ºC **[36]**, were thawed and submitted to a Colorimetric Protein Assays using a NanoDrop™ One/OneC Microvolume UV-Vis

Spectrophotometer (Thermo Scientific™) for protein quantification. The advantage of this method is the reduced amount of sample (1-2 µL) that is needed for quantification. This is very important as there were little amount of those samples available. The blank used in this quantification was ammonium bicarbonate (NH4HCO3).

4.1. Western blot

One of the most common and powerful laboratory techniques for immunological detection of proteins is the Western blot. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE)-separated proteins are electrophoretically transferred to a polyvinylidene fluoride (PVDF) membrane which is then incubated with specific antibodies, then developed to show the protein of interest.

Protein detection methods begin with complex mixtures of proteins, separating them by molecular weight via gel electrophoresis and then transferring the separated proteins to a membrane or blot. The membrane is incubated in a solution containing an antibody that is known to bind to the target protein **[39]**.

4.1.1. One-Dimensional Electrophoresis (1D) SDS-PAGE

Electrophoresis is a technique were molecules migrate through a solution due to the application of an electric field.

1 Dimension SDS-PAGE was used to separate RBC proteins by its molecular mass.

When SDS, an anionic detergent, is dissolved it has negative net charged molecules within a wide pH range. The compound works by disrupting non-covalent bonds in the proteins, and so denaturing them, in other words, causing a loss of the native conformations and shapes from the protein molecules. By binding to the proteins with high affinity and in high concentrations, the negatively charged detergent provides all proteins with a similar net negative charge and therefore a similar charge-to-mass ratio. This way, the difference in mobility of the polypeptide chains in the gel can be attributed merely to their size as opposed to both their size and charge.

All gels used in this study are NuPAGE 4-12% Bis-Tris Mini Gels (1.0mm x 10 well) (NOVEX by life Technologies). The 4% concentration is above the resolving gel and serves to homogenize the electrophoretic front so that proteins with higher / lower molecular mass migrate at the same time, while the 12% concentration effectively separates proteins from sample.

4.1.2. Electrophoretic Run

At this stage the samples were lyophilized for a volume homogenization (10µL / lane), and then a solution of 2.5µL LDS (Lithium Dodecyl Sulfate) (4X), NuPAGE® MES SDS Running Buffer (20X), Invitrogen by Thermo Fisher Scientific, 6.5µL pure water (type II), and 1µL NuPAGE Sample Reducing Agent (10X), Invitrogen by Thermo Fisher Scientific, followed by stirring at 70°C for 10 minutes to denature the sample proteins. Samples are loaded into wells in the gel. One lane is usually reserved for a marker, therefore 3µl marker (MagicMark™ XP, Western Protein Standard, Invitrogen by Thermo Fisher Scientific) was applied to the gel.

The electrophoretic run was performed at a potential of continuous 150V with a denaturing run buffer (NuPAGE® MES SDS Running Buffer (20X), NOVEX by life Technologies), 30mL of concentrated buffer with 570mL of type II water, to a final volume of 600mL. Moreover, 500µL Antioxidant (NuPAGE ™ ® Antioxidant, Invitrogen by Thermo Fisher Scientific) was added to this solution, due to the Reducing Agent that was added previously in the samples.

4.1.3. Transfer

After protein separation by SDS-PAGE, the transfer process was performed across PVDF membranes (PVDF Blotting Membrane 0.45µm 300mm×4mm, Amersham™ Hybond™, GE Healthcare by life Science).

Previously, membranes were activated with 100% (v/v) methanol (Honeywell, Riedel-de-Haën™) for 10 minutes with shaking and then washed with type II water for 5 minutes before being incubated with complete transfer buffer (Annex A1.2) for 10 minutes with stirring. Filter papers (3MM) and sponges used to pressure and increase contact between gel and membrane were also soaked in the complete   transfer   buffer.

The sandwich assembly was performed and placed in a vat with transfer buffer shown in Figure 3.1. The transfer took place at 4 ° C for 1 hour and 30 minutes at continuous 400 mA.
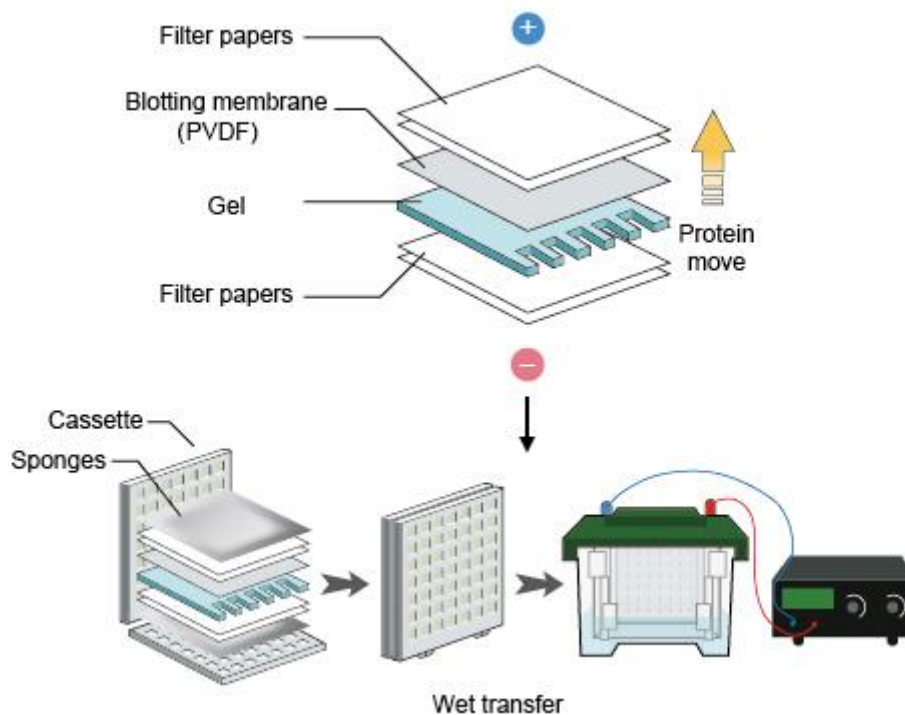

Figure 3.1 Assembly diagram for western blotting transfer (Source: **[40]**).

Following transfer, the membranes were stained with 0.1% Ponceau S in 5% acetic acid (Annex A1.3), incubated for 5 minutes with shaking to visualize the transferred bands. It is also necessary to discolour with 5% acetic acid while shaking until the background is eliminated. After digitization of the stained membranes for further analysis, the staining was removed by several consecutive washes using the Tween 20 saline phosphate buffer, defined by PBS-T (Annex A1.5), with vigorous stirring until the membrane dye completely disappeared. Membranes may be dried between 3MM paper and stored at 4 ° C until use.

4.1.4. Blocking Stage

Membranes were blocked with a 5% solution of skimmed milk powder (Nestlé) in PBS-T buffer for 1 hour with shaking to avoid non-specific binding of the test antibody.

4.1.5. Membrane Incubation with Primary Antibody

After blockage the membrane was incubated with the desired primary antibody at the chosen concentration overnight at 4 ° C and with slow agitation. The antibodies used in this study were: Anti-Peroxiredoxin-2 (Anti-PRDX2) and Anti-Catalase (Anti-CAT) from Abcam, United Kingdom; Anti-70 kilodalton heat shock protein (Anti-Hsp70) and Anti-heat shock protein 90 (Anti-Hsp90) from Santa Cruz Biotechnology, United States of America. These antibodies were all diluted in a 5% milk powder solution in PBS-T according to table 3.4.

4.1.6. Membrane Incubation with Secondary Antibody

After the stage with the primary antibody, a wash cycle was initiated consisting of two brief washes with PBS-T buffer, a 15 minute shaking wash with PBS-T buffer and three 5 minute washings with PBS-T buffer, to remove the excess of unbound primary antibody.

Secondary antibodies used were anti-rabbit IgG (Cayman Chemical, United States of America) and anti-mouse IgG (BioLegend, United States of America), which were diluted with 5% PBS-T milk powder as shown in the table 3.4. Incubation of these antibodies was performed for 1 hour with slow shaking at room temperature. After incubation, a new wash cycle of the same duration as the previous one was again performed, except for the last wash which is done for 5 minutes with 1x PBS solution (Phosphate-Buffered Saline - Annex A1.4).

Table 3.4 Primary and secondary antibody dilution.

Primary Antibody	Dilution	Secondary Antibody	Dilution
Anti-PRDX2	1:20000	Anti-Rabbit	1:75000
Anti-CAT	1:7500	Anti-Mouse	1:5000
Anti-Hsp70	1:1500	Anti-Mouse	1:15000
Anti-Hsp90	1:250	Anti-Mouse	1:15000

4.1.7. Detection Phase

This phase was performed in the darkroom, where the PVDF membrane was encased in a chemiluminescence detection solution, Enhanced ChemiLuminescence (ECL), ECL™ Western blotting detection reagents, from Amersham™, GE Healthcare. The membrane is placed in the cassette and then covered for 2 minutes with the detection solution (1mL per membrane). An medical x-ray development film (Fujifilm, Japan) was then placed over the membrane and the cassette was closed to impress the plate and remained for some time, depending on the intensity of the bands of each protein under study. After incubation, the x-ray film was revealed on a processor (Medical X-Ray Processor - Kodak).

4.1.8. Quantification of protein bands

Following obtaining the results of the development of each protein (PRDX2, CAT, Hsp70 and Hsp90), each x-ray film (plate) was scanned and the images were saved in TIFF format (16bits). Subsequently, protein bands were analysed by densitometry using the ImajeJ (https://imagej.nih.gov/ij/index.html). This software evaluates the amount of protein by optical density, allowing you to select an area in the protein / band. From the images obtained from the previously mentioned Ponceau S stained membranes, the total band was quantified, which is the normalizing protein of this study. Annex C.1 is an example of a Ponceau S-stained band membrane used for the normalization of the studied samples.

With the selected vertical bands for each sample, a graph with different peaks and areas was obtained in relation to the pixel density of each band, corresponding to the protein under study.

Data acquired from ImageJ was processed in Excel (Microsoft Office, version 2016). In this stage, the value of the band area under study was calculated in relation to the normalizing band area corresponding to the same sample and again normalized with a marker band (60 kilodalton), this means, this calculation serves to obtain the normalized values in each sample of the different proteins under study. These results were used for graphing and statistical analysis (*Wilcoxon-T* non-parametric statistical test) with support to IBM SPSS Statistics 25.

5. Membrane stripping

Stripping is the removal of primary and secondary antibodies from a western blot membrane to allow its reuse in a new Western Blot test. Due to the small amount of sample available and in order to perform four antibody incubations, membrane stripping was performed. To minimize sample loss during membrane transfer, it is highly recommended the use of PVDF membrane. ECL is also recommended as developer reagent since it is more sensitive than colorimetric reagents. The advantages of using this method are saving samples, materials, and time.

There are two protocols that differentiate in stripping intensity. In this study we used Mild stripping (Annex A1.6) that was already optimized in the laboratory with positive results.

The stripping process is simple, just wash for 20 minutes with stripping buffer, discard the buffer, rewash for 20 minutes with stripping buffer, and discard the buffer again. Two washings of 10 minutes with PBS and ending with two washes for 5 minutes with PBS-T. It is now ready for blocking stage (point 3.3.4).

RESULTS

1. Computational Protein Identification

1.1. PatternLab for Proteomics

The total number of proteins inferred was 2568 for the cytoplasmatic and 1905 for the membrane fraction. However, after applying quite stringent parameters taking into consideration the low number of samples (see Material & Methods), only 142 cytoplasmatic proteins and 44 membrane proteins were considered for further quantification (Annex E1).

1.2. MaxQuant

In this software the total number of inferred proteins was 593 cytoplasmatic proteins and 491 membrane proteins. The number is inferior in comparison to PatternLab because in the MaxQuant identification process the chosen parameters were to restrained to one unique peptide. So, regarding to this we decided to apply as the second golden rule: presence of protein in 80% in at least one cohort. Applying this rule only 114 cytoplasmatic proteins and 32 membrane proteins were considered for further quantification (Annex E2).

1.3. Venn Diagram

The Venn diagram shows the intersections identified between the Pattern-Lab and MaxQuant data (Figure 4.1). 66 (25.78%) of a total 256 proteins expressed in the cytoplasmatic fraction were equally identified by both bioinformatic platforms. From that 48 and 76 were exclusively identified by MaxQuant and PatternLab, respectively (Figure 4.1A).

In Figure 4.1B shows that 19 (25%) of a total 76 proteins expressed in the membrane fraction were equally identified by both bioinformatics platforms. From that, 13 and 25 were exclusively identified by MaxQuant and PatternLab, respectively.
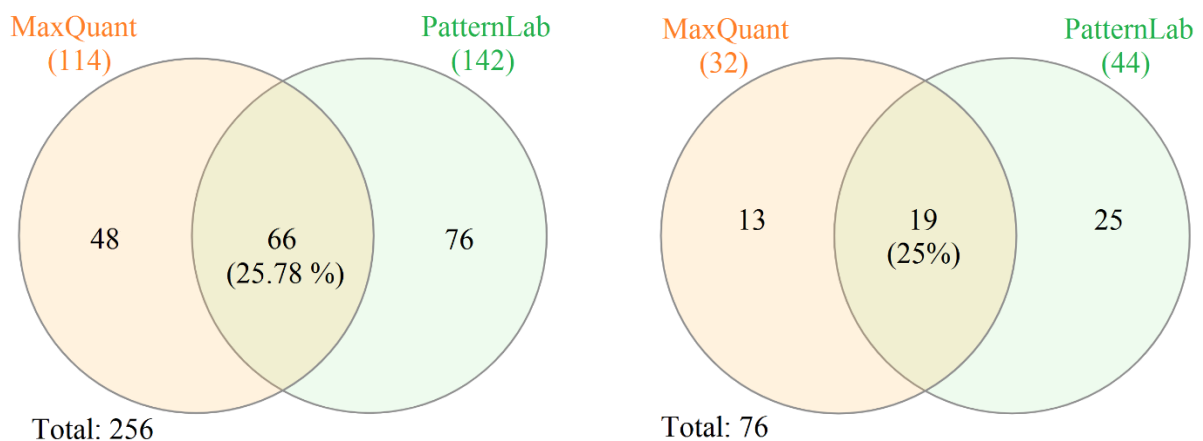

Figure 4.1 Venn Diagram with the (A) Cytoplasmatic fraction, and (B) Membrane fraction.

Table 4.1 Proteins equally identified by both Bioinformatic platforms

Acession Number	Protein name	Fraction
P31946	14-3-3 protein beta/alpha	
P62258	14-3-3 protein epsilon	Cytoplasmatic
P61981	14-3-3 protein gamma	

Acession Number	Protein name	Fraction
P27348	14-3-3 protein theta	Cytoplasmatic
P23526	Adenosylhomocysteinase	
P16157	Ankyrin-1	
P17174	Aspartate aminotransferase, cytoplasmatic	
P02730	Band 3 anion transport protein	
P31939	Bifunctional purine biosynthesis protein PURH	
P07738	Bisphosphoglycerate mutase	
P04040	Catalase	
O00299	Chloride intracellular channel protein 1	
P13716	Delta-aminolevulinic acid dehydratase	
Q16531	DNA damage-binding protein 1	
P48507	Glutamate--cysteine ligase regulatory subunit	
P78417	Glutathione S-transferase omega-1	
P04406	Glyceraldehyde-3-phosphate dehydrogenase	
P07900	Heat shock protein HSP 90-alpha	
P69905	Haemoglobin subunit alpha	
P68871	Haemoglobin subunit beta	
P02042	Haemoglobin subunit delta	
P00492	Hypoxanthine-guanine phosphoribosyltransferase	
Q04760	Lactoylglutathione lyase	
Q9BS40	Latexin	
P30740	Leukocyte elastase inhibitor	
P09960	Leukotriene A-4 hydrolase	
P00338	L-lactate dehydrogenase A chain	
P07195	L-lactate dehydrogenase B chain	
Q13228	Methanethiol oxidase	
P26038	Moesin	
Q9UNZ2	NSFL1 cofactor p47	
Q06830	Peroxiredoxin-1	
P32119	Peroxiredoxin-2	
P30041	Peroxiredoxin-6	
P18669	Phosphoglycerate mutase 1	
Q9GZP4	PITH domain-containing protein 1	
P13796	Plastin-2	
Q15102	Platelet-activating factor acetylhydrolase IB subunit gamma	
Q06323	Proteasome activator complex subunit 1	
A0A024RA52	Proteasome subunit alpha type	
P25786	Proteasome subunit alpha type-1	
P25788	Proteasome subunit alpha type-3	
P28066	Proteasome subunit alpha type-5	
O14818	Proteasome subunit alpha type-7	
P20618	Proteasome subunit beta type-1	
P49721	Proteasome subunit beta type-2	
P49720	Proteasome subunit beta type-3	

Acession Number	Protein name	Fraction
P28070	Proteasome subunit beta type-4	Cytoplasmatic
P28074	Proteasome subunit beta type-5	
P28072	Proteasome subunit beta type-6	
Q99436	Proteasome subunit beta type-7	
Q5TDH0	Protein DDI1 homolog 2	
P06703	Protein S100-A6	
P06702	Protein S100-A9	
P00491	Purine nucleoside phosphorylase	
P50395	Rab GDP dissociation inhibitor beta	
P00352	Retinal dehydrogenase 1	
P49247	Ribose-5-phosphate isomerase	
P02787	Serotransferrin	
P31948	Stress-induced-phosphoprotein 1	
P00441	Superoxide dismutase [Cu-Zn]	
P10599	Thioredoxin	
P37837	Transaldolase	
P55072	Transitional endoplasmic reticulum ATPase	
P60174	Triosephosphate isomerase	
P54725	UV excision repair protein RAD23 homolog A	
Acession Number	**Protein name**	**Fraction**
Q00013	55 kDa erythrocyte membrane protein	Membrane
P16157	Ankyrin-1	
P02730	Band 3 anion transport protein	
P08311	Cathepsin G	
P27105	Erythrocyte band 7 integral membrane protein	
Q96PL5	Erythroid membrane-associated protein	
P17931	Galectin-3	
P69905	Haemoglobin subunit alpha	
P68871	Haemoglobin subunit beta	
P02042	Haemoglobin subunit delta	
P62805	Histone H4	
P23276	Kell blood group glycoprotein	
O15439	Multidrug resistance-associated protein 4	
P00387	NADH-cytochrome b5 reductase 3	
P32119	Peroxiredoxin-2	
P11166	Solute carrier family 2, facilitated glucose transporter member 1	
Q9NP59	Solute carrier family 40 member 1	
P02549	Spectrin alpha chain, erythrocytic 1	
P11277	Spectrin beta chain, erythrocytic	

2. Protein Quantification

2.1. PatternLab

Protein quantification in PatternLab was performed by Label-free approach, using NIAF values and, using excel, a non-parametric Wilcoxon T-test statistical test.

According to all criteria and statistical tests (p-value ≤ 0.05), 3 cytoplasmatic fraction proteins and 9 membrane fraction proteins were identified as differentially expressed in SCD at crisis compared to steady-state (Table 4.2)

Table 4.2 Differentially Expressed Proteins in SCD at crisis state by PatternLab

Acession Number	Protein names	Fraction
Q08495	Dematin	Cytoplasmatic
P26038	Moesin	
P37837	Transaldolase	
Acession Number	**Protein names**	**Fraction**
P13727	Bone marrow proteoglycan	Membrane
P00918	Carbonic anhydrase 2	
P12724	Eosinophil cationic protein	
P11678	Eosinophil peroxidase	
P69905	Haemoglobin subunit alpha	
P02042	Haemoglobin subunit delta	
O95197	Reticulon-3	
P21796	Voltage-dependent anion-selective channel protein 1	
P45880	Voltage-dependent anion-selective channel protein 2	

2.2. MaxQuant

Protein quantification in MaxQuant was performed by Label-free approach, using LFQ intensity values and, using Perseus, a non-parametric Wilcoxon T-test statistical test.

According to all criteria and statistical tests (p-value ≤ 0.05), 71 cytoplasmatic fraction proteins and 28 membrane fraction proteins were identified as differentially expressed in SCD at crisis compared to steady-state (Table 4.3).

Table 4.3 Differentially Expressed Proteins in MaxQuant.

Acession Number	Protein names	Fraction
P31946	14-3-3 protein beta/alpha	Cytoplasmatic
P62258	14-3-3 protein epsilon	
P63104	14-3-3 protein zeta/delta	
F5H5V4	26S proteasome non-ATPase regulatory subunit 9	
P63261	Actin, cytoplasmatic 2	
P13798	Acylamino-acid-releasing enzyme	
E7EPV7	Alpha-synuclein	
P17174	Aspartate aminotransferase, cytoplasmatic	
P02730	Band 3 anion transport protein	
P31939	Bifunctional purine biosynthesis protein PURH	
P07738	Bisphosphoglycerate mutase	
P0DP25	Calmodulin-3	

Acession Number	Protein names	Fraction
P04040	Catalase	
O00299	Chloride intracellular channel protein 1	
P30046	D-dopachrome decarboxylase	
P13716	Delta-aminolevulinic acid dehydratase	
Q9NY33	Dipeptidyl peptidase 3	
I3L397	Eukaryotic translation initiation factor 5A	
P30043	Flavin reductase	
P09104	Gamma-enolase	
E9PKL9	GDP-L-fucose synthase	
P48506	Glutamate--cysteine ligase catalytic subunit	
P78417	Glutathione S-transferase omega-1	
P09211	Glutathione S-transferase P	
P0DMV9	Heat shock 70 kDa protein 1B	
P11142	Heat shock cognate 71 kDa protein	
P07900	Heat shock protein HSP 90-alpha	
P69905	Haemoglobin subunit alpha	
P68871	Haemoglobin subunit beta	
P02042	Haemoglobin subunit delta	
P50502	Hsc70-interacting protein	
P00492	Hypoxanthine-guanine phosphoribosyltransferase	
Q9BS40	Latexin	
P09960	Leukotriene A-4 hydrolase	Cytoplasmatic
P00338	L-lactate dehydrogenase A chain	
P07195	L-lactate dehydrogenase B chain	
Q13228	Methanethiol oxidase	
P26038	Moesin	
Q9UNZ2	NSFL1 cofactor p47	
Q32Q12	Nucleoside diphosphate kinase	
Q06830	Peroxiredoxin-1	
P32119	Peroxiredoxin-2	
P30041	Peroxiredoxin-6	
P18669	Phosphoglycerate mutase 1	
Q9GZP4	PITH domain-containing protein 1	
J3QS39	Polyubiquitin-B	
F5H345	Porphobilinogen deaminase	
P25788	Proteasome subunit alpha type-3	
P28066	Proteasome subunit alpha type-5	
P60900	Proteasome subunit alpha type-6	
O14818	Proteasome subunit alpha type-7	
A0A024RA52	Proteasome subunit alpha type	
P20618	Proteasome subunit beta type-1	
P49721	Proteasome subunit beta type-2	
P28070	Proteasome subunit beta type-4	

Acession Number	Protein names	Fraction
P28074	Proteasome subunit beta type-5	Cytoplasmatic
P28072	Proteasome subunit beta type-6	
Q5TDH0	Protein DDI1 homolog 2	
P06702	Protein S100-A9	
P00491	Purine nucleoside phosphorylase	
P50395	Rab GDP dissociation inhibitor beta	
P31948	Stress-induced-phosphoprotein 1	
P00441	Superoxide dismutase [Cu-Zn]	
P37837	Transaldolase	
P55072	Transitional endoplasmic reticulum ATPase	
E9PGT1	Translin	
H7BYY1	Tropomyosin 1	
J3KN67	Tropomyosin alpha-3 chain	
P15374	Ubiquitin carboxyl-terminal hydrolase isozyme L3	
P22314	Ubiquitin-like modifier-activating enzyme 1	
P54725	UV excision repair protein RAD23 homolog A	
Acession Number	**Protein names**	**Fraction**
Q00013	55 kDa erythrocyte membrane protein	Membrane
P63261	Actin, cytoplasmatic 2	
P16157	Ankyrin-1	
P29972	Aquaporin-1	
P02730	Band 3 anion transport protein	
A0A087WXM8	Basal cell adhesion molecule	
P35613	Basigin	
P00915	Carbonic anhydrase 1	
P08311	Cathepsin G	
P27105	Erythrocyte band 7 integral membrane protein	
P16452	Erythrocyte membrane protein band 4.2	
Q96PL5	Erythroid membrane-associated protein	
P17931	Galectin-3	
P69905	Haemoglobin subunit alpha	
P68871	Haemoglobin subunit beta	
P02042	Haemoglobin subunit delta	
P69891	Haemoglobin subunit gamma-1	
P62805	Histone H4	
P23276	Kell blood group glycoprotein	
O15439	Multidrug resistance-associated protein 4	
P00387	NADH-cytochrome b5 reductase 3	
P32119	Peroxiredoxin-2	
J3QS39	Polyubiquitin-B	
P11171	Protein 4.1	
P11166	Solute carrier family 2, facilitated glucose transporter member 1	

Acession Number	Protein names	Fraction
Q9NP59	Solute carrier family 40 member 1	Membrane
P02549	Spectrin alpha chain, erythrocytic 1	
P11277	Spectrin beta chain, erythrocytic	

2.3. Identified Proteins in this study reported as associated with SCD.

Table 4.4 shows the list of proteins identified as differentially expressed in this study that have already been described as associated with SCD.

Table 4.4 Differential Expressed Proteins previously associated with Sickle Cell Disease.

Protein	CelularComponent	Bioinformatic tool	Reference
Ankyrin-1	Membrane	MaxQuant	**[41-45]**
Catalase	Soluble	MaxQuant	**[41],[46],[47]**
Leukotriene A-4 hydrolase	Soluble	MaxQuant	**[44]**
L-lactate dehydrogenase A	Soluble	MaxQuant	**[48,49]**
L-lactate dehydrogenase B	Soluble	MaxQuant	**[48,49]**
Peroxiredoxin-1	Soluble	MaxQuant	**[41]**
Proteasome subunit beta type-1	Soluble	PatternLab	**[41]**
Dematin	Soluble	PatternLab	**[41],[42]**
Moesin	Soluble	PatternLab /MaxQuant	**[45],[50]**

Note: The computational analysis for the differential expression between Steady-state and Crisis were obtained: in MaxQuant through Perseus software with Label Free Quantification (LFQ) (Wilcoxon signed rank test, α = 0.05). For PatternLab, the same test was used (α = 0.05), in SPSS software, with Normalized Ion Abundance Factor (NIAF).

2.4. Database research

2.4.1. DAVID

Using DAVID it was possible, through GO terms, to search the biological processes related to the proteins. Table 4.5 shows 3 biological pathways identified, corresponding to the highest matches and with the lowest p-value. For example, regulation of messenger Ribonucleic Acid (mRNA) stability has 16 proteins of all identified proteins as differentially expressed.

Table 4.5 DAVID biological pathways for the differentially expressed proteins from PatternLab and MaxQuant.

Term	RT	Genes	Count	%	P-Value
regulation of mRNA stability	RT		16	16,0	2,7E-17
NIK/NF-kappaB signaling	RT		12	12,0	1,2E-13
negative regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle	RT		12	12,0	2,9E-13

Proteins with potential relevance in SCD were selected based on the UniProt functional and biological information (Table 4.6).

Table 4.6 Differential expressed Proteins potentially / hypothetically associated with Sickle Cell Disease Crisis Event.

Protein	Function (UniProt)	Celular Component	Search Engine
Hemoglobin subunit alpha **Hemoglobin subunit beta** **Hemoglobin subunit delta**	Involved in oxygen transport from the lung to the various peripheral tissues	Cytosol / Extracellular region or secreted	MaxQuant
Spectrin alpha chain, erythrocytic 1 **Spectrin beta chain, erythrocytic**	It associates with actin to form the cytoskeletal superstructure of the erythrocyte plasma membrane	Cytoskeleton	MaxQuant
Band 3 anion transport protein	Mediates electroneutral anion exchange across the cell membrane and as a structural protein	Plasma membrane	MaxQuant
Galectin-3	Involved in acute inflammatory responses including neutrophil activation and adhesion	Nucleus / Extracellular region or secreted	MaxQuant
NADH-cytochrome b5 reductase 3	Methemoglobin reduction in erythrocyte	Mitochondrion / Endoplasmic reticulum	MaxQuant
Kell blood group glycoprotein	Endothelin-3-converting enzyme activity	Plasma membrane	MaxQuant
Peroxiredoxin-1 **Peroxiredoxin-2**	Cell protection against oxidative stress by detoxifying peroxides and as sensor of hydrogen peroxide-mediated signaling events	Cytoplasm	MaxQuant
Erythroid membrane-associated protein	Cell-adhesion or receptor molecule of erythroid cells	Plasma membrane	MaxQuant
Proteasome subunit alpha type **Proteasome subunit alpha type-3** **Proteasome subunit alpha type-5** **Proteasome subunit alpha type-7** **Proteasome subunit beta type-2** **Proteasome subunit beta type-4** **Proteasome subunit beta type-5** **Proteasome subunit beta type-6**	Component of the 20S core proteasome complex involved in the proteolytic degradation of most intracellular proteins. This complex plays numerous essential roles within the cell by associating with different regulatory particles	Nucleus	MaxQuant
Stress-induced-phosphoprotein 1	Acts as a co-chaperone for HSP90AA1	Nucleus	MaxQuant
Superoxide dismutase [Cu-Zn]	Destroys radicals which are normally produced within the cells and which are toxic to biological systems	Nucleus / Mitochondrion	MaxQuant
Voltage-dependent anion-selective channel protein 1	Forms a channel through the mitochondrial outer membrane	Mitochondrion / Plasma membrane	PatternLab
Carbonic anhydrase 2	Stimulates the chloride-bicarbonate exchange activity of SLC26A6	Plasma membrane	PatternLab

Note: The computational analysis for the differential expression between Steady-state and Crisis were obtained: in MaxQuant through Perseus software with Label Free Quantification (LFQ) (Wilcoxon signed rank test, $\alpha = 0.05$). For PatternLab, the same test was used ($\alpha = 0.05$), in SPSS software, with Normalized Ion Abundance Factor (NIAF).

3. Validation

To validate the computational data obtained for cytoplasmatic fraction, four proteins PRDX2, CAT, Hsp70 and Hsp90 were selected for Western blotting analysis, taking into consideration the the availability of antibodies and samples in the laboratory. The results obtained although only statistically significant for PRDX2, validated the proteomic computational quantitative data by showing that those proteins have up-regulation tendency in SDC at crisis-state (Figure 4.2 to Figure 4.9).
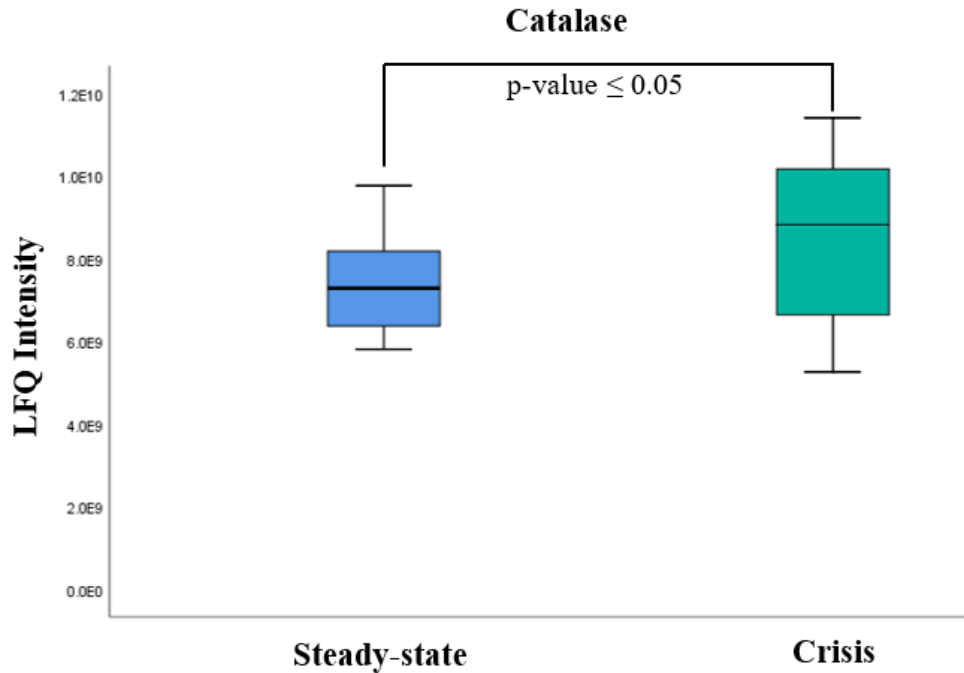
3.1. Peroxiredoxin-2 (PRDX2)
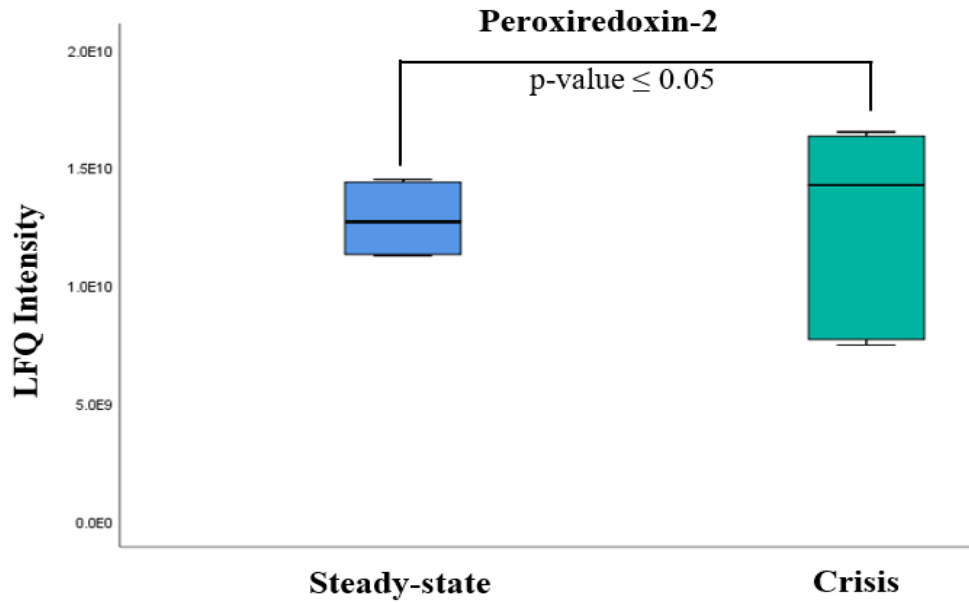
Figure 4.2 MaxQuant LFQ Intensity values with steady-state and crisis, in the cytoplasmatic fraction for PRDX2.
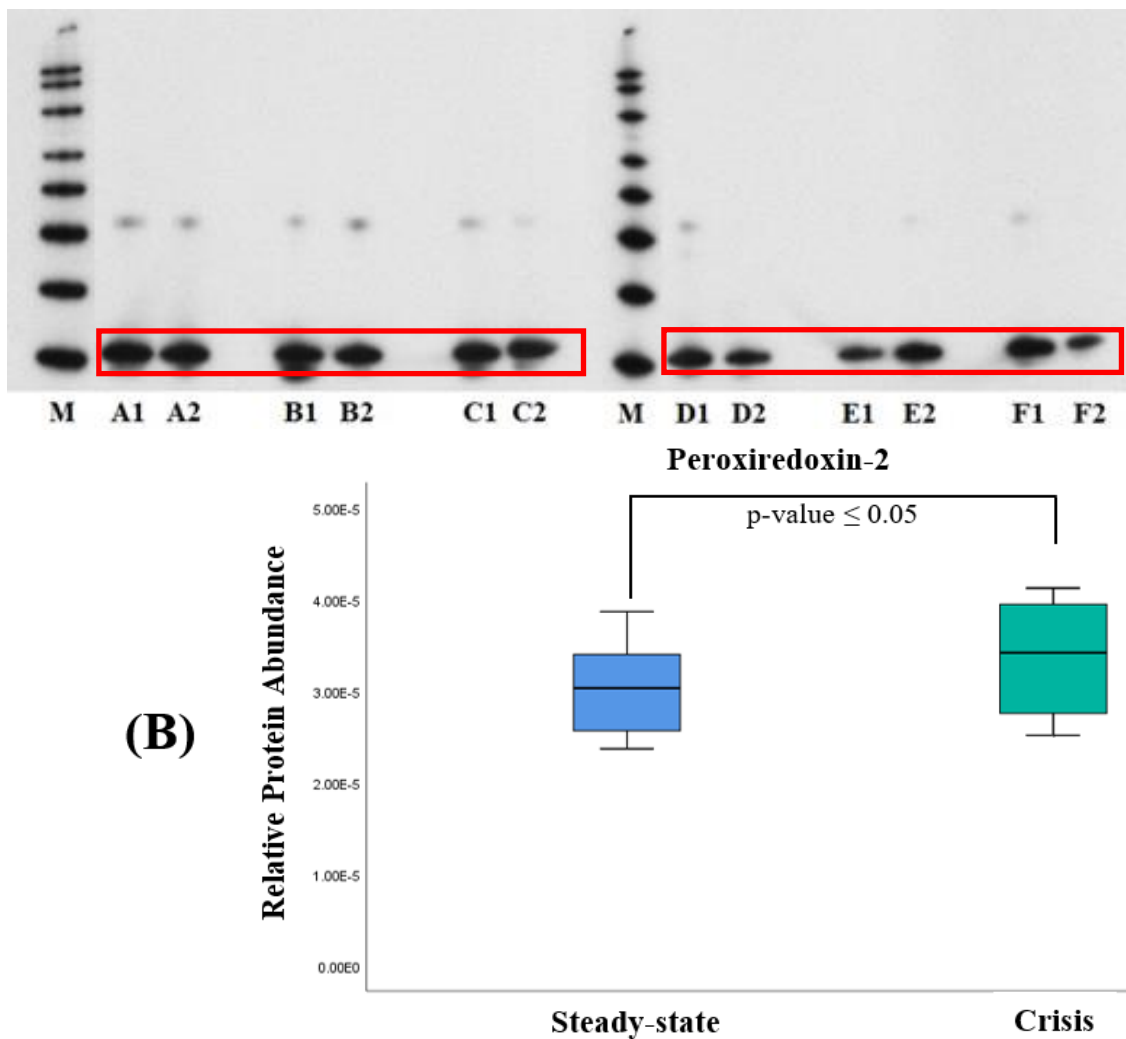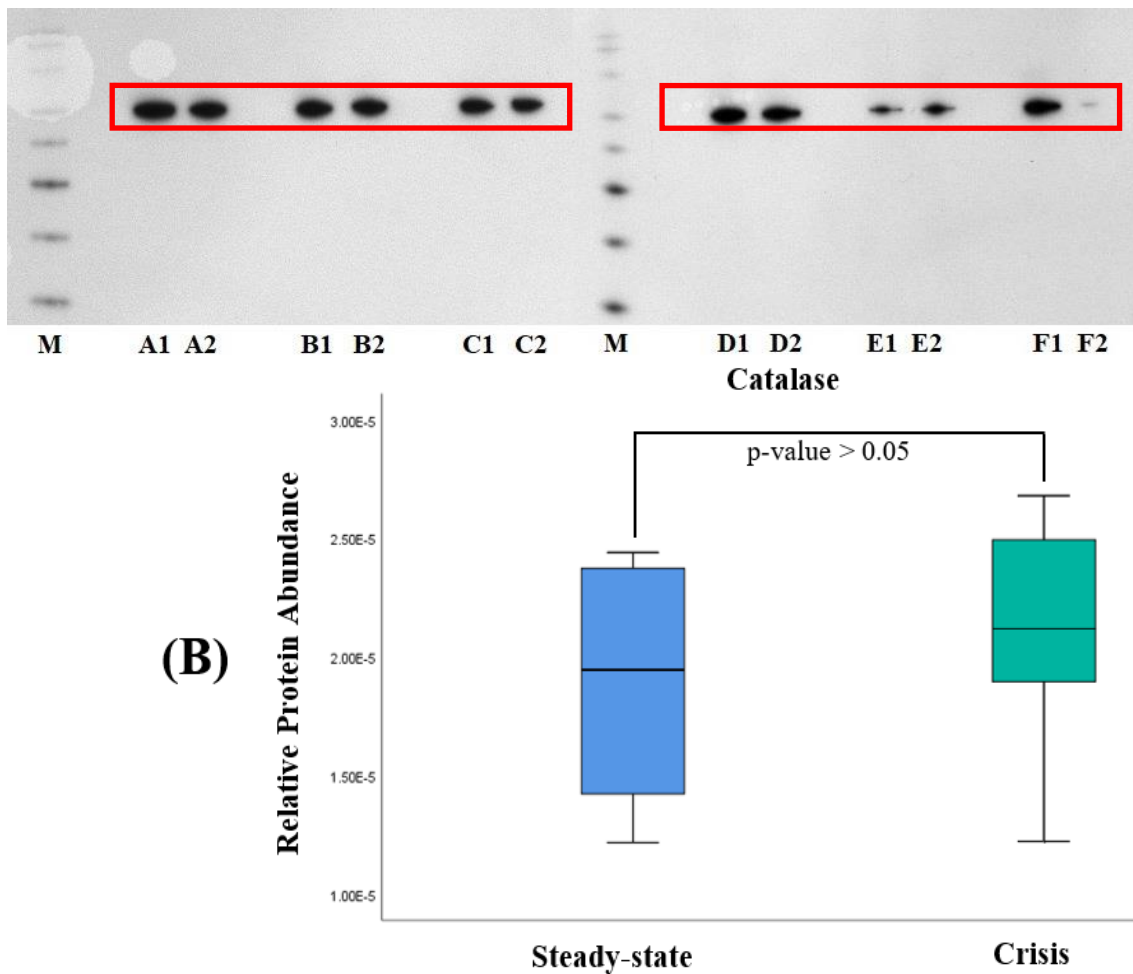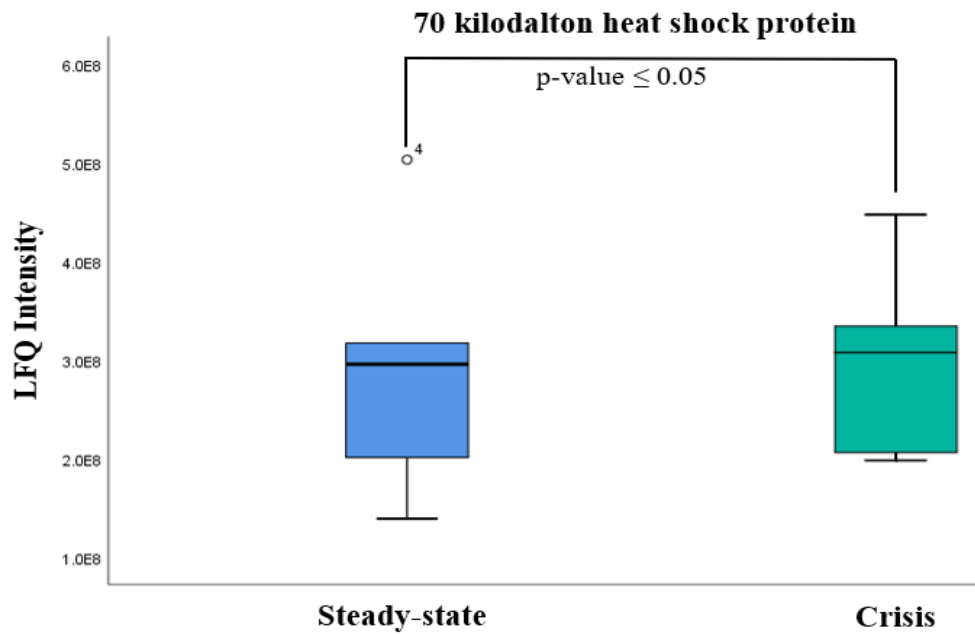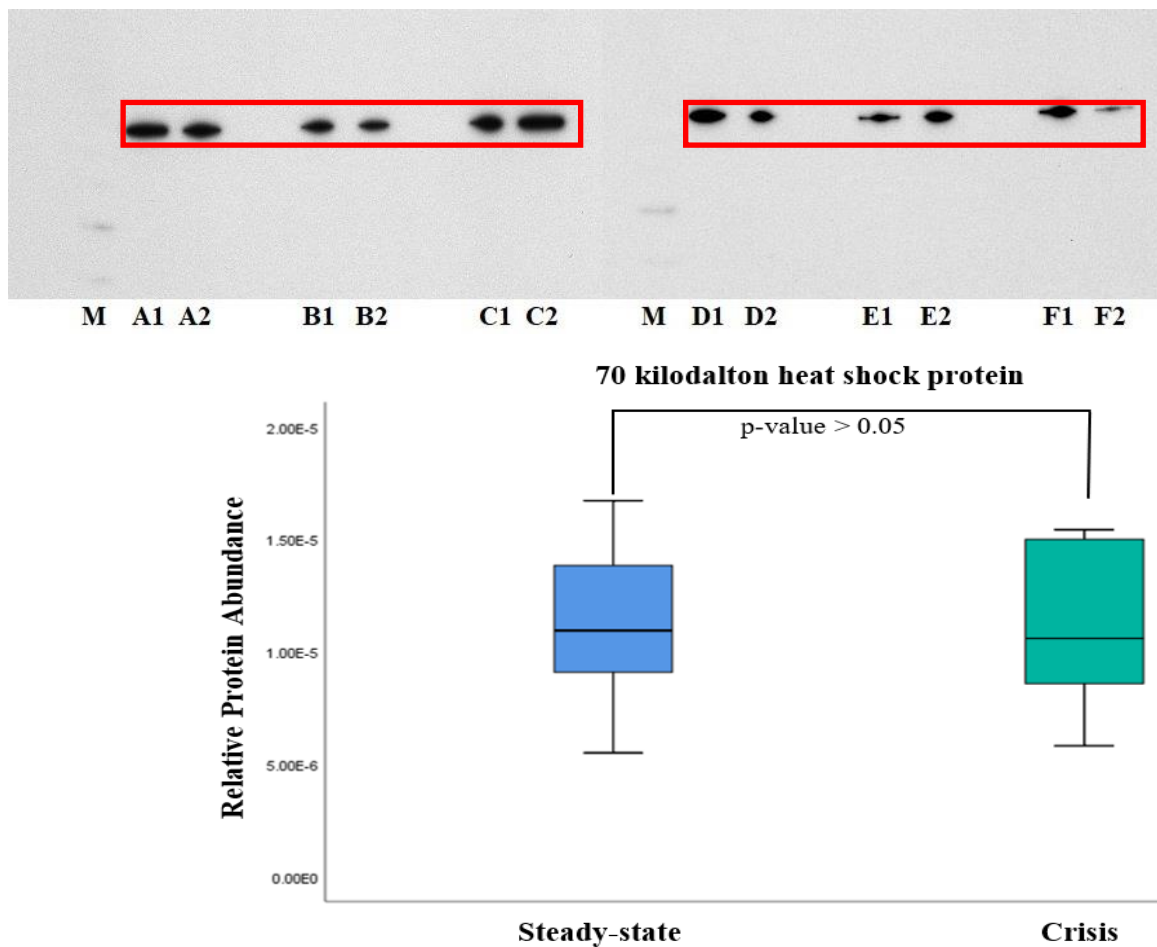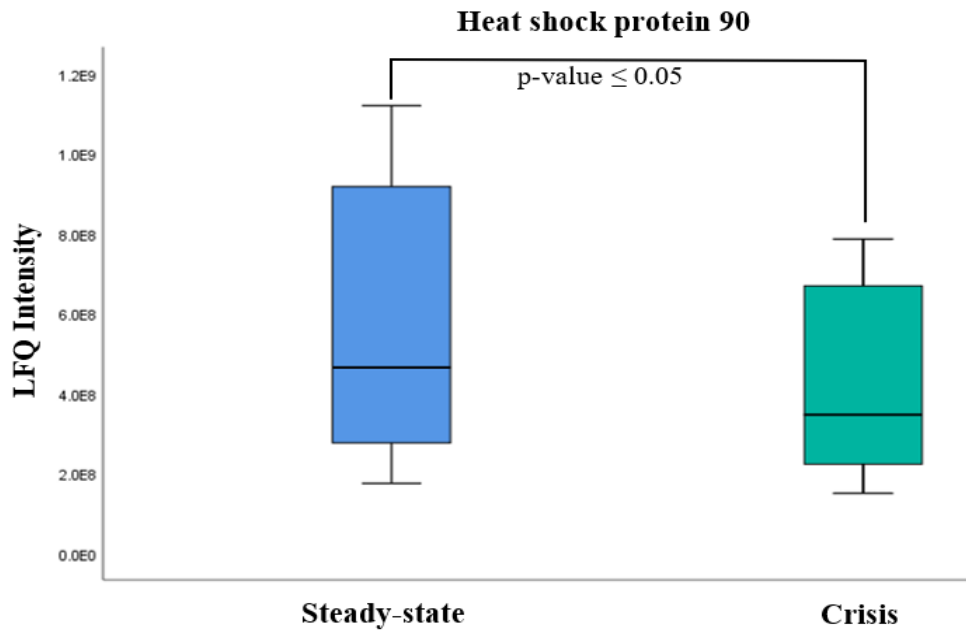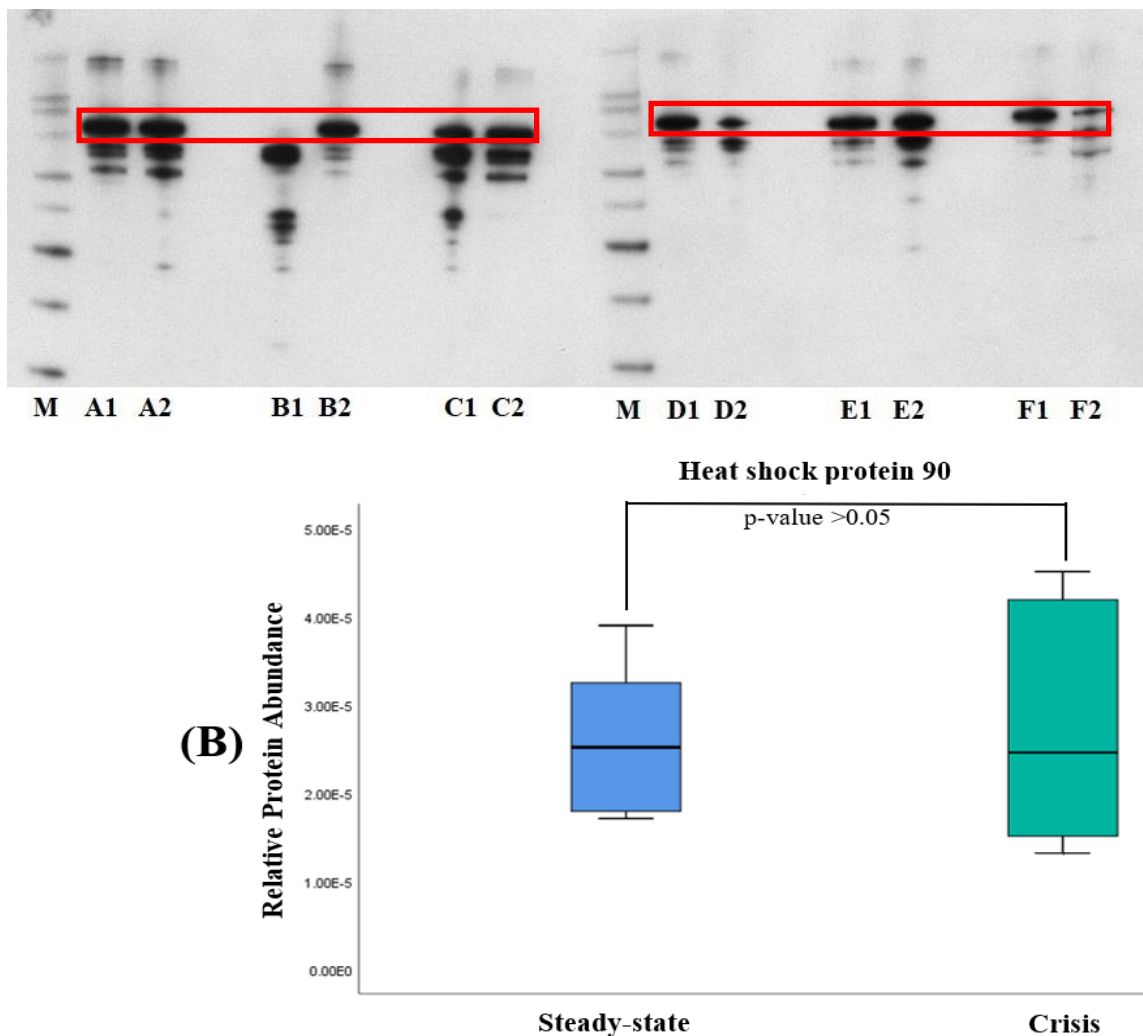

Figure 4.3 Western blot validation obtained for PRDX2 (22 kDa). (A) Western blot x-ray plate with M (MagicMark™ XP, Western Protein Standard) and 6 subjects from A to F, according to their condition, with 1 for Crisis and 2 for Steady-state. (B) Graphic with the Western blot Relative Protein Abundance obtained through ImageJ.

3.2. Catalase (CAT)

Figure 4.4 MaxQuant LFQ Intensity values with steady-state and crisis, in the cytoplasmatic fraction for CAT.

Figure 4.5 Western blot validation obtained for CAT (60 kDa). (A) Western blot x-ray plate with M (MagicMark™ XP, Western Protein Standard) and 6 subjects from A to F, according to their condition, with 1 for Crisis and 2 for Steady-state. (B) Graphic with the Western blot Relative Protein Abundance obtained through ImageJ.

3.3. 70 kilodalton heat shock protein (Hsp70)

Figure 4.6 MaxQuant LFQ Intensity values with steady-state and crisis, in the cytoplasmatic fraction for Hsp70.

Figure 4.7 Western blot validation obtained for Hsp70 (70 kDa). (A) Western blot x-ray plate with M (MagicMark™ XP, Western Protein Standard) and 6 subjects from A to F, according to their condition, with 1 for Crisis and 2 for Steady-state. (B) Graphic with the Western blot Relative Protein Abundance obtained through ImageJ

3.4. Heat shock protein 90 (Hsp90)

Figure 4.8 MaxQuant LFQ Intensity values with steady-state and crisis, in the cytoplasmatic fraction for Hsp90.

Figure 4.9 Western blot validation obtained for Hsp90 (90 kDa). (A) Western blot x-ray plate with M (MagicMark™ XP, Western Protein Standard) and 6 subjects from A to F, according to their condition, with 1 for Crisis and 2 for Steady-state. (B) Graphic with the Western blot Relative Protein Abundance obtained through ImageJ.

DISCUSSION

In this work we applied bioinformatic platforms such as Pattern-Lab and MaxQuant to analyse proteomics raw data of RBC cytoplasmatic and membrane fractions obtained from SCD patients at crisis and steady-state **[37]** in order to uncover putative candidate biomarker for SCD vaso-occlusion state. Using these bioinformatic platforms under quite stringent parameters we observed that only about 25% of 332 identified proteins in both cellular fractions were in common due probably to differences in the search machine and statistical methods used. This observation supports the importance to apply different bioinformatics platforms in parallel to increase the coverage and confidence of the obtained data. From those, 74 cytoplasmatic proteins and 37 membrane proteins were identified as differentially expressed associated with SCD crisis event.

The pathway analysis by DAVID bioinformatic platform indicated 'regulation of mRNA stability', 'NIK/NF-KappaB signaling' and 'negative regulation of ubiquitin—protein ligase activity involved in mitotic cycle' as the most significant networks as associated with the proteins identified as differentially in SCD crisis. Based on a literature survey it was interesting to confirm that some of these proteins have already been described as modulated in SCD **[41-50]**.

Validation by western blot for a group of proteins (PRDX2, CAT, Hsp70 and Hsp90), although only statically significant for PRDX2, supported the proteomics data obtained. Differences in the technology in the protein identification, specificity of the antibodies, the quality of the samples (have been storage for a long time) may explain this results.

PRDX2 is the third most abundant protein in RBC (after haemoglobin and carbonic anhydrase) **[51]** and is therefore critical for erythrocyte survival **[52]**. PRDX2 belongs to the peroxiredoxins (PRDX) family and has significant roles in the cell, such as the degradation of hydrogen peroxide (H_2O_2) to water, the reduction of organic hydroperoxides and peroxynitrite, and the protection of oxidative stress in RBC **[53, 54]**. PRDX2 is associated with antioxidant defense and can be easily inactivated by hydrogen peroxide, which consequently deactivates peroxidatic activity, decreasing their ability to act as antioxidants **[55]**. Some studies have indicated the significance of PRDX2 for erythrocyte development and maintenance in hemolytic anemia **[56]**.

CAT is a 244 KDa tetrameric protein, comprising four identical 59.7 KDa subunits, and is tightly bound to reduced NADPH (Nicotinamide Adenine Dinucleotide Phosphate) **[57]**. Catalase is one of the major enzymes that catalyzes the decomposition of ROS, such as hydrogen peroxide (H_2O_2), a free radical closely related to various pathologies. To prevent damage to cells and tissues, the catalytic activity of this enzyme is the degradation of two hydrogen peroxide molecules ($2H_2O_2$) to one oxygen molecule (O_2) and two water molecules ($2H_2O$) **[58]**.

Heat shock proteins (Hsp) such as Hsp70 and Hsp90 protect cells from different conditions of stress. After misfolding, Hsp70, the most ubiquitous and highly conserved Hsp, helps proteins adopt native conformation or regain function. Notwithstanding their protective role under stress conditions, Hsp70s are involved in fundamental cellular processes where proteins are partially or completely unfolded such as translation, translocation across membranes, presentation of substrates for degradation, assembly and disassembly of macromolecular complexes or aggregates, and in gene induction and apoptosis **[59]**. Additionally, Hsp70s participate with other cellular chaperone systems including Hsp90, Hsp60 chaperonins, small heat shock proteins and Hsp100 AAA+ disaggregase, together constituting a dynamic and functionally versatile network for protein folding, unfolding, regulation, targeting, aggregation and disaggregation, as well as degradation **[60]**. Hsp90 is a chaperone protein that helps other proteins to fold properly, stabilizes proteins against heat stress, and aids in protein degradation.

The increased levels of PRDX2, CAT, Hsp70 and Hsp90 in the RBC of SCD at crisis event as observed in this study, may suggest that patients at this state may be under increased oxidative stress that can be overcome by upregulation of these proteins.

Of the differentially expressed proteins that hypothetically may be associated with SCD, a promise candidate biomarker of crisis namely the Voltage-dependent anion-selective channel protein 1 (VDAC1) that was found decreased. A novel pathway involving RBC VDAC1 releases Adenosine triphosphate (ATP) for several systemic function including vascular caliber regulation. Our results suggest that vaso-occlusion episodes in SCD patients may be associated with decreased VDAC1 in their RBCs **[61]**.

There is already a steady-state and crisis patients study comparing the gene expression patterns between sickle cell anemia pediatric patients in both conditions **[62]**. But so far there are few studies comparing proteins as is the case with this study, which makes it better for studying the biological level of proteins involved in both steady-state and crisis in SCD.

In summary, this study indicated that SCD patients at crisis- state are under oxidative stress and the proteins such as PRDX2 and VDAC1 are promising candidates biomarkers for SCD crisis-state.

The main limitations of the study concerns to the sample size, the same set of samples used in the discovery and validation studies. In further studies a larger cohort of patients, validation study in a new set of samples using MS-based technology such as Selected reaction monitoring (SRM) / Sequential Window Acquisition of All Theoretical Mass Spectra (SWATH-MS) should be considered.

REFERENCES

[1] Aebersold R, Mann M. Mass spectrometry-based proteomics. Nature 2003; 422:198–207.

[2] Joshi K, Patil D. Proteomics. Innovative Approaches in Drug Discovery: Ethnopharmacology, Systems Biology and Holistic Targeting. Elsevier 2017; 273–294.

[3] Carvalho PC, Lima DB, Leprevost FV, Santos MDM, Fischer JSG, Aquino PF, Moresco JJ, Yates III JR, Barbosa VC. Integrated analysis of shotgun proteomic data with PatternLab for proteomics 4.0. Nature Protocols 2016; 11(1):102–117.

[4] Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 2008; 26(12): 1367–1372.

[5] Mahmood T, Yang P. Western Blot: Technique, Theory, and Trouble Shooting. North American Journal of Medical Sciences 2012; 4(9): 429–434.

[6] Galdos A, Piza ART, Maria DA, Miglino MA. *Proteómica: novas fronteiras na pesquisa. Enciclopédia biosfera* 2010; 6(11): 1–24.

[7] Misra A. Challenges in Delivery of Therapeutic Genomics and Proteomics. Elsevier Insights 2011. 387–427.

[8] Anderson J, Johansson HJ, Graham CS, Vesterlund M, Pham MT, Bramlett CS, Montgomery EN, Mellema MS, Bardini RL, Contreras Z, Hoon M, Bauer G, Fink KD, Fury B, Hendrix KJ, Chedin F, Andaloussi SEL, Hwang B, Mulligan MS, Lehtiö J, Nolta JA. Comprehensive Proteomic Analysis of Mesenchymal Stem Cell Exosomes Reveals Modulation of Angiogenesis via Nuclear Factor-KappaB Signaling. Stem cells Technology 2016; 34(3):601–613.

[9] O'Neil SE, Palviainen MJ, Ten HS, Filiou M, Gonzalez A, Hodge K, Surinova S, Penque D, Baker MS. Clinical proteomics stretch goals: EuPA 2012 roundtable report. Journal of Proteomics 2013; 88:37–40.

[10] Cash P. Proteomics: the protein revolution. Biologist (London) 2002; 49(2):58–62.

[11] Feliciano A, Torres VM, Vaz F, Carvalho AS, Matthiesen R, Pinto P, Malhotra A, Bárbara C, Penque D. Overview of proteomics studies in obstructive sleep apnea. Sleep Med 2015; 16(4):437–445.

[12] Feliciano A, Vaz F, Valentim-Coelho C, Torres VM, Silva R, Prosinecki V, Alexandre BM, Almeida A, Almeida-Marques C, Carvalho AS, Matthiesen R, Malhotra A, Pinto P, Bárbara C, Penque D. Evening and morning alterations in Obstructive Sleep Apnea red blood cell proteome. Data Brief 2017; 11:103–110.

[13] Feliciano A, Vaz F, Torres VM, Valentim-Coelho C, Silva R, Prosinecki V, Alexandre BM, Carvalho AS, Matthiesen R, Malhotra A, Pinto P, Bárbara C, Penque D. Evening and morning peroxiredoxin-2 redox/oligomeric state changes in obstructive sleep apnea red blood cells: Correlation with polysomnographic and metabolic parameters. Biochim Biophys Acta Mol Basis Dis 2017; 1863 (2):621–629.

[14] Rudy HL, Yang D, Nam AD, Cho W. Review of Sickle Cell Diseaseand Spinal Pathology. Global Spine Journal 2019; 9(7):761–766.

[15] Rees DC, Williams TN, Gladwin MT. Sickle-cell disease. The Lancet 2010; 376(9757):2018–2031.

[16] Bender MA. Sickle Cell Disease. GeneReviews 2017.

[17] Piel FB, Steinberg MH, Rees DC. Sickle Cell Disease. The New England Journal of Medicine 2017; 376(16):1561–1573.

[18] Piel FB, Patil AP, Howes RE, Nyangiri OA, Gething PW, et al. 2013. Global epidemiology of sickle haemoglobin in neonates: a contemporary geostatistical model-based map and population estimates. The Lancet 2012; 381:142–51.

[19] American Academy of Family Physicians. Available in https://familydoctor.org/condition/sickle-cell-disease. Access in September 9, 2019.

[20] Sundd P, Gladwin MT, Novelli EM. Pathophysiology of Sickle Cell Disease. Annual Review of Pathology: Mechanisms of Disease 2019; 14:263–92.

[21] Lu L, Li Z, Li H, Li X, Vekilov PG, Karniadakis GE. Quantitative prediction of erythrocyte sickling for the development of advanced sickle cell therapies. Science Advances 2019; 5(8):1-11.

[22] The Editors of Encyclopaedia Britannica. Available in https://www.britannica.com/science/sickle-cell-anemia. Access in September 12, 2019.

[23] Frimpong A, Thiam LG, Arko-Boham B, Owusu EDA, Adjei GO. Safety and effectiveness of antimalarial therapy in sickle cell disease: a systematic review and network meta-analysis. BMC Infectious Diseases 2018; 18(1):650.

[24] Cannataro M. Computational proteomics: management and analysis of proteomics data. Briefings in Bioinformatics 2007; 9(2):97–101.

[25] Colinge J, Bennett K. Introduction to Computational Proteomics. PLoS Computational Biology. 2007; 3(7): e114.

[26] Elias JE, Gygi SP. Target-Decoy Search Strategy for Mass Spectrometry-Based Proteomics. Methods Mol Biol 2010; 604:55–71.

[27] Zhang B, Chambers MC, Tabb DL. Proteomic Parsimony through Bipartite Graph Analysis Improves Accuracy and Transparency. J Proteome Res 2007; 6(9):3549–3557.

[28] Kirik U, Refsgaard JC, Jensen LJ. Improving Peptide-Spectrum Matching by Fragmentation Prediction Using Hidden Markov Models. J.ProteomeRes 2019; 18(6):2385−2396.

[29] Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, Mann M, Cox J. The Perseus computational platform for comprehensive analysis of (prote)omics data. Nature Methods 2016; 13(9):731–740.

[30] Strimbu K, Tavel J. What are biomarkers? Wolters Kluwer Health 2010; 5(6):463–466.

[31] Aronso J. Biomarkers and surrogate endpoints. British Journal of Clinical Pharmacology 2005; 59(5):491–494.

[32] Biomarkers definitions working group. Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. Clinical Pharmacology & Therapeutics 2001; 69(3):89–95.

[33] Mischak H, Ioannidis JP, Argiles A, Attwood TK, Bongcam-Rudloff E, Broenstrup M, Charonis A, Chrousos GP, Delles C, Dominiczak A, Dylag T, Ehrich J, Egido J, Findeisen P, Jankowski J, Johnson RW, Julien BA, Lankisch T, Leung HY, Maahs D, Magni F, Manns MP, Manolis E, Mayer G, Navis G, Novak J, Ortiz A, Persson F, Peter K, Riese HH, Rossing P, Sattar N, Spasovski G, Thongboonkerd V, Vanholder R, Schanstra JP, Vlahou A. Implementation of proteomic biomarkers: making it work. European Journal of Clinical Investigation 2012; 42(9):1027–1036.

[34] Penque D. Two-dimensional gel electrophoresis and mass spectrometry for biomarker discovery. Proteomics Clin Appl 2009; 3(2):155-172.

[35] Shih JL, Malhotra A. Could vitamins be helpful to patients with sleep apnea? Chest 2011; 139(2):237-238.

[36] Pacheco SA. Translating proteomics into environmental health: Biomarkers discovery for tobacco smoke-induced biological damage. Faculdade de Ciências da Universidade de Lisboa 2015. PhD final Thesis.

[37] Heberle H, Meirelles GV, Silva FR, Telles G P, Minghim R. InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. BMC Bioinformatics 2015; 16(1):169.

[38] Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Protocols 2009; 4(1):44–57.

[39] Kim B. Western Blot Techniques. Methods Mol Biol 2017; 1606:133–139.

[40] Creative Diagnostics. Available in https://www.creative-diagnostics.com/Electrophoresis-Protein-Transfer.htm. Access in September 9, 2019.

[41] Kakhniashvili DG, Griko NB, Bulla LA, Goodman SR. The Proteomics of Sickle Cell Disease Profiling of Erythrocyte Membrane Proteins by 2D-DIGE and Tandem Mass Spectrometry. Exp Biol Med 2006; 230(11):787-92.

[42] Koshino I, Mohandas N, Takakuwa Y. Identification of a Novel Role for Dematin in Regulating Red Cell Membrane Function by Modulating Spectrin-Actin Interaction. J Biol Chem 2012; 287(42):35244-50.

[43] Andolfo I, Russo R, Gambale A, Iolascon A. New insights on hereditary erythrocyte membrane defects. Haematologica 2016; 101(11): 1284–1294.

[44] Goodman SR, Pace BS, Hansen KC, D'alessandro A, Xia Y, Daescu O, Glatt SJ. Minireview: Multiomic candidate biomarkers for clinical manifestations of sickle cell severity: Early steps to precision medicin. Exp Biol Med 2016; 241(7):772–781.

[45] Lu Y, Hanada T, Fujiwara Y, Nwankwo JO, Wieschhaus AJ, Hartwig J, Huang S, Han J, Chishti AH. Gene disruption of dematin causes precipitous loss of erythrocyte membrane stability and severe hemolytic anemia. Blood 2016; 128(1):93-103.

[46] Quinn CT, Minireview: Clinical severity in sickle cell disease: the challenges of definition and prognostication. Exp Biol Med 2016; 241(7):679-88.

[47] Antwi-Boasiako C, Dankwah GB, Aryee R, Hayfron-Benjamin C, Donkor ES, Campbell AD. Oxidative Profile of Patients with Sickle Cell Disease. Med Sci 2019; 7(2): E17.

[48] Kato GJ, McGowan V, Machado RF, Little JA, Taylor J 6th, Morris CR, Nichols JS, Wang X, Poljakovic M, Morris SM Jr, Gladwin MT. Lactate dehydrogenase as a biomarker of hemolysis-associated nitric oxide resistance, priapism, leg ulceration, pulmonary hypertension, and death in patients with sickle cell disease. Blood 2019; 107(6):2279-85.

[49] Ballas SK. Lactate dehydrogenase and hemolysis in sickle cell disease. Blood 2019; 121(1):243-4.

[50] Kawaguchi K, Yoshida S, Hatano R, Asano S. Pathophysiological Roles of Ezrin/Radixin/Moesin Proteins. Biol Pharm Bull 2017; 40(4):381-390.

[51] Bayer SB, Hampton MB, Winterbourn CC. Accumulation of oxidized peroxiredoxin 2 in red blood cells and its prevention. Transfusion 2015; 55(8):1909-1918.

[52] Bayer SB, Maghzal G, Stocker R, Hampton MB, Winterbourn CC. Neutrophil-mediated oxidation of erythrocyte peroxiredoxin 2 as a potential marker of oxidative stress in inflammation. FASEB J 2013; 27(8):3315-22.

[53] Yan S, Chen S, Li Z, Wang H, Huang T, Wang X, Wang J. Biochemical characterization of human peroxiredoxin 2, an antioxidative protein. Acta Biochim Biophys Sin 2012; 44(9):759-764.

[54] Hoyle NP, O'Neill JS. Oxidation-reduction cycles of peroxiredoxin proteins and non-transcriptional aspects of timekeeping. Biochemistry 2015; 54(2):184-193.

[55] Poynton RA, Hampton MB. Peroxiredoxins as biomarkers of oxidative stress. Biochim Biophys Acta 2014; 1840(2):906-912.

[56] Romanello KS, Teixeira KKL, Silva JPMO, Nagamatsu ST, Bezerra MAC, Domingos IF, Martins DAP, Araujo AS, Lanaro C, Breyer CA, Ferreira RA, Franco-Penteado C, Costa FF, Malavazi I, Netto LES, de Oliveira MA, Cunha AF. Global analysis of erythroid cells redox status reveals the involvement of Prdx1 and Prdx2 in the severity of beta thalassemia. PLoS One 2018; 13(12):1-19.

[57] Kodydková J, Vávrová L, Kocík M, Zák A. Human catalase, its polymorphisms, regulation and changes of its activity in different diseases. Folia Biol 2014; 60(4):153-167.

[58] Huang XN, Du XY, Xing JF, Ge ZQ. Catalase-only nanoparticles prepared by shear alone: Characteristics, activity and stability evaluation. Int J Biol Macromol 2015; 90:81-8.

[59] Sharma D, Masison DC. Hsp70 Structure, Function, Regulation and Influence on Yeast Prions. Protein Pept Lett 2009; 16(6):571–81.

[60] Rosenzweig R, Nillegoda NB, Mayer MP, Bukau B. The Hsp70 chaperone network. Nature Reviews Molecular Cell Biology 2019; 20(11):665-680.

[61] Marginedas-Freixa I, Alvarez CL, Moras M, Leal Denis MF, Hattab C, Halle F, Bihel F, Mouro-Chanteloup I, Lefevre SD, Le Van Kim C, Schwarzbaum PJ, Ostuni MA. Human erythrocytes release ATP by a novel pathway involving VDAC oligomerization independent of pannexin-1. Scientific Reports 2018; 8(1):11384.

[62] Zanette DL, Santiago RP, Leite IPR, Santana SS, Guarda C, Maffili VV, Ferreira JRD, Adanho CSA, Yahouedehou SCMA, Menezes IL, Goncalves MS. Differential gene expression analysis of sickle cell anemia in steady and crisis state. Ann Hum Genet 2019; 83(5):310-317.

ANNEX

A. Solution Composition

A1.1 Incomplete Transfer Buffer (10x)	A1.2 Complete Transfer Buffer
30,2 g Tris 0,025 M; 144 g Glicine 0,192 M; Complete with type II water to a final volume of 1L.	100 mL Incomplete Transfer Buffer; 200 mL Methanol; Complete with type II water to a final volume of 1L.

A1.3 Ponceau S 0,1% in 5% Acetic Acid	A1.4 PBS 1× (Phosphate-buffered saline)
50 mL Ponceau S 0,1%; 450 mL Acetic Acid 5%.	100 mL mineral medium 10×; 900 mL type II water.

A1.5 PBS-T (Phosphate Buffered Saline with 0,1% Tween 20)	A1.6 Mild Stripping Buffer
100 mL PBS 1×; 900 mL type II water; 1 mL 0,1% Tween 20.	15 g Glycine; 1 g SDS; 10 mL Tween 20; Complete with type II water to a final volume of 1L; Adjust pH to 2.2 using hydrochloric acid.

B. Suplement tables

B.1 Post-translational modification (PTM)

Name	Delta Mass	Residues	Is Variable
Carbamidomethylation of Cysteine	57.02146	C	
Oxidation of Methionine	15.9949	M	√
Phosphorilation	79.9663	YST	√
O-Sulfonation	79.956815	YSTC	√
S-Nitrosylation	28.990164	C	√
Ubiquitination	383.228103	KSTC	√
Acetylation	42.010565	K	√
Formylation	27.994915	MK	√
N-Glycosylation	203.07882	N	√
O-Glycosylation	203.07882	ST	√

C. Suplement images

C.1 Example of a Ponceau membrane from Hsp70 and Hsp90 used for normalization.

Poster: *Sickle-cell disease investigated by computational proteomics approaches*

E- Protein Identification

E1. PatternLab for Proteomics identified proteins.

Acession Number	Protein Name	Fraction
P31946	14-3-3 protein beta/alpha	Cytoplasmatic
P62258	14-3-3 protein epsilon	
P61981	14-3-3 protein gamma	
P27348	14-3-3 protein theta	
Q9BWD1	Acetyl-CoA acetyltransferase cytosolic	
P23526	Adenosylhomocysteinase	
Q01518	Adenylyl cyclase-associated protein 1	
P02763	Alpha-1-acid glycoprotein 1	
Q9NZD4	Alpha-Haemoglobin-stabilizing protein	
P16157	Ankyrin-1	
P17174	Aspartate aminotransferase cytoplasmatic	
P53396	ATP-citrate synthase	
P02730	Band 3 anion transport protein	
P31939	Bifunctional purine biosynthesis protein PURH	
P53004	Biliverdin reductase A	
P07738	Bisphosphoglycerate mutase	
P07384	Calpain-1 catalytic subunit	
P00918	Carbonic anhydrase 2	
P04040	Catalase	
O00299	Chloride intracellular channel protein 1	
Q9Y696	Chloride intracellular channel protein 4	
O14618	Copper chaperone for superoxide dismutase	
P46109	Crk-like protein	
Q86VP6	Cullin-associated NEDD8-dissociated protein 1	
P00167	Cytochrome b5	
Q96GG9	DCN1-like protein 1	
P13716	Delta-aminolevulinic acid dehydratase	
Q08495	Dematin	
Q16531	DNA damage-binding protein 1	
P55010	Eukaryotic translation initiation factor 5	
P52907	F-actin-capping protein subunit alpha-1	
P14324	Farnesyl pyrophosphate synthase	
Q9Y3I1	F-box only protein 7	
P07954	Fumarate hydratase mitochondrial	
O75223	Gamma-glutamylcyclotransferase	
Q13630	GDP-L-fucose synthase	
P11413	Glucose-6-phosphate 1-dehydrogenase	
P48507	Glutamate--cysteine ligase regulatory subunit	
O76003	Glutaredoxin-3	
P78417	Glutathione S-transferase omega-1	

Acession Number	Protein Name	Fraction
P04406	Glyceraldehyde-3-phosphate dehydrogenase	Cytoplasmatic
P36959	GMP reductase 1	
Q9H0R4	Haloacid dehalogenase-like hydrolase domain-containing protein 2	
A0A0G2JIW1	Heat shock 70 kDa protein 1B	
P07900	Heat shock protein HSP 90-alpha	
Q9NRV9	Heme-binding protein 1	
P69905	Haemoglobin subunit alpha	
P68871	Haemoglobin subunit beta	
P02042	Haemoglobin subunit delta	
P69892	Haemoglobin subunit gamma-2	
Q6B0K9	Haemoglobin subunit mu	
P09105	Haemoglobin subunit theta-1	
P52597	Heterogeneous nuclear ribonucleoprotein F	
Q86YZ3	Hornerin	
P00492	Hypoxanthine-guanine phosphoribosyltransferase	
P01834	Immunoglobulin kappa constant	
Q14974	Importin subunit beta-1	
P14735	Insulin-degrading enzyme	
P04264	Keratin type II cytoskeletal 1	
Q04760	Lactoylglutathione lyase	
Q9BS40	Latexin	
P30740	Leukocyte elastase inhibitor	
P09960	Leukotriene A-4 hydrolase	
P00338	L-lactate dehydrogenase A chain	
P07195	L-lactate dehydrogenase B chain	
Q13228	Methanethiol oxidase	
P26038	Moesin	
P22234	Multifunctional protein ADE2	
Q8NCW5	NAD(P)H-hydrate epimerase	
Q9UNZ2	NSFL1 cofactor p47	
P61970	Nuclear transport factor 2	
J3KQ32	Obg-like ATPase 1	
Q92882	Osteoclast-stimulating factor 1	
Q06830	Peroxiredoxin-1	
P32119	Peroxiredoxin-2	
P30041	Peroxiredoxin-6	
P18669	Phosphoglycerate mutase 1	
Q9H008	Phospholysine phosphohistidine inorganic pyrophosphate phosphatase	
Q9GZP4	PITH domain-containing protein 1	
P13796	Plastin-2	
P68402	Platelet-activating factor acetylhydrolase IB subunit beta	
Q15102	Platelet-activating factor acetylhydrolase IB subunit gamma	

Acession Number	Protein Name	Fraction
Q9UQ80	Proliferation-associated protein 2G4	
Q06323	Proteasome activator complex subunit 1	
A0A024RA52	Proteasome subunit alpha type	
P25786	Proteasome subunit alpha type-1	
P25788	Proteasome subunit alpha type-3	
P28066	Proteasome subunit alpha type-5	
O14818	Proteasome subunit alpha type-7	
P20618	Proteasome subunit beta type-1	
P49721	Proteasome subunit beta type-2	
P49720	Proteasome subunit beta type-3	
P28070	Proteasome subunit beta type-4	
P28074	Proteasome subunit beta type-5	
P28072	Proteasome subunit beta type-6	
Q99436	Proteasome subunit beta type-7	
Q5TDH0	Protein DDI1 homolog 2	
P26447	Protein S100-A4	
P06703	Protein S100-A6	
P05109	Protein S100-A8	
P06702	Protein S100-A9	
P21980	Protein-glutamine gamma-glutamyltransferase 2	
P00491	Purine nucleoside phosphorylase	
P30613	Pyruvate kinase PKLR	
P50395	Rab GDP dissociation inhibitor beta	Cytoplasmatic
P00352	Retinal dehydrogenase 1	
P13489	Ribonuclease inhibitor	
P49247	Ribose-5-phosphate isomerase	
O95747	Serine/threonine-protein kinase OSR1	
P30153	Serine/threonine-protein phosphatase 2A 65 kDa regulatory subunit A alpha isoform	
Q9BRF8	Serine/threonine-protein phosphatase CPPED1	
P02787	Serotransferrin	
O75368	SH3 domain-binding glutamic acid-rich-like protein	
O43765	Small glutamine-rich tetratricopeptide repeat-containing protein alpha	
P02549	Spectrin alpha chain erythrocytic 1	
P31948	Stress-induced-phosphoprotein 1	
P00441	Superoxide dismutase [Cu-Zn]	
Q9Y490	Talin-1	
P17987	T-complex protein 1 subunit alpha	
P78371	T-complex protein 1 subunit beta	
P49368	T-complex protein 1 subunit gamma	
P50990	T-complex protein 1 subunit theta	
P40227	T-complex protein 1 subunit zeta	
Q9NYB0	Telomeric repeat-binding factor 2-interacting protein 1	

Acession Number	Protein Name	Fraction
P10599	Thioredoxin	Cytoplasmatic
P30048	Thioredoxin-dependent peroxide reductase mitochondrial	
P37837	Transaldolase	
Q00577	Transcriptional activator protein Pur-alpha	
P55072	Transitional endoplasmic reticulum ATPase	
Q99598	Translin-associated protein X	
Q6ICL3	Transport and Golgi organization protein 2 homolog	
P60174	Triosephosphate isomerase	
P23381	Tryptophan--tRNA ligase cytoplasmatic	
P45974	Ubiquitin carboxyl-terminal hydrolase 5	
Q9BSL1	Ubiquitin-associated domain-containing protein 1	
Q8IXQ3	Uncharacterized protein C9orf40	
Q9NWV4	UPF0587 protein C1orf123	
P54725	UV excision repair protein RAD23 homolog A	
P54727	UV excision repair protein RAD23 homolog B	
P38606	V-type proton ATPase catalytic subunit A	
O75083	WD repeat-containing protein 1	
P12955	Xaa-Pro dipeptidase	
Acession Number	**Protein Name**	**Fraction**
Q00013	55 kDa erythrocyte membrane protein	Mambrane
P16157	Ankyrin-1	
P20160	Azurocidin	
P02730	Band 3 anion transport protein	
P50895	Basal cell adhesion molecule	
P13727	Bone marrow proteoglycan	
P00918	Carbonic anhydrase 2	
P08311	Cathepsin G	
P31930	Cytochrome b-c1 complex subunit 1 mitochondrial	
P81605	Dermcidin	
O94919	Endonuclease domain-containing 1 protein	
P12724	Eosinophil cationic protein	
P11678	Eosinophil peroxidase	
Q99808	Equilibrative nucleoside transporter 1	
P27105	Erythrocyte band 7 integral membrane protein	
Q96PL5	Erythroid membrane-associated protein	
P17931	Galectin-3	
P69905	Haemoglobin subunit alpha	
P68871	Haemoglobin subunit beta	
P02042	Haemoglobin subunit delta	
P62805	Histone H4	
P08514	Integrin alpha-Iib	
Q14773	Intercellular adhesion molecule 4	
P23276	Kell blood group glycoprotein	

Acession Number	Protein Name	Fraction
Q08722	Leukocyte surface antigen CD47	Mambrane
Q6P1A2	Lysophospholipid acyltransferase 5	
P11279	Lysosome-associated membrane glycoprotein 1	
O15439	Multidrug resistance-associated protein 4	
P05164	Myeloperoxidase	
P00387	NADH-cytochrome b5 reductase 3	
P08246	Neutrophil elastase	
P32119	Peroxiredoxin-2	
Q92508	Piezo-type mechanosensitive ion channel component 1	
P51148	Ras-related protein Rab-5C	
O95197	Reticulon-3	
P11166	Solute carrier family 2 facilitated glucose transporter member 1	
Q9NP59	Solute carrier family 40 member 1	
P02549	Spectrin alpha chain erythrocytic 1	
P11277	Spectrin beta chain erythrocytic	
P02786	Transferrin receptor protein 1	
Q9BVK6	Transmembrane emp24 domain-containing protein 9	
P21796	Voltage-dependent anion-selective channel protein 1	
P45880	Voltage-dependent anion-selective channel protein 2	
Q9Y277	Voltage-dependent anion-selective channel protein 3	

E2. MaxQuant identified proteins.

Acession Number	Protein Name	Fraction
P31946	14-3-3 protein beta/alpha	Cytoplasmatic
P62258	14-3-3 protein epsilon	
P61981	14-3-3 protein gamma	
P27348	14-3-3 protein theta	
P63104	14-3-3 protein zeta/delta	
F5H5V4	26S proteasome non-ATPase regulatory subunit 9	
P63261	Actin, cytoplasmatic 2	
P13798	Acylamino-acid-releasing enzyme	
P23526	Adenosylhomocysteinase	
P01009	Alpha-1-antitrypsin	
E7EPV7	Alpha-synuclein	
P16157	Ankyrin-1	
P17174	Aspartate aminotransferase, cytoplasmatic	
P02730	Band 3 anion transport protein	
P31939	Bifunctional purine biosynthesis protein PURH	
P07738	Bisphosphoglycerate mutase	
K7ES02	Bleomycin hydrolase	
P0DP25	Calmodulin-3	
P00915	Carbonic anhydrase 1	

Acession Number	Protein Name	Fraction
P04040	Catalase	
O00299	Chloride intracellular channel protein 1	
J3KNF4	Copper chaperone for superoxide dismutase	
C9JVE2	DCN1-like protein	
P30046	D-dopachrome decarboxylase	
P13716	Delta-aminolevulinic acid dehydratase	
Q9NY33	Dipeptidyl peptidase 3	
Q16531	DNA damage-binding protein 1	
I3L397	Eukaryotic translation initiation factor 5A	
P30043	Flavin reductase	
P09104	Gamma-enolase	
E9PKL9	GDP-L-fucose synthase	
P48506	Glutamate--cysteine ligase catalytic subunit	
P48507	Glutamate--cysteine ligase regulatory subunit	
A0A087WUQ6	Glutathione peroxidase	
P78417	Glutathione S-transferase omega-1	
P09211	Glutathione S-transferase P	
P48637	Glutathione synthetase	
P04406	Glyceraldehyde-3-phosphate dehydrogenase	
Q9HC38	Glyoxalase domain-containing protein 4	
P0DMV9	Heat shock 70 kDa protein 1B	
P11142	Heat shock cognate 71 kDa protein	Cytoplasmatic
P07900	Heat shock protein HSP 90-alpha	
P69905	Haemoglobin subunit alpha	
P68871	Haemoglobin subunit beta	
P02042	Haemoglobin subunit delta	
P69891	Haemoglobin subunit gamma-1	
P50502	Hsc70-interacting protein	
P00492	Hypoxanthine-guanine phosphoribosyltransferase	
Q04760	Lactoylglutathione lyase	
Q9BS40	Latexin	
P30740	Leukocyte elastase inhibitor	
P09960	Leukotriene A-4 hydrolase	
P00338	L-lactate dehydrogenase A chain	
P07195	L-lactate dehydrogenase B chain	
Q13228	Methanethiol oxidase	
P26038	Moesin	
Q6XQN6	Nicotinate phosphoribosyltransferase	
Q9UNZ2	NSFL1 cofactor p47	
Q32Q12	Nucleoside diphosphate kinase	
Q9NTK5	Obg-like ATPase 1	
Q06830	Peroxiredoxin-1	
P32119	Peroxiredoxin-2	

Acession Number	Protein Name	Fraction
P30041	Peroxiredoxin-6	
P18669	Phosphoglycerate mutase 1	
Q9GZP4	PITH domain-containing protein 1	
P13796	Plastin-2	
Q15102	Platelet-activating factor acetylhydrolase IB subunit gamma	
J3QS39	Polyubiquitin-B	
F5H345	Porphobilinogen deaminase	
Q06323	Proteasome activator complex subunit 1	
Q9UL46	Proteasome activator complex subunit 2	
A0A024RA52	Proteasome subunit alpha type	
H0YMZ1	Proteasome subunit alpha type	
P25786	Proteasome subunit alpha type-1	
P25788	Proteasome subunit alpha type-3	
P28066	Proteasome subunit alpha type-5	
P60900	Proteasome subunit alpha type-6	
O14818	Proteasome subunit alpha type-7	
P20618	Proteasome subunit beta type-1	
P49721	Proteasome subunit beta type-2	
P49720	Proteasome subunit beta type-3	
P28070	Proteasome subunit beta type-4	
P28074	Proteasome subunit beta type-5	
P28072	Proteasome subunit beta type-6	
Q99436	Proteasome subunit beta type-7	Cytoplasmatic
Q5TDH0	Protein DDI1 homolog 2	
P06703	Protein S100-A6	
P06702	Protein S100-A9	
P00491	Purine nucleoside phosphorylase	
P50395	Rab GDP dissociation inhibitor beta	
P43487	Ran-specific GTPase-activating protein	
P00352	Retinal dehydrogenase 1	
J3KTF8	Rho GDP-dissociation inhibitor 1	
P49247	Ribose-5-phosphate isomerase	
P02787	Serotransferrin	
P31948	Stress-induced-phosphoprotein 1	
P00441	Superoxide dismutase [Cu-Zn]	
P10599	Thioredoxin	
E9PIR7	Thioredoxin reductase 1, cytoplasmatic	
O43396	Thioredoxin-like protein 1	
P37837	Transaldolase	
P55072	Transitional endoplasmic reticulum ATPase	
E9PGT1	Translin	
P60174	Triosephosphate isomerase	
H7BYY1	Tropomyosin 1	
J3KN67	Tropomyosin alpha-3 chain	

Acession Number	Protein Name	Fraction
E5RIW3	Tubulin-specific chaperone A	Cytoplasmatic
P54578	Ubiquitin carboxyl-terminal hydrolase 14	
P15374	Ubiquitin carboxyl-terminal hydrolase isozyme L3	
P61088	Ubiquitin-conjugating enzyme E2 N	
Q13404	Ubiquitin-conjugating enzyme E2 variant 1	
P22314	Ubiquitin-like modifier-activating enzyme 1	
P06132	Uroporphyrinogen decarboxylase	
P54725	UV excision repair protein RAD23 homolog A	
Acession Number	**Protein Name**	**Fraction**
Q00013	55 kDa erythrocyte membrane protein	
P63261	Actin, cytoplasmatic 2	
P16157	Ankyrin-1	
P29972	Aquaporin-1	
P02730	Band 3 anion transport protein	
A0A087WXM8	Basal cell adhesion molecule	
P35613	Basigin	
P00915	Carbonic anhydrase 1	
P08311	Cathepsin G	
B1AMW1	CD58 antigen,	
P27105	Erythrocyte band 7 integral membrane protein	
P16452	Erythrocyte membrane protein band 4.2	
Q96PL5	Erythroid membrane-associated protein	
P17931	Galectin-3	
Q5VZR0	Golgi-associated plant pathogenesis-related protein 1	
P69905	Haemoglobin subunit alpha	
P68871	Haemoglobin subunit beta	Membrane
P02042	Haemoglobin subunit delta	
P69891	Haemoglobin subunit gamma-1	
P62805	Histone H4	
P23276	Kell blood group glycoprotein	
O15439	Multidrug resistance-associated protein 4	
P00387	NADH-cytochrome b5 reductase 3	
P32119	Peroxiredoxin-2	
J3QS39	Polyubiquitin-B	
P11171	Protein 4.1	
P11166	Solute carrier family 2, facilitated glucose transporter member 1	
Q9NP59	Solute carrier family 40 member 1	
P02549	Spectrin alpha chain, erythrocytic 1	
P11277	Spectrin beta chain, erythrocytic	
J3KN67	Tropomyosin alpha-3 chain	
Q13336	Urea transporter 1	