UNIVERSITÀ DEGLI STUDI DI MILANO

SCUOLA DI DOTTORATO DI RICERCA IN

SCIENZE BIOCHIMICHE, NUTRIZIONALI E METABOLICHE

DOTTORATO DI RICERCA IN BIOCHIMICA, BIO/10

CICLO XXV

TESI DI DOTTORATO DI RICERCA

svolta presso Fondazione Centro San Raffaele

Centro di Genomica Traslazionale e Bioinformatica

Biomolecular NMR Laboratory, Milano, Italia

A metabolomic, structural and functional study of Polycystin1 and Nephrocystin1

Mannella Valeria

Matr. N. R08827

Docente guida: Prof. Marcello Duranti

Correlatrice: Dr.ssa Giovanna Musco

Coordinatore del Dottorato: Prof. Francesco Bonomi

ANNO ACCADEMICO 2011/2012

Index

		Pag.
1.	**ABSTRACT**	**4**
2.	**INTRODUCTION**	**9**
2.1 Cystic kidney diseases belong to ciliopathies		10
2.2 Nephronopthisis		12
2.3 Nephrocystin1 (NPHP1)		13
2.4 Autosomal Dominant Polycystic Kidney Disease (ADPKD)		14
2.5 Polycystin1 (PC1)		16
2.5.1	PC1 as "metabolic regulator"	18
2.5.2	PC1 interacts with nephrocystin1	21
3.	**AIM OF THE THESIS**	**24**
4.	**MATERIALS AND METHODS**	**27**
4.1 Metabolomic experiments		28
4.1.1	Sample preparation: Exometabolome, biofluids and tissues	28
4.1.2	NMR spectra: acquisition, metabolites identification and quantification	29
4.1.3	Statistical analysis: Principal Component Analysis (PCA) and Metabolomics Univariate and Multivariate Analysis (MUMA)	31
4.1.4	^{13}C glucose flux experiments *in vivo*: experimental procedure and NMR spectra, quantification and normalization of ^{13}C glucose and ^{13}C lactate peaks and data analysis	33
4.1.5	Cells' treatment: Glucose starvation and Oligomycin treatment	36
4.1.6	Proliferation assay	36
4.1.7	ATP content quantification	37
4.1.8	Western Blot Analysis	38
4.1.9	Histology and immunohistochemistry	38
4.2 Structure determination procedure		39
4.2.1	Generation of *Escherichia Coli* expression constructs	39
4.2.2	Expression of recombinant proteins in *E. Coli*	41
4.2.3	Purification of recombinant proteins in *E. Coli*	41
4.2.4	Circular dichroism (CD) experiments	43
4.2.5	Protein resonances assignment for structure determination	44
4.2.6	Chemical -shift Rosetta (CS- Rosetta)	45
4.2.7	Dihedral angles: 3J(HNHα) calculation and TALOS+ prediction	46
4.2.8	Determination of NOE restraints	48
4.2.9	Structure calculation	49
4.2.10	Dynamic information: ^{15}N spin relaxation measurements	50
4.2.11	Residual dipolar couplings (RDC)	53
4.2.12	NMR titrations	54
5.	**METABOLOMICS**	**56**
5.1 Metabolomic profiling of the exometabolome of WT and *Pkd1* KO cells		57
5.2 *Pkd1* KO cells display the Warburg effect		61
5.3 Additional experiments in support of metabolic dysfunction in *Pkd1* KO cells		64
5.3.1	Demonstration of the Warburg effect	64
5.3.2	Proliferation and apoptosis in *Pkd1* KO cells depend on glucose metabolism	66
5.4 Treatment with drugs		66

5.4.1 Exometabolome of MEF cells treated with drugs. 66
5.4.2 Additional experiments to test the effects of the drugs on the signaling cascade and the related biological effects. 69
5.5 *In vivo* experiments 70
5.5.1 1D ^{1}H NMR experiments of polar extracts deriving from lyophilized kidneys 71
5.5.2 ^{13} C glucose experiments: kinetic experiments and experiments after 40 minutes from glucose injection 73
5.6 Treatment with 2- deoxy- glucose (2DG) 78

6. STRUCTURE OF THE N-TERMINUS DOMAIN OF NPHP1 81

6.1 The recombinant N-terminus of NPHP1 is soluble 82
6.2 The N-terminus of NPHP1 is predicted to be a coiled coil domain 83
6.3 The N-terminus of NPHP1 is thermostable and contains α-helices 84
6.4 NMR experiments 85
6.4.1 $NPHP1_{1-115}$ chemical shifts assignment 86
6.4.2 3D structural model obtained with Chemical -Shift Rosetta (CS-Rosetta) 87
6.4.3 Hydrogen–deuterium exchange 88
6.4.4 Experimentally calculated phi angle and Talos + prediction 90
6.4.5 ^{1}H-^{1}H NOE pattern confirms the presence of α helices 92
6.4.6 Structure calculation of $NPHP1_{1-115}$ 93
6.4.7 $NPHP1_{1-115}$ displays three structured regions 98
6.5 The interaction with the coiled coil domain of PC1 (CC_PC1) 102
6.5.1 The purification of the recombinantCC_PC1 102
6.5.2 Titration of 6His-MBP-CC_PC1 into $NPHP1_{1-115}$ 104

7. DISCUSSION 106

7.1 Metabolomics 107
7.1.1 Metabolomic profiling of the exometabolome of *Pkd1* KO and WT MEF cells 107
7.1.2 -The Warburg effect occurs both in cellular model and *in vivo* 108
7.1.3 Signaling cascade 109
7.1.4 Treatments with drugs: Cells' treatment with rapamycin, metformin and AICAR; Treatment with 2-deoxy-glucose (2DG) 110
7.1.5 Conclusions and future perspectives 111
7.2 Structure of $NPHP1_{1-115}$ 112
7.2.1 $NPHP1_{1-115}$ structure 112
7.2.2 Conclusions and future perspectives 114

8. APPENDIX 115

8.1 NMR appendix 116
8.1.1 Chemical shift 116
8.1.2 Magnetization transfer 118
8.1.3 Nuclear Overhauser Enhancement (NOE) 120
8.1.4 Monodimensional (1D) NMR experiments 121
8.1.5 Bidimensional (2D) NMR experiments: ^{1}H-^{15}N HSQC, ^{1}H-^{1}H-TOCSY, ^{1}H-^{13}C HSQC and ^{1}H- ^{1}H NOESY 122
8.1.6 Three-dimensional (3D) NMR experiments: HCCH-TOCSY, HNCA, HNCACB, HN(CO)CACB, HNCO, HNHA, CC(CO)NH, HCC(CO)NH, ^{1}H-^{15}N-HSQC-NOESY, ^{1}H-^{13}C-HSQC-NOESY 126
8.1.7 ARIA (Ambiguous Restraints for Interactive Assignment) software 131

8.2 Metabolomics and metabonomics appendix **133**

8.2.1 Metabolomics and metabonomics 134
8.2.2 Mass spectrometry (MS) 136
8.2.3 Nuclear magnetic resonance (NMR) 137
8.2.4 NMR spectra: 1D NMR spectra and 2D NMR spectra 137
8.2.5 Data preprocessing: Normalization and scaling factor 141
8.2.6 Data analysis: Unsupervised analysis, supervised analysis 143
8.2.7 Identification of metabolites 144

REFERENCES **146**

1. Abstract

Nephronophthisis (NPHP) and Autosomal Dominant Polycystic Kidney Disease (ADPKD) are two genetic renal cystic diseases that lead to End Stage Renal Disease (ESRD) in childhood or adolescence and at the late-middle age, respectively[1]. At present the mechanisms at the basis of cystogenesis in both diseases are poorly understood and pharmacological therapies are still lacking.

On the one hand NPHP is characterized by kidney tubular atrophy and cysts formation occurring primarily at the cortico-medullary border, that leads to a reduction of kidneys volume. It is caused by mutations in 11 different genes. The most common form of NPHP is the juvenile one (NPHP type 1), which is caused, in most of the cases, by homozygous deletion of NPHP1 gene (2q13 chromosome). *NPHP1* gene encodes for nephrocystin1 (NPHP1), a 732 amino acids long cytoplasmatic protein. It presents a widespread expression, but the pathogenic consequences due to loss-of-function are mainly confined to the kidney. In particular, NPHP1 is localized to cell-cell junctions, cilia and cell-matrix adhesion sites and it is involved in signaling transduction, cell-cell adhesion and cell polarity[2-4].

On the other hand the hallmark of ADPKD is the formation of bilateral cysts that leads to an increase of kidneys volume weight[5,6]. The disease is caused by mutations in two genes: *Pkd1* gene (16p13.3 chromosome) which is mutated in 85% of cases and *Pkd2* (4q21 chromosome) which is mutated in the remaining part of cases. These genes encode respectively for polycystin1 (PC1), a large plasma membrane receptor 4302 amino acids long and polycystin2 (PC2), a 968 amino acids long calcium channel. The two proteins interact through their C-terminal coiled coil domains and are involved in common signaling pathways, thus explaining the same phenotype[5,7,8].

In this thesis we focused our attention on two new possible roles of PC1: the regulation of metabolism and its interaction with NPHP1. To disentangle PC1 role as “metabolic regulator” and its possible interaction with NPHP1 through its coiled coil domain, we have selected nuclear magnetic resonance (NMR) spectroscopy as primary investigation technique.

PC1 as metabolic regulator. During a routine culture of mouse embryonic fibroblasts (MEFs) our collaborators (Dr. Boletta’s group in San Raffaele Hospital, Milan, Italy) have observed that the knock out (*Pkd1* KO) MEFs for PC1 acidify the medium faster than the wild type (WT) ones. This finding rises the hypothesis that PC1 is involved in the regulation of cellular metabolism. Hence, exploiting nuclear magnetic resonance (NMR) spectroscopy, we have applied a metabolomic approach in order to investigate the metabolic pathways and the

signaling cascade deregulated by the loss of PC1. Importantly, the metabolomic analysis was performed both in cells and in mouse kidneys (WT and *Pkd1* KO) in order to compare the results both *in vitro* and *in vivo*. The analysis of the 1D ^{1}H NMR spectra of the media conditioned by the *Pkd1* KO and by the WT MEF cells (exometabolome) has highlighted some metabolic differences: the exometabolome of the *Pkd1* KO MEF cells displays an increased content of lactate, glutamate and alanine, presenting features similar to tumor cell metabolism[9-11]whereas the exometabolome of the WT MEF cells has an increased content of choline, formate, glucose, glycine and pyruvate. The two metabolites that vary mostly between the two cell lines are glucose and lactate, as highlighted by PCA analysis. In particular *Pkd1* KO MEF cells display a major consumption of glucose, which is in turn related to a major production of lactate, showing the aerobic glycolysis, also called Warburg effect, a metabolic hallmark usually observed in cancer cells.

In order to prove the occurrence of the Warburg effect *in vivo,* Ksp-Cre: $Pkd1^{flox/-}$ and $Pkd1^{flox/+}$ mice[12] were treated with uniformly labeled ^{13}C glucose[13,14]. ^{1}H-^{13}C HSQC NMR spectra on the polar extracts of kidneys and livers (used as control organ) were acquired and analyzed in order to simplify the detection of the metabolic conversion of glucose into lactate. The *Pkd1* KO kidneys displayed an increased content of glucose and lactate as compared to the WT ones, whereas livers did not display any difference. These results confirm that the aerobic glycolysis occurs also *in vivo* and is caused by the loss of functional PC1.

In order to restore the metabolic basal condition, mice were treated with 2-deoxy-glucose (2DG), a drug able to block directly the glucose metabolism[15,16]. For testing the efficacy of the treatment, Ksp-Cre: $Pkd1^{flox/-}$ and $Pkd1^{flox/+}$ mice were treated at the same time with ^{13}C glucose and 2DG and the polar extracts of kidneys and livers were analyzed through ^{1}H-^{13}C HSQC NMR spectra. The *Pkd1* KO kidneys treated with 2DG display a decrease of the lactate content, whereas the amount of glucose does not change, hence the treatment with 2DG affects the glucose metabolism but not its uptake. Moreover, from the histological point of view, the treated *Pkd1* KO kidneys show a reduction of the volume of cysts and consequently a reduction of the weight of the kidney, compared to the not treated *Pkd1* KO kidneys. *Pkd1* KO livers do not present any changes after treatment with 2DG. These data show that 2DG acts specifically on *Pkd1* KO kidneys, improving both the altered metabolism and the histology of the *Pkd1* KO kidneys, without involving other organs that, instead, express functional PC1 (eg. the liver). In conclusion, our findings indicate that PC1 acts as "metabolic regulator", as shown by the analysis of the medium conditioned *Pkd1* KO cells. In

particular *Pkd1* KO cells and *Pkd1* KO kidneys are affected by the Warburg effect, a metabolic hallmark of accelerated metabolism observed also in cancer cells. Moreover, the analysis of *Pkd1* KO kidneys of Ksp-Cre: $Pkd1^{flox/-}$ and $Pkd1^{flox/+}$ mice treated with 2DG has highlighted an improvement of the Warburg effect and the renal histology. These preliminary results are very promising for a possible therapy for ADPKD patients and will be object of further *in vivo* studies.

The coiled coil domain of PC1 does not interact with the N-terminus of NPHP1. We have recently demonstrated that one of the poly-proline motives, belonging to cytoplasmic C-terminus of PC1, is a binding partner for SH3 domain of NPHP1 and this complex is involved in the regulation of apoptosis[17]. The weak affinity of the resulting complex, as indicated by the estimated dissociation constant (K_d=0.3 +/-0.02 mM), suggests that other proteins or other domains of the same proteins could contribute to the interaction. In particular, the presence of coiled coil domains at PC1 C-terminus and at NPHP1 N-terminus, suggests the formation of a more complex macromolecular system involving interactions of both coiled coils. In order to verify and to structurally characterize at atomic level the interactions between PC1 and NPHP1 we have applied NMR spectroscopy.

The first 115 amino acids (13kDa) of NPHP1 were predicted to form a coiled domain by the Web Server PCOIL. The coiled coil domain is a α helical motif involved in homo and heteromultimerization[18]. Unexpectedly, recombinant $NPHP1_{1\text{-}115}$ behaves as a monomer, as assessed by the elution volume in size exclusion chromatography and by the overall correlation time deduced from NMR relaxation experiments. Circular dichroism experiments confirmed the presence of ~45% of α helix and indicated a high thermostability of the domain, with a melting temperature of 65°C.

The solution structure of the domain shows that the first 115 amino acids of NPHP1 present a three helix bundle fold stabilized by hydrophobic interactions. The supposed interaction between the N-terminus of NPHP1 and the coiled coil domain of PC1 (CC_PC1) was next verified performing an NMR titration with the purified recombinant proteins, recording ^{1}H-^{15}N HSQC NMR spectra of NPHP1 upon addition of unlabelled CC-PC1. CC_PC1 was produced with the tags in order to improve the solubility and to avoid aggregation[19]. The recorded ^{1}H -^{15}N HSQC spectrum does not show any chemical shift displacement, indicating, that the two proteins do not interact.

In conclusion our data show that the N-terminus of NPHP1 is a monomeric and thermostable domain characterized by the presence of three α helices that pack together to form a three helix bundle. The N-terminus of NPHP1 is not a coiled coil domain, as erroneously predicted, and does not interact with CC_PC1. Interactions between $NPHP_{1\text{-}115}$ and PC1 should be modulated by other protein domains or might involve additional proteins. Further experiments are needed to clarify this point.

2. Introduction

2.1. Cystic kidney diseases belong to ciliopathies

Cystic kidney diseases (CKD) are a group of monogenic kidney diseases that are among the most frequent lethal genetic diseases in humans[1]. CKD are divided into autosomal dominant and autosomal recessive forms. The first one is characterized by End-Stage Renal Disease (ESRD) in adulthood, whereby the most prominent example is represented by Autosomal Dominant Polycystic Kidney Disease (ADPKD); the second one, comprising nephronopthisis (NPHP) and the Autosomal Recessive Polycystic Kidney Disease (ARPKD), leads to ESRD in childhood or adolescence[1].

CKD are classified as ciliopathies, a novel class of human diseases caused by dysfunction of two physically and functionally associated organelles: the centrosome and the cilia[20]. Indeed the products of all the genes (the so called cystoproteins) mutated in CKD both in humans, in mice and in zebrafish, are expressed in primary cilia (a sensory organelles), in basal bodies or in centrosomes. The ciliary expression of cystoproteins in the ciliated cells of many organs, such as the retina, explains the involvement of other organs in patients with cystic kidney disease.

The cilium is composed of a microtubule-based core structure, the *axoneme*, composed by nine parallel doublet microtubules known as outer doublets, which elongate from the basal body, which is on turn surrounded by a ciliary membrane that is the continuation of the plasma membrane. There are two types of cilia: the primary cilium and the motile cilia[21]. The primary cilium is not motile and it is found on the apical surface of the majority of cells in the human body. It connects the physical and biochemical extracellular signals with the mechanism of cell-cycle and epithelial cell polarity[22]. Motile cilia are highly present on the epithelial cell surfaces of the trachea and oviduct, and cooperatively beat in wave-like patterns to generate fluid movement. To drive their bending motion, motile cilia have two extra microtubules in the centre of *axoneme,* termed central pair, as well as radial spokes and dynein arms attached to microtubule doublets as shown in Figure 1. The primary cilia are located in the apical surface of epithelial cells of kidneys. Their role is to perceive the flow of urine to modulate the duct morphogenesis [Figure 2].

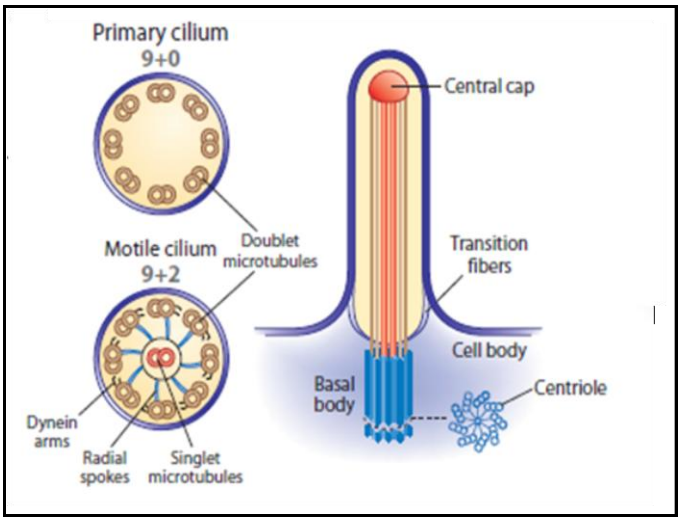


Figure 1. Architecture of cilia (primary and motile) and the basal body. [Figure from Yoder J., 2007, J Am. Soc. Nephrol.].

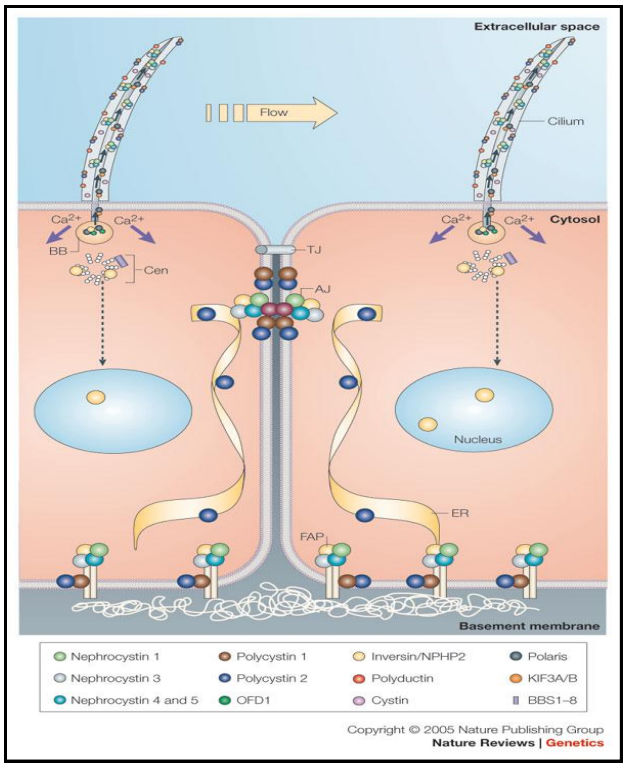


Figure 2. The primary cilium is in the apical part of epithelial cells of kidneys. The cilium responds to flow of urine and proteins associated, such as Polycystin1, Polycystin2 and Nephrocystin1, activate signaling cascade in response. [Figure from Hildebrandt F. *et al.* , 2011, N. Engl. J. Med.].

2.2. Nephronophthisis

Nephronopthisis (NPHP) was described for the first time in 1945 by Smith and Graham who called this disease "medullary cystic kidney disease". In 1951 Fanconi et al. introduced the modern definition of this disease, "familial juvenile nephronophthisis" [3]. The term "nephronophthisis" derives from the Greek and means "disintegration of nephron" describing one aspect of the histopathology of this disease. Indeed, the kidneys affected by NPHP show tubular atrophy with thickened tubular basement membrane, diffuse and severe interstitial fibrosis, development of cysts at the cortico-medullary junction. Cysts derive from collecting ducts and distal tubules[23], leading to a possible reduction of the size of the kidneys [Figure 3].
Nephronopthisis is an autosomal recessive nephropathy that affects about 1/100000 live births and it is the most common cause of the ESRD in children and adolescents. Based on the age of onset of ESRD, three distinct forms of the disease are known: infantile, juvenile and adolescent; the most common form is the juvenile one in which ESRD occurs at a mean age of 13 years. In the adolescent form the ESRD onsets at a mean age of 19 years, in the infantile one the ESRD onsets at a mean age of 1 year[3].
Patients affected by NPHP display the following symptoms: polyuria, polydipsia, secondary enuresis, growth retardation and anemia. The 10-15% of juvenile NPHP patients present also extra-renal anomalies such as retinal degeneration, cerebellar vermis aplasia, liver fibrosis and cone shaped epiphyses. NPHP is caused by mutations in 11 different genes called *NPHP 1-11*, which encode for nephrocystins, a family of cytoplasmatic proteins that are conserved among the evolution[3].

Figure 3. A section of human kidneys. On the left a normal kidney and on the right a kidney affected by NPHP. The volume of kidney affected by NPHP is reduced compared to the healthy one [Figure from Solliday M. *et al*., 2004, Atlas of Clinical Urology].

2.3. Nephrocystin1 (NPHP1)

The most common form of juvenile NPHP is caused by mutation in *NPHP1* (2q13) gene that encodes for nephrocystin1 (NPHP1)[22]. Human NPHP1 is a cytoplasmatic protein 732 amino acids long, whose function remains still poorly understood. It is supposed to have a possible role in the regulation of cell-cell adhesion, in agreement with the localization in the primary cilium and at the cell-cell junctions[4].

NPHP1 contains a putative N-terminal coiled coil domain, a Src homology 3 (SH3) domain and a nephrocystin homology domain (NHD). These domains are separated by short, poorly conserved stretches rich in glutamic acid residues [Figure 4].

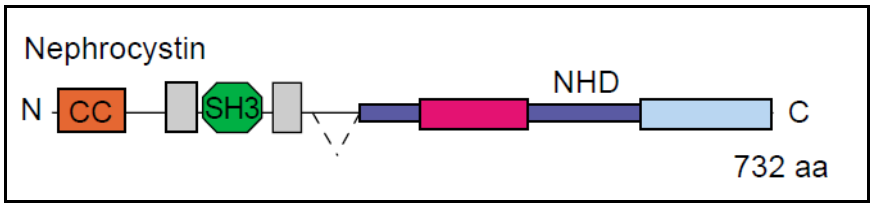


Figure 4. A schematic representation of NPHP1, a 732 aa long cytoplasmatic protein. It contains a coiled coil domain, an SH3 domain flanked by two glutamic acid rich regions (grey box) and a highly conserved C-terminal NHD domain [Figure from Saunier S. *et al.*, 2005, Curr. Opin. Genet. Dev.].

The first 115 amino acids are predicted to form a short coiled coil domain[24,25], a structure usually involved in homo/hetero association[18]. The SH3 domain spans from 157 to 207 residue and it is an important module for binding protein: in particular it recognizes proline-rich motifs[26]. The NHD domain occupies the two-third of the protein at C terminus and appears to represent a unique structural and functional protein domain for all nephrocystin proteins. This motif is involved in mediating both self-association and interaction with actin-binding protein of the filamin family[4].

Comparison of human and mouse sequences indicates that the N-terminal region of a putative coiled coil domain and the nearby SH3 domain are well conserved, showing 80% amino acid identity. The long C-terminus is better conserved displaying the 87% of amino acids identity, probably this domain also has an important function conserved throughout evolution[27].

The combination of SH3 domain with a coiled coil structure suggests that NPHP1 functions as an adapter protein belonging to a larger protein complex[4]. Independent studies have shown

that NPHP1 interacts with other nephrocystins, in particular nephrocystin4 and inversin (nephrocystin2), forming a nephrocystin complex. This complex simultaneously contacts both signaling proteins involved in the regulation of cell adhesion, such as the adaptor protein p130Cas, the protein kinase Pyk2, the activated ACDC42 kinase 1 (ACK), and proteins associated directly with the microtubular and actin cytoskeleton, such as tensin and filamin[4,28,29].

Notably, NPHP1 interacts with and is regulated by NPHP4[29]. The complex formed by NPHP1/NPHP4 interacts with the tight junction and polarity proteins including protein associated Lin 7 (PALS1), PALS1 associated tight junction protein (PATJ), and Par6 to form epithelial tight junction and to regulate epithelial morphogenesis[28]. It forms a bigger complex including further signaling and scaffolding proteins, such as Pyk2 and p130Cas. On the other hand, NPHP4 regulates the phosphorylation of tyrosines of NPHP1 by Pyk2. The interaction between NPHP1 and Pyk2 is essential to translocate Pyk2 from cytoplasm to cells matrix adhesion in response to extracellular signals[27].

2.4. Autosomal Dominant Polycystic Kidney Disease (ADPKD)

Autosomal Dominant Polycystic Kidney Disease (ADPKD) is a slowly progressive genetic disease and affects about 1-2/1000 live births[30]. It is characterized by the progressive formation and enlargement of renal fluid-filled cysts [Figure 5] leading to ERSD by late middle age. Thousands of large, spherical cysts of various sizes are formed in the cortex and medulla zone of the kidneys but also in other organs such as liver, intestine and pancreas. Cysts formation is due to an abnormal proliferation of the epithelial cells leading to the alteration of the architecture of kidney failing blood filtration[6].

The first symptoms of the disease are micro- and macro-hematuria, back pain and hypertension[6]. ADPKD arises as consequence of mutations in two genes: *PKD1* and *PKD2*, respectively encoding for polycystin1 (PC1) and polycystin2 (PC2)[30].

The gene *PKD1* (locus 16p13.3) is mutated in 85-90% of the patients and leads to ESRD around 54 years; mutations in the gene *PKD2* (locus 4q21-23) have been observed in the rest of the cases and lead to ESRD around the age of 74 years[31].

Despite the fact that the disease is inherited in autosomal dominant manner, the analysis of the epithelial cells from renal cysts indicates that cystogenesis requires mutations of both alleles, leading to homologous loss of function of either *PKD1* or *PKD2*[32]. This is known as the "two hit" model of cystogenesis [Figure 6]. The somatic "second hit" is believed to involve a single cell, which clonally proliferates, generating a cyst[32]. Hence each cyst arises as a consequence of a distinct somatic mutation event, explaining the disease's slow progression over the course of decades[33].

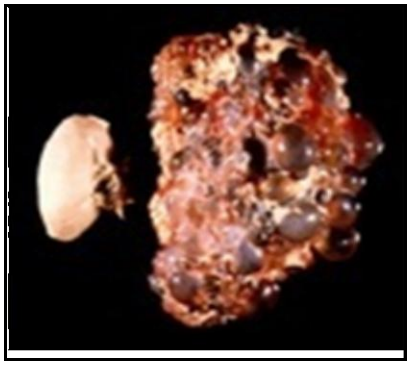


Figure 5. Comparison of normal and ADPKD kidney. On the left a normal kidney, on the right a kidney affected by ADPKD. The fluid filled cysts are evident and lead to increase in volume and weight of the organ [from PKD Fondation].

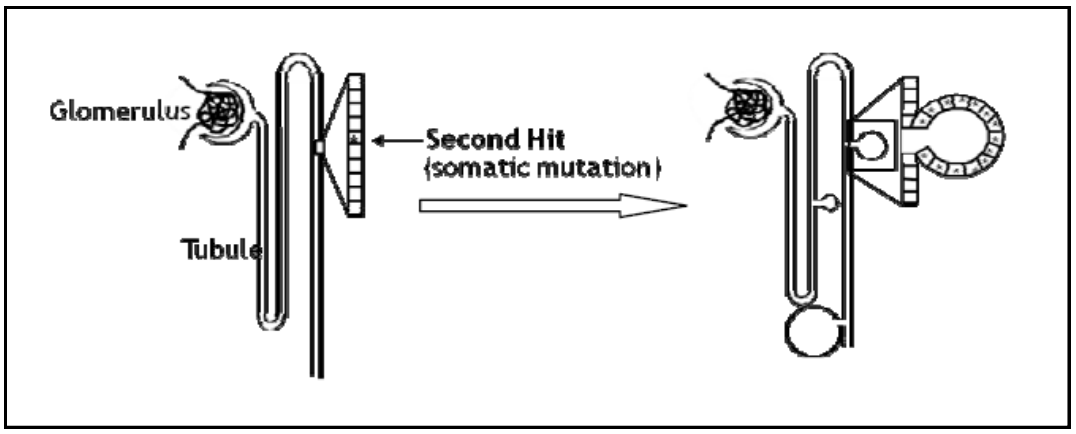


Figure 6. Comparison of healthy and cystic nephron. On the left and on the right a healthy and a cystic nephron, respectively. The somatic mutation in an heterozygous epithelial cell leads to cysts formation, resulting therefore caused by a local clonal event relating to a single cell [Figure from Qian F. *et al*.,1996, Cell.].

This model, in which cysts originate from cells that have undergone the second mutation, can also explain why only some points present cysts without involving the whole nephron.
Trans heterozygosity could also lead to disease, indeed patients with mutations in *PKD2* presents mutations also in *PKD1* gene[34,35]. The *PKD1* gene is larger than *PKD2* and, for this reason, it shows a major frequency of mutations. These two genes encode for two transmembrane proteins, in particular *PKD1* encodes for a large membrane receptor, polycystin1 (4302 amino acids) and *PKD2* encodes for polycystin2 (968 amino acids), a non selective cation channel with preference for calcium. These two proteins colocalize in the primary cilium, interact through the respective C-terminal coiled coil domains regulating the influx of calcium and the consequent signal cascade [8,36]. In the following a brief overview on the structure and function of PC1 will be presented.

2.5. Polycystin1 (PC1)

Polycystin1 (PC1) is a 4302 amino acids long membrane receptor, with a long N-terminal portion of ~ 3000 amino acids, 11 transmembrane domains and a short intracellular C-tail of ~ 198 amino acids. It is believed that PC1 plays a role in cell-cell/matrix interactions, where the long extracellular domain can homodimerize and serve as ligand. In addition PC1 functions as a mechanosensor on the primary cilium and at cell-cell junctions[5]. The N-terminus contains a novel combination of protein-protein interacting domains, including leucine-rich repeats (LRRs), a C-type lectin domain, 16 PKD repeats (IgG-like domain) and a REJ domain, whose name derives from its homology with the sea urchin Receptor for Egg Jelly[37]. PC1 molecular mass is ~ 462 kDa, however the high glycosylation level leads to a final mass of 520 kDa[38].
The protein contains a G-protein coupled receptor proteolytic site (GPS), located between the N-terminal and the first transmembrane domain, an auto-proteolytic site that leads to the cleavage of PC1 in two fragments: an N-terminal fragment (NTF, ~ 400 kDa), corresponding to the extracellular portion of protein, and a C-terminal fragment (CTF, ~ 150 kDa), composed by the rest of the protein. The cleaved NTF can be either released or remain bound to CTF. Not all PC1 proteins present in the cells undergo to the cleavage at the GPS and the full length uncleaved protein co-exists with the cleaved version in cells [39]. *In vitro* and *in vivo* studies have demonstrated that the cleavage at GPS site is essential for the correct functioning of

PC1[5]. Mice with mutations in the consensus site for cleavage HLT (cutting position is between L and T) that prevent it, survive to birth but they develop ADPKD at the first weeks of life that lead to death. The cleavage of PC1 is essential to prevent cystogenesis in the duct of distal and collector tubule [Figure 7].

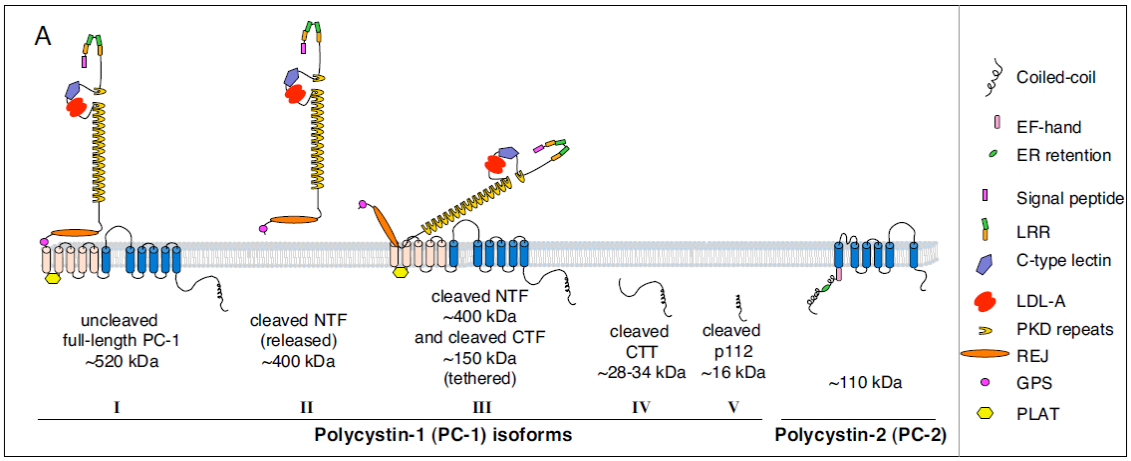


Figure 7. A schematic representation of polycystin1 (PC1). This large transmembrane receptor undergoes a series of cleavage events that generate several different species co-existing within the same cell and carrying out distinct functions. The protein exists as an uncleaved polypeptide of 4302 amino acids (aa) (I) and can be cleaved at its G-protein coupled proteolytic site, generating an N-terminal fragment (NTF) that can be released (II) or remain tethered to the C-terminal fragment (CTF) (III). Also the C-tail (CTT) is cleaved, but the cleavage sites are still unknown, releasing the CTT fragment (IV) or the last 112 aa (V) [Figure from Boletta A., 2009, Pathogenics].

The intracellular C-terminal tail is involved in signaling transduction pathways, such as Wnt pathway, calcium signaling, G protein signaling and activation of STAT1 and STAT6. Also the C terminus is cleaved in two fragments: one that contains all C terminus of ~ 28-34 kDa and another one of ~ 16 kDa, the sites of cleavage are unknown[5]. Both the fragments interact with transcription factors, including β-catenin and STAT6, respectively, that are transported in the nucleus[40]. The C-tail contains putative phosphorylation sites that undergo phosphorylation, as demonstrated by *in vitro* assays, but the function is still unknown[5].

Moreover, the C-terminus of PC1 contains a coiled coil domain that mediates the interaction with a coiled coil domain located at the C terminus of PC2[8,36], a non-selective cation channel with preference for calcium, regulating its activity. The PC1/PC2 complex regulates a number

of different biological processes including cell proliferation, apoptosis, cell migration and tubulogenesis [Figure 8].

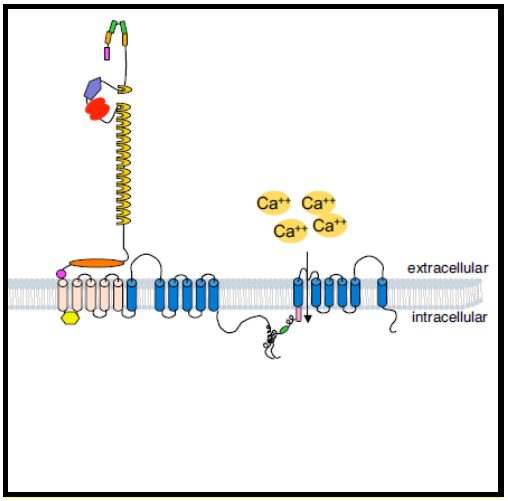


Figure 8. A schematic representation of PC1 and PC2. These transmembrane proteins interact through the coiled coil domain located in their cytoplasmatic C-terminal tail. This interaction regulates the influx of calcium from extracellular compartment. [Figure from Boletta A., 2009, Pathogenics].

2.5.1 PC1 as "metabolic regulator"

PC1 is supposed to be involved in the regulation of the cellular metabolism. This hypothesis starts from a serendipitous observation of our collaborators (Dr. Boletta's group in San Raffaele Hospital, Milan). During a routine culture of mouse embryonic fibroblasts (MEFs) they have observed that the knock out (KO) MEFs for PC1 acidify faster the medium than the wild type (WT) ones, as shown in Figure 9.

Moreover the KO cells' phenotype is quite different than WT ones, indeed they present an increase of the cellular size and proliferation rate. These phenotypic changes and the fast acidification of the medium could be related to an enhanced activity of mammalian target of rapamycin (mTOR) that is regulates negatively by PC1, as demonstrated by recent study[41].

Figure 9. The culture medium of *Pkd1*$^{-/-}$ KO MEF cell (on the right) is more acid than the one from WT (on the left) [provided by Boletta's lab].

PC1 controls the ERK-dependent phosphorylation of tuberin (Tsc2), upstream mTOR. The result is the downregulation of mTOR and of its downstream targets [41]. Whereas the loss of PC1 leads to an up regulation of the ERKs pathway and consequentially the enhanced of mTOR activity and the relative downstream targets [Figure 10].

mTOR is a serine/threonine kinase that forms two complexes complex 1 (mTORC1) and 2 (mTORC2). These complexes share some proteins but others are specific for each complex[42]. These complexes have different sensibility to inhibition due to rapamycin, a macrolide with an immunosuppressant and antiproliferative activity. In particular mTORC1 is very sensitive to rapamycin whereas mTORC2 is inhibited only by a great amount of drug or after a long-term treatment[42].

mTORC1 and mTORC2 regulate different cellular activity. mTORC1 promotes cell growth (increase of cellular size) and cell proliferation (increase of cell number) acting as a "metabolic sensor" ensuring the optimal conditions for both cell growth and proliferation. Its activity is regulated by the availability of amino acids, growth factors, energy stores and oxygen[43]. mTORC2 modulates cell survival, cell polarity and cytoskeletal organization[43]. The upstream events that control mTORC2 are still poorly understood although growth factors appear to stimulate also mTORC2[43] [Figure 11].

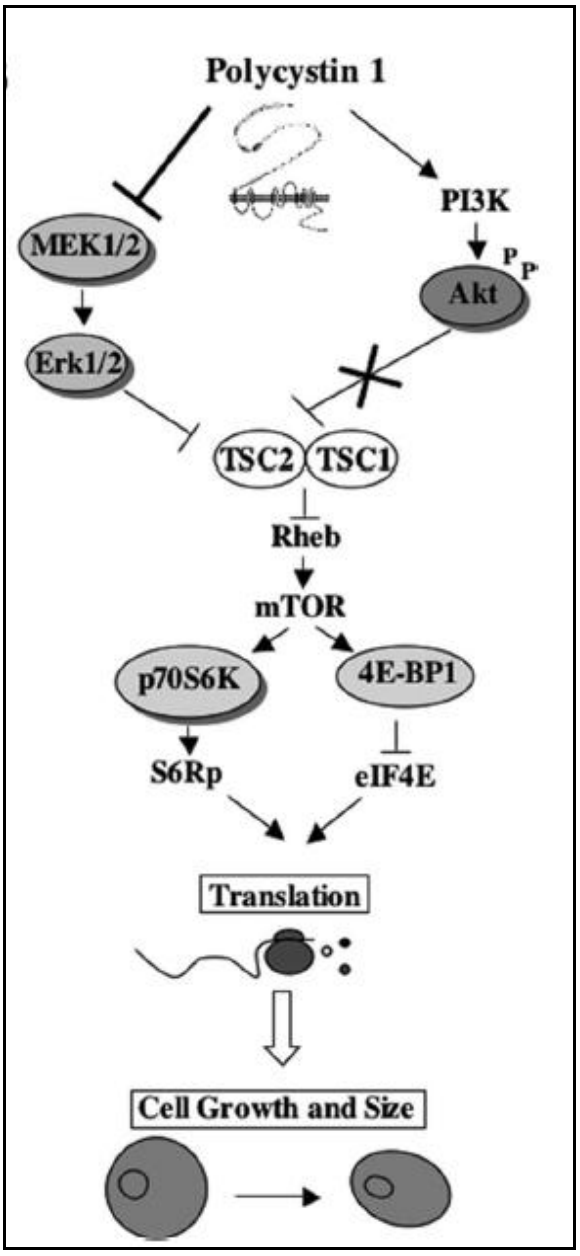


Figure 10. Model for the regulation of tuberin and mTOR pathway by PC1. PC1 induces the activation of the PI3k/Akt pathway but the pool of Akt activated in response to PC1 is not able to phosphorylate Tsc2. At the same time PC1 over expression downregulates the MEK/ERK pathway resulting in reduced phosphorylation of tuberin and thus inhibition of mTOR. In ADPKD, the absence of PC1 would result in the up regulation of ERKs responsible for upregulating the mTOR pathway [Figure from Di Stefano *et al.*, 2009, Mol. Cell. Biol.].

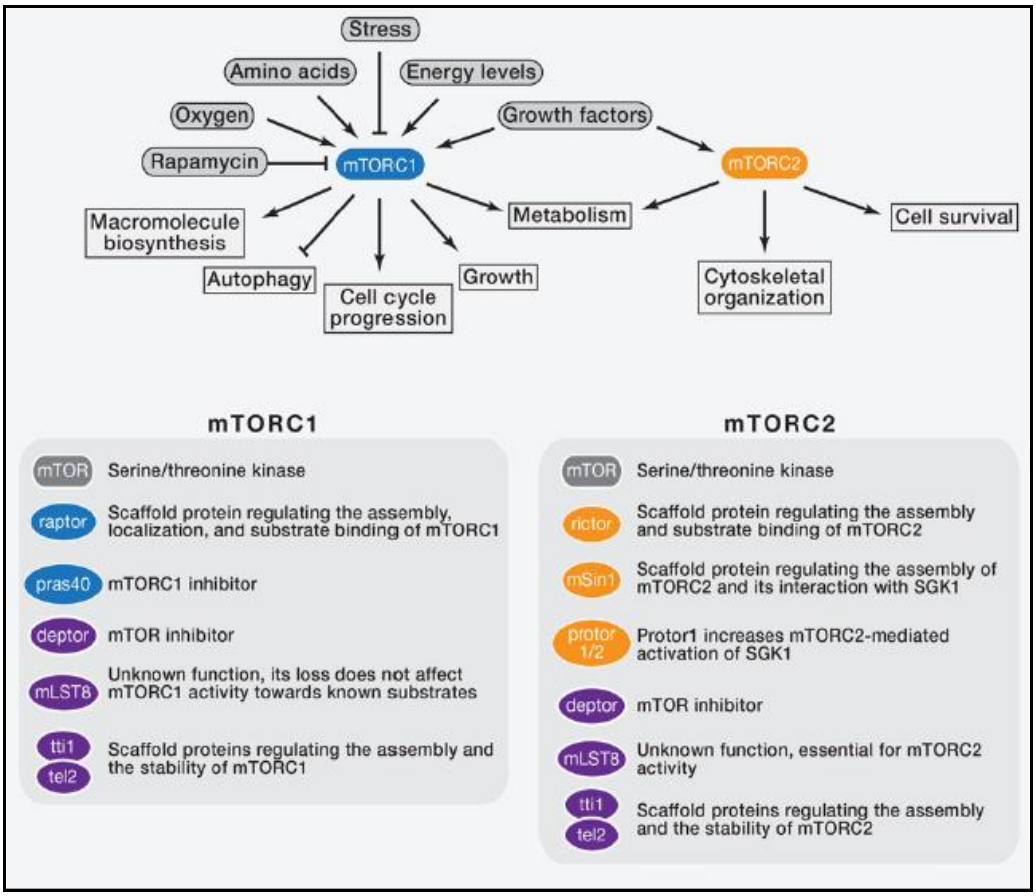


Figure 11. mTORC1 and mTORC2. The upper part of the panel describes the network of mTORC1 and mTORC2. The dark grey boxes are the upstream signal that regulates the activities of these complexes, the light grey the downstream signal regulates by them. The lower part of the panel lists the proteins and the relative functions that form the mTOR complexes [Figure adapted from Laplante M., *et al.*, 2012, Cell.]

2.5.2 PC1 interacts with NPHP1

PC1 and NPHP1 localize to identical subcellular compartment and we have recently demonstrated the interaction between these two proteins, in particular the SH3 domain of NPHP1 interact with PP2, one of the two poly proline motives present in the C-terminus of PC1. The interaction has been demonstrated both *in vitro*, through GST pull down and NMR titration and *in vivo* through co-immunoprecipitation[17].

From the biological point of view the complex appears to be involved in the regulation of apoptosis indeed studies have shown that kidneys from patients affected by ADPKD[44] or by

NPHP[17] display an increase of the apoptotic rate. NPHP1 could act downstream or in parallel with PC1for regulating apoptosis [Figure 12].

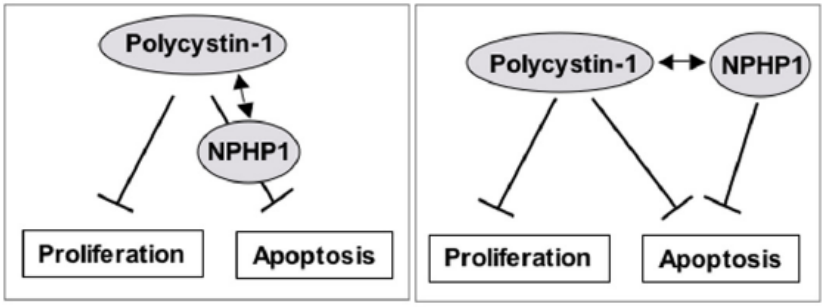


Figure 12. PC1 is able to inhibit both proliferation, acting on mTOR regulation, and apoptosis. For the inhibition of proliferation NPHP1 is not required whereas the regulation of apoptosis requires NPHP1 which could act downstream (left) or in parallel (right) to PC1 [Figure from Wodarczyk C. *et al*., 2010, PlosOne].

The affinity of the resulting complex is low with an estimated dissociation constant, measured by nuclear magnetic resonance (NMR), of about 0.3 +/- 0.2 mM. This low value indicates that the complex is not stable. For stabilizing the complex probably other proteins or other domains of the same proteins could be involved in the interaction. Indeed both proteins present a coiled coil domain, usually involved in protein-protein interaction[18]. In detail the N-terminus of NPHP1 (the first 115 amino acids) is predicted to be a coiled coil domain and the C-terminus of PC1 contains a coiled coil structure [Figure 13]. Hence, in addition to the interaction between the SH3 domain of NPHP1 and the PP2 of PC1, these two proteins could interact together through the coiled coil domains in order to stabilize the complex.

Overall it is hypothesized that the formation of the complex between NPHP1 and PC1 might uncover the link between two unrelated CKD supporting the notion that common pathogenic defects underlie to renal cysts formation.

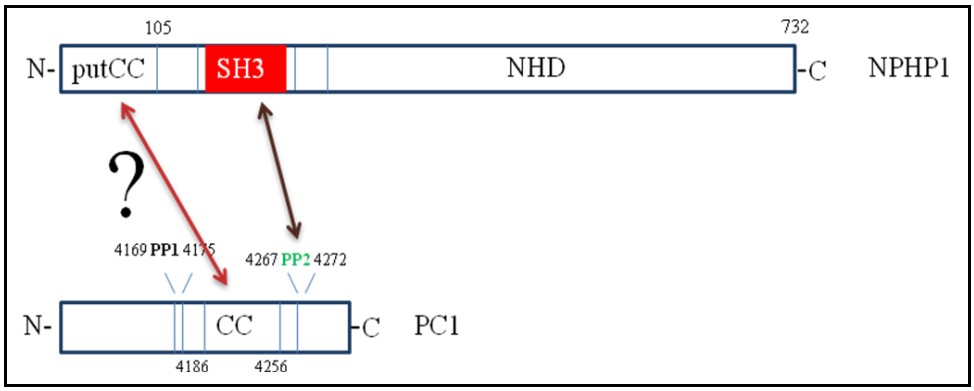


Figure 13. The scheme of the interactions between NPHP1 and PC1. NPHP1 (top) interacts via SH3 domain with one of the two poly proline domains of the PC1. The resulting complex is not stable, hence we hypothesize that the coiled coil (CC) domain of the PC1 and the putative coiled coil (putCC) domain of the NPHP1 could interact to stabilize the complex.

3. Aim of the thesis

Despite the numerous reports on ADPKD the molecular mechanisms of cystogenesis are still poorly understood, hence a complete knowledge on all pathways regulated by PC1 is still lacking. For this reason, at present a pharmacological therapy for the treatment of this genetic renal disease is still lacking.

We focused our work on two new possible roles of PC1: the regulation of metabolism and its interaction with NPHP1.

The hypothesis that PC1 could act as "metabolic regulator" stems from the serendipitous observation of our collaborators, Dr. Boletta in San Raffaele Hospital, Milan. During a routine cellular culture they have observed that *Pkd1* KO MEF cells acidify faster the medium than the WT ones, because of an increased lactate production in *Pkd1* KO cells. This effect could indicate an accelerated metabolism of these cells correlated to an increased cell size and proliferation rate. These features are linked probably to the enhanced activity of mammalian complex of rapamycin (mTOR). Indeed a previous study[41] has demonstrated that the loss of PC1 leads to an upregulation of ERKs pathway and a consequent increase of mTOR activity.

The hypothesis that the interaction between PC1 and NPHP1 might be mediated by the coiled coil motifs present in both proteins originated from our previous study[17]. We have demonstrated that one of the two poly proline motifs present at the C- terminus of PC1 interacts with low affinity with the SH3 domain of NPHP1 regulating apoptosis. This result uncovers a link between the two different ciliopathies, ADPKD and NPHP, supporting the notion that common pathogenetic defects underlie renal cyst formation. Because of the weak interaction between SH3 domain of NPHP1 and the poly proline motif of PC1 we hypothesize that the coiled coil domains of both NPHP1 and PC1 interact to stabilize the complex.

To disentangle PC1 role "metabolic regulator" and its possible interaction with NPHP1 through its coiled coil domain, we have selected nuclear magnetic resonance (NMR) spectroscopy as primary investigation technique.

We have used a metabolomic approach to study the metabolic changes induced by the absence of PC1 analyzing both the extracellular medium conditioned by the presence of KO and WT cells. To confirm *in vivo* the data obtained in cellular model we also analyzed the kidneys, deriving from mouse model of the disease.

In the second part of the thesis we applied NMR to characterize at the atomic level the three dimensional structure of the coiled coil domain of NPHP1 and to verify and characterize its interaction with the coiled coil domain of PC1.

In summary, the aims of the thesis are the following:

- investigation of the metabolic pathways deregulated by the loss of PC1 and characterization of the related signaling cascade both in cells and in kidneys in order to find a possible target for pharmacological treatment;

- determination of the structure of the putative coiled coil domain of the NPHP1;

- determination of the interaction of NPHP1 with the coiled coil domain of PC1.

4. Materials and methods

4.1. Metabolomic experimental procedure

4.1.1 Sample preparation

Exometabolome. Pkd1 KO and WT mouse embryonic fibroblasts (MEFs)[41] were grown in DMEM (Gibco) + 10% of Fetal Bovine Serum (FBS) (Euroclone). The medium was collected one day after cellular confluence, centrifuged at 4°C at 13000 rpm to precipitate the cells in order to avoid to aspirate them, and stored at -80°C. The conditional cells *Pkd1* $^{flox/flox}$ MEFs[41] were grown in the same manner previously described but they underwent the treatment with 10 μM of a recombinant TAT-Cre protein (Excellgen) for 2 hours with 100 mM of chloroquine (SIGMA) in absence of serum, in order to inactivate the *Pkd1* gene preventing any integration within the genome and for avoiding clonally effects.
For experiments aiming to investigate the effect of drug treatment on the metabolism MEFs cells were treated separately with 20 or 50 nM of rapamycin (SIGMA), 2 mM of metformin or 2 mM of AICAR. After cellular confluence, the cells were treated for 48 hours with drugs before to collect the medium.
In all experiments 530 μl of extracellular medium were mixed with 60 μl of deuterated phosphate buffer containing DSS (final concentration 150 mM) and 10 μl of NaN_3 1,2%. The final volume of sample for NMR acquisition was 600 μl.

Biofluids: urine and serum. A volume of 15 μl of serum was collected from mice Ksp-Cre: *Pkd1*$^{flox/-}$ and *Pkd1*$^{flox/+}$ [12] and immediately centrifuged at 13000 rpm at 4°C to avoid coagulation.
NMR samples were prepared mixing 10 μl of serum with phosphate buffer (final concentration 180 mM), DSS (final concentration 90 μM), used as internal standard, and 0,02% of NaN_3. The final volume of the sample was 250 μl.
Urine were taken directly from the bladder from mice Ksp-Cre: *Pkd1*$^{flox/-}$ and *Pkd1*$^{flox/+}$ [12] with a syringe and after centrifuge at 4°C at 13000 rpm to precipitate the cells.
NMR samples were prepared mixing 25 μl of urine with deuterated phosphate buffer (final concentration 180 mM), DSS (final concentration 90 μM) and 10 μl of NaN3 1,2%. The final volume of the sample was 250 μl.

Metabolites extraction from tissues. Tissue samples were snap-frozen in liquid nitrogen and stored at -80°C until extraction. The frozen tissues were weighted before and after lyophilization (24 hours). Polar and apolar metabolites were extracted from the homogenous dry tissue powder using MeOH/$CHCl_3$/H_2O (2:2:1 v/v) solvent extraction strategy[45]. The lyophilized tissues were mechanically fragmented with the help of a spatula, then chloroform and methanol were added and samples were incubated in orbital agitation for 1 hour at 4°C. After incubation chloroform and water were added and samples were centrifuged in order to separate the polar (up) and the apolar (down) phases. The polar phases were placed in new vials and then lyophilized for 24 hours.

For NMR samples the resulting powder was suspended in phosphate buffer (final concentration 150 mM), with addition of 100 mM DSS, and 5 µl of NaN3 1,2% mM. The final volume of each sample was 250 µl with 90:10 H2O:D2O ratio.

4.1.2 NMR spectra: acquisition, metabolites identification and quantification

NMR acquisition. NMR spectra were acquired using a 600 MHz spectrometer (Bruker Avance 600 Ultra Shield TM Plus, Bruker BioSpin) equipped with a triple-resonance TCI cryoprobe with a z shielded pulsed-field gradient coil. All experiments were carried out at 298 K, spectrometer temperature was calibrated using pure methanol-*d4* sample21. Sample temperature inside the spectrometer was equilibrated for 5 minutes before data acquisition. For all experiments continuous water presaturation with a RF of 35 Hz was applied during relaxation delay. Acquisition parameters for each type of sample are listed in Table 1.

Metabolites identification: Metabominer and CCPN. Metabolites were identified using two tools based on spectral reference libraries of small molecules: Metabominer[46] and CCPN Metabolomic project[47].

Metabominer is a semi-automated tool to assist the user in the identification of metabolites in a complex mixture using a comprehensive 2D spectral reference library with a detailed knowledge of the metabolite composition of the common biofluids. The input is the peak list of the spectrum, the software draws a synthetic spectrum and, selecting the library of interest, it matches the experimental peak list with the peak list of the library.

CCPN Metabolomic project, available for CCPNmr suite (it can be downloaded from the Collaborative Computing Project for NMR, http://www.ccpn.ac.uk/collaborations/metabolomics), consists of assigned 1D and 2D spectra of standard compounds according to the public database. The project is interactive, as the spectra of unknown mixtures can be easily superimposed and compared with the database spectra, thus facilitating their assignment and identification. In addition the database can be easily modified.

Sample	Experiments	Dimension	Number of scans	Receiver gain	Spectral window (Hz)	Offset (ppm)	Time domain
Not treated Exometabolome	1D ^{1}H noesy	1H	128	50.8	12019	4.72	7372
	1D ^{1}H cpmg	1H	128	50.8	10027	4.70	7372
	^{1}H^{13}C HSQC	1H 13C	48	181	9615 25655	4.70 75	2048 350
Treated Exometabolome	1D ^{1}H noesy	1H	72	50.8	7194	4.70	7372
	1D ^{1}H cpmg	13C	128	50.8	7194	4.70	7372
Urine	1D ^{1}H noesy	1H	96	71.8	12019	4.70	9830
	1D ^{1}H cpmg	1H	128	50.8	10027	4.70	7372
	J-resolved	1H 1H	24	71.8	10000 94	4.70 4.70	1024 40
	^{1}H^{13}C HSQC	1H 13C	44	181	9615 25655	4.70 75	2048 300
	^{1}H ^{1}H Tocsy	1H 1H	10	181	7212 7212	4.70 4.70	2048 512
Tissue	1D ^{1}H noesy	1H	320	10	7194	4.70	9830
	J-resolved	1H 1H	20	10	10000 94	4.70 4.70	1024 40
	^{1}H-^{1}H Tocsy	1H 1H	26	181	7212 6001	4.70 4.70	2048 512

Table 1: List of NMR experiments and relative parameters.

Spectra were processed using Topspin NMR software from Bruker and Mnova software of Mestrelab[48].

Metabolites quantification. For metabolites quantification we took advantage of the combination of (a) the algorithm called GSD (global spectrum deconvolution), available in the Mnova software package of Mestrelab[48] and (b) of a quantitative referencing strategy, known as PULCON[49]. The GSD algorithm deconvolves overlapping regions in monodimensional spectra in order to obtain separate signals of single metabolites, thus allowing measurement of their absolute integral. To convert the integral into absolute concentration an external standard, the reference, with known concentration (a tube with 0,5 mM of DSS and 1 mM sucrose) was acquired and the PULCON equation (eq.1) was applied.

$$\frac{A_{ref} \bullet cal_{ref}}{conc_{ref}} = \frac{A_{sample} \bullet cal_{sample}}{conc_{sample}} \qquad \text{eq.1}$$

$$cal = \frac{T \bullet P1}{RG \bullet NS \bullet SI}$$

where A is the peak integral, cal is the parameters of acquisition of the spectrum and conc is the concentration. The cal parameter is obtained as function of the temperature of acquisition (T), the receiver gain (RG), the number of scans (NS), the size of real spectrum (SI) and the 90° pulse length (P1).

We have acquired both technical (samples deriving from the same cell culture) and biological (samples deriving from different cell cultures) replicates of the exometabolome of *Pkd1* KO and WT MEF cells, treated or not with drugs, and of *Pkd1* $^{flox/flox}$ MEF cells. To obtain the concentration of the metabolites we have calculated the arithmetic average and the standard deviation of the technical replicates, normalized for the final number of cells in the well, and then, to account for the biological replicates and the different cellular growth, we have calculated the weighted average and the standard deviation of the weighted average.

4.1.3 Statistical analysis: Principal Component Analysis (PCA) and Metabolomics Univariate and Multivariate Analysis (MUMA)

For statistical analysis we used the metabolites concentrations and applied the Pareto scaling of the variables (eq.2) prior to perform principal component analysis (PCA), a multivariate unsupervised statistical technique.

$$\tilde{x}_{ij} = \frac{x_{ij} - \bar{x}_i}{\sqrt{s_i}} \qquad \text{eq.2}$$

The value is converted as fluctuation around zero and scaled for the root square of the standard deviation.

PCA gives a global view of the systematic variation of the data while reducing its dimensionality to few principal components (PC), which account for a large amount of the total variance between the NMR fingerprints. The final aim of PCA is to enable easy visualization of any clustering or similarity of the various samples. The results of PCA are presented in terms of score and loading plots. Samples with a similar metabolic footprint tend to cluster together in the score plot. Each PC is a weighted linear combination of the original descriptors and this information is shown in a loading plot.

To perform the statistical analysis we used a home-made R software package called *muma*: Metabolomics Univariate and Multivariate Analysis (Gaude E. *et.al.*, submitted). This user-friendly package provides a pipeline for the whole process of metabolomics statistical analysis, from data pre-processing, to dataset exploration and visualization, to the identification of potentially interesting variables (or metabolites). To this aim, we implemented well-established multivariate statistical analysis such as PCA, partial least squares discriminant analysis (PLS-DA) and orthogonal projection to latent structures (OPLS-DA) and univariate statistical techniques that are typically used in metabolomics analyses. The package assists the user in the results' interpretation with clear and ready-to-use graphical outputs. *muma* has been designed for the analysis of metabolomics data generated with different analytical platforms (NMR, MS, NIR) with a special focus on NMR data, providing dedicated tools such as Statistical TOtal Correlation Spectroscopy (STOCSY)[50] and for Ratio Analysis of NMR spectroscopy (RANSY)[51].

4.1.4 ^{13}C flux experiments *in vivo*

Experimental procedure and NMR spectra. Litters Ksp-Cre: *Pkd1*$^{flox/-}$ and *Pkd1*$^{flox/+}$ [22] at P8 were treated injecting intracutaneously 1000 mg/Kg of uniformly labeled ^{13}C- glucose (sigma Aldrich).

A litter of 8 mice (4 KO and 4 WT) was used for kinetic experiments. Mice were sacrificed 20, 40 and 60 minutes after the injection of ^{13}C glucose and kidneys, liver, serum and urine were collected. The amount of ^{13}C glucose and the relative ^{13}C lactate produced was detected in the biofluids and in the tissues.

KO and WT mice deriving from five different litters were sacrificed 40 minutes after the injection of ^{13}C-glucose, kidneys and livers were collected, washed in PBS and snap-frozen in liquid nitrogen.

KO and WT mice deriving from three different litters were used for experiments with 2-deoxy-glucose (2DG) (Sigma Aldrich). 2DG or NaCl (as control) was injected intracutaneously into mice at 500 mg/Kg at P7 and P8, the second injection was followed by an injection with 1000 mg/Kg of ^{13}C glucose finally kidneys and livers were collected.

Tissues and biofluids were treated as described in paragraph 4.1.1.

To detect the ^{13}C-glucose and the relative products, especially ^{13}C lactate, NMR spectra were acquired as described in the paragraph 4.1.2. The acquired experiments and the relative parameters are listed in Table 2.

Sample	Experiments	Dimensions	Number of scans	Receiver gain	Spectral windows (Hz)	Offset (ppm)	Time domain
Urine, serum, tissues (liver and kidney)	1D^{1}H noesy	1H	64	45.2	7194	4.70	98304
	1D^{13}C zgpg30	13C	290	181	29762	100	15360
	^{1}H^{13}C HSQC	1H	2	45.2	6602	4.70	1024
		13C			24146	80	400
	^{1}H^{13}C HSQC	1H	3	181	6602	4.71	1024
		13C			30183	100	400

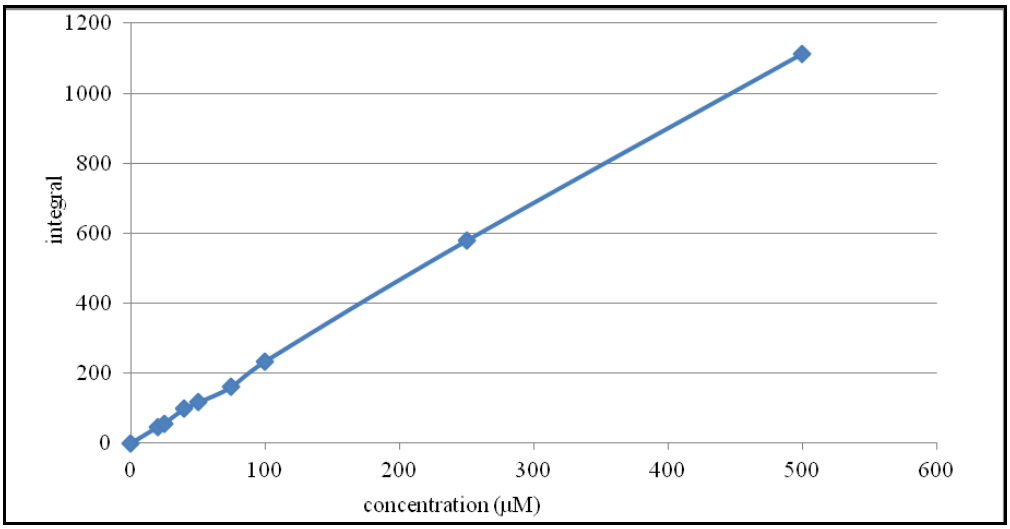
Table 2: List of acquired experiments and the relative parameters for ^{13}C glucose experiments *in vivo*

Quantification and normalization of ^{13}C glucose and ^{13}C lactate peaks. For lactate and glucose quantification we have used 2D ^{1}H-^{13}C HSQC spectra. For lactate determination the peak at 1.32 ppm and 22.91 ppm, corresponding to the methyl (CH_3) signal, was integrated. For glucose determination we have integrated the peaks corresponding to the methylene signal (C^6H_2), accounting for both α- and β-glucose (3.82, 3.78, 63.41 ppm; 3.89, 3.73, 63.47 ppm). These characteristic peaks have been chosen because they are isolated in the spectrum.

To convert the value of the peak integral into metabolite concentration we used an external standard, we therefore made a calibration curve using ^{13}C-glucose. Spectra of home-made standards of ^{13}C glucose at different concentrations: 20, 25, 40, 50, 75, 100, 250, 500 μM, were acquired. In detail, we acquired for each standard two ^{1}H-^{13}C HSQC that differ for the number of scans and the value of the receiver gain: one was acquired with a number of scans equal to 2 and the value of receiver gain of 45.2, the other one with 3 number of scans and 181 of receiver gain The calibration curves were made acquiring spectra with different parameters because the experiments performed on tissues were acquired in different conditions.

We have made two calibration curves [Figure 15] one for each set of ^{1}H-^{13}C HSQC.

A



B

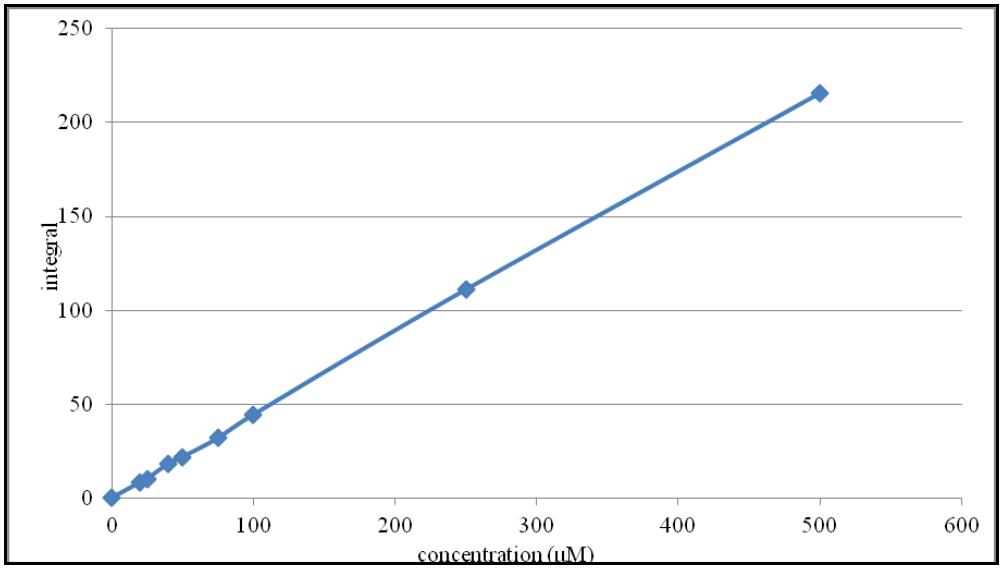


Figure 15. The calibration curves of ^{13}C glucose. The y axis represents the value of integral and the x axis the concentration of the standard. A The curve of ^{1}H-^{13}C HSQC acquired with NS 3 and RG 181, B the curve of ^{1}H-^{13}C HSQC acquired with NS 2 and RG 45.2.

The equation to convert integrals into concentrations is linear y = mx, where y represents the integral and x the concentration, m is the conversion factor. The value of m is 2.248 for ^{1}H-^{13}C HSQC acquired with NS 3 and RG 181, whereas for ^{1}H-^{13}C HSQC acquired with NS 2 and RG 45.2 m is 0.4346.

This equation can be used to convert the integral of the methylene group of the glucose in concentration but not for the methyl group of the lactate as the CH_2 and CH_3 groups have different T1 relaxation rates and different $^{1}J_{CH}$ coupling.

To find the conversion factor between CH_3 and CH_2 we made 11 standard ^{12}C-glucose and ^{12}C-lactate solutions (1:1 stoichiometry) at different concentrations (0.5, 1, 2, 3, 4, 5, 10, 20, 40, 50 and 80 mM, respectively) and $^{1}H^{13}C$ HSQC spectra were acquired. For each spectrum and therefore concentration, the ratio between the integral of the $–CH_3$ of the lactate and the $–CH_2$ of the glucose was calculated. The conversion factor was obtained from the average of the single ratio values. The obtained correction factor is 1.02.

Data analysis. Absolute integrals of the methylene group of the glucose and of the methyl group of the lactate contained in kidneys and livers (I) were measured using the Data Analysis

module of MNova. Integrals were then divided by the dry weight (DW) of the corresponding tissue ($I_{WT}=I_{WT}/DW_{WT}$; $I_{KO}=I_{KO}/DW_{KO}$) to account for the different organ weights of WT and KO mice. Moreover, to account for the genetic variability within litters, for each KO mouse the integrals (I_{KO}) of glucose and lactate peaks were normalized with respect to the integral of the corresponding WT littermate (I_{WT}). Results for both glucose and lactate were therefore expressed as peaks integral ratio between KO and WT animals ($R=I_{KO}/I_{WT}$). To obtain a single value the average and the standard deviation of five litters were calculated.
The effect of 2-deoxy-glucose (2DG) treatment was expressed as ratio (R_{2DG}) between the integrals of glucose and lactate peaks of KO mice treated with 2- deoxy-glucose (I_{KO_2DG}) and the integrals obtained from spectra performed on the corresponding KO littermates (I_{KO}) treated only with NaCl ($R_{2DG}= I_{KO\_2DG}/I_{KO}$). To obtain a single value the average and the standard deviation of five litters were calculated.

4.1.5 Cell's treatment

Glucose starvation. For glucose starvation experiments the composition of the medium is. Basic DMEM (Gibco) 2.3 g/L; Sodium Bicarbonate (Gibco) 3.7 g/L; L-glutamine (Gibco) 0.584 g/L; serum 10% (Euroclone), pen-step 1% (Gibco) with or without glucose (Sigma) 4.5 g/L .

Oligomycin treatment. Cells were treated with oligomycin (Sigma Aldrich) at the final concentration of 30 μM for 48 hours.

4.1.6 Proliferation assay

For proliferation assays immunostaining was carried out using an anti-Ki67 antibody (Novocastra) at 1:400. For cells, fluorescence was measured on triplicates of 300 cells using the microscope Axioplot (Zeiss). For kidney sections, picture at 20x were taken using the

camera Axio MRc5 (Zeiss) with the microscope Axioplane 2 (Zeiss), positive cells were counted on 6 sections for each group for each experiment using ImageJ software.

4.1.7 ATP content quantification

For ATP content evaluation, whole cell extracts of control and treated MEFs cells were prepared by suspending pellets in lysis buffer as described in Distefano et al, 2009. Intracellular ATP quantification of lysates was measured on 250 ng of protein by luciferase activity as showed in the standard protocol present in ATP determination kit (Invitrogen).

4.1.8 Western Blot Analysis

For western blot analysis, the cells were suspended in lysis buffer (250 mM sucrose; 20 mM imidazole and 1mM EDTA, pH 7.4, 0.5% Triton-X 100, supplemented with Protease Inhibitors Cocktail (Amersham) and phosphatase inhibitors (1 mM final of glycerolphosphate, sodium orthovanadate and sodium fluoride). Total lysates were quantified and Laemmli buffer was added to reach a 1X final concentration. Proteins were resolved in a SDS-PAGE gel and transferred onto PVDF membranes. Next, 5% milk in TBS-T was used for incubations with primary antibodies. HRP-conjugated secondary antibodies (Roche) were visualized using the ECL System (Amersham), which were mixed with super-Femto ECL System (Pierce) when necessary.

The primary antibodies used were: anti–p-AMPK-Thr172; anti-AMPK; anti-p-S6RP SER235/236; anti-S6Rp; anti-p-ERK T202/Y204 and anti-ERK (Cell signaling Technology) at 1:1000; anti-LC3 NB100 (Novus Biologicals), anti-actin and anti-tubulin (Sigma Aldrich) at 1:5000.

4.1.9 Histology and immunohistochemistry

After sacrifice, kidneys were removed, washed in phosphate-buffered saline (PBS), weighted and fixed in 4% Paraformaldehyde (PFA). After incubation in a sucrose in PBS gradient scale from 10% to 30% samples were incubated in 10% glycerol (Sigma) in a mixture of OCT (BIO-OPTICA) and sucrose 30%, finally embedded in OCT. Criostat sections were air-dried 1 hour, rehydrated in PBS, incubated in Harris Hematoxylin 1:10 (Sigma-Aldrich) for 2 minutes, washed, incubated in Eosin G (BIO-OPTICA) for 7 minutes, washed dehydrated and mounted in DEPEX (Sigma).

For immunohistochemistry 10 μm cryosections were washed in Tween 0.1% in PBS, fixed in PFA 4%, permeabilized in Triton 0.1% in PBS, incubated in hydrogen peroxide 0.3 % for 30 minutes, blocked for 1 hour at room temperature with 5% Normal Goat Serum (Sigma Aldrich), 3% Bovine Serum Albumine (BSA, Sigma Aldrich) in PBS, incubated over night at 4°C with the antibody (Ab) in blocking buffer diluted at 1:500. Sections were then incubated with the Dako EnVisio+System-HRP (Dako). Staining is completed by an incubation with 3.3’ diaminobenzidine tetrahydrochloride (DAB) DAB+ Substrate Chromogen System (DakoCytomation) and counterstained with Harris Hematoxylin (Sigma Aldrich) diluted 1/10 for 3 min.

4.2. Structure determination methods

4.2.1 Generation of *Escherichia Coli* expression constructs

The plasmids for expression in *Escherichia Coli* of the putative coiled coil domain of the human NPHP1 (spanning the first 115 amino acids) and the coiled coil domain of PC1 (CC_PC1) (spanning residues His 4186 - Ala 4257) were generated using a multi-step PCR-based process.

For NPHP1 the regions of interest were amplified from a pCDNA3.1 vector containing the cDNA of the entire sequence of the human NPHP1 protein (NCBI Reference Sequence: NM_000272.3).

The CC_PC1 was amplified from a pXBlue vector that contains the cDNA of the entire C-terminal tail of human PC1 (NCBI Reference Sequence: NM_001009944.2). Both plasmids were kindly provided by Dr. A. Boletta (San Raffaele Hospital, Milano, Italy).

The forward primers of both constructs contain the NcoI restriction site and the reverse primers bear a stop codon followed by the KpnI restriction site. The sequences of the primers that were used are the following:

- for NPHP1:

NPHP1 1-115F: 5’CAT GCC ATG GGT ATG CTG GCG AGA CGA CAG 3'
NPHP1 1-115R: 5' GG GGT ACC TTA GAT AGG TGC CCC AAC TTC AGT 3'

- forCC_PC1:

CC_PC1F: 5' CAT GCC ATG GGT CAC CCC TCC ACC TCC TCC 3'
CC_PC1R: 5' GG GGT ACC TTA GGC GGG CGC CCG GCT GCT 3'

The Polymerase Chain Reaction (PCR) mix was prepared with: DNA template (30 ng), Pfu buffer 10× (Promega), 10 mM of a mixture of equal quantities of the four dNTPs (Promega), 1 mM of each of the two primers (synthesized by PRIMM, Italy), 3 U of Pfu DNA polymerase (Promega) and water to a final volume of 50 µl.

The cycling parameters for PCR reaction were:

- for NPHP1

1. 95°C for 2 minutes
2. 95°C for 45 seconds
3. 61°C (i.e., primer melting temperature, Tm, plus 2-3°C) for 30"
4. 72°C for 2'
5. repeat steps 2-4 for 34 times
6. hold at 4°C.

- for CC_PC1

1. 95°C for 2 minutes
2. 95°C for 45 seconds
3. 61°C (i.e., primer melting temperature, Tm, plus 2-3°C) for 30"
4. 72°C for 2'
5. repeat steps 2-4 for 34 times
6. hold at 4°C.

The amplified DNA fragments were run in an 1% agarose gel and extracted with the QIAquick Gel Extraction kit (QIAGEN). Overnight digestion at 37°C with NcoI and KpnI (New England BioLabs) followed with: amplified fragment, buffer 1 10× (New England BioLabs), 5 U each of KpnI and NcoI and water to a final volume of 30 µl. The digested fragments were then run in an 1% agarose gel and extracted.

Also the destination vector pETM-44 (6-His-MBP tag), provided by EMBL, Heidelberg, Germany, was digested with NcoI and KpnI, run in an 0.7% agarose gel and purified in order to perform the ligation reaction with the DNA fragment. Ligation was performed overnight at 16°C. The reaction mix contained: 50 ng of the digested vector, T4 ligase buffer 10× (New England BioLabs), 200 U of T4 DNA ligase (New England BioLabs), water to a final volume of 10 µl and the digested amplified. The ng of the digested amplified to add was calculated according to the empirical relationship:

$$digested\ \ insert = 5 \times \frac{digested\ \ vector \times lengthdigested\ \ insert}{lengthdigested\ \ vector} \qquad \text{eq.3}$$

The ligation solution was used to transform DH5α *E. coli* cells. Plasmid DNA was then extracted using Plasmid Mini lits (QIAGEN) and verified by DNA sequencing by PRIMM (Milan, Italy).

4.2.2 Expression of recombinant proteins in *E. Coli*

pETM-44 vector was used for expression of both constructs. The vector allows the isopropyl β-D-1-thiogalactopyranoside- (IPTG-) induced expression in *E. Coli* strains of proteins . BL21(DE3) E. coli strain was used to express the recombinant proteins. Bacterial cells were transformed with the pETM-44 constructs and inoculated from a fresh culture (single colony) in LB medium added with kanamycin 25 mg/ml.

The N-terminal tag of the pETM-44 was exploited both to improve the solubility of the protein during expression for the presence of the maltose binding protein (MBP), and to perform the affinity chromatography on Ni2+-NTA columns (QIAGEN) in the first step of the purification, taking the advantage of the 6-His tag. The 6xHis-MBP tag can be easily removed by cleavage with HRV-3C protease (Novagen), a commercial recombinant protease from human rhinovirus type 14, which recognizes the cleavage site LEVLFQ/GP between the tags and the N-terminus of the recombinant proteins (cleavage occurs between the Q and the G residue). After HRV-3C proteolysis four residues (Gly-Pro-Met-Ala) remain at the N-terminus of the constructs.

Different conditions of induction were tested, varying the temperature of induction and the amount of IPTG, in order to find the best condition of expression. The best conditions of protein expression are the same for both constructs: 0.5 mM IPTG at 22°C over night.

4.2.3 Purification of recombinant proteins in *E.Coli*

The same protocol of purification was used for both the constructs, the only difference is that PC1 did not undergo to cleavage with HRV 3C protease.

After induction, 6×His-MBP-NPHP1 expressing cells were harvested by centrifugation, re suspended in 25 ml of lysis buffer + (Table 3) and lysed by sonication. Cell debris were removed by centrifugation at 11000 rcf for 45 minutes at 4°C and the supernatant was filtered by means of a 0.22 μm filter. The filtered solution was loaded and incubated at 4°C with a Ni^{2+}-NTA column previously equilibrated with the lysis buffer. After collecting the flow-through, the column was washed in sequence with 5 volumes of column (CV) of lysis buffer, 10 CV of wash 2, 10 CV of wash 3 and 5 CV of wash 4 in order to remove proteins that bound the Ni^{2+}-NTA resin aspecifically (buffer compositions are listed in Table 3). The 6×His-MBP-tagged recombinant protein was then eluted from the column adding small volumes of elution buffer until all the protein was detached from the resin (as checked by inspecting small volumes of the eluted solution by adding Bradford reagent). After elution the tagged protein was dialyzed over night at 4°C in dialysis buffer (50 mM Tris-HCl pH 7.5, 200 mM NaCl, 2 mM 2-β-mercaptoethanol), in order to change the buffer and create the best possible condition for the cleavage of the protease. To remove the tags, after dialysis, the tagged proteins was incubated for 16 hours at 4°C with HRV-3C at ratio 1 unit of enzyme: 0,5 mg of protein.

The solution was then loaded onto a Ni^{2+}-NTA column previously equilibrated with the dialysis buffer in order to remove 6×His-MBP-tags and HRV-3C protease (which also presents an N-terminal 6×His-tag). As last purification step, the flow-through was collected, concentrated, and loaded onto a HiLoad 16/60 Superdex 75 size exclusion chromatography column (GE Healthcare Life Science) equilibrated with a phosphate buffer suitable for recording NMR experiments (20 mM NaH_2PO_4/Na_2HPO_4 pH 6.3, 150 mM NaCl, 3 mM DTT).

All the purification steps were checked by loading samples of the protein solutions on 15% sodium dodecylsulphate polyacrylamide gel eletrophoresis (SDS-PAGE) gels. The amino acidic sequences of the both purified recombinant proteins were checked by mass spectrometry.

Uniformly ^{15}N- and ^{15}N-^{13}C- labeled protein samples were obtained by growing and inducing the bacteria in M9 minimal medium containing ($^{15}NH_4$)Cl without or with ^{13}C-glucose, respectively (Table 4). Extraction and purification of the labeled proteins were performed as previously described. The yield for both unlabeled and labeled NPHP1 1-115 was about 15 mg/l, whereas the yield of unlabeled CC_PC1 with or without the tag was about 11 mg/l.

	Lysis buffer +	Lysis buffer -	Wash 2	Wash 3	Elution buffer
Tris-HCl pH 8	20 mM	20 mM	20 mM	20 mM	20 mM
NaCl	150 mM	150 mM	1 M	150 mM	150 mM
Imidazole pH 8	10 mM	10 mM	10 mM	30 mM	300 mM
ZnCl2	100 μM	100 μM	100 μM	100 μM	100 μM
2-β-mercaptoethanol	2 mM	2 mM	2 mM	2 mM	2 mM
NP-40	0.2%	–	–	–	–
Complete EDTA-free (Roche)	1 tablet/50 ml	–	–	–	–

Table 3. Buffer compositions for NiNTA purification.

	^{15}N- labelling	^{13}C- ^{15}N-labelling
M9 salts pH 7 10×	100 ml	100 ml
D-glucose	4 g	4 g (^{13}C)
NH_4Cl	3g (^{15}N)	3g (^{15}N)
Trace elements 1000×	1 ml	1 ml
$CaCl_2$ 1 M	0.3 ml	0.3 ml
$MgSO_4$ 1 M	1 ml	1 ml
Kanamycin 25 mg/ml	1 ml	1 ml
Thiamin 1 mg/ml	1 ml	1 ml
Biotin 1 mg/ml	1 ml	1 ml

Table 4. Composition of 1 liter of M9 minimal medium for uniformly 15N- and 13C-15N-labelling.
100 ml M9 salts pH 7 10× contain: 6 g Na2HPO4, 3 g KH2PO4, 0.5 g NaCl; 125 ml traces elements 1000× contain: 12.5 g EDTA pH 7.5, 2.10 g FeCl3●6H2O, 210 mg ZnCl2, 30 mg CuCl2●2H2O, 25 mg CoCl2●6H2O, 25 mg H3BO3, 4 mg MnCl2●6H2O

4.2.4 Circular Dichroism (CD) experiments

CD experiments were recorded on a Jasco J-815 instrument equipped with a Peltier temperature control system using a square cuvette with 1 mm path length. Experiments were

performed with a sample containing 15 uM of unlabelled NPHP1 1-115 in 20 mM NaH_2PO_4/Na_2HPO_4 pH6.3, 150 mM NaF, 0,3 mM DTT.

CD spectra were performed at 293 K. To measure thermal unfolding curves, sample was heated at 1°C/min from 283 K to 369 K recording CD signal at two constant wavelengths (209 nm and 222 nm).

Spectra were analyzed using Spectra Manager ® Jasco and secondary structure content was estimated using the same program.

4.2.5 Protein resonances assignment for structure determination

Spectral assignment was done using the following spectra: ^{1}H-^{15}N HSQC, HNCA, HNCACB, HN(CO)CACB, HNCO, HNHα, ^{1}H-^{13}C HSQC, CC(CO)NH, HCC(CO)NH, (HB)CB(CGCD)HD, HCCH-TOCSY and ^{1}H-^{1}H TOCSY. In detail the ^{1}H-^{15}N HSQC, HNCA, HNCACB, HN(CO)CACB, HNCO and HNHα spectra were employed to assign the backbone, whereas of ^{1}H-^{13}C HSQC, CC(CO)NH, HCC(CO)NH, (HB)CB(CGCD)HD, HCCH-TOCSY and ^{1}H-^{1}H TOCSY experiments were used to assign the side chains. The acquisition parameters are listed in Table 5.

All NMR spectra were acquired on a Bruker Avance 600 MHz spectrometer equipped with inverse triple-resonance cryoprobe and pulsed field gradients, with the exception of HCCH-TOCSY acquired on 900 MHz at 298K. Experiments were performed using a concentration of protein samples between 0,5 and 0,9 mM in a buffer containing 20 mM NaH2PO4/Na2HPO4 pH 6.3, 150 mM NaCl, 3 mM dithiotreithol (DTT). D2O percentage was 10% (v/v) or 100% (v/v).

Spectra were processed using NMRPipe[52]and NmrDraw or Topspin NMR software from Bruker and then analyzed with CCPNmr software[53].

	^{1}H				^{15}N			^{13}C			
	dim	**TD**	**SW (Hz)**	**off (ppm)**	**TD**	**SW (Hz)**	**off (ppm)**	**TD**	**SW (Hz)**	**off (ppm)**	**Mx (ms)**
^{1}H-^{15}N HSQC	D	2048	7211	4.7	200	1824	116.5				
^{1}H-^{13}C HSQC	D	2048	6009	4.71				180	21127	72	
^{1}H-^{1}H TOCSY (H_2O)	D	2048	7211	4.71							60
	I	480	7211	4.71							
^{1}H-^{1}H TOCSY (D_2O)	D	2048	6602	4.7							60
	I	480	6602	4.7							
HCCH-TOCSY	D	1024	10822	4.71				276	9054	65	12
	I	50	10822	4.71							
HNCA	D	2048	8417	4.7	90	1946	118	118	4829	51	
HNCACB	D	2048	7212	4.7	80	1825	116.5	162	10563	41	
HN(CO)CACB	D	2048	6602	4.7	92	1825	116.5	162	9809	41.5	
HNCO	D	2048	8418	4.7	64	1946	118	64	2717	174	
HNHA	D	1024	6602	4.71	84	1825	116.5				
	I	160	6602	4.71							
CC(CO)NH	D	2048	7212	4.7	70	1825	116.5	170	10865	40	12
HCC(CO)NH	D	1024	6602	4.7	56	1307	117.5				12
	I	204	6602	4.7							
(HB)CB(CGCD)HD	D	1024	3597	4.71	200	10564	35				

Table 5: Acquisition parameters of the NMR experiments for NPHP1 1-115 resonance assignment. Abbreviations are: "dim": dimension ("D" and "I" being direct and indirect, respectively), "TD": time domain, SW: spectral width ,"off": offset ,"mx": mixing time.

4.2.6 Chemical-Shift Rosetta (CS- Rosetta)

The chemical shifts are strongly influenced by local conformation, so they represent a helpful tool for molecular structure determination by NMR. The secondary chemical shifts (Δδ, the differences between the observed chemical shifts and their corresponding random coil values) can be used to predict the protein secondary structure[54].

Chemical-Shift Rosetta or CS-Rosetta is a rather innovative program for protein structure determination that allows structure determination of protein using only chemical shift information. This program combines the well established ROSETTA structure prediction program[55]with the empirical relation between chemical shift and structure.

Rosetta is based on the concept that 3D structure of a protein is directly related to its amino acid sequence. Hence for each small segment of the query protein, Rosetta selects two hundred fragments from the crystallographic structural database similar for amino acids sequence and consequently representative of the possible conformations of the peptide segment. Then these fragments are assembled with a Monte Carlo based process to search for compact, low energy folds. The refinement protocol of the Rosetta full atoms employs Monte Carlo minimization coupled with a detailed all-atom force field in order to search the low energy structures. CS-Rosetta improves the Rosetta performance adding the chemical shifts, the structural information obtained experimentally[56].

4.2.7 Dihedral angles: 3J(HNHα) calculation and TALOS+ prediction

*3J(HNHα) calculation NMR acquisition.*The HNHα is a 3D spectrum in which magnetization is transferred from ^{1}H and ^{15}N and then selectively to the backbone 1Hα via three-bonds J-coupling 3J(HNHα)[57]. Magnetization is then transferred back via ^{15}N to ^{1}H for detection. Hence for each amide group the spectrum displays one diagonal peak and one cross peak, corresponding to the resonance of the amide proton and of the α proton of the same residue, respectively.

The experiment is designed to determine accurately the value of the 3J(HNHα) coupling constant for each protein residue using the following relation:

$$\frac{Intensity\ cross\ peak}{Intensity\ diagonal\ peak} = -\tan^2\left(2\pi\varepsilon \cdot 3JHN, H\alpha\right) \qquad \text{eq.4}$$

in which the peaks intensities ratio is a function of the HN-Hα magnetization transfer time (ε) and of the 3JHN, Hα value.

The value of 3JHN, Hα coupling constants can be converted into the values of the corresponding phi dihedral angles of the protein backbone according to the Karplus relationship [Eq 5; Figure 16][57]. The conversion was calculated by means of CCPNmr software.

$$3JHN,H\alpha = A\cos^2(|\varphi - 60^\circ|) - B\cos(|\varphi - 60^\circ|) + C \qquad \text{eq.5}$$

Where the optimal values of the constant A, B and C for the protein are 6.51, -1.76 and 1.60 respectively.

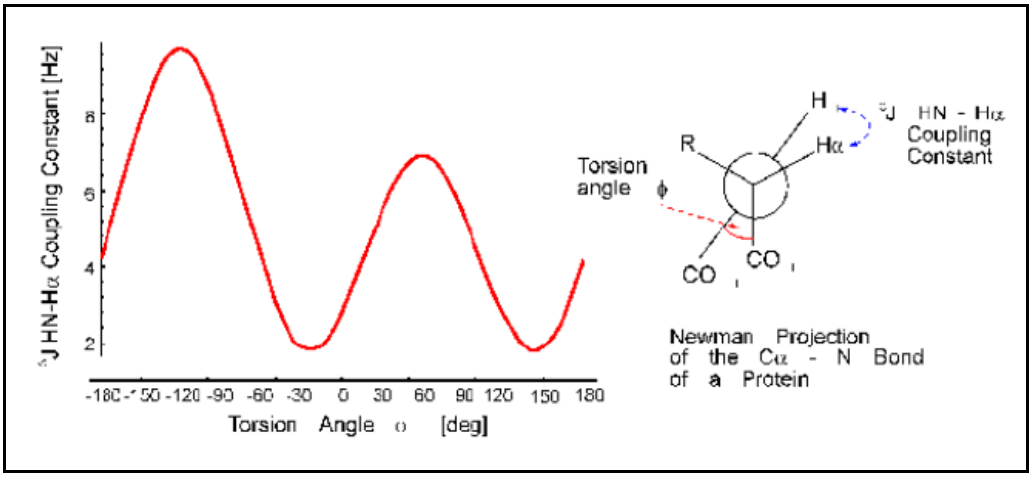


Figure 16: The Karplus curve. The 3JHN-Hα is linked to the phi angle.

Talos + software. Talos+ is an improved version of the TALOS software (Torsion Angle Likelihood Obtained from Shift and sequence similarity), generally used for prediction of the dihedral angles phi and psi of a given protein sequence from chemical shifts of ^{1}Hα, ^{13}Co, ^{13}Cα, ^{13}Cβ and ^{15}N and ^{1}H of the amide[55].

The protein is divided in the overlapping triplets and only the phi and psi angles of the central residue are predicted. An artificial neural network (ANN) analyzes the chemical shifts and the sequence and compares this information with a database of 200 protein structures with known secondary shifts to obtain a prediction of psi and phi angles and an estimate of the likelihood of the prediction depending on the best 10 matches found in the database [Figure 17].

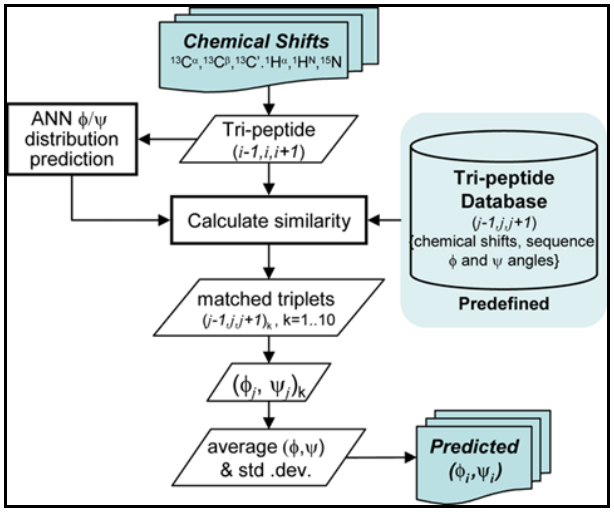


Figure 17: The flowchart of Talos+. (Figure from http://spin.niddk.nih.gov/NMRPipe/talos/).

Talos + reports a secondary structure prediction, generated by ANN, and an estimated backbone order parameters S2 to prevent assignment of torsion angles to regions that are dynamically disordered. In addition, Talos+ prediction can be easily implemented in ARIA2.1.3 software to calculate the protein structure using dihedral angles as constrains.

4.2.8 Determination of NOE restraints

The NOE restraints are the most important data for three dimensional structure determination of a protein by NMR. They give a range of distances between ^{1}H closely atoms (less 5-6 Å). They are determined by the intensities of the cross peaks of NOESY spectra, such as ^{1}H-^{1}H NOESY, ^{1}H-^{15}N NOESY and ^{1}H-^{13}C NOESY. These interactions drop off very fast with the distance as they depend on the inverse of the sixth power of the distance ($1/r^{6}$). Nuclear Overhauser Enhancement (NOE) relies on dipolar interactions[58].

^{1}H-^{1}H-NOESY (2D) in H2O and D2O, ^{1}H-^{13}C-HSQC-NOESY for ^{13}C aliphatic atoms and ^{1}H-^{15}N-HSQC-NOESY experiments were performed at 298 K on a Bruker Avance 600 and 900 MHz spectrometers equipped with inverse triple-resonance cryoprobe and pulsed field gradients. Protein samples were 0.7-0.8 mM in NMR buffer. D2O concentration was 10%

(v/v) or 100% (v/v). Acquisition parameters are listed in Table 6. NOE cross-peaks were manually assigned.

		^{1}H			^{15}N			^{13}C			
	Dim	**TD**	**SW (Hz)**	**Off (ppm)**	**TD**	**SW (Hz)**	**Off (ppm)**	**TD**	**SW (Hz)**	**Off (ppm)**	**Mx**
^{1}H-^{1}H-NOESY (H_2O)	D	2048	7212	4.7							100
	I	480	7212	4.7							
^{1}H-^{1}H-NOESY (D_2O)	D	2048	6602	4.71							120
	I	480	6602	4.71							
^{1}H-^{13}C-HSQC-NOESY (^{13}C aliphatic atoms)	D	2048	10776	4.7				70	18116	40	100
	I	230	10776	4.7							
^{1}H-^{15}N-HSQC-NOESY	D	1024	10823	4.7	50	1961	117.5				100
	I	256	10823	4.7							

Table 6: Acquisition parameters of the NMR experiments for NPHP1 1-115 NOEs assignment. Abbreviations are: "dim": dimension ("D" and "I" being direct and indirect, respectively), "TD": time domain, SW: spectral width, "off": offset, "mx": mixing time.

4.2.9 Structure calculation

The standard procedure to determine the solution structure of a protein from NMR data is to minimize an objective function which incorporates experimental data (e.g., intensities of assigned NOE cross-peaks, dihedral angles, residual dipolar couplings) and physical knowledge (physical and chemical properties of protein nuclei and chemical bonds). The task is accomplished by software such as CYANA[59] and ARIA[60-62]. These programs calculate the protein structure based on the conformational restraints using the experimental data and using a dynamic force field which describes the protein physical and chemical properties. We used ARIA 2.1.3.[62](see Appendix) to calculate the solution structure of the N- terminus of NPHP1.

4.2.10 Dynamic information: ^{15}N spin relaxation measurements

NMR spectroscopy is a powerful tool not only to solve the 3D structure of biomolecules but also to provide their dynamic information, including their internal motions and the rotation diffusion (τc) in solution. The dynamic information of the protein is achieved by means of ^{15}N spin relaxation measurement[63]. For this type of investigation three NMR experiments are generally acquired: the longitudinal and the transverse relaxation, to obtain T1 and T2 values respectively, and of heteronuclearNOE (hetNOE). T1 and hetNOE describe high frequency motions, in the time scale of ps-ns, whereas the T2 describes slow motions in the time scale of μs-ms.

All the experiments were performed using ^{15}N-labelled NPHP1 1-115 using sensitivity enhanced pulse field gradient sequences[64]. Heating effects were minimized combining a "dummy" irradiation scheme with an efficient pre-scan ^{1}H saturation sequence[65].

T1 and T2 relaxation delays are listed in Table 7. T1 and T2 values were obtained using the fitting routine implemented in NMRView, using the equation 6:

$$I = I_0 e^{-t/T} \qquad \text{eq. 6}$$

where I is the peak intensity, I_0 is the extrapolation of the peak intensity for null t and T is T1 (or T2). Both T1 and T2 experiments use a 2.5 seconds recycle delay.

The HetNOE experiment is based on bidimensional ^{1}H -^{15}N HSQC experiment.

In the ^{1}H -^{15}N HSQC experiment the magnetization is transferred from hydrogen of an amidic group to nitrogen directly bound, via J-coupling. Chemical shift is evolved on the nitrogen and magnetization is transferred back to hydrogen for detection. Hence in this spectrum all amidic groups of the protein are visible (see appendix).

T1	T2
250*	72
500	12*
2000	244
50*	44

900	112*
1700	28
150	144
250	12
1400	56
350	112
50	200
1100	–
700	–

Table 7. List of delays used for the ^{15}N relaxation measurements. Delays are expressed in ms. The asterisk indicates that the delay has been repeated twice for error estimation.

The basic pulse sequence for ^{1}H-^{15}N HSQC comprises three blocks and then the detection of the signal [Figure 18]. Block A comprises a INEPT (Insensitive Nuclei Enhances by Polarization Transfer) module, whose purpose is to transfer nuclear spin polarization between J-coupled nuclei, from the more sensitive one the hydrogen, to less sensitive one, the nitrogen. The time delay $\Delta/2$ is set as near as possible to $1/(4J)$, where J is the scalar coupling between ^{1}H and ^{15}N, $^{1}J_{HN}$ value (about 90 Hz). In this manner only magnetization of amide protons is transferred to the adjacent ^{15}N nucleus. During block B, a ^{15}N frequency labeling is achieved by incrementing the variable delay t1, which leads to a generation of the indirect dimension of the 2D spectrum (F1 frequency). Block C comprises a reverse INEPT module. Nuclear spin polarization is again transferred from ^{15}N to ^{1}H. This enables data acquisition in which 1H magnetization is directly detected during t2, which corresponds to F2 frequency in 2D spectrum.

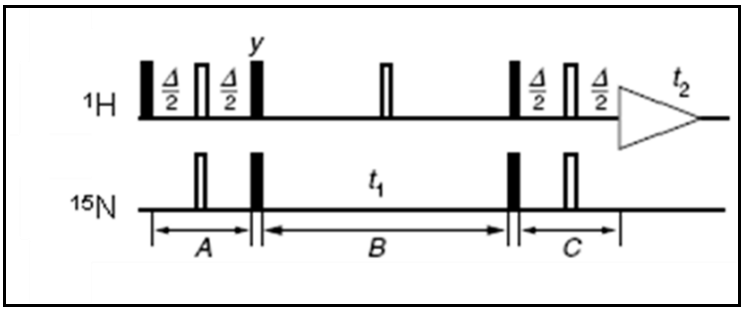


Figure 18. Basic pulse sequence for ^{1}H-^{15}N HSQC experiment.

HetNOE experiment is performed acquiring two ^{1}H-^{15}N HSQC in interleaved fashion were recorded: one is acquired as a normal HSQC spectrum, used as reference for the intensity of the peaks, and the other one is acquired with a presaturation of the ^{1}H at the beginning of the pulse sequence in order to give rise to the heteronuclear NOE effect between the ^{1}H and the ^{15}N nuclei [Figure 19].

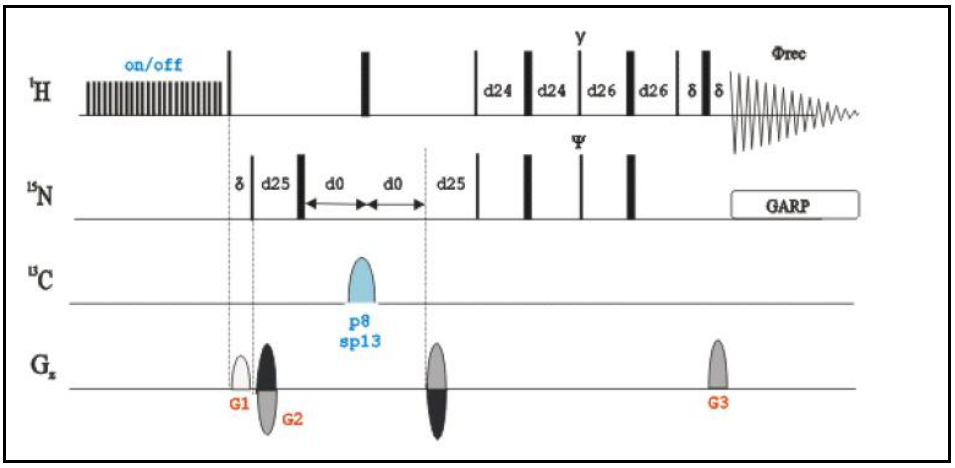


Figure 19. Pulse sequence to record 2D ^{1}H ^{15}N HSQC spectra used to measure the hetNOE for each protein residue. ^{1}H presaturation is activated at the beginning of ^{1}H ^{15}N HSQC experiment in which the heteronuclear NOE effect works between the ^{1}H and ^{15}N nuclei. The presaturation is not active in the ^{1}H ^{15}N HSQC reference spectrum . G1, G2 and G3 are gradients.

The hetNOE value for each residue is calculated as the ratio of the volume of the HN peak in the ^{1}H presaturated spectrum (Isat) and the volume of the same HN peak in the reference spectrum without ^{1}H presaturation (Iref):

$$hetNOE = \frac{Isat}{Iref} \qquad \text{eq.7}$$

The intensity of the HN peak in the presaturated spectrum depends on the flexibility of the residue, the intensity is smaller for residues that belong to flexible regions, in the time scale of ps-ns.

The value of hetNOE can range between -1 and 1 and it is plotted for each residue of the protein in order to visualize the flexible regions and the structured ones of the proteins. The structured regions have an hetNOE value between 0.8 and 1 whereas the flexible regions have a hetNOE value < 0.5.

4.2.11 Residual dipolar couplings (RDC)

Residual dipolar couplings (RDCs) measurement allows to investigate internuclear vectors' orientations, irrespectively of the distance between them[66]. In order to acquire this information for proteins in solution, a partial alignment of the molecules with the field must be obtained, leading to an additional coupling to the scalar coupling.

The ^{15}N-^{1}H RDC measurements were performed using the axially compressed acrylamide gel method described in[67]. A 0.25 ml acrylamide solution (final percentage 5%) was prepared. Polymerization was started adding 0.1% w/v ammonium persulfate and 0.5% w/v tetramethylethylenediamine (TEMED). Immediately after starting the polymerization, the solution was injected into a 3.5 mm plastic tube. As polymerization was completed, the gel was removed from the tube, extensively and repeatedly washed with MilliQ water and dried overnight at 37°C. The dried gel was then placed into an NMR shigemi sample tube and 0.31 ml of NMR buffer solution containing 0.3 mM of NSD1-P5C5 was added. After gel had absorbed the protein solution and swollen, a vertical pressure was exerted on the plunger to reduce its height (which was approximately reduced by a third). A sensitivity enhanced HSQC experiment that separates the upfield and downfield components in the indirect dimension into different subspectra with a sequence similar to the generalized TROSY scheme and decoupling of the N-nucleus during detection [68]was performed identical protein solution in the presence and in the absence of the acrylamide gel. Data were inspected with MODULE[69]. The Q factors, which quantify the average agreement between the experimental RDCs and the theoretical values calculated from a given structure[70], have been calculated using PALES[71].

4.2.12 NMR titrations

The NMR chemical-shift mapping method is a highly sensitive tool to prove binding and to map binding sites in protein-ligand interactions[72]. It requires ^{15}N-labeled protein in order to record 2D ^{1}H-^{15}N HSQC spectra of the free protein, then stepwise additions of the unlabeled titrant are performed and followed by collecting ^{1}H-^{15}N HSQC spectra after each addition. If binding occurs, the chemical environment of the residues that constitute and/or are in close proximity to the protein binding site will change. As a consequence, during titration these residues show chemical shift perturbations of their backbone amide groups. The type of the observed perturbation depends on the rate of binding with respect to the time scale of NMR chemical shift:

- if binding is in fast exchange, the exchange between the bound and the free form is faster than the difference in chemical shifts. Only one set of protein signals is visible, and their positions typically shift according to the ratio between the bound and the free species. Therefore, the chemical shifts move from the free form of the protein to the position of the bound state, which is reached once the protein sample has been completely saturated with ligand. The chemical shift displacement can be fit to a mathematical model and used to calculate the affinity (K_D) of the interaction.

- If binding is in slow exchange, residues affected show reduction of their peak intensities, while new peaks, belonging to the protein in the bound form, appear at different chemical shifts, with an increasing intensity upon ligand addition.

- If binding is in intermediate regime, whereby the rate of exchange between the free and the bound states is comparable to the difference in chemical shifts between the two states. Consequently the behavior is more complicated, as it entails a mixture of signal shifts, decreasing signals and newly appearing signals. It results in very broad signals which makes the analysis very difficult.

When the binding is in the fast exchange, the HN peaks' chemical shift displacement is measured using the following equation:

$$CSD = \sqrt{\frac{\Delta\delta^2 HN + \frac{\Delta\delta^2 N}{25}}{2}} \qquad \text{eq. 8}$$

The hypothesis that the N-terminus of NPHP1 1-115 domain interacts with the coiled coil domain of PC1 (CC_PC1) was tested by NMR titration. The ^{1}H- ^{15}N HSQC of the free ^{15}N N-terminus of NPHP1 concentrated 0.3 mM was recorded at 298K and then the not labeled ligand, the 182 µl of 6His-MBP- CC_PC1 concentrated 0.4 mM, was added and the another ^{1}H-^{15}N HSQC was recorded. We have tested only the stochiometry 1:1.

5. Metabolomics

5.1. Metabolomic profiling of the exometabolome of WT and *Pkd1* KO cells

To find the metabolic hallmarks of *Pkd1* KO and WT cells we compared the medium conditioned by the presence of cells. We have analyzed the metabolic content of the media after one day of conditioning with WT and *Pkd1* KO MEF cells, respectively. Importantly *Pkd1* KO cells display a different phenotype as compared to the WT cells: they show an increased cell size, they are double in volume than the WT ones, and they have an increased proliferation rate, as shown by the growth curve [Figure 20]. In order to facilitate the comparison between the two cell types we decided to perform the metabolomic analysis collecting the media after one day conditioning, where both kind of cells have similar number of cells and they are in proliferate phase.

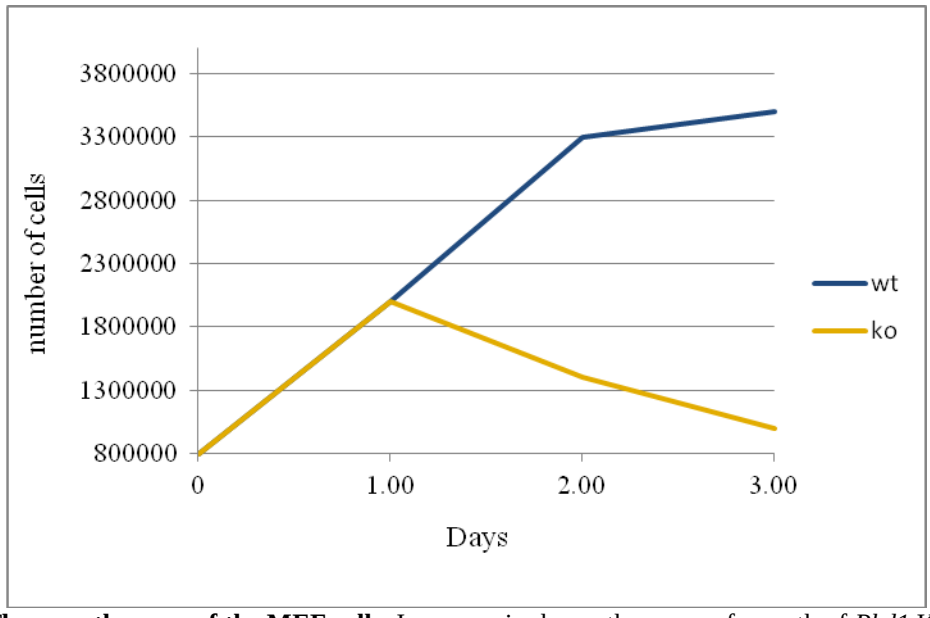


Figure 20. The growth curve of the MEF cells. In orange is shown the curve of growth of *Pkd1* KO cells, they begin to die starting from the second day of growth, probably because of contact inhibition, in blue is shown the curve of growth of WT cells, these cells show proliferation during three days.

We next recorded the 1D ^{1}H NMR spectra of the media conditioned with the cells, the representative spectra of media conditioned with of WT (blue) and *Pkd1* KO (orange) cells are shown in Figure 21.

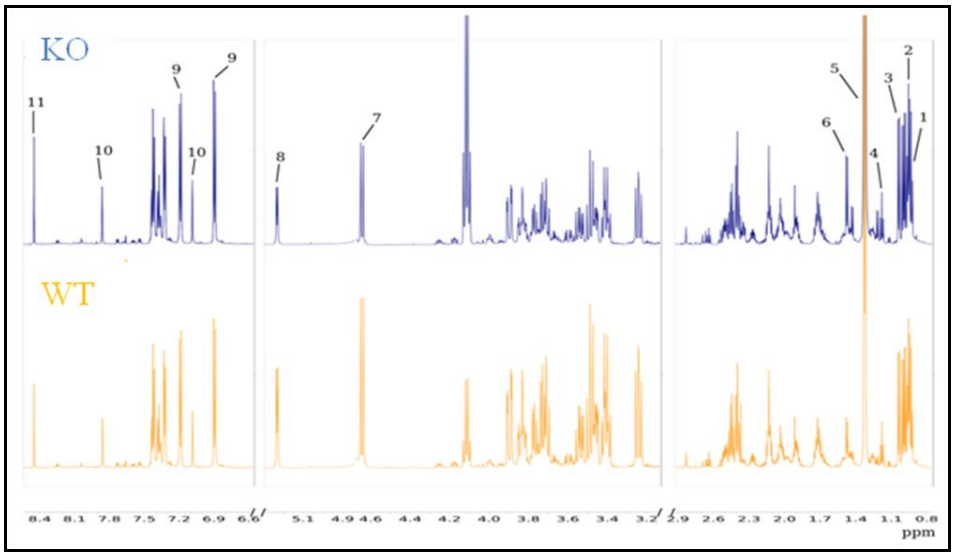


Figure 21. The metabolic profile of the medium conditioned by the presence of *Pkd1* KO cells (top), the spectrum blue, and the one conditional on the presence of WT cells (bottom) in orange. The numbers indicate some assigned metabolites: 1= leucine; 2= valine; 3= isoleucine; 4= ethanol; 5= lactate; 6= alanine; 7=β-glucose; 8= α-glucose; 9= tyrosine; 10= histidine; 11=formic acid.

Using Metabominer[46], 2D spectra ^{1}H-^{13}C HSQC and 2D ^{1}H-^{1}H TOCSY, and CCPN Metabolomic project[47] and the PULCON equation[49] (see paragraph 4.1.2) we have identified and quantified 22 metabolites in the medium (Table 8). The medium conditioned by *Pkd1* KO cells shows more content of lactate, glutamate and alanine compared to the medium conditioned by WT cells, whereas the medium conditioned by WT cells shows more content of glucose, glycine, pyruvate, choline and formate [Figure 22 A, B].

The *Pkd1* KO cells display a metabolism similar to cancer cells producing lactate[9], glutamate[10,11] and alanine[9] in the medium. Indeed, similarly to cancer cells, they display an accelerated metabolism as indicated by production of lactate, at the same time the production of alanine is necessary to secrete the surplus of carbon and nitrogen formed by the fast metabolism[9], whereas the glutamate supplies as stimulus to enhance the cellular growth[10,11]. On the other hand the *Pkd1* KO cells do not display a major production of glycine as tumor cells. The production of glycine could derive from serine degradation, a metabolic pathway deregulated in tumors. The secretion of glycine by tumor cells is necessary to protect them from the immune response[73].

Metabolite	**Peak used for quantification (ppm)**	**WT (mM·10^{-6}/cell)**	***Pkd1* KO (mM·10^{-6}/cell)**	**t-test (WT/*Pkd1* KO)**
Glucose	5.24-5.23 (d) alpha 3.26-3.23 (t) beta	13.54 (0.20)	6.27 (0.08)	***
Lactate	4.13-4.10 (q)	5.31 (0.10)	12.80 (0.25)	***
Pyruvate	2.37 (s)	0.0998 (0.0019)	0.033 (0.0002)	***
Succinate	2.40 (s)	0.2628 (0.0021)	0.2252 (0.0078)	ns
Choline	3.20 (s)	0.0108 (0.0002)	0.0095 (0.0002)	**
Formate	8.45 (s)	0.0947 (0.0018)	0.0901 (0.0006)	***
Ethanol	3.68-3.66 (q)	0.440 (0.003)	0.399 (0.013)	ns
Alanine	1.48-1.47 (d)	0.23 (0.01)	0.279 (0.002)	***
Arginine	1.67-1.60(m)	0.325 (0.003)	0.251 (0.009)	ns
glutamate	2.36-2.33 (m)	0.132 (0.002)	0.246 (0.007)	***
Glutamine	2.21-2.06 (m)	1.900 (0.020)	1.800 (0.050)	ns
Glycine	3.56 (s)	0.384 (0.007)	0.277 (0.005)	***
Histidine	7.09 (s)	0.155 (0.002)	0.122 (0.003)	ns
Isoleucine	1.01-1.00 (d) 0.95-0.92 (t)	0.474 (0.003)	0.358 (0.006)	ns
Leucine	0.97-0.95 (t)	0.386 (0.003)	0.319 (0.012)	ns
Lysine	3.07-3.04 (m)	0.500 (0.002)	0.428 (0.006)	ns
Methionine	2.65-2.63 (t)	0.1018 (0.0005)	0.0902 (0.0012)	ns
Phenylalanine	7.44-7.43-7.41 (m)	0.281 (0.002)	0.225 (0.001)	ns
Threonine	3.59-3.58 (t)	0.700 (0.009)	0.583 (0.004)	ns
Tryptophan	7.74-7.72 (d)	0.064 (0.003)	0.059 (0.002)	ns
Tyrosine	7.20-7.18 (m or d)	0.286 (0.002)	0.220 (0.002)	ns
Valine	1.04-1.03 (d) 0.99-0.98 (d)	0.435 (0.005)	0.348 (0.009)	ns

Table 8. The quantification of the 22 identified metabolites in the medium conditional on the presence of *Pkd1* KO and the WT MEF cells. In the first column are listed the identified metabolites, the second column shows the chemical shifts of the peaks referred to DSS in which s=singlet, d=double, t=triplet, q=quartet ; m=multiplet denote the peak multiplicity; α and β indicates the two glucose anomers; the third and fourth columns show the quantification of the metabolites identified in the medium conditioned by WT and KO cells, respectively, the standard deviation is reported in parenthesis. The last column shows the T-test between KO and WT(* $p<0.05$; ** $p<0.001$; *** $p<0.001$). The metabolites in bold are significant in the T-test analysis.

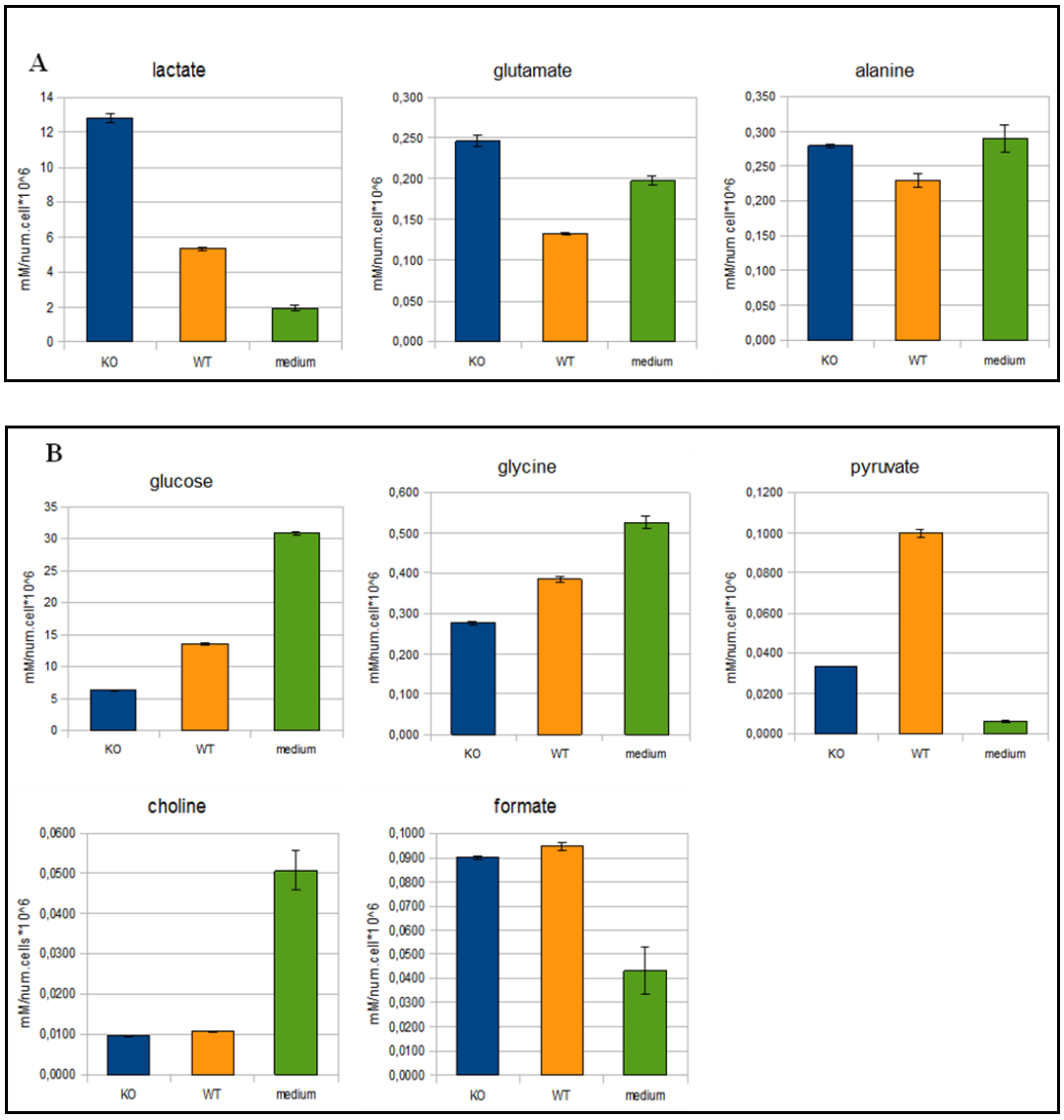


Figure 22. The concentrations of the metabolites that result significant forthe T-test analysis. The metabolites that present a major amount in the exometabolome conditioned by A. *Pkd1* KO cells, and B. WT ones.

Hence, considering the similarity between the cancer cells and the *Pkd1* KO cells, we also investigated the concentration of glutamine, another key metabolite of the cancer cells. The amount of glutamine in the medium conditioned by *Pkd1* KO cells does not display any difference compared to the medium conditioned with the WT cells [Figure 23], indicating that the metabolism of *Pkd1* KO cells share only some similarities with cancer cells.

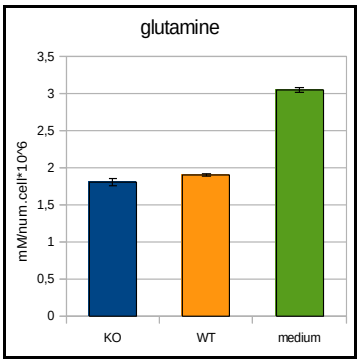


Figure 23. The concentration of glutamine measured in the medium conditioned by *Pkd1* KO cells. The blue bar, by WT ones, the orange bar and the medium alone, the green one.

5.2. *Pkd1* KO cells display the Warburg effect

The concentrations of the 22 metabolites are next used for an unsupervised statistical analysis, the principal component analysis (PCA), in order to highlight which metabolites contribute mostly to differentiate the two groups of cells [Figure 24]. The score plot displays a well defined difference between the two cell lines, showing a separation along the first principal component (PC1). The loading plot shows that glucose and lactate contribute mostly to separation in PC1. Glucose and lactate are anticorrelated, i.e. *Pkd1* KO cells show an increased production of lactate in the medium, whereas WT cells show a higher content of glucose in the medium. In other words *Pkd1* KO cells have a high consumption of glucose (decrease of glucose in the medium) and a high production of lactate, as compared to the WT ones.

Among the different metabolites glucose and lactate are the two molecules that mostly contribute to the difference between WT and *Pkd1* KO cells [Figure 24].

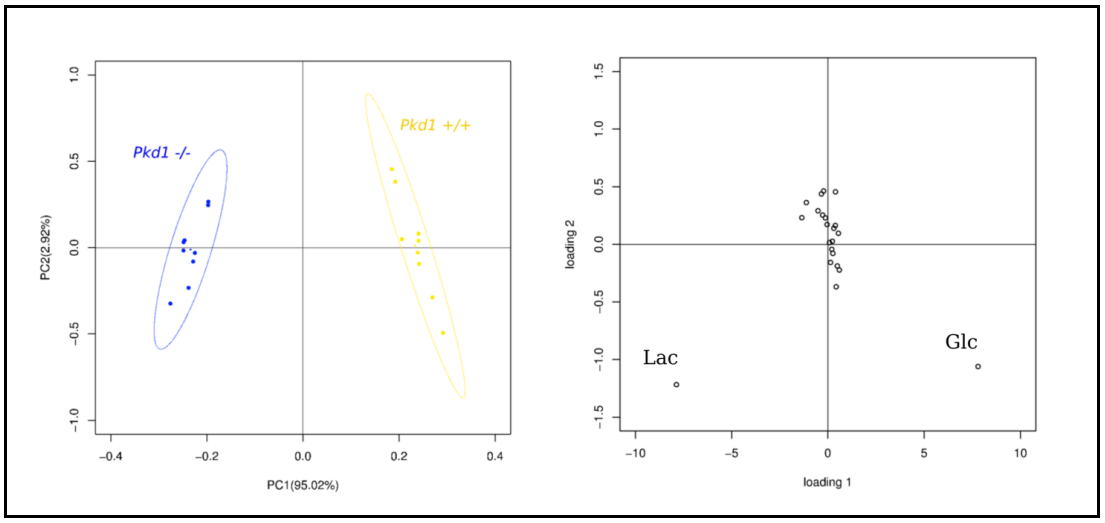


Figure 24. PCA analysis. On the left the score plot: the *Pkd1* KO and the WT cells are very well separated by the first component (PC1); on the right the loading plot: glucose and lactate contribute to the separation between the two cell lines in opposite manner.

From this result we have hypothesized that the *Pkd1* KO cells are affected by the aerobic glycolysis, the so called Warburg effect, the metabolic hallmark of cancer cells[74]. We hypothesize that the aerobic glycolysis derives from the loss of functional PC1. As the C- tail of PC1 regulates negatively mTORC1[41] it is conceivable that the loss of functional PC1 leads to an enhanced activity of mTORC1 that has as metabolic effect on the aerobic glycolysis. As the analyzed MEF cells derive from two different mice, we cannot exclude a priori that the observed metabolic change might to be attribute to a clone-to-clone variation effect, to exclude this possibility we have performed and analyzed the 1D ^{1}H NMR spectra on the medium conditioned by MEF cells isolated from a mouse model carrying Floxable alleles of the *Pkd1* gene. MEF cell lines (*Pkd1*$^{flox/flox}$) were isolated and immortalized. To achieve the inactivation of the *Pkd1* alleles without introducing clonal effects due to genomic integration of the Cre gene, a cell-permeable TATCre fusion protein capable of efficiently inducing the recombination of *loxP* sites and that is subsequently degraded by the cells[41] were used. Treatment with the Cre recombinase resulting in a minimum of 70 to 80% inactivation of the *Pkd1* allele, as assessed by PCR analysis of the genomic DNA. The resulting *Pkd1*$^{flox/flox}$ MEF population is not affected by clone-to-clone variability, but, at the same time, it is not totally knock-out for *Pkd1* gene.

In order to evaluate the consumption of glucose in the medium conditioned with the different cell types we have expressed the amount of glucose as consumption using the following equation:

$$Consumption = \frac{[glu\cos e\ medium] - [glu\cos e\ remaining\ cells]}{final\ number\ of\ cells} \qquad \text{eq.9}$$

We have calculated the glucose consumption and the lactate production also for the *Pkd1* KO , WT MEF cells, and *Pkd1* $^{flox/flox}$ MEF cells [Figure 25 A, B].

Taken together all these experiments clearly indicate that the Warburg effect is due to the loss of PC1.

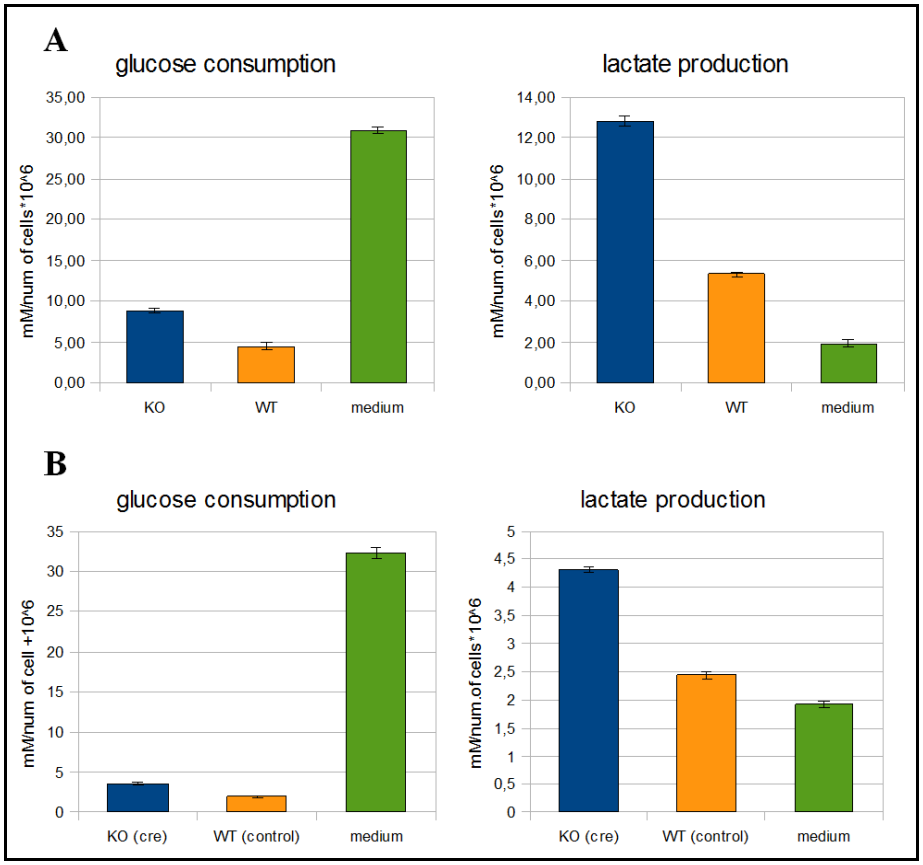


Figure 25. Glucose consumption and lactate production. A *Pkd1* KO and WT MEF cells and B *Pkd1*$^{flox/flox}$ MEF cells.

5.3. Additional experiments in support of metabolic dysfunction in *Pkd1* KO cells

The metabolomic analysis on the different cellular models reveals that the loss of PC1 leads to the aerobic glycolysis as *Pkd1* KO MEF cells show a major consumption of glucose and major production of lactate. To support these data and to study the molecular mechanism associated to the loss of PC1, our collaborators have performed additional experiments, that have been briefly summarized in the following.

5.3.1 Demonstration of Warburg effect

As expected the *Pkd1* KO MEF cells display more ATP content compared to the WT ones [Figure 26 A]. Treating cells with oligomycin, a drug able to block the ATP-synthase and hence the production of ATP through oxidative phosphorylation, the intracellular amount of ATP decreases in both cell lines, but importantly the difference in ATP content between the *Pkd1* KO and WT cells is not restored [Figure 26 B]. This result confirms that the *Pkd1* KO cells are affected by Warburg effect.

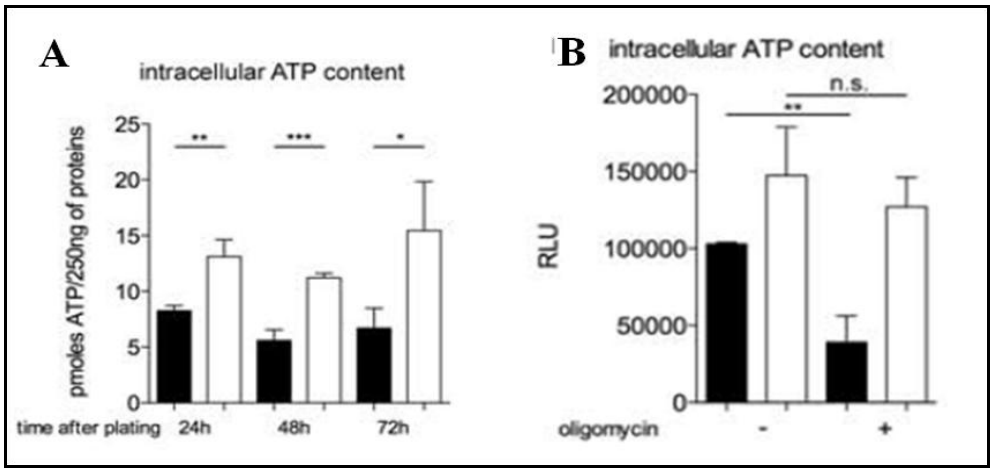


Figure 26. The quantification of the amount of ATP. A. The amount of ATP measured 24, 48 and 72 hours respectively after cellular plating. B. The amount of ATP measured in the absence or in presence of oligomycin, respectively. Results deriving from *Pkd1* KO and WT cells are shown in white and black bars, respectively.

Moreover, real-time PCR analysis reveals that *Pkd1* KO cells display a transcriptional signature of glycolytic enzymes [Figure 27], showing an increase of the transcript of lactate dehydrogenase A (LDHA) and pyruvate kinase M2 (PKM2). Whereas the transcript of the hexokinase I (HKI), that phosphorylates the glucose, does not display a significant difference.

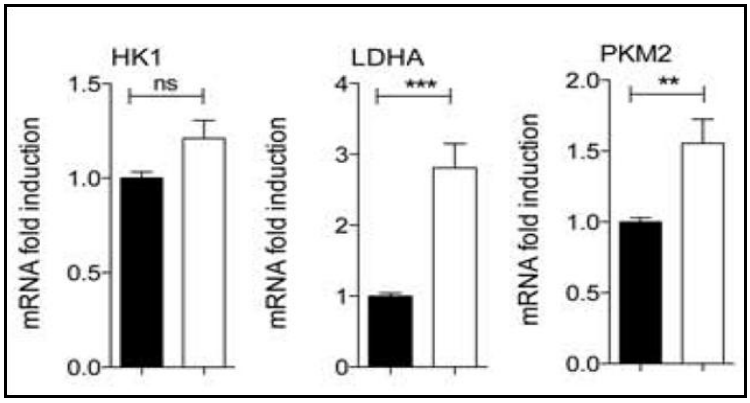


Figure 27. The transcription signature of *Pkd1* KO (white bar) and WT (black bar) cells.

In addition AMPK is, at the same time, inhibited [Figure 28 A] in *Pkd1* KO cells by the major intracellular content of ATP and by LKB1[75], that is regulated by ERKs pathway [Figure 28 B].

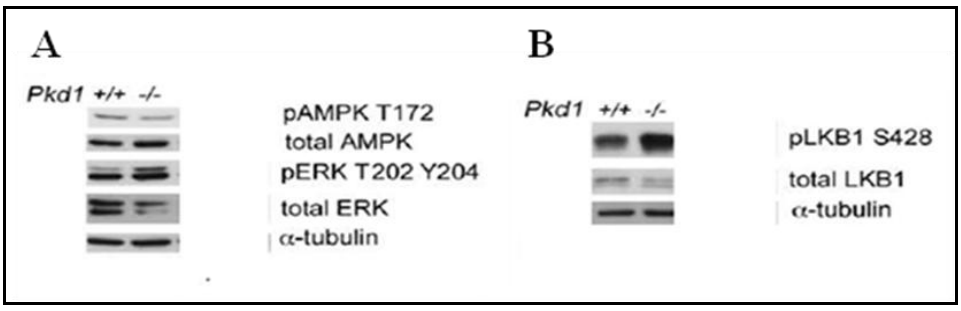


Figure 28. AMPK is inhibited by the loss of functional PC1. A. The *Pkd1* KO cells show a decrease of phosphorylation of AMPK and a related increase of the phosphorylation of ERK. B. The increase of phosphorylated LKB1 in *Pkd1* KO cells.

5.3.2 Proliferation and apoptosis in *Pkd1* KO cells depend on glucose metabolism

Pkd1 KO cells present an increased proliferation and apoptosis[76,77] which are dependent on mTORC1[41] and on glucose metabolism. Indeed glucose deprivation reduces proliferation [Figure 29 A] and sensitizes *Pkd1* KO cells to apoptosis [Figure 29 B].

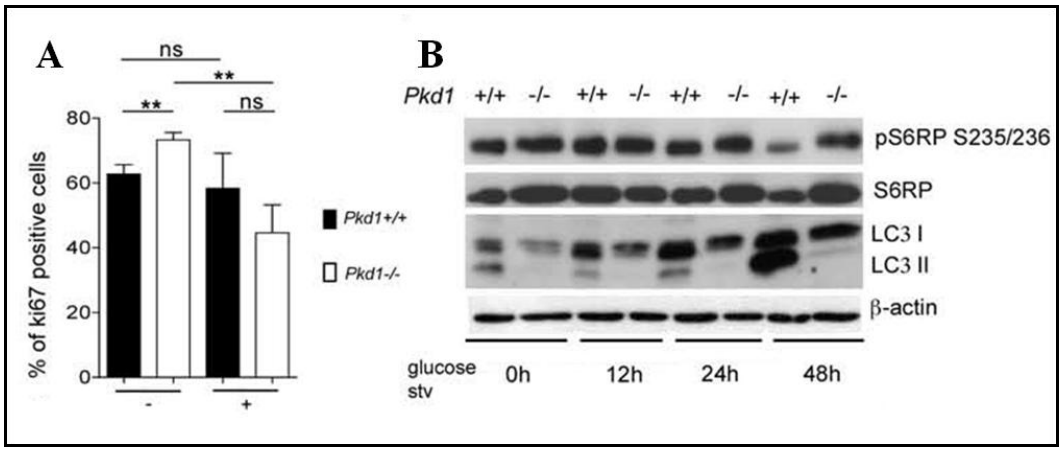


Figure 29. **The role of glucose starvation in** A. proliferation rate measured using Ki-67 protein as marker and B. in the activation of autophagy. S6RP is the S6 ribosomal protein, a downstream target of mTOR pathway, its major phosphorylation indicates an increased activity of mTOR. LC3 is a marker of autophagy and in particular the amount of LC3 II correlates well with the number of autophagosomes and it is used to monitor the autophagic activity of the cells.

5.4. Treatment with drugs

5.4.1 Exometabolome of MEF cells treated with drugs

The up regulation of ERK pathway due to the loss of PC1 in *Pkd1* KO cells leads to an enhanced activity of mTORC1 that induces the aerobic glycolysis and a production of ATP that inhibits AMPK. On the other hand the up regulation of ERK pathway phosphorylates LKB1, activating it and consequently inhibiting AMPK.

Each step of this proposed signaling cascade can be targeted by a drug, approved by FDA. For example AMPK is activated by metformin[78] and AICAR[79], mTORC1 activity is inhibited by rapamycin[42] [Figure 30].

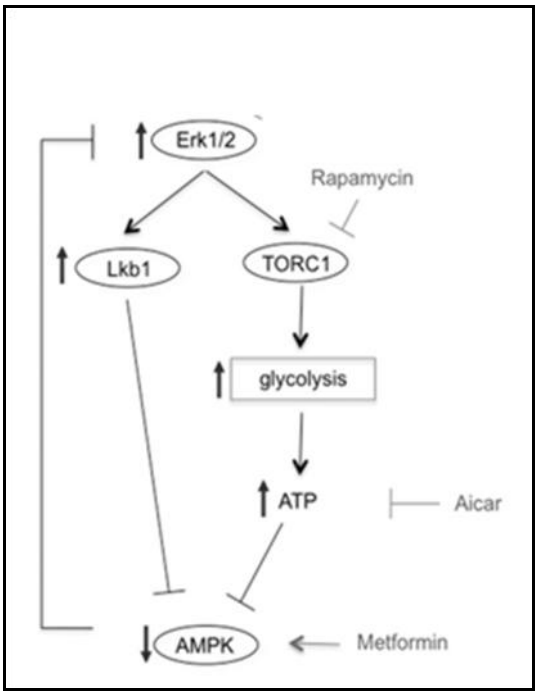


Figure 30. The scheme of the proposed signaling cascade associated to the loss of PC1. In Figure are highlighted the tested drugs and their relative targets.

To investigate the effects of each drug on the glucose metabolism, different experiments have been performed. MEF cells were treated for 48 hours with rapamycin, AICAR and metformin, singularly. The medium is then collected to perform 1D ^{1}H NMR experiments and then we have quantified and compared the consumption of glucose [Figure 31 A] and the production of lactate [Figure 31 B] for *Pkd1* KO and WT cells not treated or treated with each drug singularly. The amount of glucose consumption and lactate production of WT cells does not display any changes upon treatment with different drugs. The *Pkd1* KO cells do not revert their metabolic phenotype after drug treatment as they maintain an increase of glucose consumption and lactate production. This effect is emphasized principally by treatment with metformin, as expected because the lactic acidosis is one of the side effect of the metformin[78].

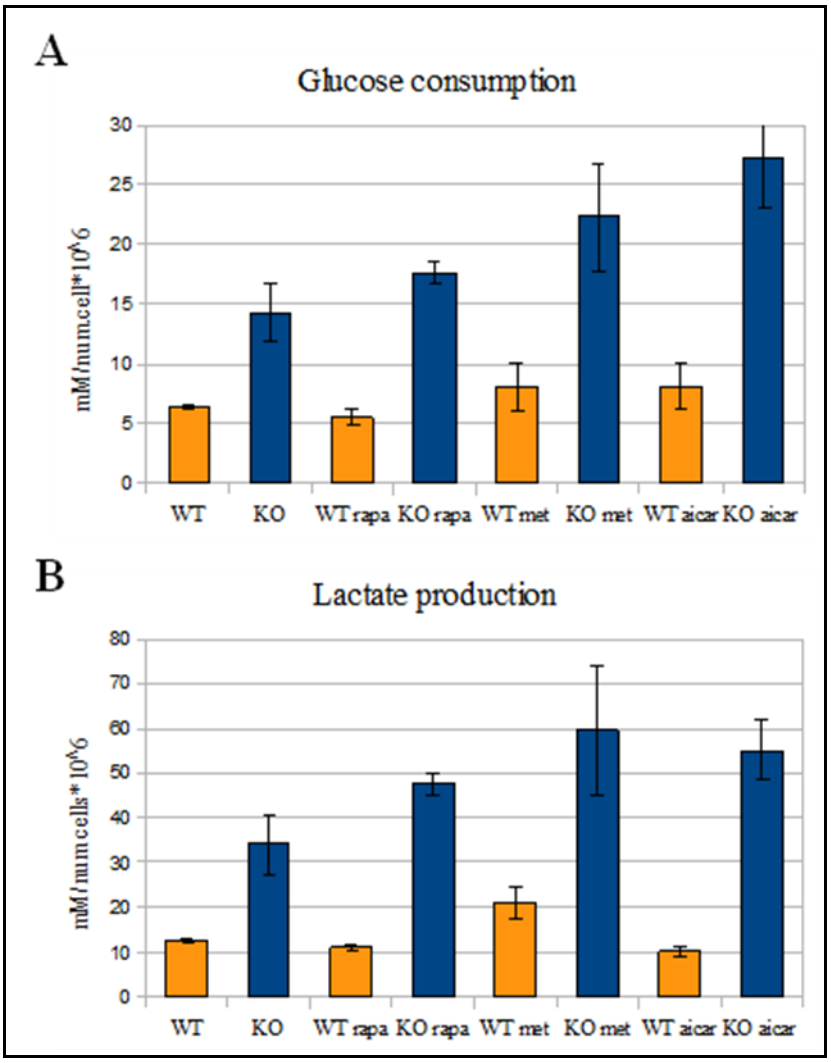


Figure 31. Comparison between not treated and treated *Pkd1* KO and WT MEF cells. A. Glucose consumption and B. the lactate production of WT (orange) and *Pkd1* KO cells (blue) in different condition: starting from left, not treated, treated with rapamycin (rapa), with metformin (met) and treated with AICAR.

Also cells treated with AICAR display the same phenotype of the cells treated with metformin, even though lactic acidosis is not known as side effect of the treatment with AICAR. Overall the treatment with AICAR results very toxic for cells indeed they show an high death rate.

5.4.2 Additional experiments to test the effects of the drugs on the signaling cascade and the related biological effects.

The treatment with AICAR, metformin or rapamycin is not able to revert the metabolic phenotype of the *Pkd1* KO cells. Our collaborators have therefore performed additional experiments *in vitro* in order to verify the effect of these drugs on the signaling cascade and to the related biological effect.

Treating *Pkd1* KO cells with rapamycin the autophagy [Figure 32 A] and cell survival under glucose deprivation are partially restored [Figure 32 B].

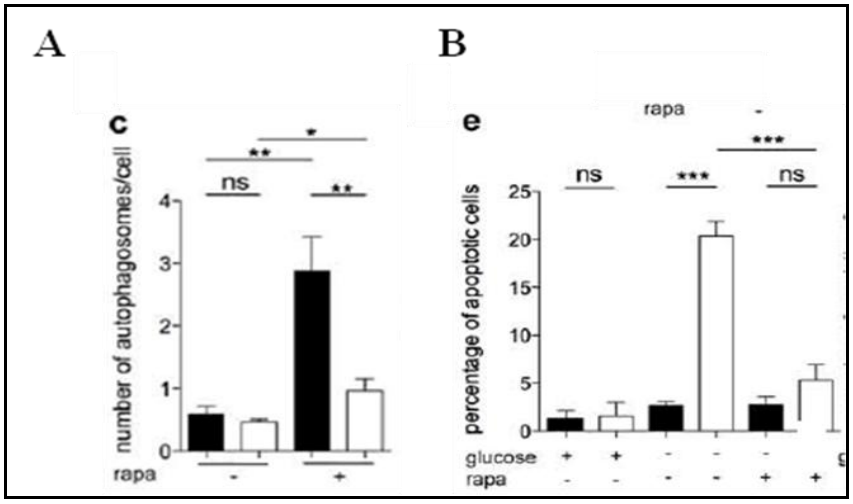


Figure 32. Effect of the treatment with rapamycin.A. Rapamycin partially restores the autophagy in *Pkd1* KO cells (white bar) B. Rapamycin decreases the apoptotic level of *Pkd1* KO cells under glucose starvation conditions.

Treatment with metformin and AICAR [Figure 33 A, B] activates AMPK and, unexpectedly, inhibits ERKs pathway.

These data suggest the existence of a negative feed-back whereby AMPK can regulate ERKs activity. Thus, the signaling cascade related to the loss of PC1 is possibly regulated in a circular manner suggesting that acting at any level of the cascade might restore the basal condition [Figure 34].

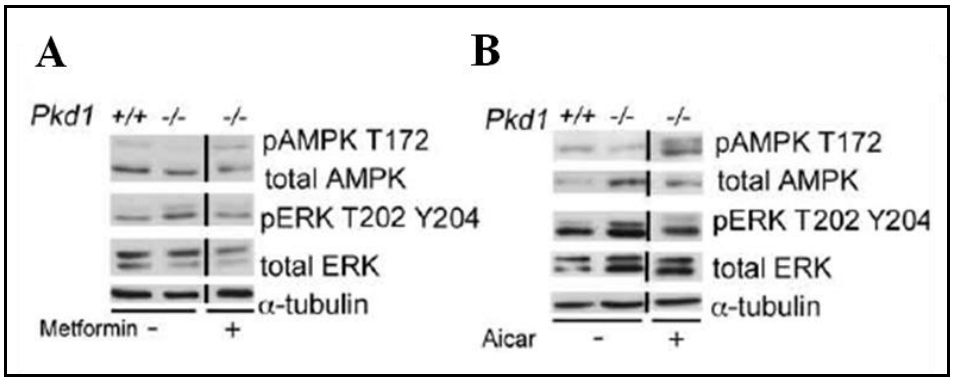


Figure 33. Effect of the treatment with metformin and AICAR. A. The effect of metformin and B. AICAR on AMPK and the ERKs pathway.

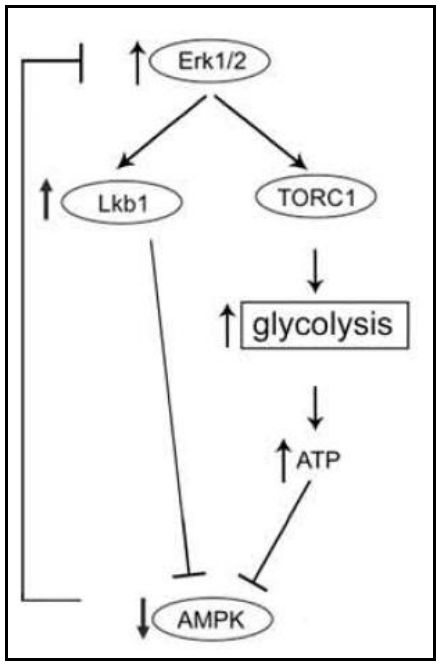


Figure 34. A proposed model of signaling cascade due to the loss of PC1.

5.5. *In vivo* experiments

In order to verify that the Warburg effect occurs also *in vivo* we have performed NMR experiments on the polar extract of the whole kidney deriving from of Ksp-Cre: $Pkd1^{flox/-}$ and $Pkd1^{flox/+}$ mouse model[12] in which PC1 is inactivated only in kidneys.

The $Pkd1^{flox/-}$ KO kidneys and the $Pkd1^{flox/+}$ WT kidneys are quite different: the KO kidney is bigger than the WT one and it weights about ten times (340 mg) more than the WT one (34 mg) [Figure 35 A]. After lyophilization the weight of the KO kidney decreases of about ten

times, whereas the weight of the WT kidneys decreases of about three fold compared to the weight of the corresponding fresh tissue [Figure 35 B], highlighting that the increase of renal weight is due to the fluid contained in the cysts. The lyophilized KO kidney is more fragile as compared to the WT one because of the loss of the renal architecture.

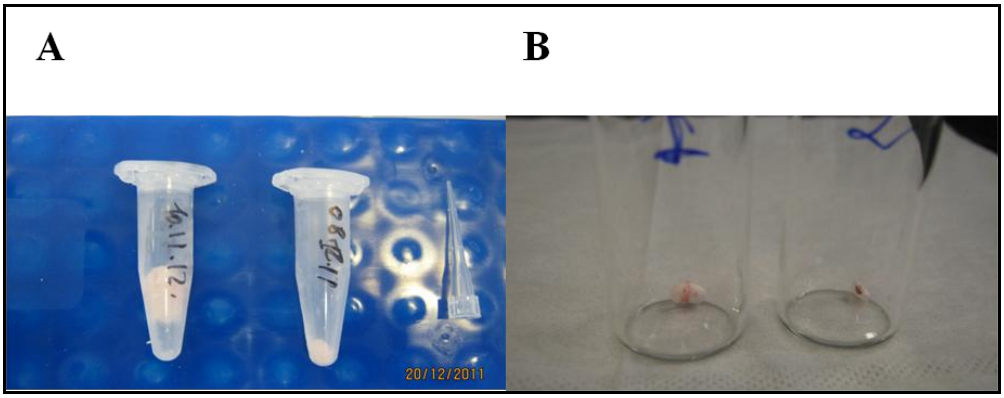


Figure 35. The mouse kidneys A before lyophilization and B. after lyophilization**.** We can appreciate that the KO kidney *Pkd1*$^{flox/-}$, the left one in both pictures, is bigger than the WT kidney *Pkd1*$^{flox/+}$, on the right.

5.5.1 1D ^{1}H NMR experiments of polar extracts deriving from lyophilized kidneys

We have analyzed the polar metabolites extracted from three WT and three KO kidneys deriving from three different litters. We have performed 1D ^{1}H NMR experiments [Figure 36] and analyzed the content of glucose and lactate.

The amount of these two metabolites is expressed as the ratio integral/dry weight to account for the difference in weight of the two kidneys. The KO kidneys show more glucose [Figure 37 A] as compared to the WT. However the measured lactate content [Figure 37 B] is affected by a high error. Because of these high errors this experiment was judged insufficient to confirm that the Warburg effect occurs also *in vivo*. From these experiments we can only hypothesize that the KO kidneys have more glucose content than WT ones.

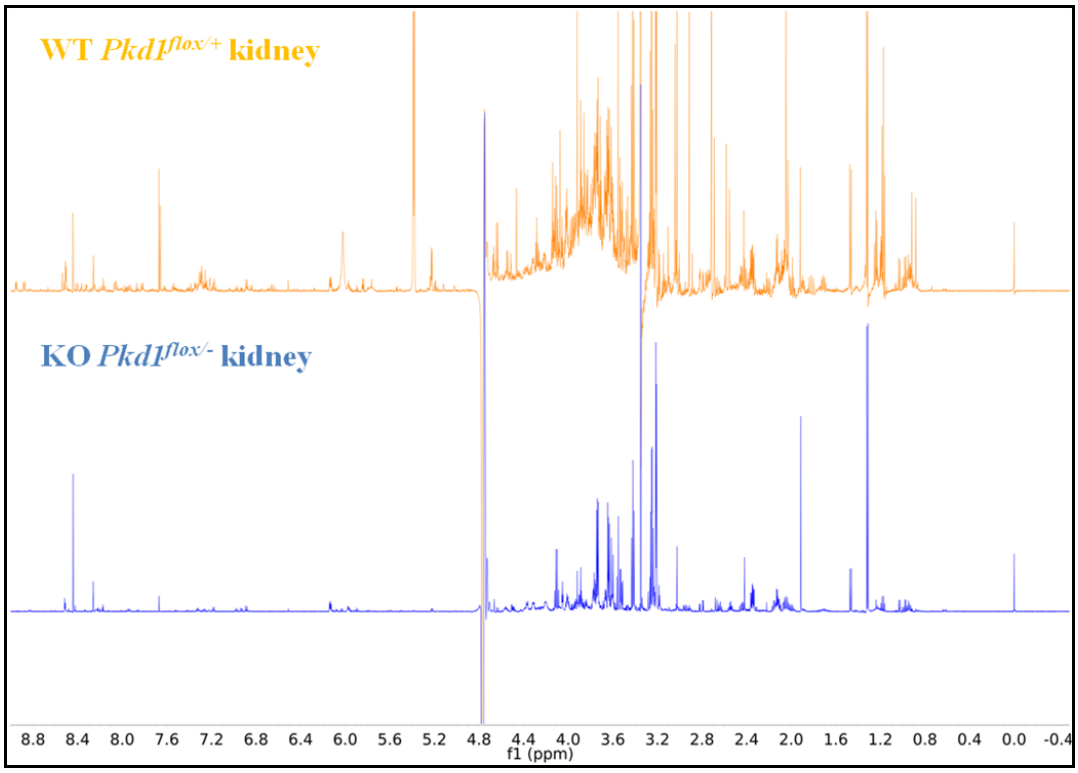


Figure 36. The representative kidney 1D ^{1}H NMR spectra. *Pkd1*$^{flox/+}$ WT kidney (top) and *Pkd1*$^{flox/-}$ KO kidney (bottom).

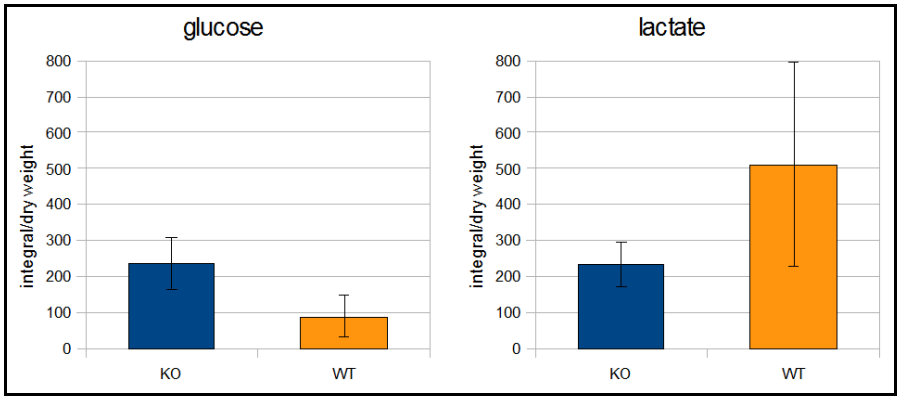
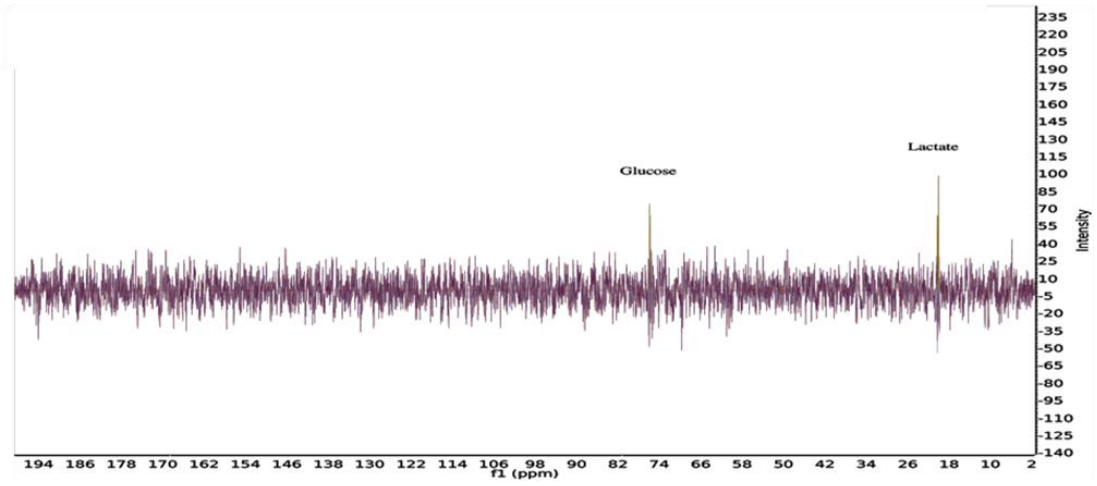


Figure 37. The amount of A. glucose and B. lactate in kidneys. The blue bar represents the KO kidneys and the orange one the WT ones.

5.5.2 ^{13}C glucose experiments

The analysis of the 1D ^{1}H NMR spectra acquired on the polar extracts of the whole kidney does not confirm the Warburg effect *in vivo.* The high error that affects the measurements is probably related to the genetic variability between the litters. Hence, following a protocol previously reported in literature[13,14] we have decided to perform experiments using uniformly labeled ^{13}C glucose in order to follow the path of glucose and its conversion into lactate. We have performed a preliminary kinetic experiment in order to set the best experimental conditions, including: the amount of glucose to inject, the time point in which to sacrifice the mouse, the type of spectra to acquire and how to analyze the data. In order to follow the glucose pathway and its conversion into lactate we have collected and analyzed serum, urine, kidneys and liver, as control organ.

We have first tried to acquire the 1D ^{13}C spectra [Figure 38]. Unfortunately, the 1D ^{13}C spectra could not be used because the glucose and lactate signals of the WT kidneys were comparable to the noise signal, their intensities were affected by big errors and could not be used for quantification [Figure 38 top]. To increase sensitivity we have acquired ^{1}H ^{13}C HSQC on the same samples and we have used the corresponding peaks to analyze the glucose and lactate content.



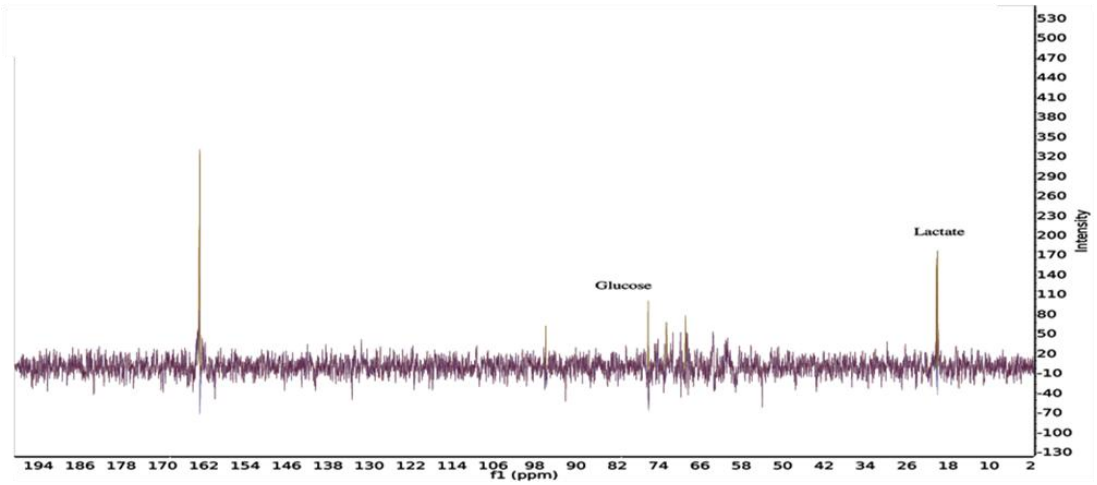


Figure 38: Representative kidney 1D ^{13}C spectra. WT (top) and KO (bottom) of 1D ^{13}C kidney spectrum collected after 40 minutes from ^{13}C glucose injection.

Kinetic experiments. For choosing the time point in which to sacrifice the mice, we have performed a kinetic experiment following the amount of ^{13}C glucose and ^{13}C lactate in serum, urine, livers (organ control) and kidneys at three different time points.

We used a litter composed by 8 littermates: 4 KO and 4 WT mice. Mice were sacrificed after 20, 40 and 60 minutes after ^{13}C-glucose injection (after 40 minutes 2 mice were sacrificed to have a replicate of this time point).

We have acquired ^{13}C-^{1}H HSQC experiments on serum, urines, kidneys and livers and we have measured the concentration of glucose and lactate (see paragraph 4.1.4).

Lactate was not present in urine and serum, therefore in biofluids we could only detect the glucose amount [Figure 39]. The urine deriving from KO mice are more diluted as compared to WT urines because of the loss of renal function, consistently the amount of glucose is very small as compared to WT. Monitoring the decrease of the amount of glucose in the bloodstream during the time we can appreciate that the KO mice have a quicker glucose consumption as compared to the WT ones. This effect could be an indication of the accelerated metabolism that affects the KO mice.

In kidneys and livers we have detected both glucose and lactate. The KO organs show more glucose and lactate and the difference is more relevant 40 minutes after ^{13}C glucose injection. In detail, the difference in the content of glucose and lactate is significant in the kidneys but not in livers [Figure 40 A and B].

Hence we have decided to perform the following experiments sacrificing the mice 40 minutes after ^{13}C glucose injection.

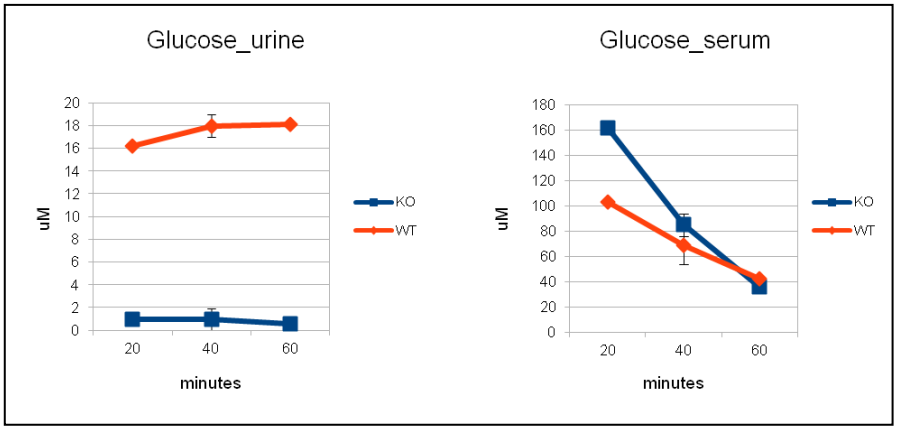


Figure 39. Glucose quantification. The concentration of glucose in urine (on the left) and in serum (on the right).

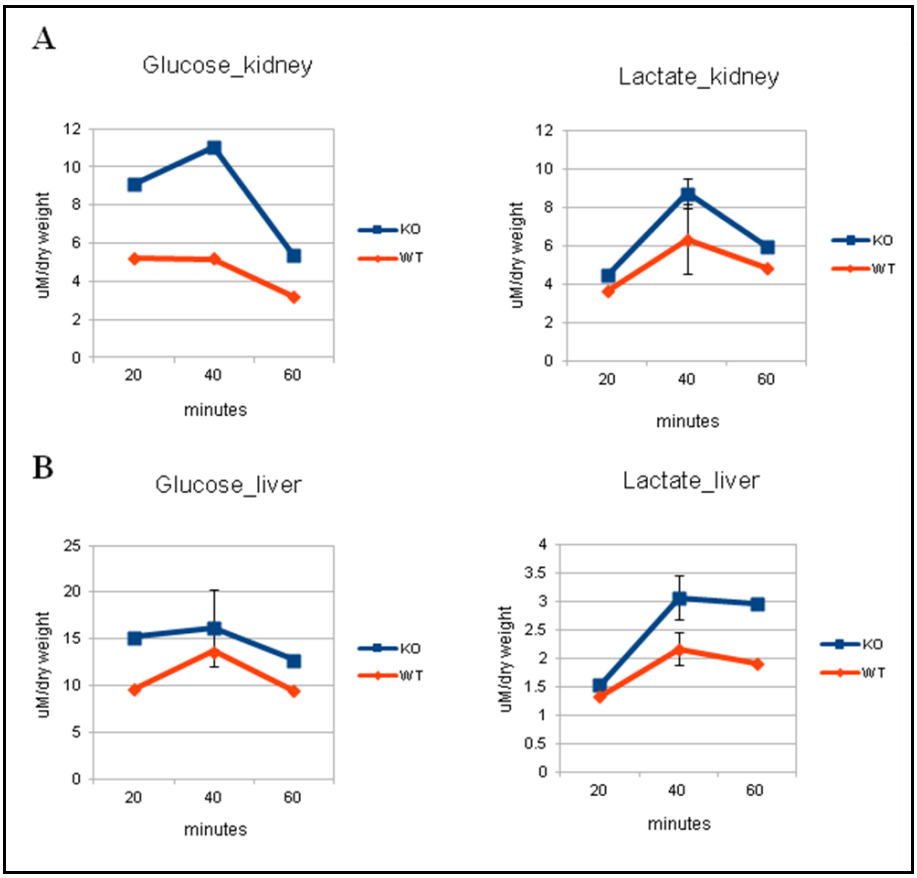


Figure 40. Glucose and lactate quantification. The measured concentration of glucose and lactate in A. kidneys and B. liver.

^{13}C glucose experiments after 40 minutes from glucose injection. Based on the kinetic experiments we have decided to perform all the following experiments only on the kidneys and livers collected after 40 minutes from glucose injection.

To remove the differences due to the genetic variability between the five litters used for the experiments and to appreciate only the difference due to the loss of PC1, the WT of each litter was used as internal standard, expressing the amount of glucose and lactate of the KO kidney as ratio with the corresponding amount of WT kidney ($R=I_{KO}/I_{WT}$). Next, we have calculated the average and the standard deviation both for kidneys and for livers (see paragraph 4.1.4).

Analyzing the ^{1}H ^{13}C HSQC spectra performed on the polar extracts of the kidneys we have confirmed that the Warburg effect occurs also *in vivo* and it is caused by the loss of PC1. Indeed the *Pkd1* KO kidneys contain more glucose [Figure 41 A] and lactate [Figure 41 B) than the WT ones, whereas the *Pkd1* KO livers do not show significant differences in glucose and lactate content.

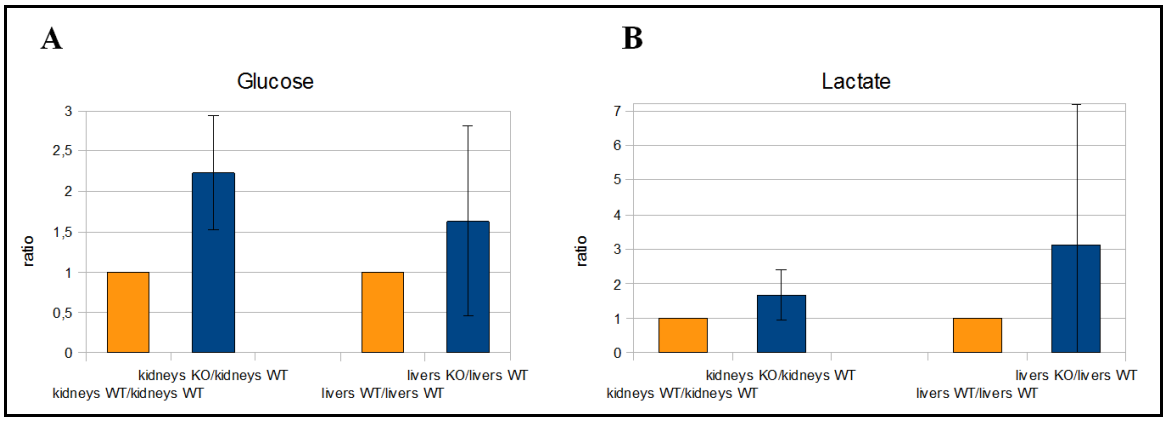


Figure 41. Glucose and Lactate quantification A. The amount of glucose in kidneys on the right and liver on the left. B. The lactate content in kidneys on the right and liver on the left. In orange the ratio between the WT organs and in blue the ratio between KO and WT organs.

To support these data our collaborators have measured the renal ATP content [Figure 42 A] confirming that KO kidneys have more ATP than the WT ones, and through the real time PCR analysis they have confirmed that the up regulation of glycolytic enzymes occurs also in kidneys showing an increase of the transcripts of hexokinase I (HK I) and pyruvate kinase (PMK2), whereas the transcript of the lactate dehydrogenase A (LDHA) does not display a significant difference [Figure 42 B].

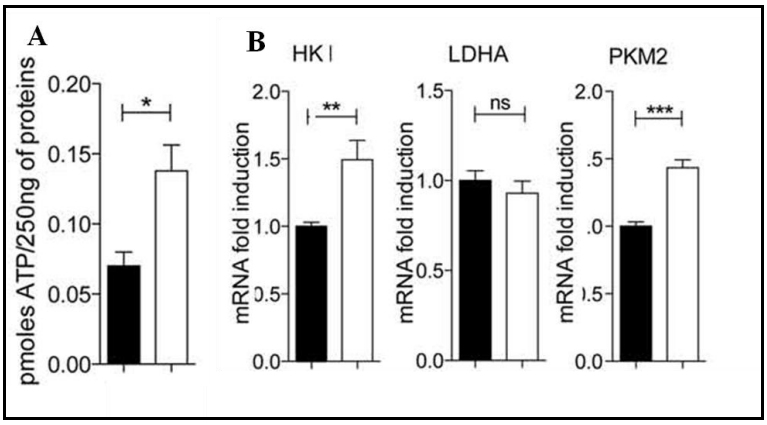


Figure 42. The Warburg effect occurs also *in vivo.* A. The ATP content in kidneys. B. The real time PCR analysis on the glycolytic enzymes.

Through Western blot analysis [Figure 43] they have also verified that the up regulation of ERKs pathway occur also in kidneys with the related enhanced activity of mTORC1 and the inhibited activity of AMPK, as previously demonstrated in the cellular model [Figure 34].

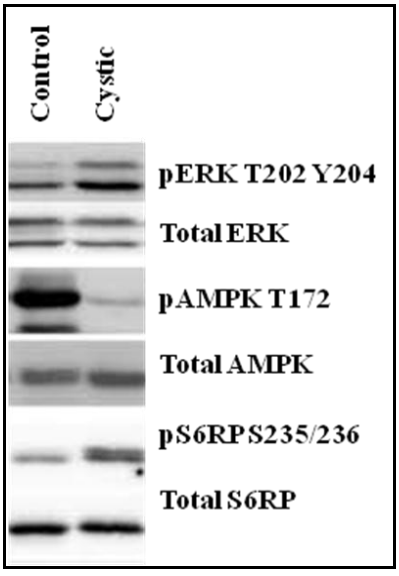


Figure 43. The proposed signaling cascade occurs also *in vivo.* Western blot analysis confirms that the hypothetic pathway discovered in cellular model occurs *in vivo.*

5.6. Treatment with 2-deoxy glucose (2DG)

Through ^{13}C glucose NMR experiments we have demonstrated that the Warburg effect and the related signaling cascade associated to the loss of PC1 also occurs *in vivo*.

We next wondered whether it was possible to block the Warburg effect blocking glucose metabolism using 2-deoxy-glucose (2DG).

2DG is a drug able to block the glucose metabolism being an analog of the glucose that cannot be metabolized [Figure 44 A]. It is a competitor of the glucose and it is phosphorylated by the hexokinase 1 but it is not metabolized by the next enzyme involved in the metabolism of glucose, for example the phosphoglucose isomerase for glycolysis, and the glucose 6-phosphate dehydrogenase for the pentose phosphate pathway [Figure 46 B]. Generally, cells treated with 2-DG increase the number of glucose transporters on the cellular membrane in order to enhance the uptake of glucose to survive[15,16]. The treatment leads to an increase in the intracellular glucose content and a decrease of the metabolic products of glucose In order to observe changes in the glucose metabolism upon treatment with 2DG we have analyzed the polar extracts of kidneys and livers deriving from three different litters treated at the same time with 2-deoxy-glucose (2DG) and uniformly labeled ^{13}C glucose performing the ^{1}H ^{13}C HSQC NMR experiments.

In order to check the efficacy of the treatment independently from the genetic variability we have used the untreated *Pkd1* KO organs as internal standard ($R_{2DG}= I_{KO\_2DG}/I_{KO}$) (see paragraph 4.1.4). Notably the treatment with 2DG is effective only in kidneys and not in the liver (control organ). 2DG does not affect the glucose content but only the lactate content that decreases in kidneys whereas the livers does not appear to be affected by the treatment [Figure 45 A and B].

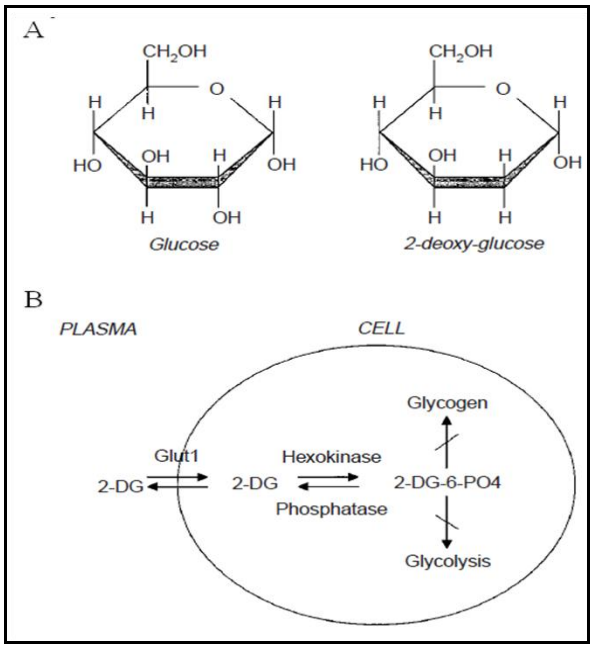


Figure 44. Structure and metabolism of 2DG. A. Structure comparison of glucose and 2-deoxy- glucose. 2DG and glucose differ in the second carbon. B. Schematic representation of 2-DG action. 2DG enters in the cell through the glucose transporter and it is phosphorylated by hexokinase. Due to low levels of intracellular phosphatase, 2-DG-PO_4 is trapped inside the cell. 2-DG-PO_4 is unable to undergo further metabolism. High intracellular levels of 2-DG-PO_4 cause allosteric and competitive inhibition of hexokinase. This results in inhibition of glucose metabolism. [Figure from Aft RL et al. British Journal of Cancer, 2002].

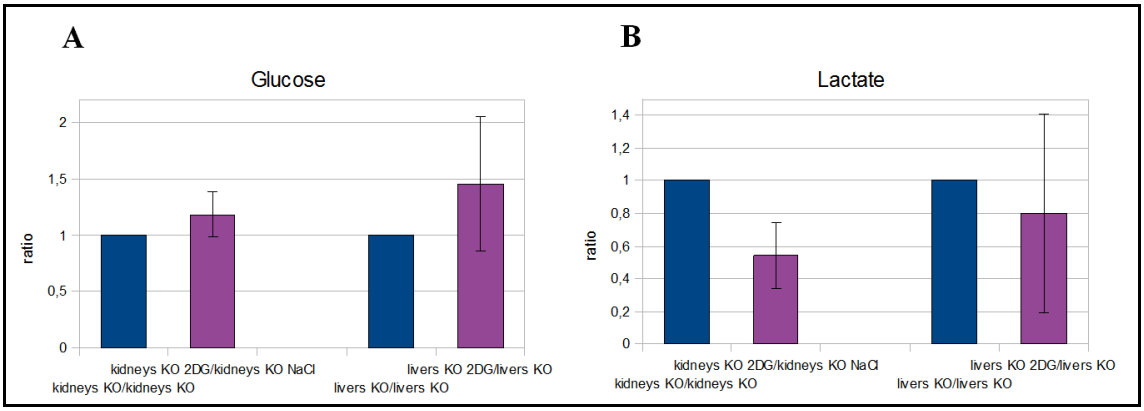


Figure 45. Quantification of glucose and lactate. A. The glucose content for kidneys, on the right and liver, on the left. B. the lactate content for kidneys and liver. In blue the ratio between the KO organs not treated with 2DG, in violet the ratio between the KO not treated organs with the KO organs treated with 2DG.

In addition our collaborators have highlighted that the treatment with 2DG improves the histology of the kidneys, as they present a decrease in cystic volume in the weight of the organ [Figure 46 A] and in the index of proliferation of cystic lining epithelium [Figure 46 B].

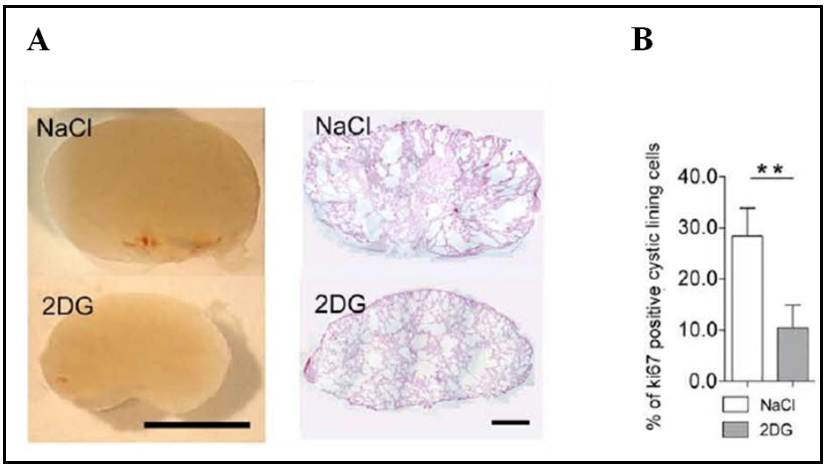


Figure 46. Effect of 2DG treatment on the renal histology. A. The histology of KO kidneys not treated (top) and treated with 2DG (bottom). B. The measure of the proliferation index.

6. Structure of the N-terminus of NPHP1

6.1 The recombinant N-terminus of NPHP1 is soluble.

To study the putative coiled coil domain of the human NPHP1 (spanning the first 115 amino acids of the NPHP1) the corresponding DNA sequence was cloned into pETM44 and the recombinant domain was expressed in *E. Coli* and purified as described in paragraph 5.1. The construct has the following sequence, where the amino acids belonging to the tag and that remain after the protease cleavage are highlighted in italics:

*GPMA*MLARRQRDPLQALRRRNQELKQQVDSLLSESQLKEALEPNKRQHIYQRCIQLK QAIDENKNALQKLSKADESAPVANYNQRKEEEHTLLDKLTQQLQGLAVTISRENITEV GAPT

The recombinant protein was soluble and folded, as shown by the good peak dispersion of the 1D and 2D NMR spectra [Figure 47] moreover the sharp peaks of 1D NMR spectrum showed that the protein was also monomeric.

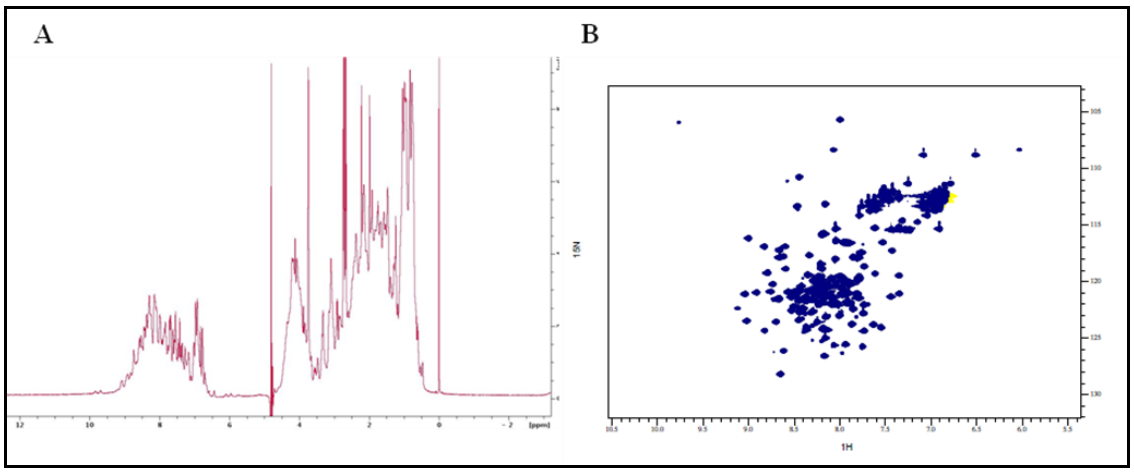


Figure 47. NMR spectra of the putative coiled coil domain of NPHP1. A. The 1D ^{1}H NMR spectrum and B. the 2D ^{1}H-^{15}N HSQC of the putative coiled coil domain of NPHP1 concentrated 0.4 mM, recorded at 298K in PBS 20mM, 3mM DTT.

6.2 The N-terminus of NPHP1 is predicted to be a coiled coil domain

The first 115 amino acids (13kDa) of NPHP1 were predicted to be a coiled domain by the Web Server PCOIL [Figure 48]. The coiled coil domain is a structural motif in which two or more alpha helices are coiled together. It is characterized by the repetition (3-4 times) of 7 amino acids, whose positions are defined as a, b, c, d, e, f, g. In particular, position a and d are occupied by hydrophobic amino acids, e and g by charged amino acids, these positions are important for structure stabilization[18]. This kind of domain is usually involved in protein-protein interaction[18,24,25,80]. In particular COILS predicts for the N-terminus of NPHP1 the presence of a coiled coil from residues Leu 14-Glu 34; Ile 54-Glu75 and Asn 83-Ala 104.
Surprisingly, the recombinant protein does not show any sign of aggregation and dimerization as assessed form gel filtration elution volume, which corresponds to the volume associated to a protein of 13kDa. Moreover 1D NMR 11-echo experiments indicate that the overall correlation time of the protein is 15 ms, which is in line with a molecular weight of 13 kDa , being in monomeric for in solution.

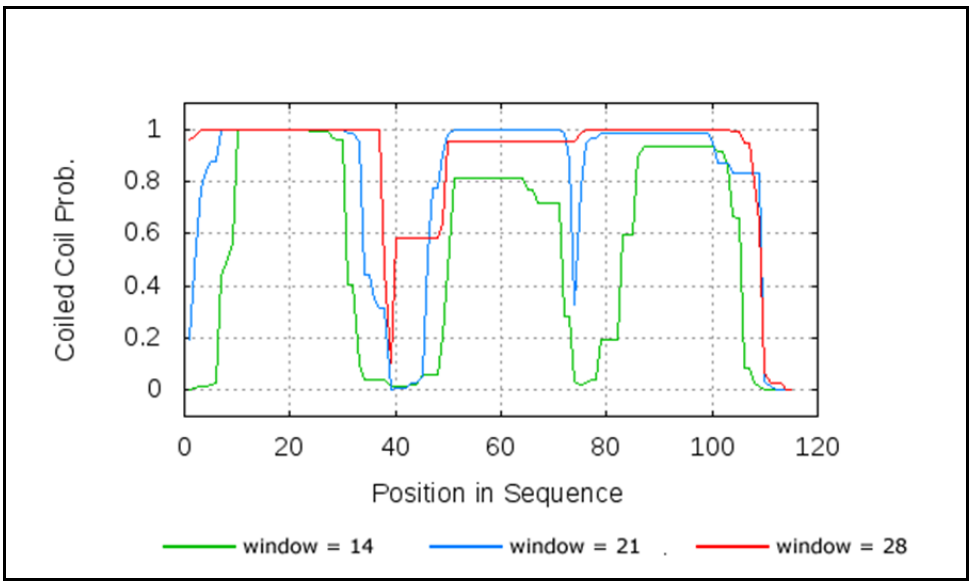


Figure 48. COILS web server prediction. COILS compares the sequence with a database of known coiled coils and derives a similarity score. Then the program calculates the probability that the sequence will adopt a coiled coil conformation. The graph show the coiled coil probability over the sequence. COILS uses sliding windows of 28 (red), 21 (blue) and 14 (green) amino acids. The most accurate prediction is for window of 21 amino acids.

6.3 The N-terminus of NPHP1 is thermostable and contains α-helices

We have performed circular dichroism (CD) experiments on the N-terminus of NPHP1 ($NPHP1_{1-115}$) in order to characterize the secondary structure of this domain and its thermal stability.

The CD is a very useful technique to study the protein folding and to characterize its secondary structure, to determine whether protein-protein or protein-ligand interactions alter the conformation of protein and to investigate the conformational stability of a protein in denaturing conditions (such as temperature, pH and denaturants). This technique is based on the measurement of the differences in the absorption of left-handed polarized light versus right-handed polarized light which arises from structural asymmetry in the "far-UV" spectral region (190-250 nm). At these wavelengths the chromophore is the peptide bond, and the signal arises when it is located in a regular, folded environment. Alpha-helix, beta-sheet, and random coil structures each give rise to a characteristic shape and magnitude of CD spectrum.

The CD spectrum of the N-terminus of NPHP1 shows a typical α helix profile [Figure 49] with two minima at 209 nm and 222 nm and a maximum at 192 nm.

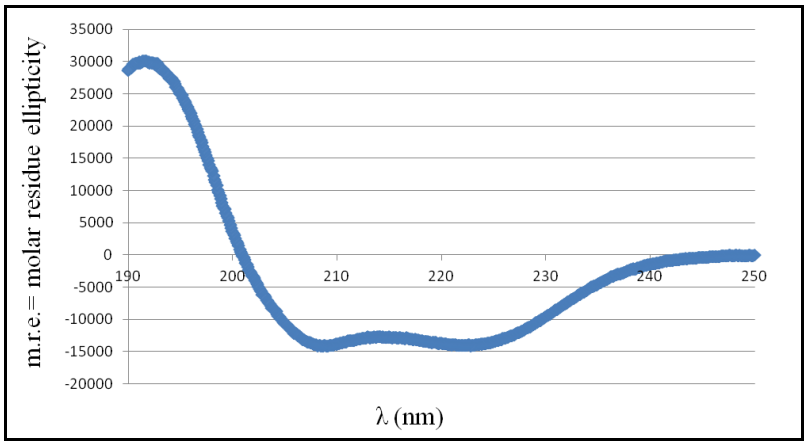


Figure 49. The CD spectrum of the N-terminus of NPHP1. CD experiment was performed using 20μM of protein in solution with 20mM phosphate buffer, 150 mM NaF and 0.3 mM DTT.

This domain contains ~ 45% of α helix, as estimated from the molar residual ellipticities value at 209 nm and 222 nm (-14000 deg mol^{-1} cm^{-2}) using the following equation[81]:

$$\%alphahelix = \left(-[\theta]_{222nm} + 3000\right)/39000 \qquad \text{eq.10}$$

Finally, the melting curve obtained following the changes of the molar ellipticities value at 209 nm shows that the domain is extremely stable with a high melting temperature of ~ 65°C [Figure 50].

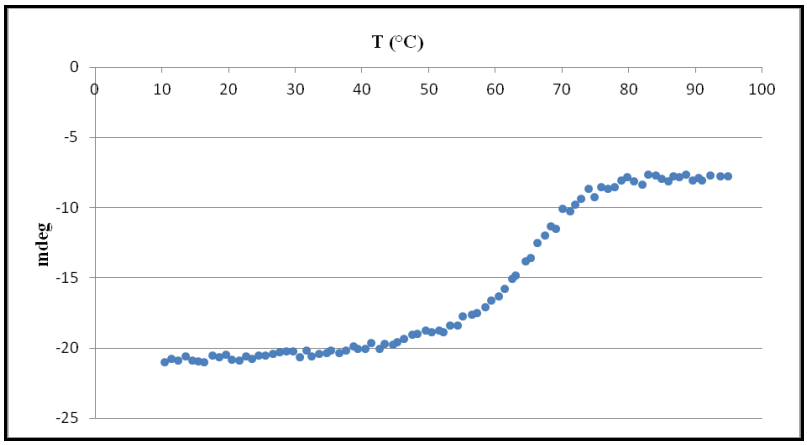


Figure 50. The melting curve of the putative coiled coil domain of NPHP1. CD experiment was performed using 20µM of protein in solution with 20mM phosphate buffer, 150 mM NaF and 0.3 mM DTT.

Taken together these results show that the N-terminal domain of NPHP1 has a high α helical content, as secondary structure, and it is thermostable.

6.4. NMR experiments

We have solved the structure of $NPHP1_{1\text{-}115}$ using classical multidimensional- heteronuclear NMR experiments (see paragraph 4.2.5)

6.4.1 NPHP1$_{1\text{-}115}$ chemical shift assignment.

^{1}H, ^{15}N and ^{13}C resonances of the NPHP1$_{1\text{-}115}$ were assigned through 2D and 3D homo- and heteronuclear experiments (as described in paragraph 4.2.5 and appendix). We started with the peak picking of the ^{1}H-^{15}N HSQC spectrum and analyzing the HN strips in the 3D HNCA, HNCACB and HN(CO)CBCA in order to assign the backbone resonances [Figure 51].

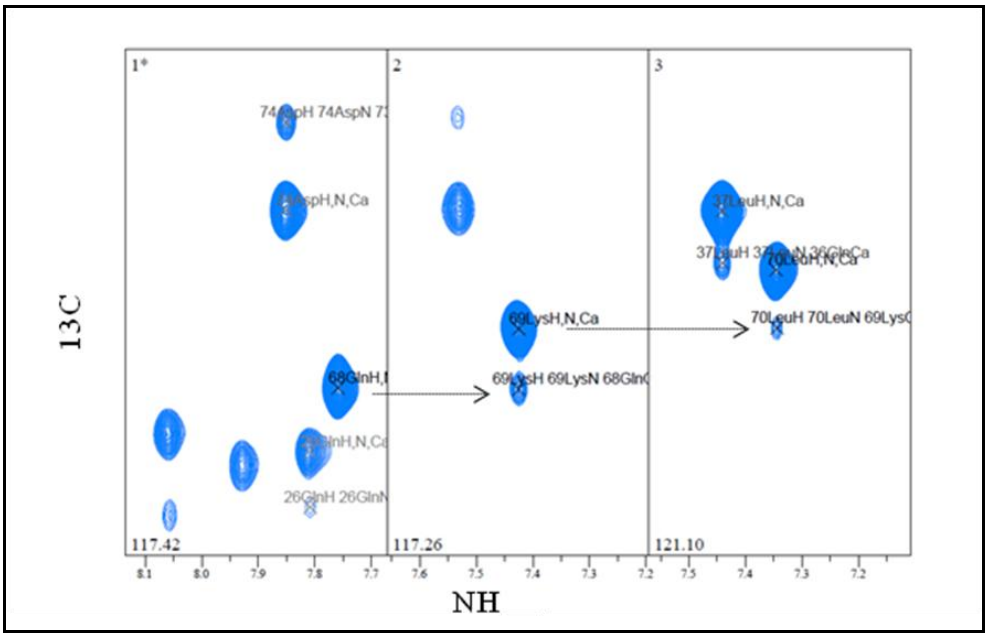


Figure 51. Representative strips showing backbone assignment. **A**mide strips of HNCA NMR spectrum for Gln 68, Lys 69 and Leu 70 used to assigned the chemical resonance of the ^{13}Cα. The amide nitrogen is coupled both to the Cα of its own residue and to the Cα of the preceding residue, both these magnetization transfers occur and peaks for both Cα nuclei are visible in the spectrum. However the coupling to the directly bonded Cα is stronger, hence the corresponding peaks will appear with stronger intensity.

Because of the high α-helical content peaks clustered in the central part of the HSQC experiments making the assignment procedure quite difficult because of peaks overlap. Despite these difficulties we have assigned ~98.3 % of the backbone, ~98% of the ^{1}H side chain and ~9% of the NH groups belonging to the side chains. We used backbone chemical shifts to obtain the Chemical Shift Index (CSI) plot of the construct [Figure 52]. The CSI procedure allows to compare the ^{1}Hα, ^{13}Cα, ^{13}Cβ and ^{13}CO chemical shifts of the protein residues with ^{1}Hα, ^{13}Cα, ^{13}Cβ and ^{13}CO chemical shift of the same residues in random coil

conformation. The comparison allows to assign a value (named chemical shift index) ranging from -1 to 1, representing the propensity of each residue to be in α-helix or in β-strand, respectively.

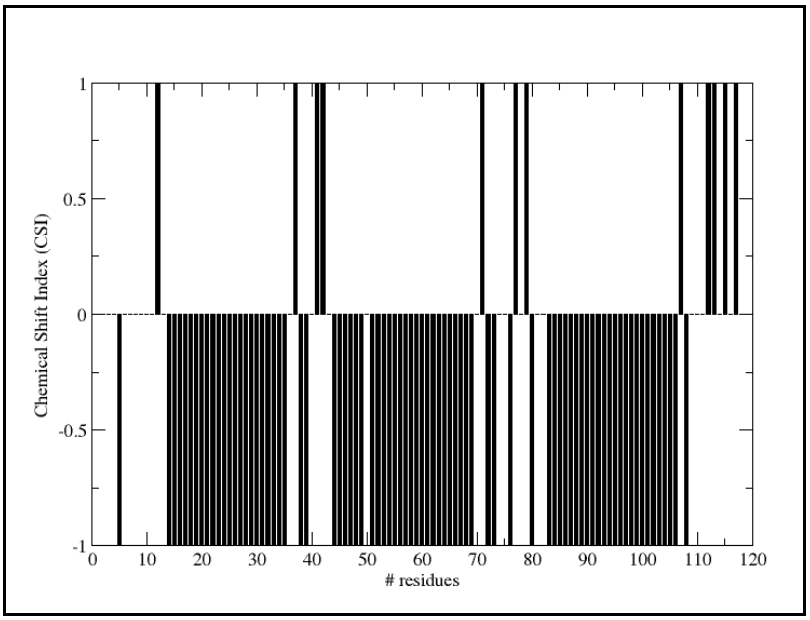


Figure 52. The CSI plot made by means of CCPNmr. The chart confirm that the domain shows only α helices as secondary structure.

In agreement with circular dichroism spectra, the CSI confirms that the N- terminus of NPHP1 contains only α-helices as elements of secondary structure, in detail the first helix spans residues Leu 14 and Glu 36 residues, the second one Asn 44 – Lys 69 and the third one Asn 83 – Thr 108.

6.4.2 3D structural model obtained with CS Rosetta

In order to obtain a structural model of the $NPHP1_{1-115}$ we have used Chemical-Shift Rosetta (CS-Rosetta), an innovative software for protein structure determination that allows structure determination of protein using only chemical shift information[55].

The model of the first 115 amino acids of the human NPHP1 shows, as expected, only α helices as secondary structures. The domain is composed by three helices: the first between

the Pro13 to the Gln 36, the second from Pro 43 to Asn 69 and the third between Tyr 82 to Asn 111, connected by short loop, almost completely in agreement with the result obtained with the CSI.

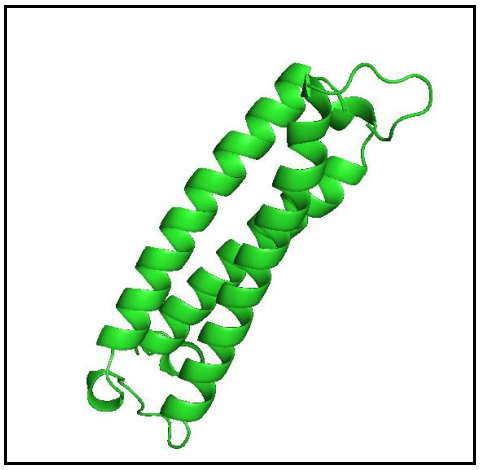


Figure 53. The 3D model of NPHP1$_{1\text{-}119}$ obtained with CS-Rosetta. The NPHP1$_{1\text{-}115}$ displays a helix bundle fold.

6.4.3. Hydrogen–deuterium exchange

To identify the backbone amide protons involved in hydrogen bonds the ^{15}N labeled sample was lyophilized and re-suspended in 100% of D_2O. A series of ^{1}H-^{15}N HSQC were recorded at different time after the re-suspension in D_2O. The amide hydrogens exposed at the solvent exchange with deuterons very fast becoming invisible in the spectrum in a short time after re-suspension with D_2O, whereas the hydrogen atoms involved in H-bonds exchange more slowly with deuterons, remaining visible in the spectrum after long time after the re-suspension in D_2O.

After one hour from the re-suspension in D_2O in the ^{1}H-^{15}N HSQC [Figure 54] 21 HN are still visible. They belong to the following residues: Leu 24, Gln 26, Gln 27, Val 28, Asp 29, Leu 31, Leu 32, Ile 54, Leu 56, Lys 57, Gln 58, Ala 59, Asp 61, Glu 62, Asn 63, Lys 64, Leu 96, Thr 97, Leu 100 and Leu 101.

These residues, which are more protected from solvent exchange, are presumably involved in the formation of hydrogen bonds forming the α helix. This information will be used as restraint in the structure calculation.

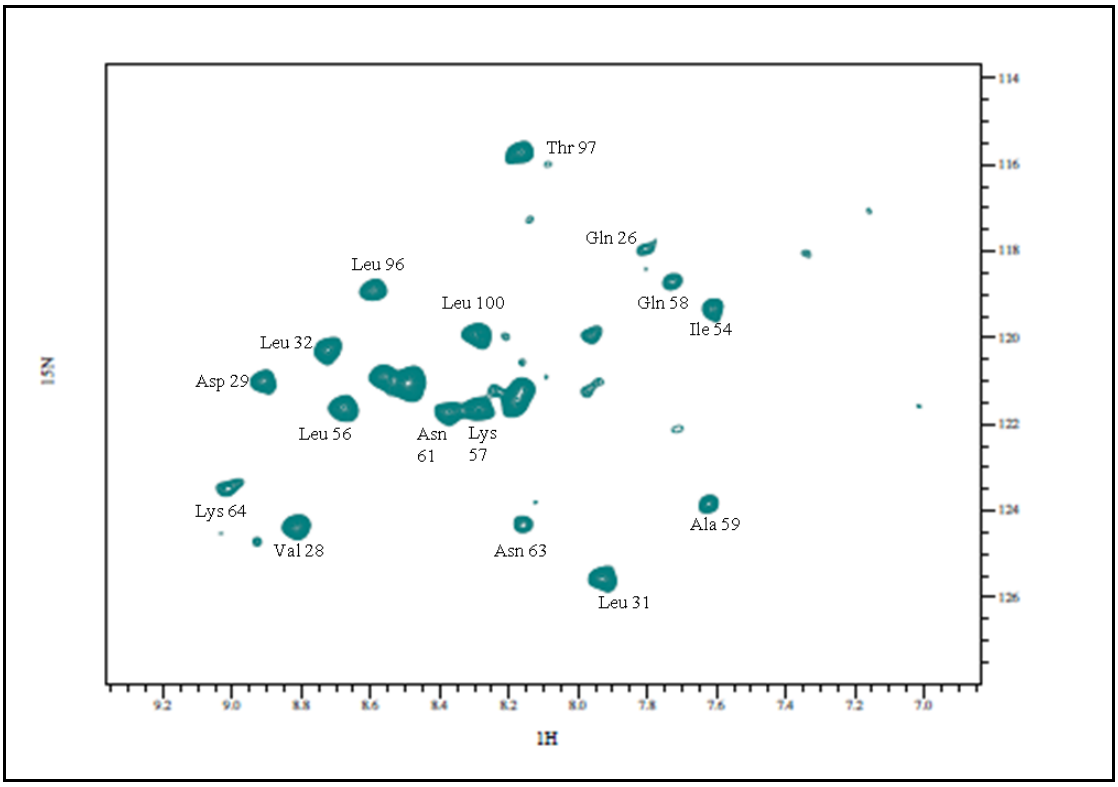


Figure 54. The ^{1}H-^{15}N HSQC recorded after 1 hour from re-suspension in D_2O. 21 amide HN are still visible in the spectrum and represent the NH groups involved in the H-bonds.

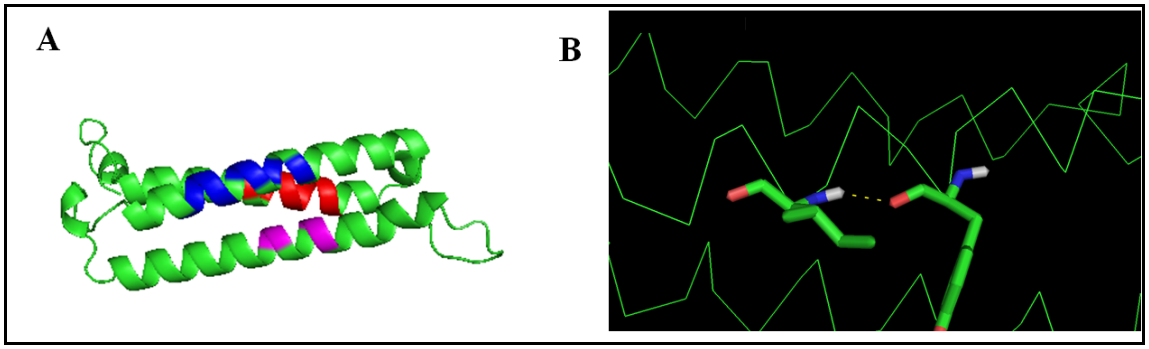


Figure 55. The amino acids involved in H-bonds. A. The residues that do not exchange with D_2O are highlighted in the 3D structure. In red the residues belonging to the first helix, in blue to the second one and in magenta to the third. B. An example of H-bond formed in the structure. The H-bond takes place between the amide -HN group of the Ile 54 and the –C=O of the Tyr 50.

6.4.4 Experimentally calculated *phi* angle and Talos + prediction

The 3D HNHα experiment correlates the $^{1}H_N$ and ^{15}N atoms of a residue to its $^{1}H\alpha$ atom. This spectrum is useful to extract the $^{3}J(HNH\alpha)$ coupling constants from the ratio of the volume of the cross peak ($^{1}H_N$-$^{1}H\alpha$) with the corresponding diagonal peak ($^{1}H_N$-$^{1}H_N$). The Karplus equation (paragraph 4.2.7, equation 5) correlates the *phi* angle of each residue with the corresponding $^{3}J(HNH\alpha)$ (see paragraph 4.2.7). The values of $^{3}J(HNH\alpha)$ can therefore give information on the local backbone conformations of the single residues: $^{3}J(HNH\alpha)$ values < 4.8 Hz and values >8.5 Hz indicate that the corresponding residue adopt an α-helical and conformation β strand conformation, respectively. The $^{3}J(HNH\alpha)$ value could not be calculated for each residues because some HN strips contained superimposed peaks belonging to different $^{1}H_N$ due to the similarity of $^{1}H_N$ chemical shifts because of the high level of overlapping signals. Overall we obtained only 38 reliable $^{3}J(HNH\alpha)$ values.

In addition we have also employed the software Talos + (see paragraph 4.2.7) in order to estimate the *phi* dihedral angle from the $^{1}H_\alpha$, $^{13}C'$, $^{13}C_\alpha$, $^{13}C_\beta$ and the amide ^{15}N and ^{1}H chemical shifts. The software was able to make a good estimate for 93 residues (78%). The predicted and the calculated *phi* angles are listed in table 9.

	Talos +			**HNHα**					**Talos +**			**HNHα**		
	Phi	***Δphi***	**Q**	**^{3}J**	**E**	***Phi***			***Phi***	***Δphi***	**Q**	**^{3}J**	**E**	***Phi***
Ala 4	-92.8	80.7	dyn	-	-	-		Ile60	-60.8	6.16	good	-	-	-
Met 5	-85.0	76.2	dyn	-	-	-		Asp61	-62.7	4.15	good	-	-	-
Leu 6	-101	28.8	dyn	6.64	0.039	-161,-79		Glu62	-66.6	4.24	good	-	-	-
Ala 7	-89.9	25	dyn	-	-	-		Asn63	-64.7	4.98	good	-	-	-
Arg8	-88.5	30.2	dyn	-	-	-		Lys64	-62	5.32	good	4.23	0.64	-61
Arg 9	-110	63.9	dyn	-	-	-		Asp65	-61.3	7.85	good	-	-	-
Gln 10	-108	64.3	dyn	-	-	-		Ala66	-69.2	6.25	good	4.3	0.52	-62
Arg 11	-94.4	90.9	dyn	-	-	-		Leu67	-64.7	9.85	good	-	-	-
Asp 12	-80.9	28.8	good	-	-	-		Gln68	-67.3	7.92	good	4.44	0.42	-177,-63
Pro 13	-62.8	8.9	good	-	-	-		Lys69	-93.9	17.33	good	7.37	0.27	-155,-85
Leu 14	-65.6	2.97	good	-	-	-		Leu70	-76.9	54.82	warn	4.78	0.48	-175,-65
Gln 15	-64.3	4.72	good	-	-	-		Ser71	-92.3	27.42	good	-	-	-
Ala 16	-67.8	5.94	good	-	-	-		Lys72	-64	4.71	good	5.8	1.5	-73
Leu 17	-67	4.77	good	-	-	-		Ala73	-69.5	8.82	good	4.95	0.3	-173,-67
Arg 18	-62.7	4.45	good	3.07	1.06	-51		Asp74	-68.6	9.26	good	-	-	-
Arg 19	-66.8	6.03	good	-	-	-		Glu75	-68.4	10.22	good	-	-	-
Arg 20	-66.1	5.57	good	-	-	-		Ser76	-95.4	21.81	good	-	-	-
Asn 21	-65.9	6.76	good	3.67	0.83	-56		Ala77	-81.1	22.64	good	5.8	0.16	-167,-73

Gln 22	-65.8	6.05	good	-	-	-		Pro78	-62.7	7.61	good	-	-	-
Glu 23	-66.6	6.03	good	-	-	-		Val79	-101.6	19.42	good	-	-	-
Leu 24	-67.6	5.38	good	-	-	-		Ala80	-67.3	12.28	warn	3.9	0.46	-59
Lys 25	-61.4	4.13	good	-	-	-		Asn81	-83.4	22.57	good	7.3	0.51	-156,-84
Gln 26	-61.7	3.19	good	-	-	-		Tyr82	-65.6	7.09	good	-	-	-
Gln 27	-64.1	4.5	good	-	-	-		Asn83	-62.3	4.99	good	3.9	0.62	-59
Val 28	-67.8	5.25	good	4.98	0.74	-67		Gln84	-65.6	5.29	good	-	-	-
Asn 29	-64.2	6.11	good	2.63	0.98	-47		Arg85	-62.8	6.48	good	-	-	-
Ser 30	-67.9	5.7	good	-	-	-		Lys86	-64.8	6.51	good	-	-	-
Leu 31	-66	3.77	good	4.18	0.59	-179,-61		Glu87	-68	7.59	good	-	-	-
Leu 32	-64.2	5.94	good	3.78	1.04	-57		Glu88	-65.1	6.39	good	-	-	-
Ser 33	-65.3	5.15	good	4.28	0.28	-179, -61		Glu89	-68.3	6.58	good	-	-	-
Glu 34	-64.8	3.97	good	-	-	-		His90	-65.3	4.47	good	-	-	-
Ser 35	-68.8	10.2	good	4.06	0.56	-60		Thr91	-67.3	5.04	good	4.1	0.7	-60
Gln 36	-94.2	16.1	good	8.14	0.28	-149,-91		Leu92	-62.6	4.34	good	3.7	0.86	-57
Leu 37	-72.8	12.1	warn	6.13	0.30	-165, -75		Leu93	-61.7	4.27	good	-	-	-
Lys 38	-56.8	2.97	good	-	-	-		Asp94	-65.6	6.37	good	3.5	0.21	-55
Glu 39	-67.7	6.09	good	4.75	0.29	-175,-65		Lys95	-64.5	6.99	good	-	-	-
Ala 40	-74.8	12.4	good	-	-	-		Leu96	-62.6	4.82	good	3.5	1.24	-55
Leu 41	-76.8	12	warn	7.5	0.25	-154,-86		Thr97	-68.6	8.15	good	-	-	-
Glu 42	-73.6	7.68	good	5.2	0.26	-171,-68		Gln98	-63.8	1.7	good	-	-	-
Pro 43	-53.8	7.72	good	-	-	-		Gln99	-65.9	3.45	good	-	-	-
Asn 44	-66.5	5.83	good	5.21	0.34	-171,-69		Leu100	-64.3	5.24	good	-	-	-
Lys 45	-66.3	5.59	good	-	-	-		Gln101	-64.1	5.27	good	-	-	-
Arg 46	-62.6	2.94	good	-	-	-		Gly102	-68.2	5.09	good	-	-	-
Gln 47	-62.3	4.03	good	4.20	0.38	-179,-61		Leu103	-64.9	3.92	good	4.97	0.51	-173,-67
His 48	-67.3	6.38	good	-	-	-		Ala104	-61	2.93	good	-	-	-
Ile 49	-62.8	4.72	good	-	-	-		Val105	-65.4	8.38	good	-	-	-
Tyr 50	-65.1	3.51	good	-	-	-		Thr106	-68.8	9.24	good	-	-	-
Gln 51	-62.9	4.12	good	4.12	0.57	-60		Ile107	-74.5	18.85	good	-	-	-
Arg 52	-67.3	6.4	good	-	-	-		Ser108	-72.1	19.27	good	-	-	-
Cys 53	-60.7	5.66	good	-	-	-		Arg109	-76.4	14.02	dyn	-	-	-
Ile 54	-63.7	8.58	good	4.54	0.57	-63		Glu110	-77.1	27.26	dyn	-	-	-
Gln 55	-63.2	2.93	good	-	-	-		Asn111	-94.43	24.7	dyn	7.62	0.07	-153,-87
Leu 56	-64.3	4.11	good	-	-	-		Ile112	-111.2	63.66	dyn	-	-	-
Lys 57	-59.2	4.45	good	-	-	-		Thr113	-121-3	58.15	dyn	-	-	-
Gln 58	-64.3	4.13	good	4.59	0.39	-176,-64		Glu114	-116.2	63.56	dyn	-	-	-
Ala 59	-63.7	3.05	good	4.68	0.44	-176,-64		Val115	-123	66.02	dyn	-	-	-

Table 9. List of the *phi* angles predicted by TALOS+ and calculated by HNHα. TALOS+ columns indicate: φ: predicted φ angle, $D\varphi$: predicted error angle, Q: quality ("good", "warn" not good prediction keep attention, and "dyn" for dynamic residues). HNHα columns are: ^{3}J: calculated value of the scalar coupling $^{3}J_{HN,H\alpha}$, E: estimated error, φ: predicted φ angle.

6.4.5 ^{1}H-^{1}H NOE pattern of NPHP1$_{1-115}$ confirms the presence of α-helix

To determine the solution structure of NPHP1$_{1-115}$ by NMR after having assigned the domain resonances (^{1}H, ^{13}C, ^{15}N) we performed and analyzed NOESY 2D and 3D spectra (^{1}H-^{1}H NOESY, ^{1}H-^{15}N NOESY and ^{1}H-^{13}C NOESY) to collect interproton NOE restraints (paragraph 4.2.8). NOE restraints are the most important experimental data for the calculation of a NMR solution structure, as they give a range of distances between two closely spaced ^{1}H atoms (<5 A).

A chart made by means of CCPNmr summarizes the secondary structure and the assigned ^{1}H-^{1}H NOEs of the domain [Figure 56]. The sequential NOEs between Hα$_i$ and H$^N_{(i+1)}$ and between Hβ$_i$ and H$^N_{(i+1)}$ indicated as $d_{\alpha N}$ and $d_{\beta N}$, respectively, are found over all the protein sequence, whereas the sequential NOEs between H$^N_{(i)}$ and H$^N_{(i+1),}$ d_{NN}, are missing at the N-terminus region indicating the flexibility of the region. The regular secondary structure such as α-helix or β-strands, forces protons in a well defined spatial relationship, thus generating a specific pattern of NOEs. Analysis of the NOE patterns (NOEs between Hα$_i$ and the H$^N_{i+3}$ and between Hα$_i$ and Hβ$_{i+3}$ indicated as $d_{\alpha N(i,\ i+3)}$ and $d_{\alpha\beta(i,i+3)}$ respectively) reveals that the domain adopts α-helical conformation. This pattern of NOEs is found all along the structure: in the stretches comprising the residues 14 and 35, 45 and 69 and 88 and 108, that form three α-helices. This result is in agreement with the model obtained with CS-Rosetta and with the secondary structure predicted by CSI.

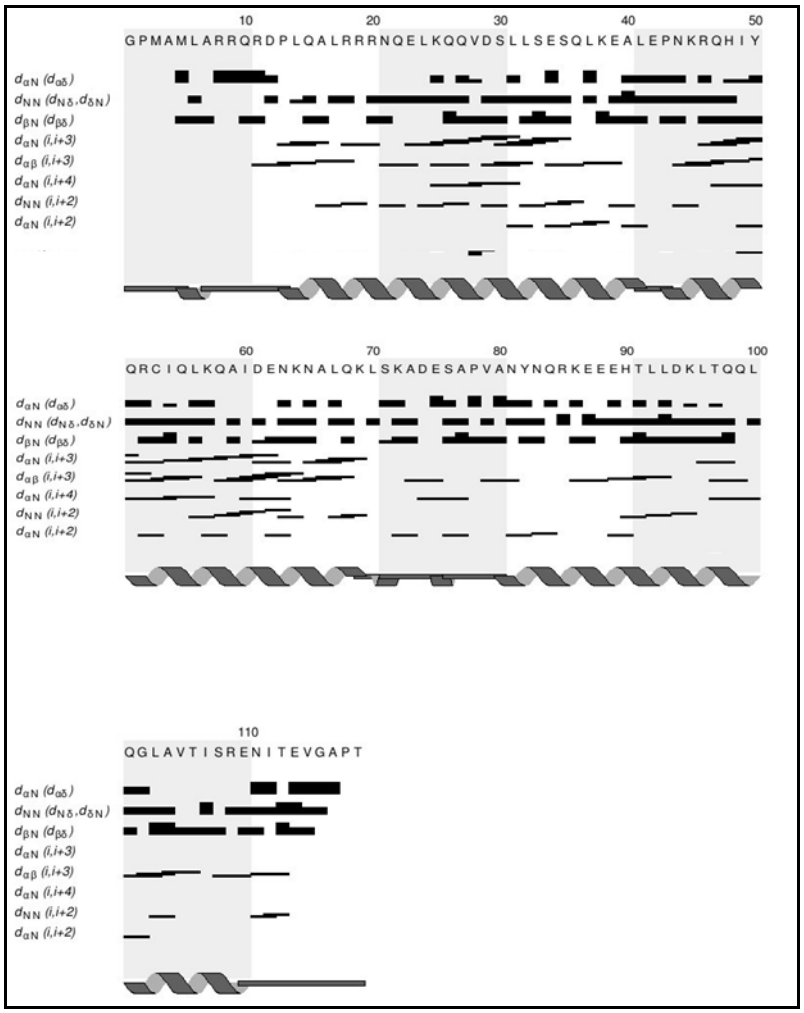


Figure 56. The chart of the pattern of the NOEs made by CCPNmr. In this chart the NOEs assigned and the predicted secondary structure by CSI are summarized.

6.4.6 Structure calculation

We next calculated with ARIA 2.1.3 software[62] the structure of the NPHP1$_{1\text{-}115}$ using 1561 manually assigned NOEs restraints, 93 dihedral angles predicted by Talos + and 22 H-bonds as further restraints. The restraints used for final structure calculation are listed in table 10.

NOE distance restrains	Unambiguous	Ambiguous
Intrasequential (0)	445	162
Sequential (| *i-j* | = 1)	307	109
Short (| *i-j* | = 2-3)	214	74
Medium (| *i-j* | = 4-5)	71	22
Long (| *i-j* | > 5)	124	31
Total	1163	398
Dihedral angles restrains	93	
H-bonds restrains	22	

Table 10. Restrains used to calculate the 20 lowest energy structures of the NPHP1. The NOEs are classified according to the distances between the residues.

The best ensemble of structures achieved up to now for $NPHP1_{1-115}$ is reported in Figure 57. It represents the 20 lowest energy structures calculated by ARIA2.1.3 software after refinement in water. The statistics associated to the bundle of the structures is summarized in table 11.

RMSD (Å)	
Residues 14-108 (excluding N- and C-termini)	1.336+/-0.271
Only helices (14-32; 44-69; 83-108)	0.890 +/- 0.180
Structural quality	
% residues in most favored regions	84.1
% residues in additionally allowed regions	12.9
% residues in generously allowed regions	1.7
% residues in disallowed regions	1.4

Table 11. Statistics of the 20 lowest energy structures of the N-terminus of NPHP1. RMSD is calculated on N, Cα and C' atoms. Ramachandran quality parameters are assessed using the PROCHECK-NMR software[82].

The $NPHP1_{1-115}$ adopts a well defined tertiary structure, it is a right-handed 3-stranded α helix bundle composed by three antiparallel α helices: α1, Leu14 – Leu32, α2, Asn44 –Lys69, α3, Asn83- Ser108, connected by short linkers: loop1: Leu 33- Pro 43; loop2: Leu 70- Tyr 82. The N- and the C- termini comprising residues Gly1- Pro13 and Arg109- Thr119 respectively are flexible. The second helix does not show any curvature, whereas the first and the third

helix show a kink at the residue Arg 19 and the Glu 99 respectively, for this reason the three helices are not coaxial. The backbone RMSD calculated considering the residues belonging to the three α-helices is 0.890 +/- 0.180 Å, we have excluded the N- and C-termini and the interhelical regions. The RMSD calculated overall the structure, excluding the N- and the C-termini, is reported in table 11.

To compare the NMR ensemble of the structure and the model obtained with CS-Rosetta we have also calculated the RMSD between them considering only the three α helices. The RMSD calculated on the backbone is 1.76+/- 0.18. Hence the model obtained with CS-Rosetta is very similar with the solution structure.

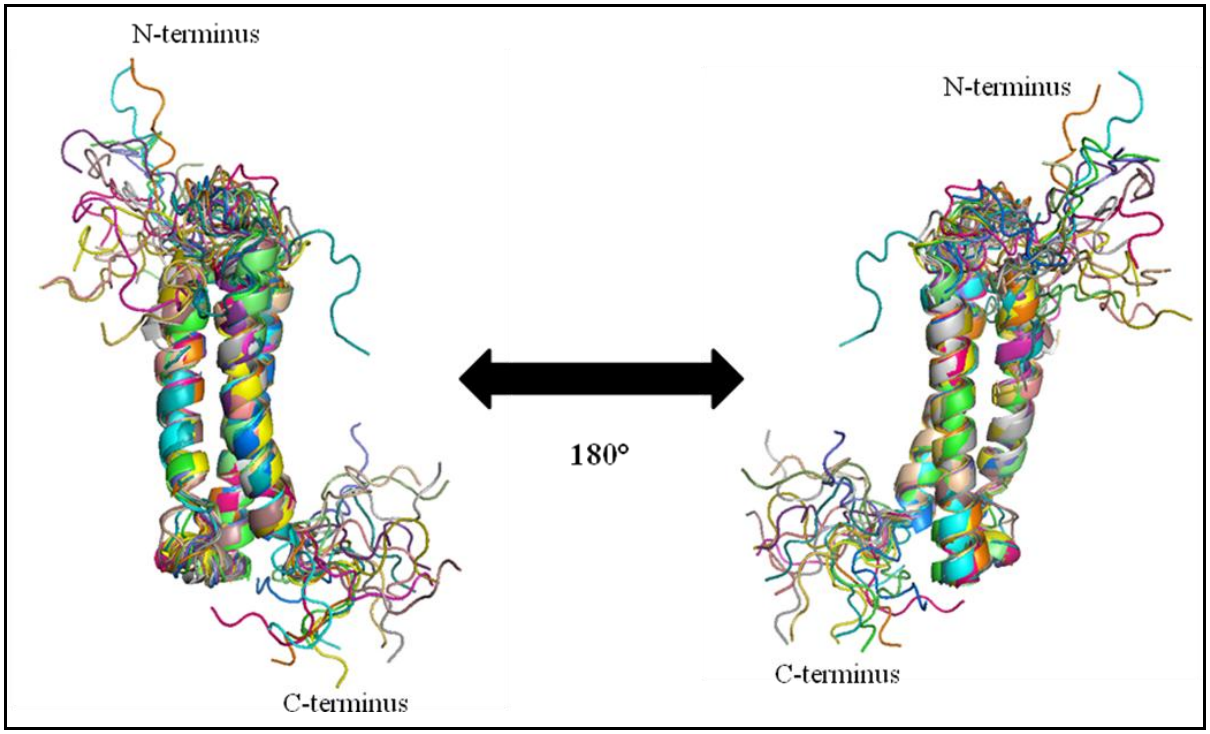


Figure 57. The NPHP1$_{1\text{-}115}$ structure. Cartoon representation of the NMR ensemble of the 20 lowest energy structure of NPHP1$_{1\text{-}115}$. Superimposition has been performed on the N, Cα and C' atoms of the residues belonging to the α-helices. The structure in the right side of the panel is related to the former by a 180° rotation.

The protein core is formed by hydrophobic contacts between the side chains of a significant number of hydrophobic residues pointing towards the interior of the helix bundle. In particular hydrophobic contacts occur between the Leu 103-Leu 31, Leu 24- Ile 60, Leu 17-Leu 67, Leu 100-Ile 49, Leu 100-Val 28 [Figure 58A]. The structure is further stabilized by polar contacts that occur between the second and the third helix . These polar contacts

involve the following residues: Lys 57-Glu 101, Lys 64-Asp 94; His 90-Gln 68 [Figure 58B]. However, the majority of the polar side chains of the polar residues are exposed to the solvent thus creating a highly charged surface [Figure 59]. In particular the charges are randomly distributed on the surface of the structure.

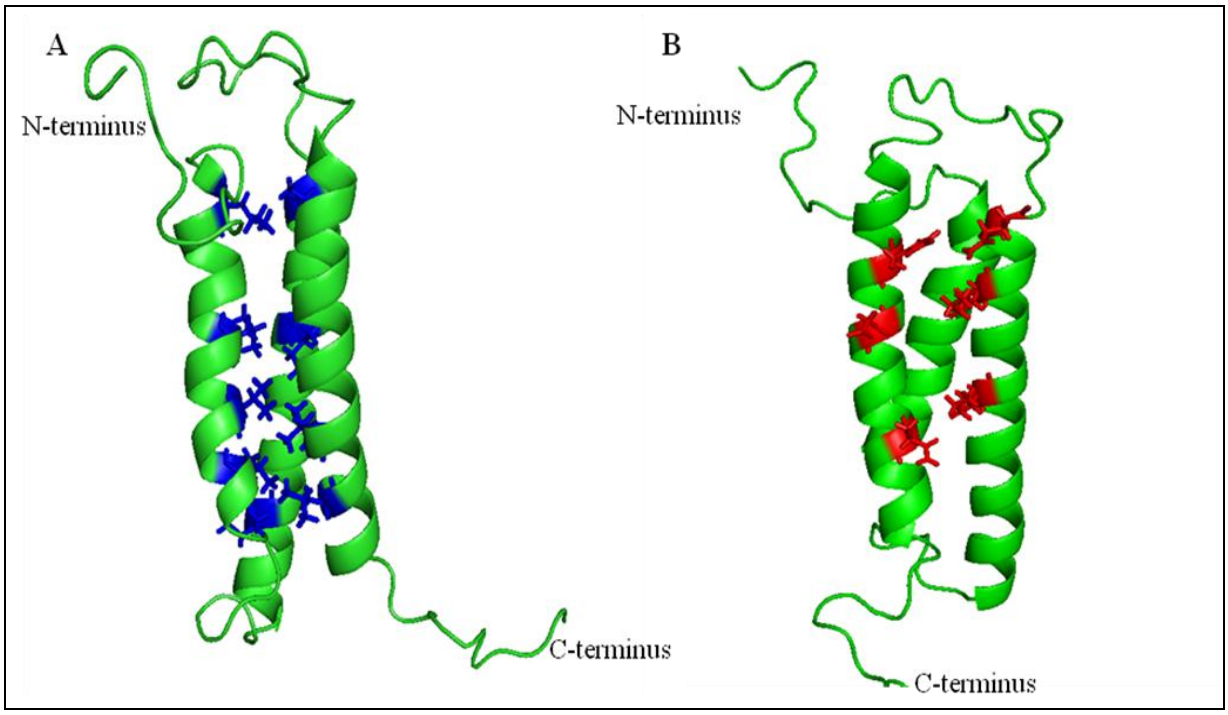


Figure 58. The contacts that stabilize the structure. A. The hydrophobic core of the structure. B. The polar contact that stabilize the structure between the α2 and the α3.

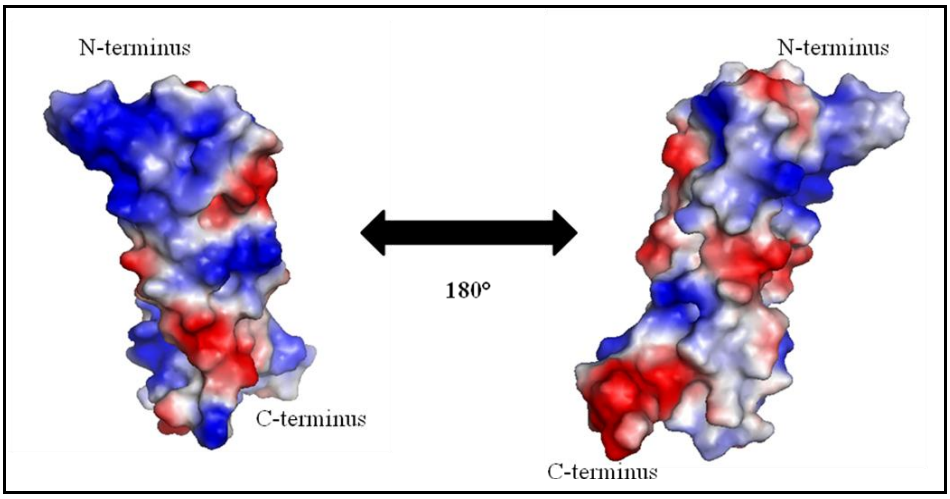


Figure 59. Electrostatic surface of NPHP1$_{1-115}$. The charge is randomly distributed on the surface . The structure on the right side is related to the left by a 180° rotation. Blue and red indicate positive and negative electrostatic surface potentials, respectively.

To gain insights into the relative orientation of the three helices, we performed residual dipolar coupling (RDC) measurements on a ^{15}N-labelled NPHP1$_{1-115}$ sample. We considered only the RDC values for residues that belong to the structured regions. The RDC values of 46 residues satisfied this requirement.

We used PALES [71] to calculate the Q factors of the 20 lowest energy structures for the 46 RDC values experimentally calculated and the structure #3 displays the best agreement with these RDC values (Q factor: 0.55). The measured RDCs were compared to the values back calculated from NPHP1$_{1-119}$ structure#3 using MODULE[69]. As shown in Figure 60, the two set of data set are overall in agreement.

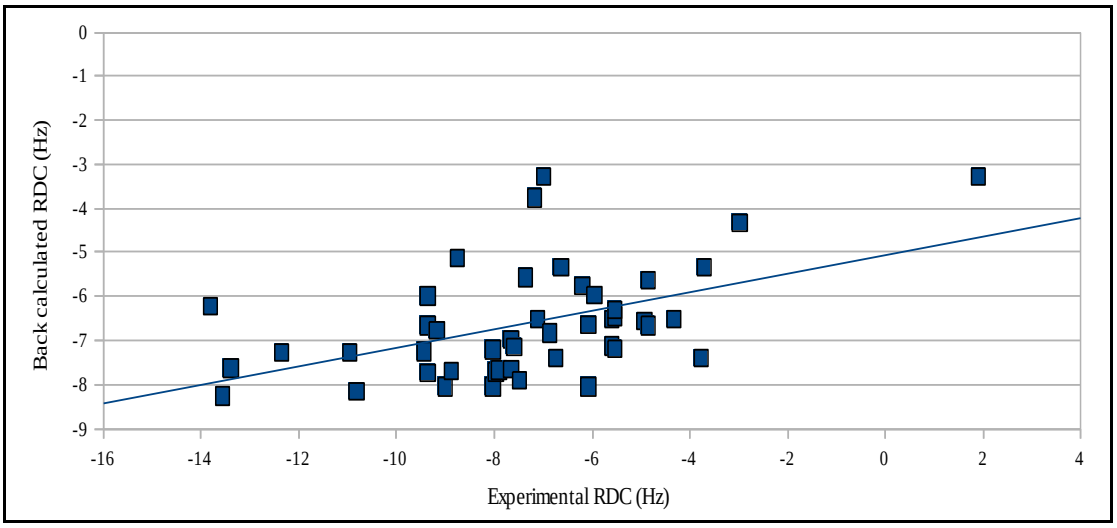


Figure 60. Comparison of experimentally measured RDC values and RDC values back-calculated from the calculated structures. Plot of the measured RDC versus the RDC values back-calculated from structure #3.

A search using the DALI server [83] for NPHP1$_{1\text{-}115}$ was performed to compare our calculated structures with those present in the Protein Data Bank. Unfortunately from this search we did not obtain any structure similar to the NPHP1$_{1\text{-}115}$.

6.4.7 NPHP1$_{1\text{-}115}$ displays three structured regions

In order to obtain information on the local backbone dynamics we performed heteronuclear relaxation NMR experiments including T1 and T2 relaxation and heteronuclearNOE (hetNOE) experiment. Results deriving for heteronuclear NOE (hetNOE) and T1 relaxation give information on the local backbone flexibility in the ns-ps time-scale, as they are sensitive to the high-frequency motions whereas the T2 relaxation rate is also influenced by slow motion in the μs-ms time-scale [84].

The hetNOE value for each HN group of the protein backbone, ranges from -1 to 1, they are calculated as described in Eq. 6 in chapter 4.2.10, [Figure 61]. Residues in highly structured regions (including α1, α2, α3) move in a coherently way with the protein rotational diffusion tensor and have an hetNOE value between 0.8 and 1. On the contrary, residues in the flexible

regions (including res 1-14, 37-43, 70-87 and 109-119), such as loops and N- and C-termini, have typically hetNOE value <0.5.

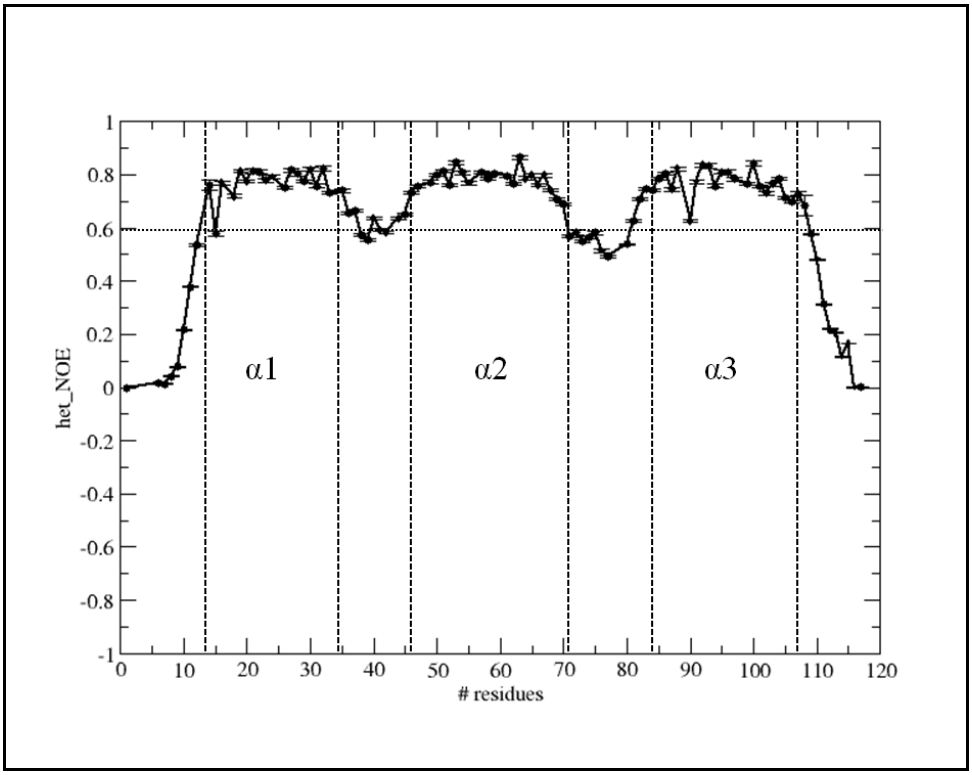


Figure 61. The NPHP1$_{1\text{-}115}$ heteronuclear NOEs for local backbone flexibility. The first 14 residues and the 11 residues corresponding to the N- and the C-terminus respectively show hetNOE value <0.5, it indicates the high backbone mobility at the ns-ps time-scale. Some flexibility is also found in the residues at the interhelical connecting loops (residues Leu 37- Pro 43, Leu 70- Tyr 82). Whereas the hetNOE value >0.6 of the three helices α1, α2 and α3 indicates a restricted dynamic.

The T1 relaxation rate is expressed as R1 (1/T1), as a frequency. Hence high R1 values indicate a high dynamics of the HN backbone whereas low values indicate a restricted dynamics of the HN backbone [Figure 62]. Residues in highly structured regions, the three helices α1, α2, α3 in our case, present a low value of R1 relaxation rate. On the contrary, residues in the flexible regions (including res Gly 1-Pro 13, Leu 37- Pro 43, Leu 70- Tyr 82 and Arg 109- Thr 119), such as loops and N- and C-termini, have a high value of R1 relation rate.

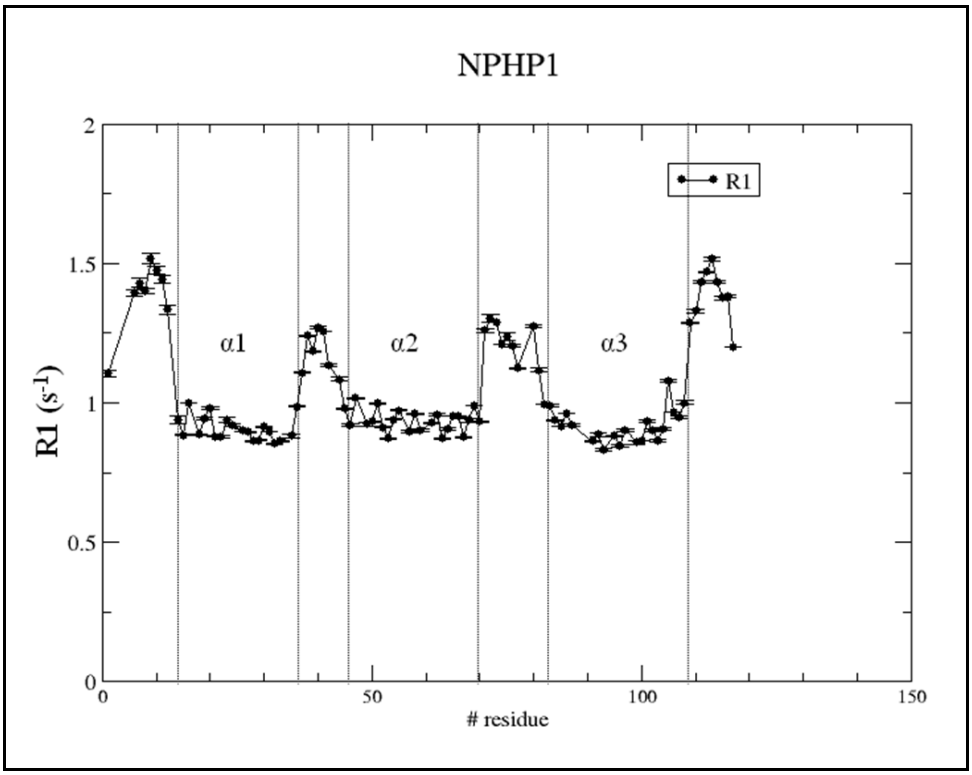


Figure 62. The NPHP1$_{1\text{-}115}$ R1 relaxation rate. The first 14 residues and the last 11 residues corresponding to the N- and the C-terminus respectively show a high R1 value such as the residues at the interhelical connecting loops (residues Leu 37- Pro 43, Leu 70- Tyr 82), whereas the three helices α1, α2 and α3 show a low value of R1 relaxation rate.

Like T1 relaxation rate, also T2 relaxation rate is expressed as R2 (1/T2). High R2 values are usually indicative of motions in this slow time scale or chemical exchange. High R2 value are observed for the three helices (α1,α2 and α3) whereas the N- and C- terminus and the interhelical loops have a low R2 value [Figure 63].

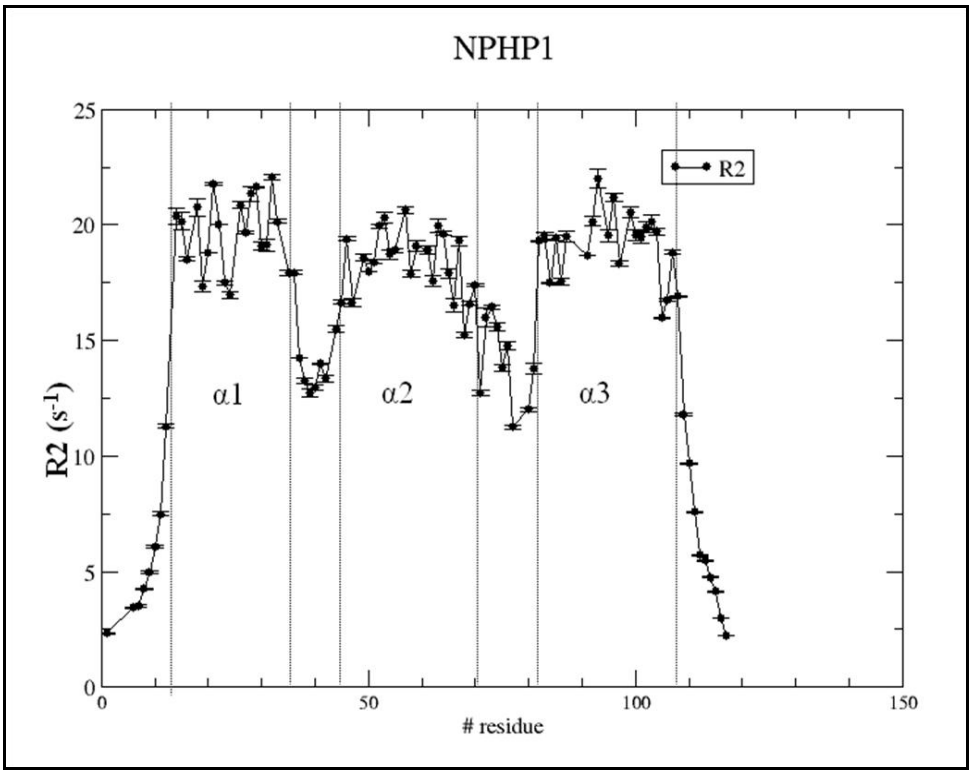


Figure 63. The NPHP1$_{1\text{-}115}$ R2 relaxation rate. The first 14 residues and the last11 residues corresponding to the N- and the C-terminus respectively show a low R2 value such as the residues at the interhelical connecting loops (residues Leu 37-Pro 43, Leu 70-Tyr 82), whereas the three helices α1, α2 and α3 show a high value of R2 relaxation rate.

Through T1 and T2 relaxation measurements and heteronuclearNOE (hetNOE) NMR experiments we have investigated the backbone dynamics of the NPHP1$_{1\text{-}119}$ confirmed the presence of three structured regions, corresponding to the α helices (α1: Leu14 – Leu36, α2: Asn44 –Lys69, α3: Asn83- Ser108), connected by two short linkers (loop1: Leu37- Pro43; loop2: Leu70- Tyr82) that display flexibility as the N- and the C- termini comprising residues Gly1- Pro13 and Arg109- Thr119, respectively.

The average value of the ^{15}N R2/R1 ratio over a protein backbone allows the determination of the molecular rotational correlation time (τ_C, the average time for a molecule to reorient about one radian) under two conditions:

- the presence of motions in ps-ns time scale are not significant (conventionally residues presenting hetNOE<0.65 are excluded from this analysis)

- R2 is not significantly lengthened by chemical or conformational exchange[85].

We thus calculated the average value of R2/R1 considering the 101 residues that satisfy the aforementioned conditions, and computed the τ_C according to empirical equation 11.

$$\tau_c = \frac{\sqrt{\left(6\langle R2/\ R1\rangle - 7\right)}}{2\omega_N} \qquad \text{eq.11}$$

where ω_N and <R2/R1> indicate the ^{15}N resonance frequency and the average R2/R1 ratio, respectively. For $NPHP1_{1\text{-}119}$, <R2/R1>= 20.382, so the calculated τ_C is 14.056 +/ -2.457 ns, which is in agreement with the expected rotational correlation time of a protein of this dimension.

6.5 The interaction with the coiled coil domain of PC1

6.5.1 The purification of the recombinant coiled coil domain of PC1

In order to investigate the interaction between $NPHP1_{1\text{-}115}$ and the coiled coil domain of PC1 (CC_PC1) by NMR we first produced the CC_PC1 (spanning residues His 4186 to Ala 4257) with recombinant techniques. The CC_PC1 DNA sequence was cloned in pETM44 plasmid and purified from *E. Coli* (paragraph 4.2.1). The purified construct has the following sequence, where the amino acids belonging to the tag and that remain after the protease cleavage are highlighted in italics:

*GPMG*HPSTSSSQLDGLSVSLGRLGTRCEPEPSRLQAVFEALLTQFDRLNQATEDVYQL EQQLHSLQGRRSSRAPA

We tried to purify the domain without the tag but the protein was aggregated, for this reason we were not able to reach high protein concentrations. Figure 64 shows a 1D NMR spectrum, the broad line width of the spectrum confirmed that the protein is aggregated and not suitable for further structural studies.

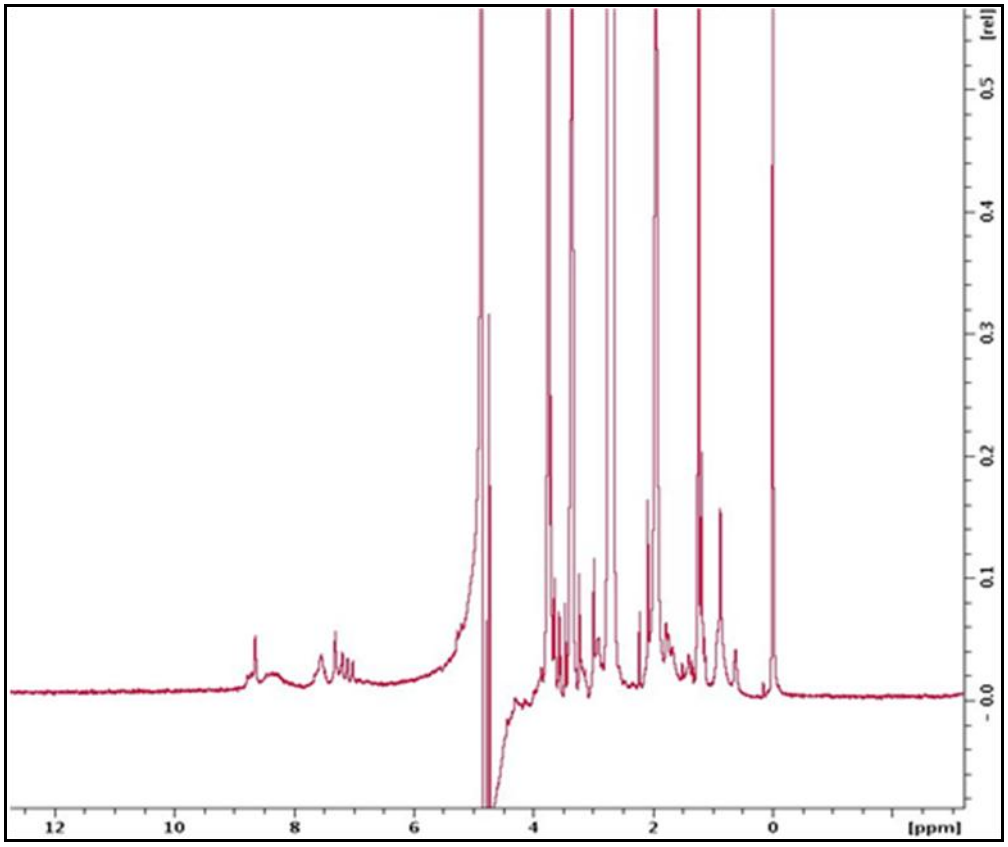


Figure 64. The 1D ^{1}H NMR spectrum of CC_PC1. CC_PC1 used for spectrum was concentrated 0.3 mM and dissolved in PBS 20mM, 3mM DTT, spectrum was recorded at 298K.

In a previous study aiming to investigate by NMR the interaction between the coiled coil domain of polycystin 2 and the coiled coil domain of PC1[19] the authors have used the recombinant coiled coil domain of PC1 with the His-MBP tag. We therefore decided to purify the protein with MBP tag, in order to improve its solubility and to avoid aggregation. Unfortunately, also this construct displays aggregation. We recorded the 1D ^{1}H spectrum that displays more signal dispersion but the peaks are broad due to the high molecular weight of the MBP [Figure 65].

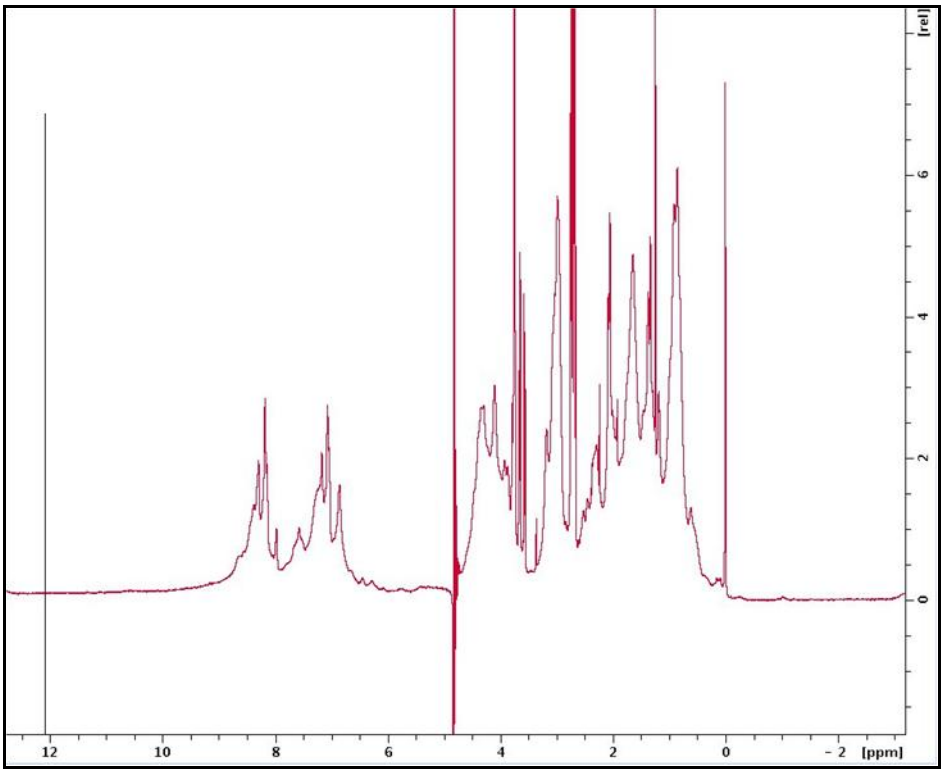


Figure 65. The 1D ^{1}H NMR spectrum of His-MBP-CC_PC1. His-MBP-CC_PC1 used for spectrum was concentrated 0.3 mM and dissolved in PBS 20mM, 3mM DTT, spectrum was recorded at 298K.

6.5.2 Titration of 6His-MBP-CC_PC1 into NPHP1$_{1\text{-}115}$

Despite the low sample concentration and the high aggregation state of the protein we have tried to perform a titration with ^{15}N labeled NPHP1$_{1\text{-}115}$. to prove binding between 6His-MBP-CC_PC1 and NPHP1$_{1\text{-}115}$. We recorded a spectrum of ^{15}N NPHP1 in the presence of PC1 (1:1 stoichiometry) [Figure 66, the black and the red spectra represent NPHP1$_{1\text{-}115}$ in absence or presence of PC1 respectively].

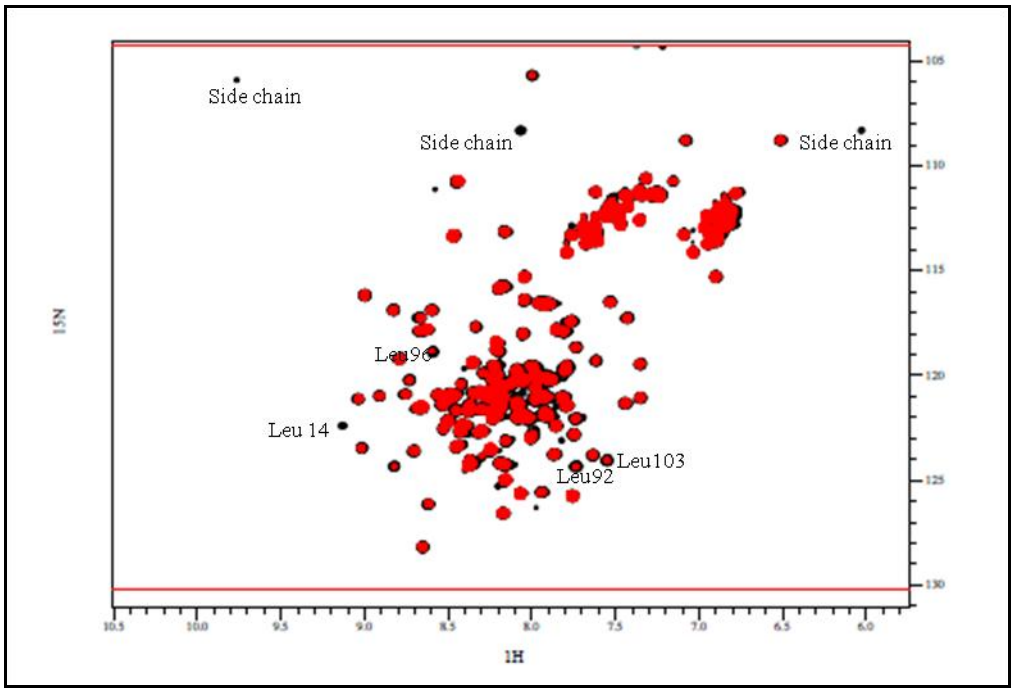


Figure 66. The NPHP1$_{1\text{-}115}$ appears to do not bind the CC_PC1. The ^{1}H-^{15}N HSQC spectra of ^{15}N NPHP1$_{1\text{-}115}$ alone (black) and after the addition of a 1-fold molar excess of CC_PC1(red).

Because of the low concentration of the recombinant CC_PC1 we have added 182 µl of ligand concentrated 0.4 mM to a solution containing a volume of 250 µl of the ^{15}N NPHP1$_{1\text{-}119}$ concentrated 0.3 mM leading to a dilution of the sample. Addition of 6His-MBP-CC_PC1 to ^{15}N NPHP1$_{1\text{-}115}$ did not cause any chemical shift displacement, suggesting the absence of interaction between the two domains.

However few NPHP1 residues (Leu 14, Leu 92, Leu 96 and Leu 103) showed decrease of intensity upon addition of CC_PC1, we cannot therefore completely exclude the hypothesis of the interaction between CC_PC1 and NPHP1$_{1\text{-}115.}$ The interaction will be object of further investigations. In particular we will perform GST pull down between the entire C- tail of PC1 and NPHP1$_{1\text{-}115}$ and we will optimize the purification of the CC_PC1 using a longer construct in order to stabilize it.

7. Discussion

7.1 Metabolomics

7.1.1 Metabolomic profiling of the exometabolome of *Pkd1* KO and WT MEF cells.

In this thesis we have used solution state NMR to analyze the metabolome of the medium conditioned by *Pkd1* KO and WT MEF cells. The analysis has highlighted some metabolic differences between the two exometabolomes confirming the hypothesis that PC1 acts as "metabolic regulator".

Importantly, the metabolic profiling of the *Pkd1* KO MEF cells' exometabolome revealed similarities with tumor metabolism, as *Pkd1* KO MEF cells produce high levels of lactate, alanine and glutamate, similarly to what observed in cancer cells[9]. The production of lactate derives from a major consumption of glucose, hence the *Pkd1* KO cells switch to aerobic glycolysis. This phenomenon will be discussed in the next paragraph, as it is been the main object of this thesis.

The high secretion of alanine observed in *Pkd1* KO MEF is most probably due to the high level of alanine aminotransferase (ALT), that converts pyruvate into alanine, and can be explained as a metabolic solution found by the cells to eliminate the large excess of carbon and nitrogen, toxic for the cells, deriving from the accelerated metabolism[9]. In this context, to support this hypothesis, it would be interesting to investigate by transcriptomics whether *Pkd1* KO cells display higher transcripts of ALT, as observed in tumor cells[9].

Conversely, the high content of secreted glutamate observed in *Pkd1* KO MEF cells' exometabolome might be related to a stimulus to promote cell survival, growth and proliferation through metabotropic glutamate receptors (mGluRs), such as the G-proteins-coupled-receptors, that leads to a stimulation of ERKs pathway and phosphaditilinositole 3'-kinase[10,11]. The high level of glutamate, which is usually present in low levels in the extracellular medium, might derive from the conversion of extracellular glutamine, which is generally a high abundant nutrient of the extracellular medium. Glutamine enters into the cell and is converted into glutamate by a mitochondrial enzyme, the glutaminase (GLS), glutamate can accumulate inside the or can be secreted outside the cell. This process, called glutaminolysis, is a hallmark of the tumor metabolic pathway[9]. But in this case the metabolic profiling of the medium conditioned by *Pkd1* KO cells displays only a major production of glutamate whereas the *Pkd1* KO cells do not show a major consumption of glutamine. Also in this case it would be interesting to verify whether *Pkd1* KO cells, similarly to tumor cells,

display high levels of GLS transcript. Taken together, these metabolic findings indicate that the *Pkd1* KO MEF cells are affected by an accelerated metabolism related to their increased proliferation rate, once again suggesting similarities between the metabolic dysfunctions occurring both in tumor cells and in *Pkd1* KO MEF.

7.1.2 The Warburg effect occurs both in cellular models and *in vivo*

The most relevant metabolic difference distinguishing *Pkd1* KO from WT MEF cells consists in the presence of increased levels of aerobic glycolysis, the Warburg effect, as highlighted by the increased lactate production and by the increased glucose consumption observed in *Pkd1* KO MEF cells as compared to WT cells.

The Warburg effect is a metabolic hallmark of cancer and proliferative cells, which adopt this metabolic strategy to sustain the high energetic demand required by the increased proliferation rate. Usually, cells of differentiated tissues metabolize glucose to produce ATP through oxidative phosphorylation, whereas in presence of oxygen cells of proliferative or tumor tissues produce ATP through glycolysis converting glucose into lactate, the principal end product of this pathway [Figure 67]. Similarly to tumor and proliferative cells *Pkd1* KO MEF cells display an increased proliferative rate thus requiring more energy production as compared to the WT cells.

In order to prove that the Warburg effect occurs also *in vivo* we analyzed the glucose and lactate content in kidneys deriving from Ksp-Cre: $Pkd1^{flox/-}$ and $Pkd1^{flox/+}$ mice[12]. Unfortunately, the analysis of 1D ^{1}H NMR spectra of the polar extracts of WT and KO kidneys did not confirm our hypothesis, as the measurements were affected by high errors generated by the high biological variability between the analyzed litters. We therefore decided to treat Ksp-Cre: $Pkd1^{flox/-}$ and $Pkd1^{flox/+}$ mice with uniformly labeled ^{13}C glucose in order to better follow the conversion of glucose into lactate. The use of stable isotope is an experimental procedure widely applied in order to trace a specific metabolite in cells and in animal models[13,14]. This technique is very useful to confirm a specific metabolic pathway associated to a particular disease and to verify whether a drug produces the expected effects. The analysis of the polar extracts of the kidneys deriving from Ksp-Cre: $Pkd1^{flox/-}$ and $Pkd1^{flox/+}$ mice treated with uniformly labeled ^{13}C glucose allowed us to confirm that the

Warburg effect arises also *in vivo.* In fact *Pkd1* KO kidneys displayed an increased amount of glucose and lactate compared to WT kidneys.

Finally, the analysis of the transcripts of the glycolytic enzyme through real-time PCR analysis, performed by our collaborators, has revealed that *Pkd1* KO cells and *Pkd1* KO kidneys displayed a transcriptional signature of the glycolytic enzymes, thus supporting our metabolomic findings.

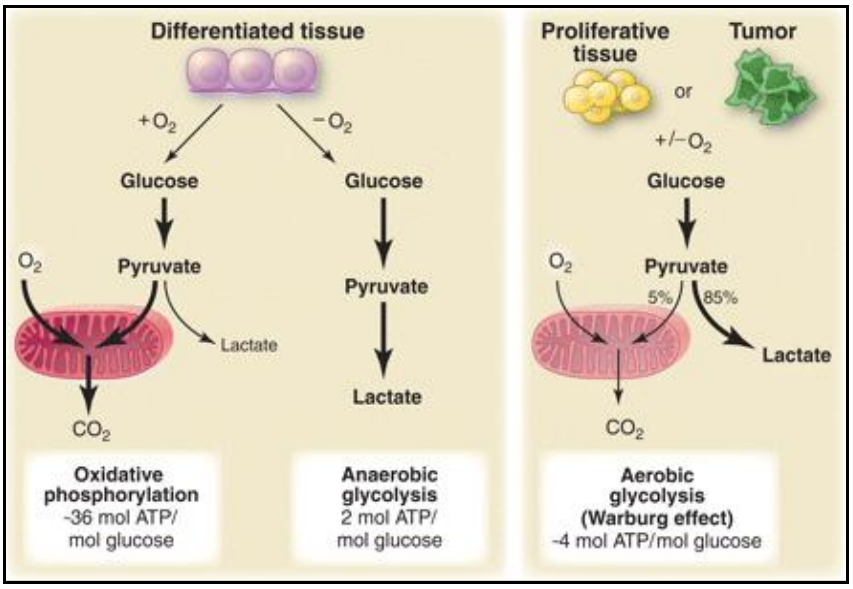


Figure 67. The fate of the glucose in differentiated (on the left) and in proliferative or tumor tissues (on the right). Glucose in differentiated tissues is used to produce ATP through the oxidative phosphorylation, whereas proliferative and tumor tissues produce ATP through the aerobic glycolysis. [Figure from Vander Heiden, et al. Science, 2009].

7.1.3 Signaling cascade

The observation that *Pkd1* KO MEF cells and *Pkd1* KO kidneys are affected by the Warburg effect inspired further experiments to investigate the deregulated signaling cascade deriving from loss of functional PC1. These experiments were performed by our collaborators. A previous study [41] showed that the loss of PC1 enhances the activity of mTORC1 through the up-regulation of the ERKs pathway. The enhanced mTORC1 activity causes the increase of the cell size and the cellular proliferative rate thus inducing the Warburg effect. The

accelerated metabolism in absence of functional PC1 increases also the intracellular ATP content, inhibiting AMP-activating protein kinase (AMPK), an enzyme that acts as metabolic regulator detecting the AMP:ATP ratio to maintain the energy homeostasis. In addition, AMPK is also regulated upstream by phosphorylation of LKB1[75]in an ERKs pathway dependent manner. Hence, summarizing the Warburg effect, due to the loss of PC1, depends on the up-regulation of ERK pathway. ERKs pathway act in a dual manner: inhibiting the LKB1-AMPK axis on the one hand while activating the mTORC1-glycolytic cascade on the other. Enhanced metabolic rates further inhibit AMPK.

7.1.4 Treatment with drugs

We next wondered whether each step of the signaling cascade generated by the loss of PC1, including also the Warburg effect, could be targeted and restored to basal levels using FDA approved drugs. We have therefore treated in separated experiments *Pkd1* KO and WT MEF cells with rapamycin, to inhibit the mTORC1 activity and with metformin and with AICAR, to activate AMPK. We also treated Ksp-Cre: $Pkd1^{flox/-}$ and $Pkd1^{flox/+}$ mice with 2-deoxy-glucose (2DG) in order to test its effects on aerobic glycolysis.

Cells' treatments with rapamycin, metformin and AICAR. We observed that the separated treatments with rapamycin, metformin and AICAR were able to partially restore the deregulated signaling cascade involving ERKs, mTORC1 and AMPK in *Pkd1* KO cells, whereas they did not have effect on the WT cells, however the analysis of the exometabolome of the treated *Pkd1* KO and WT MEF cells revealed that these drugs were not able to revert the metabolic phenotype. At variance to our expectations the exometabolome of the treated *Pkd1* KO MEF cells displayed a higher consumption of glucose and a higher production of lactate compared to the exometabolome of not treated *Pkd1* KO MEF cells. The treatments appears therefore to induce an enhanced aerobic glycolysis. This effect was partially expected for the treatment with metformin, as lactic acidosis is a known side effect of this drug[78]. Conversely, based on the inhibitory activity of rapamycin on mTORC1, we expected that this drug could partially restore the physiological levels of the aerobic glycolysis. However from our results we conclude that rapamycin in this context does not influence the Warburg effect.

Concerning AICAR, we were expecting that its inhibition role in the signaling cascade might also restore the aerobic glycolysis at physiological levels, however our results did not provide any evidence of its restoring influence. Interestingly, both rapamycin and AICAR lead to increased levels of lactate.

Treatment with 2-deoxy-glucose (2DG). We tested *in vivo* the efficacy of 2DG, an analog of glucose that blocks directly the glucose metabolism, acting as a competitor of the glucose. After simultaneous treatment of Ksp-Cre: $Pkd1^{flox/-}$ and $Pkd1^{flox/+}$mice with 2DG and ^{13}C glucose we analyzed the glucose and lactate content in liver and kidney. After treatment with 2DG we observed an improvement of the glucose metabolism in *Pkd1* KO kidneys, which showed a decrease in lactate content compared to not treated *Pkd1* KO kidneys; the glucose content in WT and KO kidneys mice remained very similar. This result indicates that the treatment with 2DG blocks the glucose metabolism without interfering with the glucose uptake. Moreover, *Pkd1* KO kidneys treated with 2DG displayed also an improvement of the kidney histology displaying a reduction of cysts' volume and consequently a reduction of ~ 10% of the kidneys' weight (the weights of not treated and 2DG treated kidney's were ~ 210 mg, ~ 190 mg, respectively). These preliminary results on the efficacy of 2DG treatment in ADPKD mouse model are very encouraging and suggest that glucose metabolism could be a promising target for a pharmacological therapy.

7.1.5 Conclusions and future perspectives

Metabolomics is widely used to discover disease biomarkers and to understand which metabolic pathways are deregulated in a particular disease[86,87]. Through this approach we have confirmed the hypothesis that PC1 acts as "metabolic regulator" identifying the Warburg effect, as the most relevant metabolic change that occurs for the loss of PC1 both in cellular model and in mouse kidneys. Moreover the metabolomic approach allowed us to test the effect of the rapamycin, metformin and AICAR treatment on the aerobic glycolysis in cellular model. Importantly NMR metabolomics allowed us to trace the effects of 2DG treatment *in vivo*. In particular 2DG is the only drug able to partially restore the physiological levels of aerobic glycolysis showing also an improvement of the renal histology.

These results allow to expand our knowledge on the molecular mechanisms regulated by PC1and to get new insights into the complex mechanism underlying cystogenesis, thus opening new perspectives in novel therapeutic strategies for ADPKD.

The study on the involvement of PC1 in the metabolic regulation is still at the beginning. In order to have a complete vision on the metabolic pathways regulated by PC1 we will investigate and compare the endometabolome, the entire set of intracellular metabolites, in *Pkd1* KO and WT MEF cells. In addition we will compare the endometabolome of the cells not treated and singularly treated with AICAR, metformin and rapamycin in order to investigate the metabolic changes that occur under these pharmacological treatments.

The analysis of the exometabolome of the *Pkd1* KO cells has highlighted some metabolic similarities with the tumor metabolism, we will therefore study a cellular model for the renal carcinoma and the kidneys deriving from a mouse model of renal carcinoma and compare them with the cellular model and the mouse model of ADPKD in order to elucidate the similarities and the differences between the two diseases at metabolic level and at the level of the associated signaling cascade.

7.2 Structure

7.2.1 NPHP1$_{1\text{-}115}$ structure

In this thesis we have exploited the solution state NMR also to solve the structure of the N-terminus of NPHP1 and to investigate a possible interaction between the N-terminus of NPHP1 and the coiled coil domain of PC1.

The first 115 amino acids of NPHP1 were predicted to be a coiled coil domain based on the analysis of its primary sequence, they were therefore expected to mediate homo-oligomerization. Unexpectedly, the recombinant NPHP1$_{1\text{-}115}$ construct did not display oligomerization features, behaving as monomer in solution, as assessed from gel filtration elution volume and from the value of the rotational correlation time τ_c, 14.056 +/- 2.457 ns, which corresponds to the values associated to a protein of 13kDa.

In addition NPHP1$_{1\text{-}115}$ showed high thermostability with a melting temperature of 65°C. Its secondary structure is characterized by a high α helical content (~45%), as estimated from the

analysis of its CD spectrum. In line with the estimated secondary structure content, the solution structure of $NPHP1_{1-115}$ displays a three- helix bundle fold. In detail, the N-terminus of NPHP1 shows a right-handed helix bundle, composed by three antiparallel α helices connected by short linkers, whereas the N- and the C- termini display high motility. The three helices are not completely coaxial, as the first and the third helix show a kink at residues Arg 19 and the Glu 99, respectively.

The structure is mainly stabilized by hydrophobic interactions with a significant number of side hydrophobic chains (Leu 103-Leu 31, Leu 24- Ile 60, Leu 17- Leu 67, Leu 100-Ile 49, Leu 100-Val 28) pointing towards the interior of the helix bundle. Most of the side chains of the polar residues are solvent exposed, therefore the surface of the protein is highly charged showing a random distribution of the charges. Only a few polar residues interact between the second and the third helix in order to stabilize the structure (Lys 57-Glu 101, Lys 64-Asp 94; His 90-Gln 68).

The function of this domain is still unknown. Helix bundles, similarly to coiled coil domains are also involved in protein-protein interactions, in particular they usually interact with cytoskeleton proteins such as actin. This is in agreement with the known role of NPHP1 in the organization of a protein complex where it regulates the actin cytoskeleton at sites of epithelial cell-cell adhesion through the binding between the NHD domain with filamin family protein of actin- binding protein[4].

NPHP mutations do not involve this domain[88]; however they are located at the C-terminus of the three helix bundle in a non structured region, suggesting that pathological mutations do not affect the protein fold, but are rather related to a functional role.

The supposed interaction between the coiled coil domain of PC1 (CC_PC1) and the helix bundle of NPHP1 could not be demonstrated in this thesis. The ^{1}H-^{15}N HSQC recorded after the addition of CC_PC1 with the tag to $NPHP1_{1-115}$ did not display any chemical shift displacement. However we observed the decrease in intensity of some peaks which could suggest a possible binding (specific or aspecific). For this reason we cannot totally exclude a possible interaction between the two domains and further experiments are needed to clarify this point.

7.2.2 Conclusions and future perspectives

In conclusion from these structural studies we learned that $NPHP1_{1-115}$ is not a coiled coil domain but a stable monomeric three helix bundle. This is the first experimental evidence that highlights the presence of this structural element at the N-terminus of NPHP1.

We could not verify a direct binding between CC_PC1 the N terminus of NPHP1 and further investigations are needed to clarify this aspect. To this aim we will perform GST pull down experiments using the helix bundle of NPHP1 and the entire C-tail of PC1 which contains the coiled coil domain. In addition, we will try to improve the purification of the CC_PC1 using a longer construct (spanning residues His 4186 -Ala 4272) that comprises the coiled coil structure and the second poly proline motif, in order to stabilize and improve the solubility of the recombinant CC_PC1.

It is known that the second poly proline motif of PC1 and the SH3 domain of NPHP1 interact with a low affinity[17], therefore, in order to test if the presence of the CC_PC1 and the $NPHP1_{1-115}$ can improve the affinity of the complex, we will titrateof the first 220 amino acids of NPHP1, which comprise both the helix bundle and the SH3 domain, into the longer PC1 construct, which comprises the coiled coil structure and the second poly proline motif (spanning residues Pro 4267 - Arg 4272), using isothermal titration calorimetry (ITC) as technique.

Part of the results of this thesis have been accepted for publication in Nature Medicine "Defective Glucose Metabolism in Polycystic Kidney disease Identifies A Novel Therapeutic Paradigm". Rowe I., Chiaravalli M., Mannella V., Ulisse V., Quilici G., Song X., Xu H., Pema M., Mari S., Qian F., Pei Y., Musco G., Boletta A.

8. Appendix

8.1 NMR appendix

8.1.1 Chemical shift

Nuclei present a fundamental quantomechanical property, the nuclear spin whose value can be a multiple of ½ and can be positive or negative. Nuclei presenting non zero nuclear spin show a nuclear magnetic moment and can be investigate by nuclear magnetic resonance (NMR).
If an external magnetic field is applied, called B_0 and conventionally described as directed along the z axis, they will align with B_0 or opposite to it: this phenomenon is called the Zeeman effect. Spins inside a static magnetic field will precess around B_0. The frequency of this precession, called Larmor frequency of the spin (ω_0), is dependent on the applied magnetic field B_0 and the γ, the gyromagnetic ratio, a property of each particular type of nucleus, according to the following equation:

$$\omega_0 = - \gamma B_0 \qquad \text{eq.12}$$

The chemical shift represents the dependence of nuclear magnetic energy level on the chemical environment. Electrons in the molecular orbitals surround nuclei and create small local magnetic fields which slightly shield nuclei from the external magnetic field. Consequently, the total magnetic field experienced by each nucleus includes the local magnetic fields induced by the currents of electrons [Figure 68].

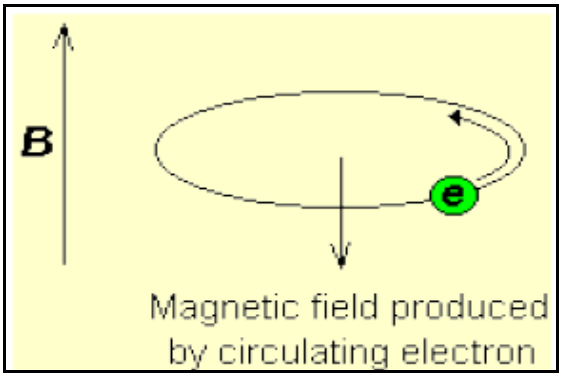


Figure 68. Electrons in the molecular orbitals surround nuclei and create small local magnetic fields which slightly shield nuclei from the external magnetic field.

The electron distribution of the nuclei with the same nuclear spin (e.g. ^{1}H, $^{1}3C$, ^{15}N) usually varies according to the local geometry (binding partners, bond lengths, angles between bonds, etc.) thus changing the local magnetic field around each nucleus. This is reflected into the slightly different energy levels for chemically different spin . The variation of resonance frequencies of the same kind of nucleus due to variations in the electron distribution is called chemical shift.

The size of the chemical shift of each nucleus (ν_i) in the molecule is given with respect to a reference frequency or reference sample (ν_{DSS}), usually DSS (2,2-dimethyl-2-silapentane-5-sulfonic acid) for proteins, to which is assigned the chemical shift of zero, expressed in Hertz (Hz), whereas the denominator (ν_0) represents the operating frequency of the magnet, expressed in Megahertz (MHz). Chemical shift (δ) is calculated from:

$$\delta = \frac{(\nu_i - \nu_{DSS})}{\nu_0} \qquad \text{eq.13}$$

Since the numerator is usually in Hertz(Hz), and the denominator in MegaHertz(MHz), chemical shift is expressed in ppm (parts per million). The resonance frequencies depend on the applied field but the chemical shift is independent of it. On the other hand the resolution of a NMR spectrum will increase with applied magnetic field, resulting in better lines (peaks) separation.

Different chemical groups have different chemical shifts due to their dependence on the chemical environment. This is one of the major information sources in NMR spectra [Figure 69].

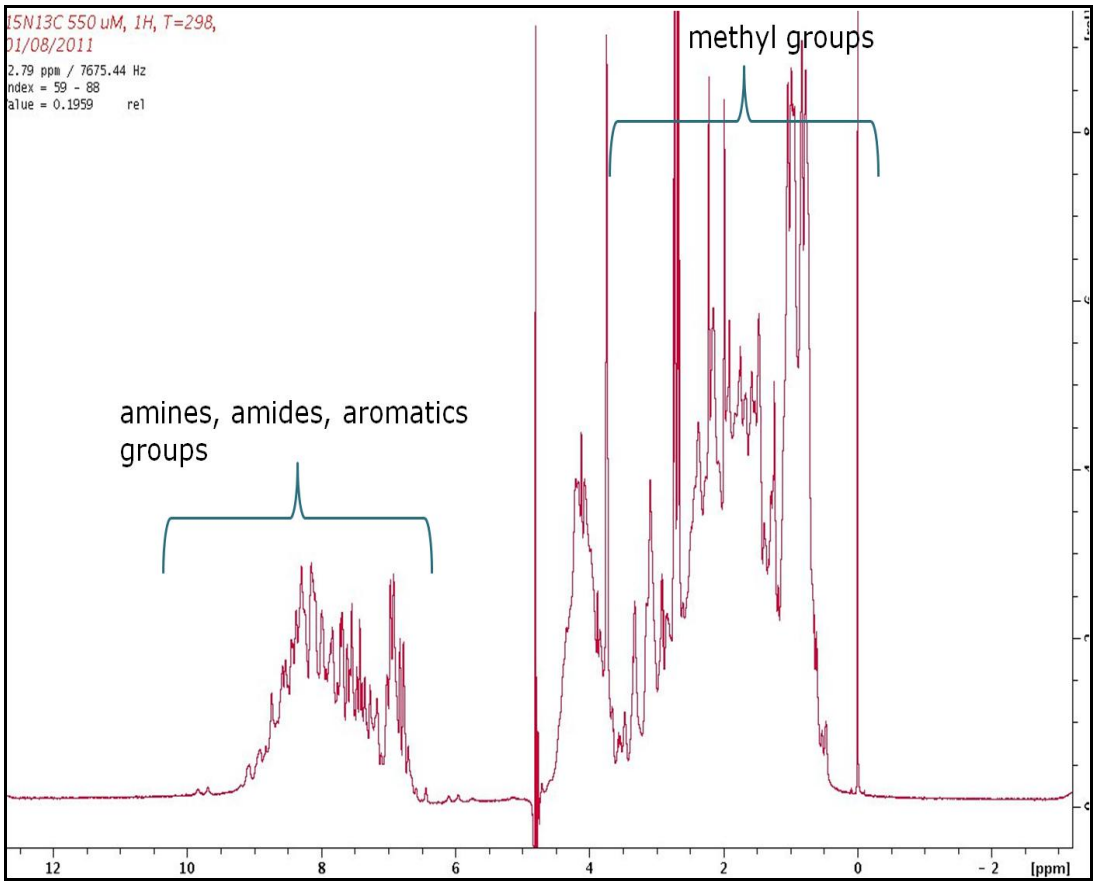


Figure 69. 1D ^{1}H protein spectrum.

8.1.2 Magnetization transfer

Nuclear spins interact through two different mechanism: scalar coupling (also named spin-spin coupling) and dipolar interactions.

Scalar coupling originates from the interaction between bond electrons, thus it occurs only between atoms covalently bound and it decreases with the number of chemical bonds that separate the two nuclei. It depends on the coupling constant J, that is expressed in Hertz (Hz), and does not depend on the applied magnetic field.

Dipolar interactions, instead, occur between nuclei that are close in the space, they can be detected only for nuclei which are less than 5 Å apart. These interactions drop off very fast with the distance, as they depend on the inverse of the sixth power of the distance ($1/r^6$). Nuclear Overhauser Enhancement (NOE) relies on dipolar interactions.

The magnetization can be transferred through these two mechanisms between nuclei of the same kind (homonuclear experiments) or between nuclei of different kinds (heteronuclear

experiments). In the homonuclear experiments the magnetization is transferred, through scalar coupling or dipolar interactions, among nuclei of the same type, generally ^{1}H. Radio frequency pulses are applied along only one channel of the NMR spectrometer. In heteronuclear experiments, magnetization is transferred among different types of nuclei, generally ^{1}H, ^{15}N and ^{13}C nuclei, always through scalar coupling. In this case two or more channels are used (one is always dedicated to ^{1}H and one or two for the heteronucleus/i) and the pulse sequence contains one or more pulse block, generally the INEPT (Insensitive Nuclei Enhances by Polarization Transfer) block, that transfer magnetization through the heteronuclear J coupling from 1H nuclei to low γ heteronuclei (i.e. ^{15}N and ^{13}C).

An example is the ^{1}H-^{15}N HSQC. The basic pulse sequence of the ^{1}H-^{15}N HSQC comprises three blocks before detection. Block A is the INEPT (Intensive Nuclei Enhances by Polarization Transfer) pulse block that creates the ^{1}H antiphase magnetization and transfers it from ^{1}H nuclei to low γ heteronuclei through heteronuclear scalar coupling. The time delay Δ/2 is set as near as possible to 1/(4J) to maximize the magnetization transfer, where J is the scalar coupling among the two nuclei (in this case J_{HN} for the ^{1}H-^{15}N HSQC). Block B is a spin echo block in which ^{15}N nuclei magnetization evolves with its chemical shift. The evolution time t1 is incremented at each repetition of the pulse sequence through the same increment of t1, in order to generate the second (indirect) dimension. ^{1}H coupling and ^{1}H chemical shift evolution during step B is removed through a 180° pulse applied only on the ^{1}H channel at mid evolution time. Block C is simply a reversal of the initial INEPT sequence (that is, all pulses and delays are applied in reverse order, with the phase of the 90° pulses inverted) except that the first pulse has been removed. The double 90° pulse applied to both nuclei at the beginning of step C transfers magnetization from ^{15}N back to ^{1}H as an antiphase magnetization, which is then refocus during the last 2(Δ/2) delay.

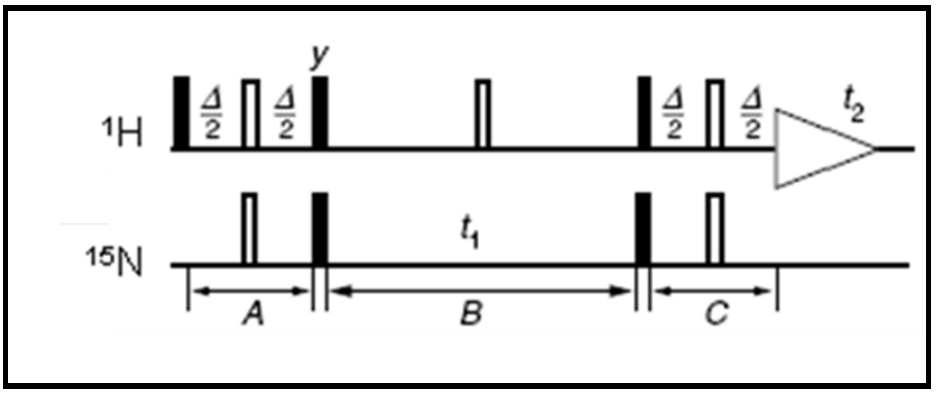


Figure 70. The basic pulse sequence of ^{1}H-^{15}N HSQC. Block A represented the INEPT, block B the evolution of magnetization on the heteronuclei, block C the reverse INEPT.

The NMR experiments can be divided in two classes according to the mechanism of nuclear spin interaction that is employed to transfer magnetization during the mixing time. In the NMR experiments used to assign the molecular resonances, magnetization is transferred through scalar coupling, hence the achieved 2D and 3D spectra show peaks for nuclei that are chemically bound to each others. To solve the structure of a molecule we need information on the spatial relationships among nuclei and 2D and 3D NMR experiments are employed in which magnetization is transferred between nuclei through dipolar interactions. The 2D and 3D NOESY spectra show only peaks for nuclei less than 5-6Å apart, irrespectively if they are chemically bound or not.

8.1.3 Nuclear Overhauser Enhancement (NOE)

The Nuclear Overhauser Enhancement (NOE) effect is due to dipolar interactions between nuclei. Dipolar interactions drop off very fast with the distance, hence only a proton-proton distance (^{1}H-^{1}H NOE) can occur with a distance 5-6Å. The intensity of an NOE peak depends on the inverse of the sixth power of the average distance between them (<r>) and the correlation function $f(\tau_c)$, which describes the modulation of the dipolar coupling by a stochastic process with an effective correlation time τ_c (eq. 14)

$$NOE \propto (1 / \langle r^6 \rangle) f(\tau_c) \qquad \text{eq.14}$$

^{1}H-^{1}H NOE distances in proteins can be classified as:

- intra-residue: between the ^{1}H protons of the same residue;

- sequential: between the backbone $^{1}H^{N}$ protons or between $^{1}H^{N}$ and aliphatic protons in consecutive residues;

- medium-range: all non sequential distances with a segment of five consecutive residues;

- long-range: all other inter-residue distances among protons.

The secondary structures, α-helices and β-strands, force protons in well defined spatial relationships. Consequently some ^{1}H-^{1}H NOE patterns are indicative of the stretches of secondary structure in protein. For example α-helices are characterized by proximity between the residues i and i+3 and the residues i and i+4, therefore in the ^{15}N-HSQC-NOESY spectrum NOEs between the HN_i and HN_{i+2} and HN_{i+3}, HN_i and $H\alpha_{i+3}$ would be visible. In β strands short-medium range distances are usually absent because the individual strands consist of almost fully extended polypeptide segments. Residues folded in β-sheets are rather characterized by short-sequential and long-range backbone ^{1}H-^{1}H NOEs.

8.1.4 Monodimensional (1D) NMR experiments

1D ^{1}H experiment is the simplest NMR experiment and it consists of two sections: preparation and detection [Figure 71A]. During preparation the sample is subjected to an appropriate radiofrequency pulse (or series of pulses), while during detection the resulting signal is recorded. In the simplest case the preparation is a 90° pulse which rotates the equilibrium magnetization Mz onto the y axis (My). After this pulse each spin precesses with its own Larmor frequency around the z axis and induces a signal in the received coil, named free induction decay (FID). Usually the experiment is repeated several times and the data are summed up to increase the signal to noise ratio. After summation the data are Fourier transformed to yield the final 1D spectrum [Figure 71B].

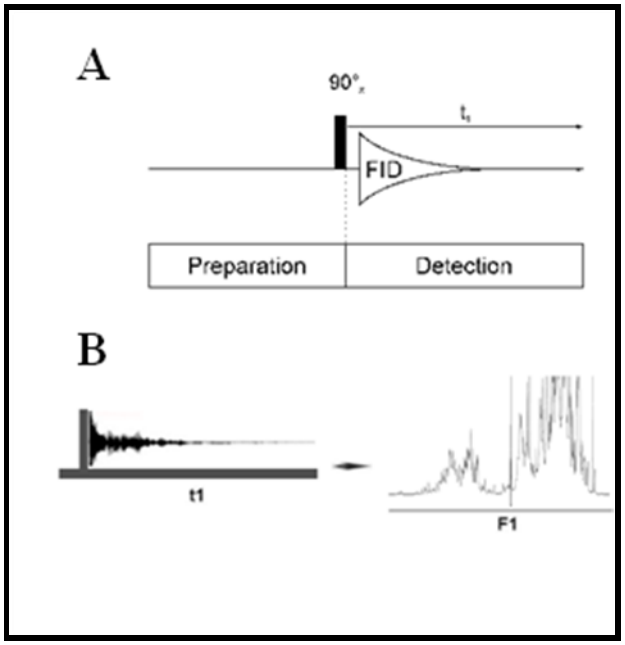


Figure 71. 1D NMR spectrum A. Schematic pulse sequence of 1D spectrum B**.** FID, a time-domain function, is transformed in a frequency-domain function through the Fourier transform.

Proteins contain a large number of hydrogen atoms, so their 1D spectra contain peaks heavily overlapping [Figure 69]. Multidimensional spectra spread these signals over another dimension and allow to correlate the frequencies of different nuclei.

8.1.5 Bidimensional (2D) NMR experiments

Bidimensional (2D) spectra contain two additional blocks with respect to monodimensional spectra: the evolution and mixing time [Figure 72A]. The pulse sequence is repeated *n* times, incrementing the *t1* time delay of the evolution block at each repetition. The *n* FIDs generated by the pulse sequence repetitions will show a signal with a decaying sinusoidal trend depending on *t1* (as opposed to the sinusoidal trend of each FID, which depends on *t2*). The Fourier transformation (FT) can be applied twice to a function depending on two time variables, generating a bidimensional spectrum with two independent frequency dimensions, Ω1 and Ω2, corresponding to *t1* and *t2*, respectively [Figure 72B]. The correlation between

the two dimensions is built during the mixing time by means of adequate radio frequency pulses which drive the magnetization transfer among interacting nuclear spins.

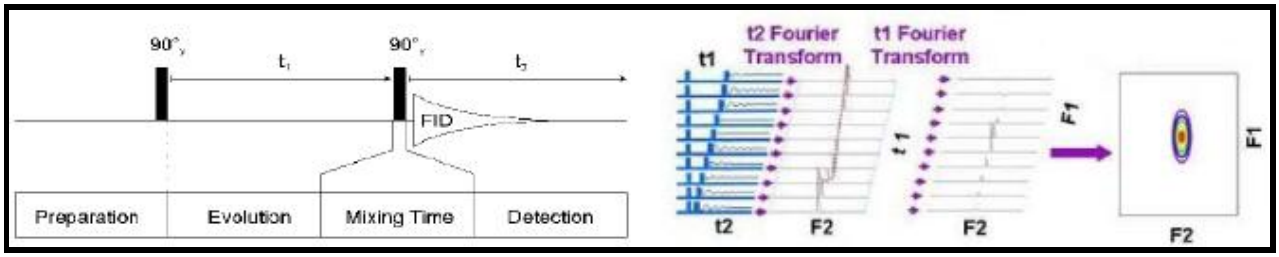


Figure 72. 2D NMR spectrum A. Schematic pulse sequence of 2D spectrum NMR B**.** Schematic representation of the FT on a series of FIDs depending on two time variables. 2D spectra are usually presented as contour plots, as shown in the box after the white arrow.

^{1}H ^{15}N HSQC. ^{1}H ^{15}N HSQC is a bidimensional heteronuclear spectrum that correlates the hydrogen of an amide group to the attached nitrogen. Magnetization in transferred via J-coupling J_{HN}=90Hz from hydrogen to nitrogen. Chemical shift evolves on the nitrogen and magnetization is then transferred back to hydrogen for detection [Figure 73]. The position of the peak inside the spectrum (chemical shift) depends on the chemical environment of each HN.

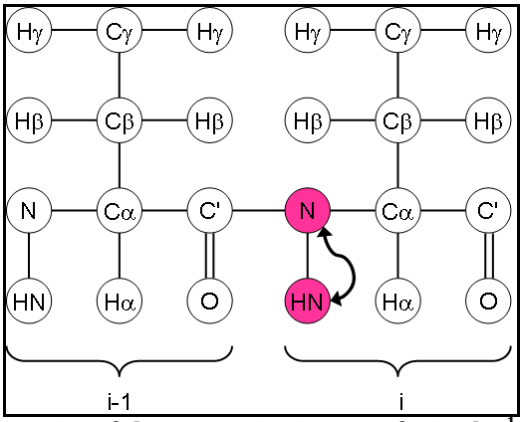


Figure 73. A schematic representation of the magnatization transfer in the ^{1}H-^{15}N HSQC: in pink the atoms observed and in black arrow how the magnetization is transferred.

For this reason this spectrum is considered the protein fingerprint. In this spectrum are also visible the peaks of the side chains of glutamine, asparagine and tryptophan [Figure 74].

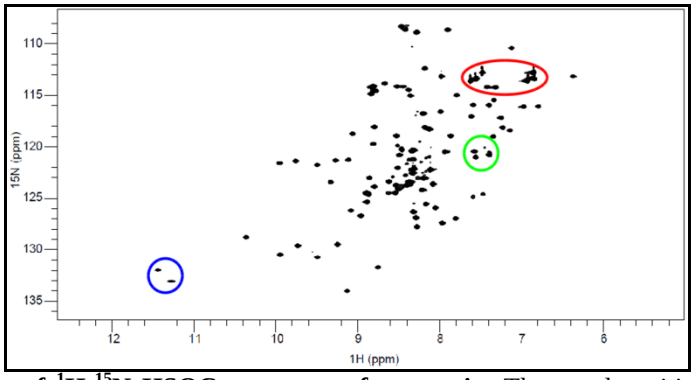


Figure 74. An example of ^{1}H-^{15}N HSQC spectrum of a protein. The usual position of the side chains of histidine and tryptophan are highlighted with the blue circle, the side chains of asparagine and glutamine in red one, each group HN group display two peaks with the same ^{15}N resonance and two different ^{1}H chemical shifts. The arginine side chains Nε-Hε are highlights with the red circle.

^{1}H-^{1}H TOCSY. It is a 2D homonuclear experiment. Magnetization is transferred along the ^{1}H nuclei belonging to the same molecule or aminoacidic residue, through homonuclear scalar coupling and then it is detected from all the ^{1}H nuclei. The sample is not labelled and this spectrum is useful to recognize the hydrogen belonging to the same spin system. This spectrum is acquired both in H_2O and in D_2O solvent. The spectrum acquired in D_2O results to be simple as protons of HN groups exchange with 2H of the solvent and thus disappear ingfrom the spectrum when thery are not proteceted.

^{1}H-^{13}C HSQC. This spectrum is the carbon equivalent of the ^{1}H-^{15}N HSQC. Magnetization is transferred from hydrogen to carbon and then back to the hydrogen for detection [Figure 75]. All ^{1}H ^{13}C group are observed.

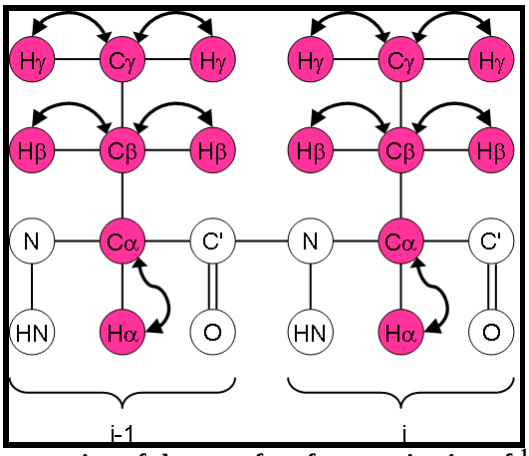


Figure 75. The schematic representation of the transfer of magnetization of ^{1}H-^{13}C HSQC: in pink the atoms observed and in black arrow how the magnetization is transferred.

^{1}H-^{1}H NOESY. On the contrary of the previous described spectra, in ^{1}H-^{1}H NOESY magnetization is transferred between all proton through the dipolar coupling interactions among the nuclear spins that are close in the space [Figure 76]. This spectrum is helpful both to assign resonances and to get structural restraints between protons.

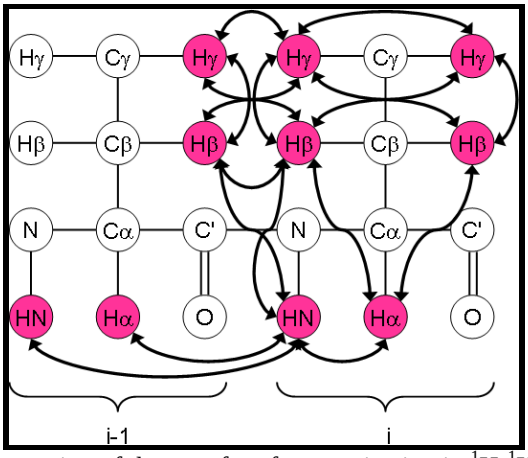


Figure 76. The schematic representation of the transfer of magnetization in ^{1}H-^{1}H NOESY: in pink the atoms observed and in black arrow how the magnetization is transferred.

8.1.6 Three -dimensional (3D) NMR experiments

The pulse sequence for a 3D experiment can be easily constructed from the 2D one inserting an additional indirect evolution time and a second mixing period between the first mixing period and the direct data acquisition. Each of the different indirect time periods (t1, t2) is incremented separately [Figure 77].

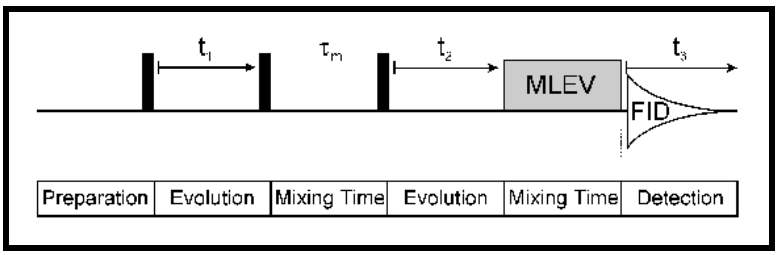


Figure 77. The schematic pulse sequence of 3D spectrum.

A 3D spectrum can be seen as a 2D spectrum which was extended into a third dimension in order to have a spread of the peaks.

Most NMR spectra used for triple resonance backbone assignment have a ^{1}H, ^{15}N and ^{13}C dimension each. Several other types of spectra as 3D NOESY and HCCH- TOCSY spectra have two ^{1}H dimensions and one ^{15}N or ^{13}C dimension.

HNCA. Magnetization in transferred from ^{1}H to ^{15}N and then via the $J_{C\alpha}$ J-coupling ($J_{NC\alpha}$=20.8Hz) to the 13Cα and then back again to ^{15}N and ^{1}H for detection [Figure 78]. Chemical shift evolves for ^{1}HN as well ^{15}NH and $^{13}C\alpha$, resulting in a 3D spectrum. The amide nitrogen is coupled both to the Cα of its own residue and to the Cα of the preceding residue, both these magnetization transfers occur and peaks for both Cα nuclei are visible in the spectrum. The coupling to the directly bonded Cα is stronger, hence the corresponding peaks will appear with stronger intensity.

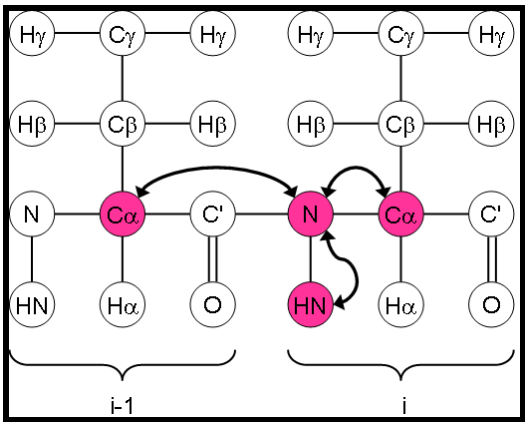


Figure 78. The schematic representation of the magnetization transfer in HNCA: in pink the atoms observed and in black arrow how the magnetization is transferred.

HNCACB. Magnetization is transferred from $^{1}H\alpha$ and $^{1}H\beta$ to the $^{13}C\alpha$ and $^{13}C\beta$ respectively, and then from the $^{13}C\beta$ to the $^{13}C\alpha$. From $^{13}C\alpha$ the magnetization is transferred to $^{15}N_i$ from both $^{13}C\alpha_i$ and $^{13}C\alpha_{i-1}$. Thus for each NH four peaks are visible: the Cα and the Cβ from the same residue as the NH which display a strong intensity and the Cα and the Cβ of the preceding one which display weak intensity.

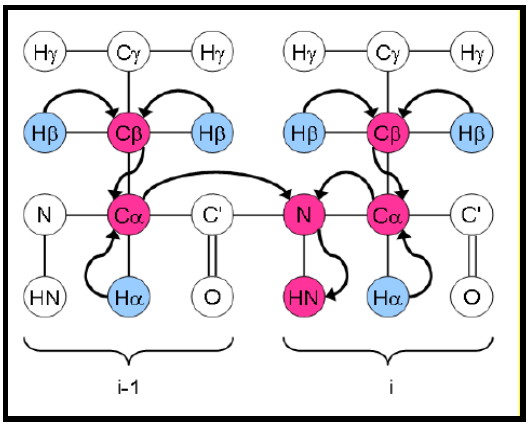


Figure 79. The schematic representation of the magnetization transfer in HNCACB: in pink the atoms observed, the atoms in which the magnetization flows but they are not observed and in black arrow how the magnetization is transferred.

HN(CO)CACB. The magnetization is transferred from the Hα and Hβ to $^{13}C\alpha$ and $^{13}C\beta$ respectively and then from the $^{13}C\beta$ to $^{13}C\alpha$. From here it is transferred to ^{13}CO, then to ^{15}NH and finally to ^{1}HN for detection. Chemical shift evolves simultaneously on the $^{13}C\alpha$ and $^{13}C\beta$

so they appear in one dimension. Chemical shift evolves in the other two dimension the ^{15}NH and ^{1}HN, but it does not evolve on ^{13}CO. For each NH group the peaks of the Cα and Cβ of the preceding residue are visible.

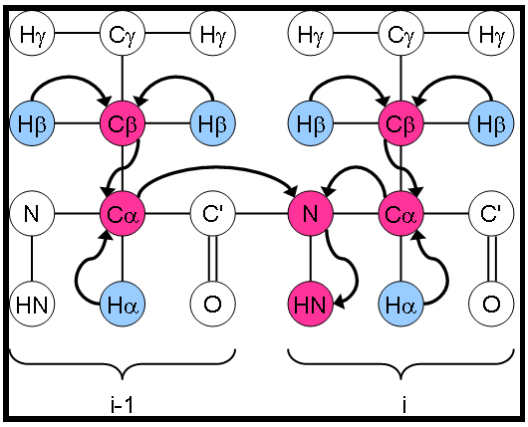


Figure 80. The schematic representation of the magnetization transfer in HN(CO)CACB: in pink the atoms observed, in blue the atoms in which the magnetization flows but they are not observed and in black arrow how the magnetization is transferred.

HNCO. Magnetization is transferred from ^{1}H to ^{15}N and then selectively to the carbonyl ^{13}C via J-coupling J_{HN-CO}=15 Hz, and then back for detection. For each HN group there is only one CO peak visible belonging to the preceding residue. This spectrum is used to obtain the CO chemical shifts which are employed in software as TALOS+ to help the prediction of dihedral angles and of the secondary structures.

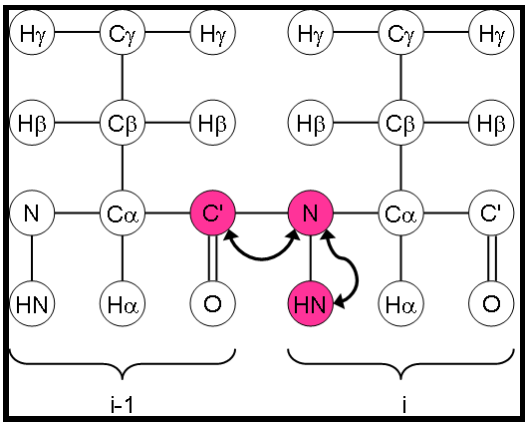


Figure 81. The schematic representation of the magnetization transfer in HNCO: in pink the atoms observed and in black arrow how the magnetization is transferred**.**

HNHα. The HNHα is a 3D spectrum in which magnetization is transferred from ^{1}H and ^{15}N and then selectively to the backbone $^{1}H\alpha$ via three-bonds J-coupling $^{3}J(HNH\alpha)$[57]. Magnetization is then transferred back via ^{15}N to ^{1}H for detection. Hence for each amide group the spectrum displays one diagonal peak and one cross peak, corresponding to the resonance of the amide proton and of the α proton of the same residue, respectively.

CC(CO)NH. Magnetization is transferred also in this case through scalar coupling from the side-chain ^{1}H nuclei to the attached ^{13}C nuclei. Then isotropic ^{13}C mixing is used to transfer magnetization between the ^{13}C nuclei. From here, magnetization is transferred to ^{13}CO, and then to $^{15}N^{H}$ and finally to $^{1}H^{N}$ for detection. Chemical shift evolves simultaneously on all side-chain ^{13}C nuclei, as well as $^{15}N^{H}$ and $^{1}H^{N}$ nuclei, resulting in a 3D spectrum. For each NH group there is one peak for every ^{13}C nucleus of the side chain of the preceding residue.

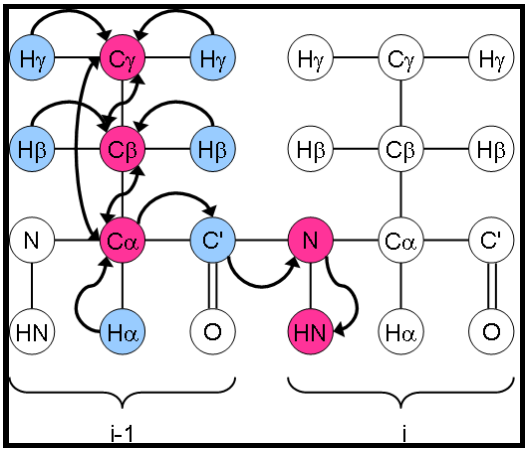


Figure 83. The schematic representation of the magnetization transfer in CC(CO)NH: in pink the atoms observed, in blue the atoms in which the magnetization flows but they are not observed and in black arrow how the magnetization is transferred.

HCC(CO)NH. Magnetization is transferred from the ^{1}H nuclei to their attached ^{13}C nuclei. Then isotropic ^{13}C mixing is used to transfer magnetization between the ^{13}C nuclei. From here, magnetization is transferred to ^{13}CO, then to $^{15}N^{H}$ and finally to $^{1}H^{N}$ for detection. Chemical shift evolves simultaneously on all side-chain ^{1}H nuclei, as well as $^{15}N^{H}$ and $^{1}H^{N}$ nuclei, resulting in a 3D spectrum with one nitrogen and two hydrogen dimensions. For each

HN group the ^{1}H chemical shifts from the whole side chain of the preceding residue are visible.

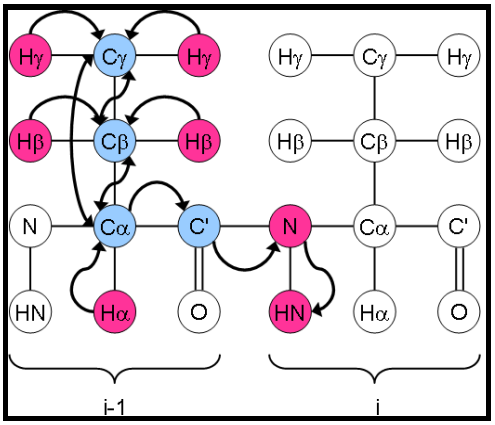


Figure 84. The schematic representation of the magnetization transfer in HCC(CO)NH: in pink the atoms observed, in blue the atoms in which the magnetization flows but they are not observed and in black arrow how the magnetization is transferred.

^{1}H-^{15}N-HSQC-NOESY (3D). Magnetization is transferred through dipolar interactions between all protons using the NOE effect (paragraph 9.1.3). Then magnetization is transferred to neighbouring ^{15}N nuclei and back to ^{1}H for detection. For each NH group, NOE peaks of $^{1}H^{N}$ atom to all other ^{1}H atoms close by are visible.

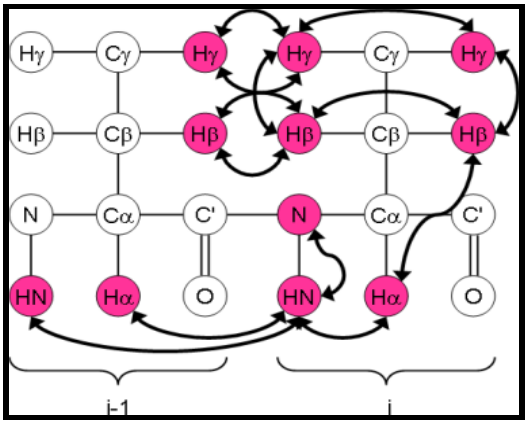


Figure 85. The schematic representation of the magnetization transfer in ^{1}H-^{15}N-HSQC-NOESY: in pink the atoms observed and in black arrow how the magnetization is transferred.

^{1}H-^{13}C-HSQC-NOESY (3D). Magnetisation is transferred between all protons exploiting the NOE effect (paragraph 9.1.3). Then magnetisation is transferred to neighbouring ^{13}C nuclei and back to ^{1}H for detection. Transfer either occurs to/from the aliphatic ^{13}C nuclei or to/from aromatic ^{13}C nuclei (but not both) depending on the ^{13}C frequency used during the pulse sequence. For each CH group, NOE peaks of the $^{1}H^{C}$ atom to all other ^{1}H atoms nearby are visible. A double set of peaks usually arises from a CH_2 group in which the two ^{1}H have different chemical shifts.

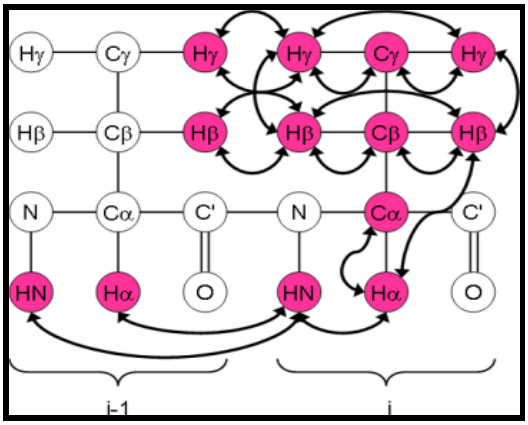


Figure 86. The schematic representation of the magnetization transfer in ^{1}H-^{13}C-HSQC-NOESY: in pink the atoms observed and in black arrow how the magnetization is transferred.

8.1.7 Structure calculation

The standard procedure to determine the solution structure of a protein from NMR data is to minimize an objective function which incorporates experimental data (intensities of assigned NOE cross-peaks, dihedral angles, residual dipolar couplings) and physical knowledge (quantified by a molecular dynamics force field). We used ARIA 2.1.3 [62] to calculate the solution structure of $NPHP1_{1\text{-}115}$.

Distances are derived from NOE cross-peak volumes (V) to generate restraints for structural calculations by converting this information to a target distance D using the following equation:

$D = (\alpha^{-1} V)^{-1/6}$ eq. 15

The calibration factor α depends on the properties of the system under investigation and on the experimental setup, and cannot be measure directly. It is estimated for each spectrum by ARIA during a calibration step. Rather than a definite value, the distance is restrained to an interval comprised in a range defined by a lower distance bound, L=max(0, D–Δ), and an upper distance bound, U=D+Δ (where $\Delta=0.12D^2$).

ARIA introduces the concept of ambiguous distance restraint (ADR), which allows to combine more than one assignment in a single restraint according to equation 16

$$d_{ambig} = \left(\sum_i d_i^{-6}\right)^{-1/6} \quad \text{eq. 16}$$

where the overall distance d_{ambig} is treated as having contributions from all possible assignments, each contributing with a distance d_i.

The assignment possibilities are filtered consistently with the molecular conformations calculated. Each cycle of ARIA iterative procedure is constituted by three steps:

- independent calibration of each NOESY spectrum in order to calculate the calibration factor α

- detection of wrong assignments and noise peaks restrains are subjected to violation analysis. Each restraint distance's bounds (L and U) are compared to the corresponding distances found in the ensemble. "Violation" occurs when the distance found in the structures deviates from the bounds by more than a user-defined *violation tolerance* in a percentage of structures higher than the user-defined *violation threshold*

- calculation of the new ensemble of structures. Restraints from different spectra are merged and multiple structures compatible with restraints and empirical information are produced by means of simulated annealing. The temperature is increased in order to allow the sampling of high energy conformations aiming to overcome energetic barriers and escape local minima. Subsequently, the temperature is slowly reduced searching for the conformation corresponding to the global energetic minimum.

From the second iteration on, prior the calibration the program performs an analysis of the

previously calculated structures. The conformers with lowest energy ("converged conformers") are selected and used for the further analyses. Finally, the best structures of the final iteration are subjected to a refinement in a shell of water molecules via a short trajectory with a full MD force field.

Finally ARIA uses the programs WHAT IF, PROCHECK and PROSA II to evaluate the quality of both the last interaction structure ensemble and the solvent-refined ensemble, For every program separate report files are stored.

8.2.Metabonomics and Metabolomics appendix

8.2.1 Metabonomics and metabolomics

Metabonomics and metabolomics are the youngest "omics" sciences that study the metabolome[89]. The metabolome is constituted by about 2200 small molecules with low molecular weight (<1000 Da) involved in highly complex and organized biochemical networks. They include metabolic substances and products such as lipids, small peptides, vitamins and other protein cofactors[90]. Moreover the metabolome is dynamic and it responses more quickly than genome, trascriptome and proteome to changes caused by drugs, environment and disease, but at the same time it can complement information derived from others " –omics" sciences to have a global view on the system biology [90].

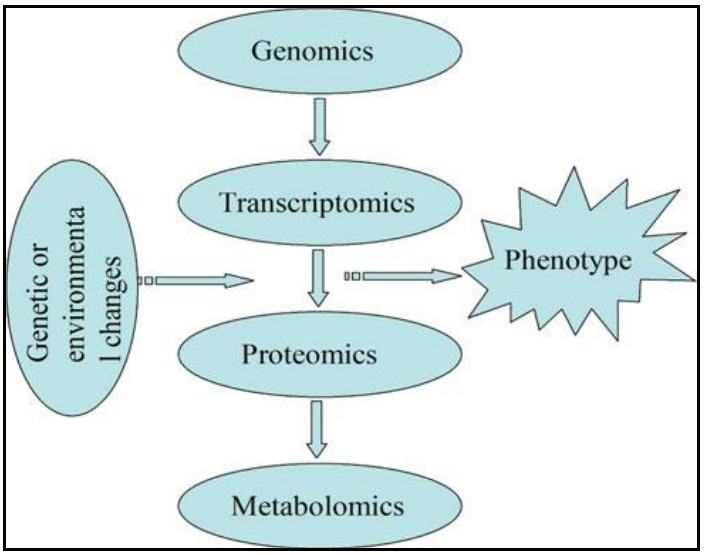


Figure 87. The relationship between the genomics, transcriptomics, proteomics and metabolomics. [Figure from Ma Y. et al. Mol. Biol. Rep. 2012]

Nowadays the terms metabolomics and metabonomics are often interchangeable referring to multi-component analysis of metabolites in a biological system. Originally the two definitions were slightly different: metabolomics was defined as the global analysis of all metabolites in a sample, while metabonomics comprised the analysis of metabolomic responses to drugs or disease[91].

Through metabolomics we can study biofluids, cells or tissues in order to discover biomarkers for a particular disease or to understand the behavior of a particular tissue and/or biofluid under defined stimuli. For these reasons the data must be reproducible, hence the accurate and correct manipulation and storage of biological sample are essential and they represent a critical step[92]. Moreover, another essential and critical step in metabolomics studies is the experimental design. During this step the experimental conditions and the adequate sample size are decided in order to minimize the variability due to sample manipulation.
The study of the metabolome can be usually divided in four steps [93]:

- Metabolic profiling: identification and quantification of a selected number of predefined metabolites related to a specific metabolic pathway;

- Metabolic fingerprinting: global and rapid analysis to classify the sample, usually applied as a screening tool to discriminate samples belonging to different classes;

- Metabolic footprinting: analysis of the (exo)metabolites secreted/excreted by an organism, such as conditional medium in the presence of cells;

- Metabolite target analysis: qualitative and quantitative analysis of one, or several, metabolites belonging to a specific metabolic pathway.

The pipeline to design, perform and analyze metabolomic data is described in the following[94][Figure 88].
The two most widely used analytical tools for metabolomic studies are: mass spectrometry (MS) and nuclear magnetic resonance (NMR). These two complementary techniques have different experimental and technological features and applications depending on the specific questions and focus of the experiment [92].

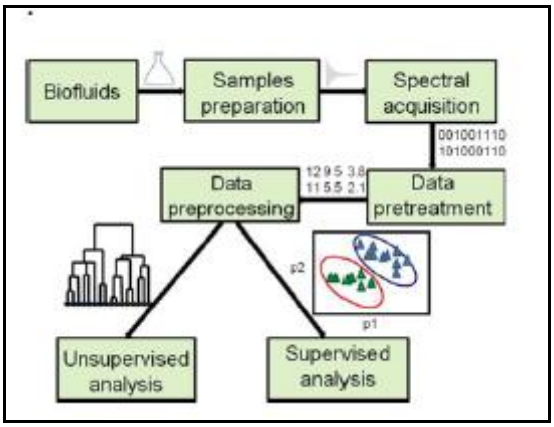


Figure 88. Schematic workflow for metabolomic analysis. [Figure from Smolinska A., et al., Analytica Chimica Acta 2012].

8.2.2 Mass spectrometry

Mass spectrometry (MS) is more sensitive than NMR but the metabolites must be separated from biological fluid before detection, typically by using high-performance liquid chromatography (HPLC) or gas-chromatography (GC) . Recently, technological advances in separation, ion sources and mass analyzers have increased the sensitivity, selectivity, specificity and speed of metabolites detection and identification by MS. Manipulation of samples can be misleading because compound moreover information could be lost because this technique disrupts the structures and the interactions of molecular complexes[82].

The global analysis of metabolites by MS should consider: the efficient and unbiased extraction of metabolites from the sample, the separation or fractionation of the analytes by chromatography, the ionization of the molecules, detection of signals and identification of them.

There are many software developed to analyze MS spectra that are capable to filter noise, detect peaks (find the peaks corresponding to the same compounds, mass-to-charge ratio, m/z), alignment and normalization (adjusting the intensities within each sample run by reducing the systematic error) [92].

8.2.3 Nuclear magnetic resonance (NMR)

NMR spectroscopy is a non-destructive, non-invasive, highly selective and fast method able to analyze samples without previous extensive pre-processing and separation, but it is less sensitive than MS. The results of NMR measurements have proven highly replicable across laboratories and instruments [92]. NMR can provide measurements for different types and sizes of both polar and non-polar molecules simultaneously at a particular time point providing a metabolic "snapshot". In addition the instruments are highly versatile and users can obtain spectral information for different nuclei, typically for ^{1}H, ^{13}C, ^{15}N, ^{31}P nuclei, both in liquids and in solids using high-resolution solution NMR spectroscopy or high resolution magic angle spinning (HRMAS), respectively. The employment of HRMAS for the metabolic analysis on the whole cells and tissues avoids losses due to the extraction protocol compared with the solution NMR.

8.2.4 NMR spectra

1D spectra. Generally the metabolomic samples, such as urine, blood, tissue and cells, contains both low and high molecular weight components and an high amount of water.
The large NMR signal of the water has to be eliminated using appropriate standard NMR solvent suppression methods, generally the most used methods is the presaturation. In Figure 89B the schematic representation of a pulse sequence of a 1D spectrum with presaturation during preparation.

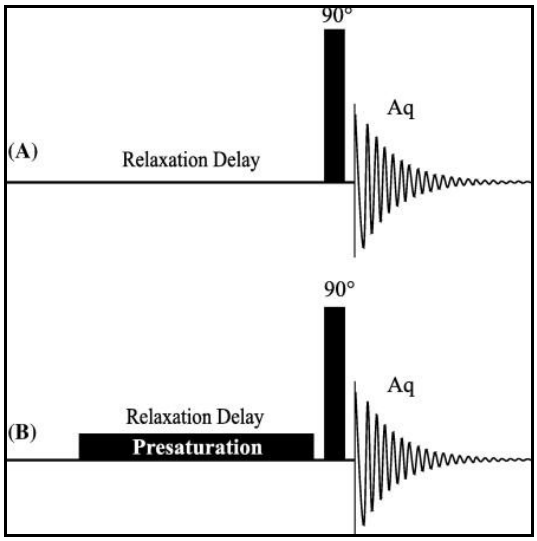


Figure 89. The schematic pulse sequence of the 1D ^{1}H spectrum A without presaturation B with presaturation**.**

For the presence of the low and high molecular weight components in metabolomic samples the monodimensional spectrum presents a wide range of signal line widths [Figure 90]: broad lines from proteins and lipoproteins signals strongly contribute to ^{1}H NMR spectra, with high signal overlap deriving resulting in overcrowded spectra.
Often spectra are acquired through edited pulse sequences on the basis of molecular diffusion or NMR relaxation in order to enhance the contribution of macromolecules or the signal of small molecules[95]. Generally for each sample three 1D ^{1}H NMR spectra are acquired:

- 1D nuclear Overhauser effect with presaturation (1D NOESY-presat): unedited spectrum in which all metabolites are detected [Figure 90 A]

- 1D Carr-Purcell-Meiboom-Gill (CPMG) spin-echo sequence: edited for T2 relaxation, to allow detection of only small molecules whereby the broad signals due to the contribution of lipids and proteins are removed [Figure 90 B]

- 1D diffusion spectrum: edited for the diffusion coefficient of the molecule in which only the macromolecules, as lipids and proteins, are detected, removing signals from small molecules[Figure 90 C].

In the metabolomic samples a reference compounds is always added as chemical shift reference: 3-trimethylsilylpropionic acid (TSP) or 2,2-dimethyl-2-silapentane-5- sulfonate sodium salt (DSS) for samples in aqueous solvent, tetramethylsilane (TMS) for ones in organic solvents[95].

The peak integral of each molecule is related both to the concentration of the substance in the sample and to the number of protons that gives rise to the peak[95]. Hence the absolute concentration can be calculated from the peak integral in several manners:

- Using the internal standard with a known concentration;

- Using an external standard with known concentration.

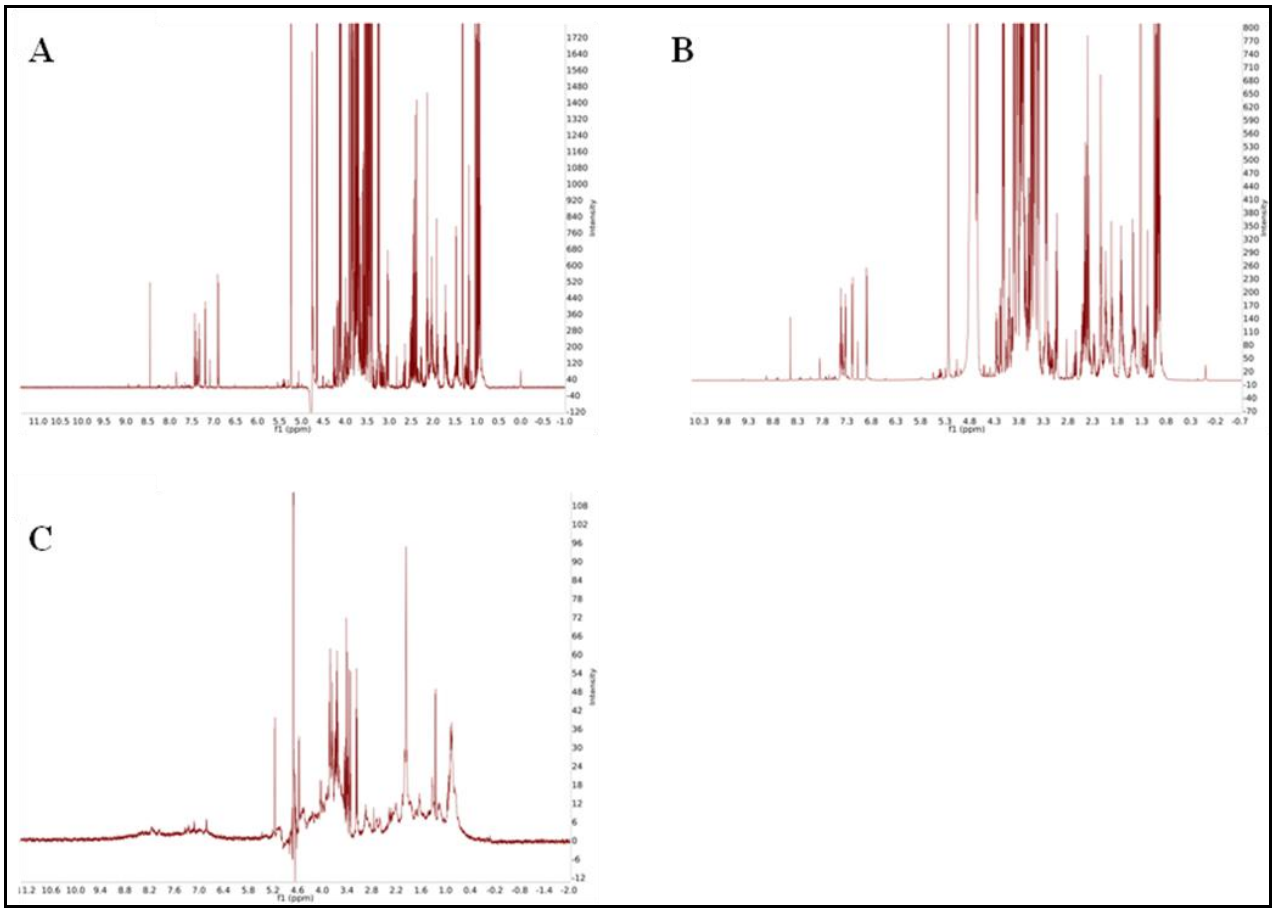


Figure 90. The 1D ^{1}H NMR spectra generally acquired for metabolomics analysis. A. 1D ^{1}H Noesy **B.** 1D ^{1}H cpmg edited for T2 **C.** 1D ^{1}H diffusion edited for the diffusion coefficient.

2D NMR spectra.To identify the molecules of a metabolic sample the 1D spectra are not sufficient, therefore 2D spectra are acquired to increase the signal dispersion and elucidate the connectivity between signals[95]. The 2D experiments generally acquired are:

- J-resolved: this experiment yields the information on the multiplicity and coupling patterns of resonance [Figure 91 A];

- TOtal Correlation Spectroscopy (TOCSY): this experiments provides spin-spin coupling connectivity identifying the hydrogens belonging to the same spin system. [Figure 91 B];

- Heteronuclear Single Quantum Coherence (HSQC): the spectrum can be used to detect the hydrogens and heteroatomes(^{15}N and ^{13}C) linked by one bond, usually for metabolomic analysis the ^{1}H ^{13}C HSQC is acquired. The magnetization is transferred from ^{1}H to ^{13}C though chemical bonds, hence in the spectrum all carbons connected to a proton are displayed [Figure 91 C].

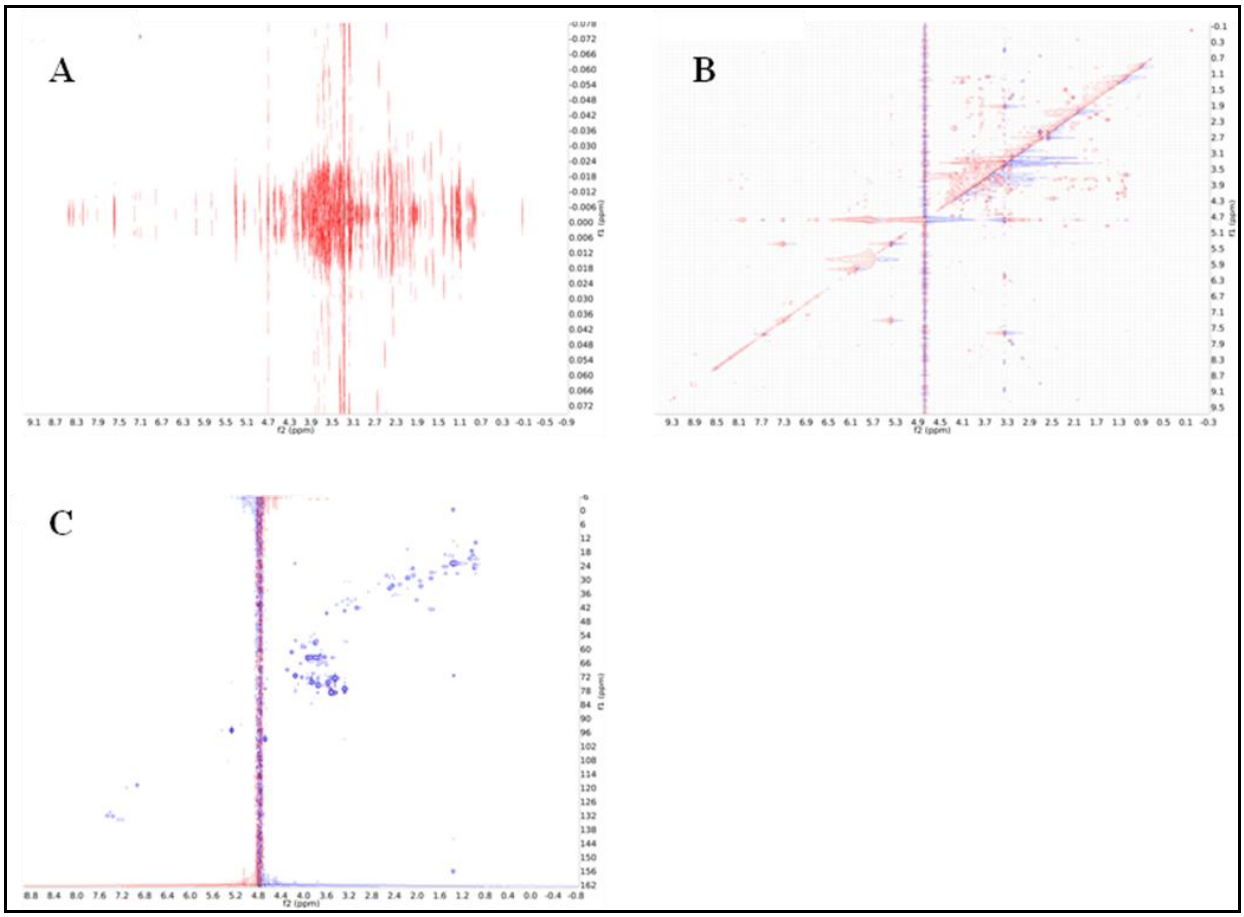


Figure 91. The 2D NMR spectra generally acquired for metabolomics analysis. A. 2D J-resolved **B.** 2D ^{1}H-^{1}H TOCSY **C.** 2D ^{13}C- ^{1}H HSQC.

8.2.5 Data preprocessing

Data preprocessing is an intermediate step between row spectra and data analysis. The main objective of data preprocessing is to transform the data in such a way that the samples in the dataset are comparable, to facilitate data analysis. This step, like the experimental design, is crucial[94].

Spectra from MS or NMR analysis are processed in order to remove the noise deriving from spectra acquisition through baseline removal. In NMR and MS spectra the baseline distortions can be corrected in different ways through mathematical methods based on polynomial fitting.

Usually spectra overlays do not show perfect alignment, this phenomenon is often caused by instrumental factors, changes of the pH and temperature. The MS spectra are generally aligned on the time scale choosing a linear variables for each spectrum, while the alignment of the NMR spectra is set on an internal reference signal. This type of alignment is often not sufficient as it can only remove the global shift but not the local one, for this reason there are many software that implement the local alignment of the spectra, as MestreLab software [48].

A further preprocessing step of NMR spectra is the binning. In binning the NMR spectra are divided into segments (so called bins or buckets) and the total area within each bin is calculated to represent the original spectrum. Therefore, some minor peaks shifts can be removed by spectral binning. There are many types of spectral binning, the most common type is an equidistant binning of 0.04 ppm. This indicates that every spectrum is divided into evenly spaced integral regions with spectral width of 0.04 ppm. The disadvantage of this equidistant binning is the lack of flexibility of the boundaries. If a peak is split into two bins, the peak frequency may significantly influence the data analysis. To prevent the peaks being split by the boundaries of bins, methods which are based on non-equidistant spacing have been proposed, such as adaptive-intelligent binning (AI-binning) [96]and Dynamic adaptive binning[86]. These methods take into account peak positions and thus focus on a better peak definition. Therefore, it is possible to obtain binning where one bin covers only a complete peaks.

Normalization and scaling transformation. This step of preprocessing tries to account for the variations of the overall concentrations of the samples. The main aim is to make all the samples comparable with each other by for instance removing or minimizing the total amount of material for samples or metabolite dilution.

The MS spectra are generally normalized for an internal known standard added to sample before or after extraction[95], whereas the most common method to normalize the NMR spectra is the "integral normalization". This standard method normalizes the individual spectra to constant total integrated intensity across the whole profile[94].

The concentration of metabolites in biofluids can span different orders of magnitude. Moreover, the variation in metabolites level is often linked to concentration in such a way that metabolites with highest concentrations strongly influence data analysis. For this reason it is very important to scale the metabolites concentration before analysis to avoid a bias from the most abundant metabolites.

There are many scaling methods, the most commonly used are the mean-centering, the autoscaling and the pareto scaling[97].
The mean-centering method converts. all the concentrations into fluctuations around zero instead. The method cannot be sufficient if the data contain sub populations with different variability, for this reason the method is usually used in combination with other scaling methods.

$$\tilde{x}_{ij} = x_{ij} - \bar{x}_i \qquad \text{eq. 17}$$

Autoscaling, also called unit or unit variance scaling, uses the standard deviation as the scaling factor through equation 18. After the application of this scaling all metabolites have a standard deviation equal to one.

$$\tilde{x}_{ij} = \frac{x_{ij} - \bar{x}_i}{s_i} \qquad \text{eq. 18}$$

Pareto scaling is very similar to autoscaling, the square root of standard deviation is used as the scaling factor. In this way the large fold changes are decreased more than the small fold changes, consequently the large fold changes are less dominant compared to the raw data.

$$\tilde{x}_{ij} = \frac{x_{ij} - \bar{x}_i}{\sqrt{s_i}} \qquad \text{eq. 19}$$

8.2.6 Data analysis

Multivariate statistical methods are used to analyze the great amount of data deriving from the metabolomic analysis, in order to provide the maximum information out of the system under study. Analysis can be performed with supervised and unsupervised methods.

Unsupervised analysis. Usually, this kind of analysis is used as a method to explore and discover the overall structure of the data without any additional information, finding trends and groups in the data. The most used methods is the principal component analysis (PCA).
PCA converts the multidimensional data space into a low-dimensional model plane. This technique expresses most of the variance in a dataset using a smaller number of factors, the so called principal components (PCs), which account for a sequentially minor proportions of the total variance between analyzed samples[96].
Each PC is orthogonal to the other PCs and therefore contains different information. The variation in spectral data is described by a few PCs compared to the number of original variables, thus enabling the detection of trends and groupings in the data.
The result of PCA analysis is displayed through two plots: the score and the loading plot.
The score plot displays the samples as points in the new plane and shows how they are related to each other. Hence those points which cluster together have similar profiles, on the contrary, objects that lie far away are characterized by different properties.
The loading plot describes the way in which the old variables are linearly combined to new variables (PCs) and indicates which variables have the greatest contribution in the new reference system.

Supervised analysis. Differently from unsupervised methods, supervised techniques require additional information on experimental details (e.g. samples classification). Several supervised techniques have been developed for driving statistical analysis when experimental perturbations cannot affect the total variance of the system or when technical noise covers experimental information.
The most largely used supervised multivariate techniques is partial least square (PLS) analysis and its classification strategy derivative, PLS discriminant analysis (PLS-DA)[98]. But these statistical techniques were not used in this metabolomic work.

8.2.7 Identification of metabolites

Analyzing the 1D and 2D NMR spectra and the MS spectra it is also possible to identify the metabolites contained in a biofluid or tissue.

The identification of the metabolites usually occurs comparing the acquired spectra with the data stored in public databases of known metabolites [99][Table 13]. Nowadays all the public databases contain structural and chemical information of the metabolite and the relative NMR and/or MS spectrum, but some of them are specific for a particular biological system, such as human, yeast or bacteria. They contain also biological information, including the metabolic pathways in which the metabolites are involved. In particular the Human Metabolome DataBase (www.hmdb.ca) [100]contains structural, chemical, biological and clinical information of the metabolites detected in NMR and MS spectra,

Database	Content	Tools
A. Reference spectral database		
NIST 08	EI-MS and MS/MS spectra, retention time	MS search software, AMDIS, MS interpreter utility
GMD	GC-MS spectra, retention time, experimental methods and protocols	compound name converter
MassBank	GC, ESI, and FAB-MS, MS/MS spectra, chemical structure and name, separation method, experimental conditions	3D visualization
METLIN	LC/MS, LC/MS/MS, FT-MS spectra, chemical structures, retention time, link to KEGG, chemical formula	3D visualization, JME molecular visualization tool
MMCD	NMR and LC MS spectra, physical properties, links to chemical and metabolomic databases, chemical structure	PubChem structure editor, SESAME, Google™
B. Compound-, species-, and compound-specific metabolite profile database		
HMDB	Metabolites found in the human body, NMR, GC-MS and MS/MS data, physical, clinical and physicochemical properties, chemical structure, concentration data, pathway information, experimental methods	Data Extractor, ChemSketch, BLAST search, MetaboCard, MS and NMR spectral search utility, MetaboLIMS
CSFMetabolome	CSF metabolites, NMR, GC-MS, and LC/FT-ICR-MS data, physical properties, associated disease, concentration data, chemical structure	Similar tools in HMDB
DrugBank	Drug, drug targets, physicochemical, clinical, and pharmacological properties, links to other databases, 3D and chemical structure	DrugCard, ChemAxon, MarvinSketch, HMDB Tools, BioSpider, PolySearch
LMSD	Lipids, chemical structure and formula, physical properties, links to other lipid and pathway databases, experimental data and protocols	ChemDraw, MS prediction tools, MarvinView, LMSD tools package
PubChem	Chemical compounds and structures, links to *Entrez* databases, bioactivity data from high throughput screening programs	PubChem Structure Editor, *Entrez*, PubChem3D, PubChem Download Facility
C. Metabolic pathway database		
KEGG	Metabolic and regulatory pathways, protein–protein interactions, genes of completely and partially sequenced genome, drugs, glycans, small molecules, reactions, functional hierarchies	KEGG Atlas, KegHier, KegArray, KegDraw, KegTools, KEGG2, KEGG API
BioCyc	Pathway and genome database of organisms with completely or partially sequenced genome, database of chemical compounds, links to other databases	Pathologic, Genome browser, Pathway Tools, BLAST search
MetaCyc	Metabolic pathways of organisms determined from experiments, enzymes, genes, chemical compounds, biochemical reactions, links to other databases	Pathologic, Genome browser, BLAST search, Pathways Tools, Google™
HumanCyc	Human metabolic pathways, genes, metabolic enzymes, links to other databases	Pathologic, Genome browser, Pathways Tools, BLAST search
Reactome	Human metabolic pathways, reactions, proteins, links to other databases	SkyPainter, PathFinder, BioMart
D. Metabolomic LIMS database		
SetupX	Experimental design to data reporting, raw data, links to other databases	BinBase, Google™
Sesame LIMS	Experimental protocols, data, sample details, laboratory resources, links to other databases	Tools for accessing and visualization of data, and collaborative analysis
MetaboLIMS	NMR and MS experimental data, metabolite database, sample information	BioSpider, PolySearch
MeMo	Yeast metabolomic experiment, links to databases specific to yeast	SEMEDA

Table 13. List of the metabolomic database available. Except NIST08 and MeMO, the other database are free. [Table from Go E. P. 2010, J. Neuroimmune Pharmacol.]

REFERENCES

1. Hildebrandt F, Benzing T, Katsanis N. Ciliopathies. *N Engl J Med*. 2011;364(16):1533-1543.

2. Hurd TW, Hildebrandt F. Mechanisms of nephronophthisis and related ciliopathies. *Nephron Exp Nephrol*. 2011;118(1):e9-14.

3. Wolf MT, Hildebrandt F. Nephronophthisis. *Pediatr Nephrol*. 2011;26(2):181-194.

4. Donaldson JC, Dise RS, Ritchie MD, Hanks SK. Nephrocystin-conserved domains involved in targeting to epithelial cell-cell junctions, interaction with filamins, and establishing cell polarity. *J Biol Chem*. 2002;277(32):29028-29035.

5. Boletta A. Emerging evidence of a link between the polycystins and the mTOR pathways. *Pathogenetics*. 2009;2(1):6.

6. Wilson PD. Molecular mechanisms of polycystic kidney disease. *Biochim Biophys Acta*. 2011;1812(10):1201.

7. Qian F, Watnick TJ, Onuchic LF, Germino GG. The molecular basis of focal cyst formation in human autosomal dominant polycystic kidney disease type I. *Cell*. 1996;87(6):979-987.

8. Tsiokas L, Arnould T, Zhu C, Kim E, Walz G, Sukhatme VP. Specific association of the gene product of PKD2 with the TRPC1 channel. *Proc Natl Acad Sci U S A*. 1999;96(7):3934-3939.

9. DeBerardinis RJ, Cheng T. Q's next: The diverse functions of glutamine in metabolism, cell biology and cancer. *Oncogene*. 2010;29(3):313-324.

10. Sharma MK, Seidlitz EP, Singh G. Cancer cells release glutamate via the cystine/glutamate antiporter. *Biochem Biophys Res Commun*. 2010;391(1):91-95.

11. Rzeski W, Ikonomidou C, Turski L. Glutamate antagonists limit tumor growth. *Biochem Pharmacol*. 2002;64(8):1195-1200.

12. Wodarczyk C, Rowe I, Chiaravalli M, Pema M, Qian F, Boletta A. A novel mouse model reveals that polycystin-1 deficiency in ependyma and choroid plexus results in dysfunctional cilia and hydrocephalus. *PLoS One*. 2009;4(9):e7137.

13. Fan TW, Lane AN, Higashi RM, et al. Altered regulation of metabolic pathways in human lung cancer discerned by (13)C stable isotope-resolved metabolomics (SIRM). *Mol Cancer*. 2009;8:41-4598-8-41.

14. Lane AN, Fan TW, Higashi RM, Tan J, Bousamra M, Miller DM. Prospects for clinical cancer metabolomics using stable isotope tracers. *Exp Mol Pathol*. 2009;86(3):165-173.

15. Aft RL, Zhang FW, Gius D. Evaluation of 2-deoxy-D-glucose as a chemotherapeutic agent: Mechanism of cell death. *Br J Cancer*. 2002;87(7):805-812.

16. Sottnik JL, Lori JC, Rose BJ, Thamm DH. Glycolysis inhibition by 2-deoxy-D-glucose reverts the metastatic phenotype in vitro and in vivo. *Clin Exp Metastasis*. 2011;28(8):865-875.

17. Wodarczyk C, Distefano G, Rowe I, et al. Nephrocystin-1 forms a complex with polycystin-1 via a polyproline motif/SH3 domain interaction and regulates the apoptotic response in mammals. *PLoS One*. 2010;5(9):e12719.

18. Mason JM, Arndt KM. Coiled coil domains: Stability, specificity, and biological implications. *Chembiochem*. 2004;5(2):170-176.

19. Giamarchi A, Feng S, Rodat-Despoix L, et al. A polycystin-2 (TRPP2) dimerization domain essential for the function of heteromeric polycystin complexes. *EMBO J*. 2010;29(7):1176-1191.

20. Ishikawa H, Marshall WF. Ciliogenesis: Building the cell's antenna. *Nat Rev Mol Cell Biol*. 2011;12(4):222-234.

21. Berbari NF, O'Connor AK, Haycraft CJ, Yoder BK. The primary cilium as a complex signaling center. *Curr Biol*. 2009;19(13):R526-35.

22. Saunier S, Salomon R, Antignac C. Nephronophthisis. *Curr Opin Genet Dev*. 2005;15(3):324-331.

23. Eley L, Moochhala SH, Simms R, Hildebrandt F, Sayer JA. Nephrocystin-1 interacts directly with Ack1 and is expressed in human collecting duct. *Biochem Biophys Res Commun*. 2008;371(4):877-882.

24. Lupas A. Coiled coils: New structures and new functions. *Trends Biochem Sci*. 1996;21(10):375-382.

25. Lupas A, Van Dyke M, Stock J. Predicting coiled coils from protein sequences. *Science*. 1991;252(5010):1162-1164.

26. le Maire A, Weber T, Saunier S, et al. Solution NMR structure of the SH3 domain of human nephrocystin and analysis of a mutation-causing juvenile nephronophthisis. *Proteins*. 2005;59(2):347-355.

27. Benzing T, Gerke P, Hopker K, Hildebrandt F, Kim E, Walz G. Nephrocystin interacts with Pyk2, p130(cas), and tensin and triggers phosphorylation of Pyk2. *Proc Natl Acad Sci U S A*. 2001;98(17):9784-9789.

28. Delous M, Hellman NE, Gaude HM, et al. Nephrocystin-1 and nephrocystin-4 are required for epithelial morphogenesis and associate with PALS1/PATJ and Par6. *Hum Mol Genet*. 2009;18(24):4711-4723.

29. Liebau MC, Hopker K, Muller RU, et al. Nephrocystin-4 regulates Pyk2-induced tyrosine phosphorylation of nephrocystin-1 to control targeting to monocilia. *J Biol Chem*. 2011;286(16):14237-14245.

30. Torres VE, Harris PC, Pirson Y. Autosomal dominant polycystic kidney disease. *Lancet*. 2007;369(9569):1287-1301.

31. Torres VE, Harris PC. Autosomal dominant polycystic kidney disease: The last 3 years. *Kidney Int*. 2009;76(2):149-168.

32. Pei Y, Watnick T, He N, et al. Somatic PKD2 mutations in individual kidney and liver cysts support a "two-hit" model of cystogenesis in type 2 autosomal dominant polycystic kidney disease. *J Am Soc Nephrol*. 1999;10(7):1524-1529.

33. Chapin HC, Caplan MJ. The cell biology of polycystic kidney disease. *J Cell Biol*. 2010;191(4):701-710.

34. Koptides M, Deltas CC. Autosomal dominant polycystic kidney disease: Molecular genetics and molecular pathogenesis. *Hum Genet*. 2000;107(2):115-126.

35. Koptides M, Mean R, Demetriou K, Pierides A, Deltas CC. Genetic evidence for a trans-heterozygous model for cystogenesis in autosomal dominant polycystic kidney disease. *Hum Mol Genet*. 2000;9(3):447-452.

36. Qian F, Germino FJ, Cai Y, Zhang X, Somlo S, Germino GG. PKD1 interacts with PKD2 through a probable coiled-coil domain. *Nat Genet*. 1997;16(2):179-183.

37. Qian F, Boletta A, Bhunia AK, et al. Cleavage of polycystin-1 requires the receptor for egg jelly domain and is disrupted by human autosomal-dominant polycystic kidney disease 1-associated mutations. *Proc Natl Acad Sci U S A*. 2002;99(26):16981-16986.

38. Boletta A, Qian F, Onuchic LF, et al. Biochemical characterization of bona fide polycystin-1 in vitro and in vivo. *Am J Kidney Dis*. 2001;38(6):1421-1429.

39. Sandford R, Mulroy S, Foggensteiner L. The polycystins: A novel class of membrane-associated proteins involved in renal cystic disease. *Cell Mol Life Sci*. 1999;56(7-8):567-579.

40. Low SH, Vasanth S, Larson CH, et al. Polycystin-1, STAT6, and P100 function in a pathway that transduces ciliary mechanosensation and is activated in polycystic kidney disease. *Dev Cell*. 2006;10(1):57-69.

41. Distefano G, Boca M, Rowe I, et al. Polycystin-1 regulates extracellular signal-regulated kinase-dependent phosphorylation of tuberin to control cell size through mTOR and its downstream effectors S6K and 4EBP1. *Mol Cell Biol*. 2009;29(9):2359-2371.

42. Laplante M, Sabatini DM. mTOR signaling in growth control and disease. *Cell*. 2012;149(2):274-293.

43. Lieberthal W, Levine JS. Mammalian target of rapamycin and the kidney. II. pathophysiology and therapeutic implications. *Am J Physiol Renal Physiol*. 2012;303(2):F180-91.

44. Woo D. Apoptosis and loss of renal tissue in polycystic kidney diseases. *N Engl J Med*. 1995;333(1):18-25.

45. Ching Yu Lin, Huifeng Wu, Ronald S. Tjeerdema, Mark R. Viant. Evaluation of metabolite extraction strategies from tissue samples using NMR metabolomics. *Metabolomics*. 2007;3(1):55.

46. Xia J, Bjorndahl TC, Tang P, Wishart DS. MetaboMiner--semi-automated identification of metabolites from 2D NMR spectra of complex biofluids. *BMC Bioinformatics*. 2008;9:507-2105-9-507.

47. Chignola F, Mari S, Stevens TJ, et al. The CCPN metabolomics project: A fast protocol for metabolite identification by 2D-NMR. *Bioinformatics*. 2011;27(6):885-886.

48. Cobas C, Seoane F, Domínguez S, Sykora S. A new approach to improving automated analysis of proton NMR spectra through global spectral deconvolution (GSD). *Spectroscopy Europe*. 2010;23(1):26.

49. Wider G, Dreier L. Measuring protein concentrations by NMR spectroscopy. *J Am Chem Soc*. 2006;128(8):2571-2576.

50. Maher AD, Fonville JM, Coen M, Lindon JC, Rae CD, Nicholson JK. Statistical total correlation spectroscopy scaling for enhancement of metabolic information recovery in biological NMR spectra. *Anal Chem*. 2012;84(2):1083-1091.

51. Wei S, Zhang J, Liu L, et al. Ratio analysis nuclear magnetic resonance spectroscopy for selective metabolite identification in complex samples. *Anal Chem*. 2011;83(20):7616-7623.

52. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR*. 1995;6(3):277-293.

53. Vranken WF, Boucher W, Stevens TJ, et al. The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins*. 2005;59(4):687-696.

54. Wishart DS. Interpreting protein chemical shift data. *Prog Nucl Magn Reson Spectrosc*. 2011;58(1-2):62-87.

55. Shen Y, Lange O, Delaglio F, et al. Consistent blind protein structure generation from NMR chemical shift data. *Proc Natl Acad Sci U S A*. 2008;105(12):4685-4690.

56. Bonvin AM, Rosato A, Wassenaar TA. The eNMR platform for structural biology. *J Struct Funct Genomics*. 2010;11(1):1-8.

57. Vuister G.W. BA. Quantitative J correlation: A new approach for measuring homonuclear three-bond J(HNH.alpha.) coupling constants in 15N-enriched proteins. *Journal of the American Chemical Society*. 1993;115(17):7772.

58. Levitt MH. Spin dynamics: Basics of nuclear magnetic resonance. *John Wiley & Sons, New York*. 2008.

59. Güntert P. Automated NMR structure calculation with CYANA. *Methods Mol.Biol.* 2004;278:353.

60. Bardiaux B, Malliavin T, Nilges M. ARIA for solution and solid-state NMR. *Methods Mol Biol*. 2012;831:453-483.

61. Bardiaux B, Bernard A, Rieping W, Habeck M, Malliavin TE, Nilges M. Graphical analysis of NMR structural quality and interactive contact map of NOE assignments in ARIA. *BMC Struct Biol*. 2008;8:30.

62. Rieping W, Habeck M, Bardiaux B, Bernard A, Malliavin TE, Nilges M. ARIA2: Automated NOE assignment and data integration in NMR structure calculation. *Bioinformatics*. 2007;23(3):381-382.

63. Morin S. A practical guide to protein dynamics from 15N spin relaxation in solution. *Prog Nucl Magn Reson Spectrosc*. 2011;59(3):245-262.

64. Farrow NA, Muhandiram R, Singer AU, et al. Backbone dynamics of a free and phosphopeptide-complexed src homology 2 domain studied by 15N NMR relaxation. *Biochemistry*. 1994;33(19):5984-6003.

65. Yip GN, Zuiderweg ER. Improvement of duty-cycle heating compensation in NMR spin relaxation experiments. *J Magn Reson*. 2005;176(2):171-178.

66. Chen K, Tjandra N. The use of residual dipolar coupling in studying proteins by NMR. *Top Curr Chem*. 2012;326:47-67.

67. Sass HJ, Musco G, Stahl SJ, Wingfield PT, Grzesiek S. An easy way to include weak alignment constraints into NMR structure calculations. *J Biomol NMR*. 2001;21(3):275-280.

68. Cordier F, Dingley AJ, Grzesiek S. A doublet-separated sensitivity-enhanced HSQC for the determination of scalar and dipolar one-bond J-couplings. *J Biomol NMR*. 1999;13(2):175-180.

69. Dosset P, Hus JC, Marion D, Blackledge M. A novel interactive tool for rigid-body modeling of multi-domain macromolecules using residual dipolar couplings. *J Biomol NMR*. 2001;20(3):223-231.

70. Cornilescu G, Marquardt JL, Ottiger M, Bax A. Validation of proteinStructure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. *J.Am.Chem.Soc.* 1998;120:6836.

71. Zweckstetter M. NMR: Prediction of molecular alignment from structure using the PALES software. *Nat Protoc*. 2008;3(4):679-690.

72. Rajagopal P, Waygood EB, Reizer J, Saier MH,Jr, Klevit RE. Demonstration of protein-protein interaction specificity by NMR chemical shift mapping. *Protein Sci*. 1997;6(12):2624-2627.

73. Wheeler MD, Ikejema K, Enomoto N, et al. Glycine: A new anti-inflammatory immunonutrient. *Cell Mol Life Sci*. 1999;56(9-10):843-856.

74. Koppenol WH, Bounds PL, Dang CV. Otto warburg's contributions to current concepts of cancer metabolism. *Nat Rev Cancer*. 2011;11(5):325-337.

75. Zheng B, Jeong JH, Asara JM, et al. Oncogenic B-RAF negatively regulates the tumor suppressor LKB1 to promote melanoma cell proliferation. *Mol Cell*. 2009;33(2):237-247.

76. Grantham JJ, Geiser JL, Evan AP. Cyst formation and growth in autosomal dominant polycystic kidney disease. *Kidney Int*. 1987;31(5):1145-1152.

77. Woo D. Apoptosis and loss of renal tissue in polycystic kidney diseases. *N Engl J Med*. 1995;333(1):18-25.

78. Kuller LH. Metformin use among individuals at risk for type 2 diabetes. *Curr Diab Rep*. 2012;12(3):265-273.

79. Corton JM, Gillespie JG, Hawley SA, Hardie DG. 5-aminoimidazole-4-carboxamide ribonucleoside. A specific method for activating AMP-activated protein kinase in intact cells? *Eur J Biochem*. 1995;229(2):558-565.

80. Lupas AN, Gruber M. The structure of alpha-helical coiled coils. *Adv Protein Chem*. 2005;70:37-78.

81. Morrow JA, Segall ML, Lund-Katz S, et al. Differences in stability among the human apolipoprotein E isoforms determined by the amino-terminal domain. *Biochemistry*. 2000;39(38):11657.

82. Dettmer K, Aronov PA, Hammock BD. Mass spectrometry-based metabolomics. *Mass Spectrom Rev*. 2007;26(1):51-78.

83. Holm L, Kaariainen S, Rosenstrom P, Schenkel A. Searching protein structure databases with DaliLite v.3. *Bioinformatics*. 2008;24(23):2780-2781.

84. Kay LE, Torchia DA, Bax A. Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: Application to staphylococcal nuclease. *Biochemistry*. 1989;28(23):8972-8979.

85. Barbato G, Ikura M, Kay LE, Pastor RW, Bax A. Backbone dynamics of calmodulin studied by 15N relaxation using inverse detected two-dimensional NMR spectroscopy: The central helix is flexible. *Biochemistry*. 1992;31(23):5269-5278.

86. Anderson PE, Mahle DE, Doom TE, Reo NV, DelRaso NJ, Raymer ML. Dynamic adaptive binning: An improved quantification techniquefor NMR spectroscopic data. *Metabolomics*. 2011;7(2):179.

87. Kell DB. Metabolomic biomarkers: Search, discovery and validation. *Expert Rev Mol Diagn*. 2007;7(4):329-333.

88. Schermer B, Hopker K, Omran H, et al. Phosphorylation by casein kinase 2 induces PACS-1 binding of nephrocystin and targeting to cilia. *EMBO J*. 2005;24(24):4415-4424.

89. Ma Y, Zhang P, Yang Y, Wang F, Qin H. Metabolomics in the fields of oncology: A review of recent research. *Mol Biol Rep*. 2012;39(7):7505-7511.

90. Xia JF, Liang QL, Hu P, Wang YM, Luo GA. Recent trends in strategies and methodologies for metabonomics. *Chinese Journal of Analytical Chemistry*. 2008;37(1):136.

91. Nicholson JK, Lindon JC. Systems biology: Metabonomics. *Nature*. 2008;455(7216):1054-1056.

92. Cuperlovic-Culf M, Barnett DA, Culf AS, Chute I. Cell culture metabolomics: Applications and future directions. *Drug Discovery Today*. 2010;15(15/16):610.

93. Fiehn O. Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. *Comp Funct Genomics*. 2001;2(3):155-168.

94. Smolinska A, Blanchet L, Buydens LM, Wijmenga SS. NMR and pattern recognition methods in metabolomics: From data acquisition to biomarker discovery: A review. *Anal Chim Acta*. 2012;750:82-97.

95. Beckonert O, Coen M, Keun HC, et al. High-resolution magic-angle-spinning NMR spectroscopy for metabolic profiling of intact tissues. *Nat Protoc*. 2010;5(6):1019-1032.
96. De Meyer T, Sinnaeve D, Van Gasse B, et al. NMR-based characterization of metabolic alterations in hypertension using an adaptive, intelligent binning algorithm. *Anal Chem*. 2008;80(10):3783-3790.
97. van der Berg R, Hoefsloot H, Westerhuis J, Smilde A, van der Werf M. Centering, scaling, and transformations: Improving the biological information content of metabolomics data.. *BMC Genomics*. 2006;8(7):142.
98. Boulesteix AL, Strimmer K. Partial least squares: A versatile tool for the analysis of high-dimensional genomic data. *Brief Bioinform*. 2007;8(1):32-44.
99. Go EP. Database resources in metabolomics: An overview. *J Neuroimmune Pharmacol*. 2010;5(1):18-30.
100. Wishart DS, Knox C, Guo AC, et al. HMDB: A knowledgebase for the human metabolome. *Nucleic Acids Res*. 2009;37(Database issue):D603-10.