UNIVERSITÀ
DEGLI STUDI
FIRENZE

DOTTORATO DI RICERCA
INTERNATIONAL DOCTORATE IN STRUCTURAL BIOLOGY

CYCLE XXVIII

COORDINATOR Prof. Claudio Luchinat

New strategy of protein expression in mammalian cells for in-cell NMR

Settore Scientifico Disciplinare CHIM/03

Candidate
Dott. Erica Secci

Tutor
Prof. Lucia Banci

Coordinator
Prof. Claudio Luchinat

November 2012 – 2015

This thesis has been approved by the University of Florence, the University of Frankfurt and the Utrecht University

Universiteit Utrecht

CERM Firenze

FRANKFURT AM MAIN

Table of contents

1. Introduction

1.1 Aim of the research

The research activities of my doctorate have been devoted to the application of in-cell NMR for the characterization of proteins directly inside human cells. Within this context, I have also addressed the development of a methodology to perform sequential labeling for in-cell NMR in human cells. The in-cell NMR approach, which consists in observing one or more selectively labeled protein(s) or nucleic acid in living cells through high resolution NMR experiments, is a powerful tool to obtain structural and functional information *in situ*, thus overcoming the limitation of studying macromolecules isolated from the other cellular components. Protein folding, post-translational modifications and interactions with specific partners or cellular elements can be studied in the physiological environment in which the protein of interest operates.

My research work also addressed the *in vitro* characterization of the phosphorylation pattern of the non-structural protein 5A domain 3 of Hepatitis C virus by casein kinase 2.

1.2 In-cell NMR spectroscopy

The cellular interior is packed with macromolecules that differ in their shape, location, activity, and ability to interact. A detailed understanding of the biological function of macromolecules, such as proteins, DNA, RNA, or their complexes, often requires structural knowledge. Biophysical methods for the structural characterization of biomolecules are often confined to non native, *in vitro* experimental setups. X-ray crystallography and high-resolution electron microscopy are intrinsically excluded from *in vivo* approaches because of their requirement for pure samples and crystalline or vitrified specimens. Due to the samples requirements, the effects of cellular properties on biomolecular conformation and dynamics, such as viscosity, molecular crowding, and interaction with other macromolecules, as well as the influence of the concentration

of ions and other small molecules cannot be analyzed [1-4]. Nuclear magnetic resonance (NMR) spectroscopy, the only other method for structural investigations at atomic level, allows for the direct and selective observation of NMR-active nuclei within any NMR-inactive environment. The noninvasiveness of NMR spectroscopy, and its ability to study biological macromolecules at physiological temperatures, led to the establishment of the in-cell NMR approach in 2001 that allows for the investigation of the structural behavior of isotopically labeled proteins inside living bacterial cells [5, 6]. Since then, the approach has been extended to several other cellular systems, including *Xenopus laevis* oocytes, yeast, insect cells and mammalian cells, to study the conformation of biological macromolecules in their natural environment [7-11].

The ability to detect changes in the conformation of a macromolecule between the *in vitro* situation and the interior of living cells is based on the sensitivity of the NMR chemical shift to be affected by the structural properties. Changes in the conformation, but also binding of small or large molecules and post-translational modifications, translate in NMR chemical shifts.

Detection of a particular macromolecule in the crowded intracellular environment requires that this macromolecule is labeled with (i.e. enriched in) NMR-active isotopes. Labeling with the isotopes ^{13}C (1.1% natural abundance) and ^{15}N (0.36% natural abundance) has proven to be the method of choice to filter out the background signals arising from all the hydrogen atoms of the cell, and enable the extraction of information on the macromolecule of interest. In particular, 2D [^{1}H, ^{15}N]-HSQC, which measures the chemical shifts of the amide hydrogen and nitrogen atoms, easily and efficiently provides information about the conformation of the protein backbone. Two strategies can be pursued to produce a cell sample containing the labeled protein of interest: delivery of a heterologously expressed, purified protein inside the cell or overexpression of the protein directly in cells. In the case of protein delivered into cell system, the cellular background signals contain only signals due to the ^{13}C and ^{15}N natural abundance. If, however, labeling occurs within the cellular system also used for the NMR spectroscopic experiments, not only the macromolecule of interest, but also other cellular components will be labeled. To increase the labeling selectivity, the expression of the protein has to occur as quickly as possible and the cell growth must be slow down

expression time. ^{15}N-labeling produces only a small number of background signals, whereas ^{13}C-labeling creates strong background peaks that make unambiguous identification of resonance of the macromolecules of interest difficult. The higher background level in the case of ^{13}C-labeling respect to ^{15}N-labeling is due to the higher number of C-H groups in biomolecules than of N-H groups and to the fact that N-H groups exchange with water while C-H groups do not. Amino acid type selective labeling of proteins can also apply for in-cell NMR experiments. Unfortunately, not all amino acids can be used for selective labeling procedures. Good candidates are those at the end of a biosynthetic pathway, which therefore are not needed as precursors for other amino acid synthesis.

A significant challenge in the application of in-cell NMR is the inherent insensitivity of NMR spectroscopy and the consequent need of high effective concentration (10–250 μM) of the macromolecules under study. These concentrations are typically several orders of magnitude higher than the native concentrations of the macromolecule of interest (0.5–1 μM).

To observe of a solution [^{1}H, ^{15}N]-HSQC spectrum of the protein of interest, it is necessary that the macromolecule tumbles freely in solution with a sufficient short correlation time. Several investigations have shown that viscosity reduces the rotational diffusion by a factor of 2–3 in most eukaryotic cell types and of 10-11 in *Escherichia coli* (*E. coli*) cells relative to measurements in water, resulting in an increase of the apparent molecular weight of the biomolecule roughly by the same factor [6, 12-14]. However, interactions with other large cellular components, such as macromolecular complexes, nucleic acids, or membranes, can dramatically slow down the average tumbling rate, thus resulting in broad or even undetectable resonance lines [15, 16]. Intrinsically disordered proteins (IDPs) show often weaker interactions with their surroundings and can compensate the concomitant reduction in their tumbling rate with their inherently high internal flexibility [17-20]. In contrast, the backbone resonances of many globularly folded proteins are not detectable, thus suggesting that their tumbling rates are significantly slower *in vivo*, most likely as a result of nonspecific interactions with other cellular components [15, 16, 21-23]. One way to visualize these proteins in the cellular environment is to perform selective labeling of methyl groups, which can be detected more readily because of their fast internal rotation [24]. Besides weak nonspecific interactions with other cellular components, the loss of signals in in-cell

NMR experiments can also be caused by the formation of large specific complexes, degradation, or conformational exchange. Specific- complex formation is in most cases unlikely, as the macromolecule under NMR investigation is typically several orders of magnitude higher than the cellular concentration of a specific binding partner. The degradation of macromolecules was mainly observed in in-cell NMR experiments on nucleic acids [25, 26]. Finally, signal loss can also occur as a result of conformational exchange processes [27].

Another critical parameter that influences the applicability of in-cell NMR experiments is the survival time of the cells in the NMR tube; the cells have to remain viable at least for the time period of the measurements. The high cellular density inside the NMR tube can cause oxygen starvation and limits the amount of available nutrients. If the experimental sensitivity and the protein concentration are sufficiently high, the NMR spectra can be measured relatively quickly (in less than one hour). Cells can be kept alive for longer times in the NMR instrument if the sample holder is modified to allow for the constant exchange of the medium to supply fresh nutrients as well as oxygen [28].

Since the first in-cell NMR experiments [5], *E. coli* has been one of the most used cell type, due to the advantages of working with bacteria (for example, their rapid growth rates, robustness and high recombinant protein expression levels) (Fig. 1a). Several applications for structural and functional studies of proteins overexpressed in bacteria have been reported: studies of macromolecular crowding [15, 29], protein structure and folding [27, 30-34], protein-protein interactions [35, 36] and interactions with drugs [9, 37], nucleic acids [38], or other cellular components [15, 23]. The structure of TTHA1718 protein was solved from in-cell NMR data obtained in *E. coli* cells [31]. Bacterial cells could be used as a model for the cytoplasm, but they lack compartmentalization, opposite to eukaryotic cells, and are unable to do post-translational modifications that are crucial for the function of many eukaryotic proteins.

Therefore, in-cell NMR on eukaryotic cells is a necessary step forward. Methods have been developed to observe proteins in live eukaryotic cells by NMR, which rely either on the insertion of pure, isotopically labeled proteins (produced in *E. coli*) from outside the cells, or on the overexpression of the protein directly inside the cells. The first eukaryotic cells used in in-cell NMR experiments were *Xenopus laevis* oocytes [7]. The

main advantage of *X. laevis* oocytes is their large size (ca. 1 mm in diameter), which allows microinjection of the protein or nucleic acid of interest into the citosol (Fig. 1b). Advantages of this approach for protein insertion in the cell are a better control on the intracellular protein concentration, the ability to isotopically label specific sites of the macromolecule, and the ability to introduce modifications, such as paramagnetic spin labels, which are difficult to introduce through the overexpression mode. *X. laevis* oocytes have been used to study phosphorylation of folded and intrinsically disordered proteins and to investigate the conformation of nucleic acids [18, 25, 26, 39-41].

When working with cultured mammalian cells, microinjection is not an option, owing to the small size of these cells compared to oocytes and therefore other methods had to be developed to enable the required high intracellular concentration to be attained. One procedure consists in tagging, through disulfide bonds, the protein of interest with a cell-penetrating peptide (CPP) that delivers the system into the cytoplasm (Fig. 1c). The protein is subsequently released from the CPP by endogenous enzymatic activity or by autonomous reductive cleavage [9, 42]. Although the mechanism of protein delivery with CPPs is not fully understood, a key molecular determinant is net charge [43]; as long as the CPP carries a sufficient amount positive charge to override the net charge of the cargo, its detailed amino-acid composition seems to be of secondary importance. Therefore, constructs with a negative or zero net charge can be efficiently delivered by CPPs into mammalian cells only by rational alteration of the net charge. Another method to insert an isotopically labeled protein into the cytosol was developed by Ogino *et al*. They took advantage of the pore-forming toxin streptolysin O to insert the protein thymosin β4 in human 293F cells (Fig. 1d). One attractive feature of this approach, in contrast to the CPP strategy, is that no modifications to the protein are required for intracellular sample delivery [8]. In-cell NMR spectra are thus uncompromised by CPP tags or by additional amino acids that remain present after intracellular tag removal. One disadvantage is the requirement for mammalian cells that grow in suspension and for highly soluble target proteins. Because the amount of delivered sample depends on passive diffusion through the toxin pore, and not on an active uptake process like the CPP case, the protein concentration gradient across the cell membrane directly relates to the overall transduction efficiency. Recently, protein electroporation has been developed, which consists in applying an electric pulse to cells so to increase permeability of extracellular membrane by pore formation, and to insert the purified

protein of interest into the cytosol (Fig. 1e) [44, 45]. This method is suitable to deliver native proteins, without requirements for engineered protein tags or targeting sequences, and not require any treatment of cells with toxic compounds, which decrease cell viability and signal the activation of damage response pathways. Shekhtman *et al*. used this technique to probe weak protein (quinary) interactions inside the cytoplasm of HeLa cells.

In addition to deliver the protein of interest into cytoplasm, it is possible to produce in-cell NMR sample also by overexpressing a labeled protein directly inside eukaryotic cells. In-cell NMR in the yeast *Pichia pastoris* was applied to study the influence of metabolic changes on protein structure and dynamics at atomic resolution (Fig. 1f) [11]. Hamatsu *et al.* have shown that insect cells are also amenable to in-cell protein NMR by exploiting a baculovirus expression system, reporting both [^{1}H, ^{15}N]-HSQC spectra and three dimensional spectra of four small globular proteins, including GB1 and calmodulin (Fig. 1g) [10]. In order to study the various maturation steps of a human protein right after protein biosynthesis, including post-translational modifications, and investigate the interactions with specific partners it is necessary to overexpress the protein of interest directly inside human cells [46]. Banci *et al.* have developed a method to overexpress one or more proteins in human embryonic kidney cells (HEK293T) by transient transfection. Protein overexpression levels could be controlled by varying the amount of transfected DNA (Fig. 1h). To reach high transfection efficiency, plasmids were complexed with polyethylenimine (PEI) prior to transfection.

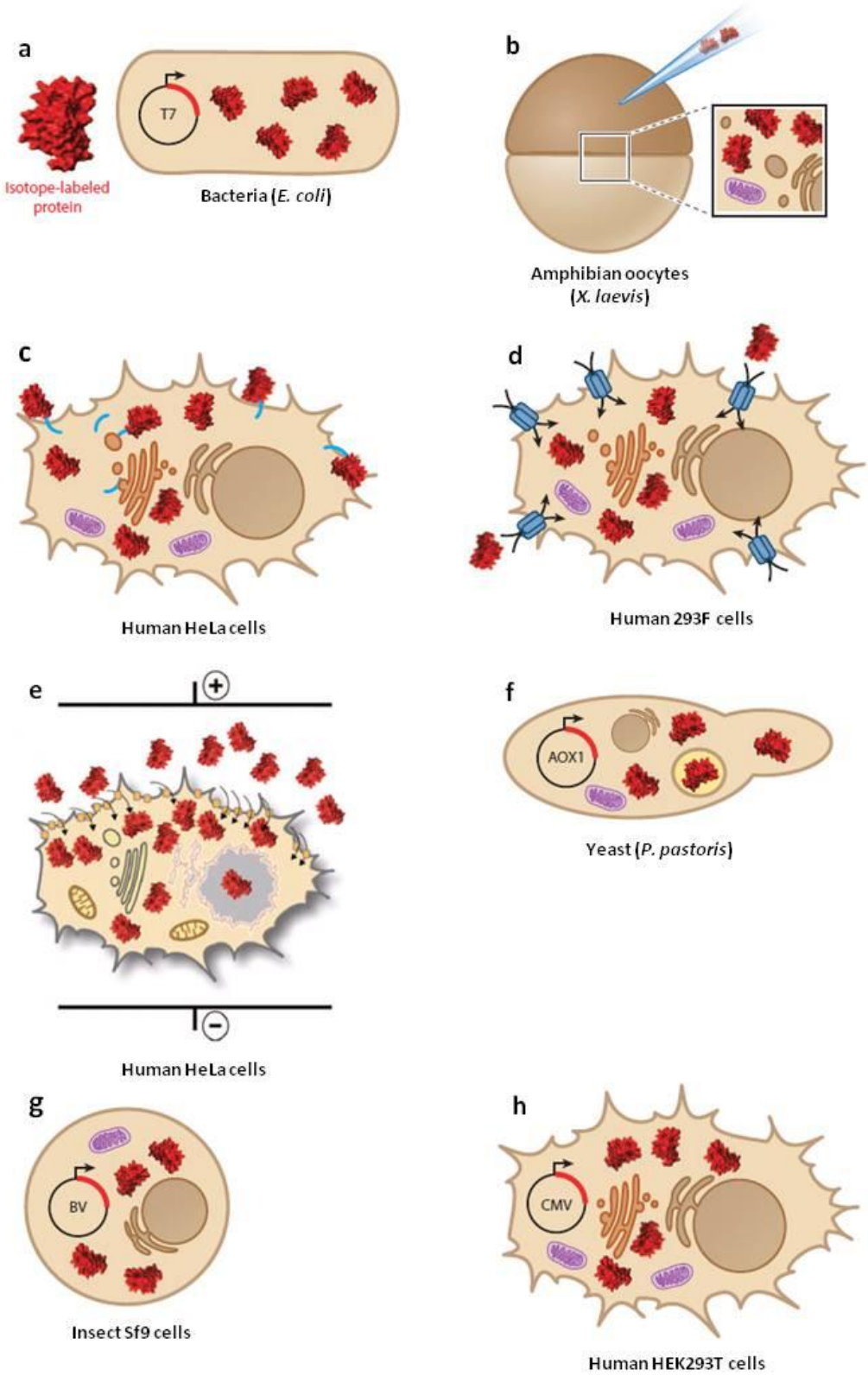


Figure 1 Overview of prokaryotic and eukaryotic in-cell NMR systems. (a) Accumulation of isotope-labeled protein inside *Escherichia coli* by T7 promoter–driven overexpression. (b) Delivery of isotope-labeled proteins by microinjection into *Xenopus laevis* oocytes. (c) Transduction of cell-penetrating peptide (CPP)-tagged (*cyan*) cargo proteins into cultured mammalian cells. (d) Delivery of isotope-labeled proteins into mammalian cells through bacterial pore-forming toxins (*blue*). (e) Delivery of isotope-labeled protein into cultured human cells by electroporation. (f) AOX1-driven protein production in the yeast *Pichia pastoris.* (g) Baculovirus (BV)-infected insect Sf9 cells. (h) Cytomegalovirus (CMV) promoter– driven protein overexpression and isotopic labeling in human HEK293T cells. Cells are not drawn to scale. Adapted from [44, 47].

1.3 Mia40

Mia40 (*Mitochondrial intermembrane space import and assembly protein 40*) is an oxidoreductase, located in the mitochondrial intermembrane space (IMS), that imports and traps proteins in the IMS through oxidative protein folding [48, 49]. The IMS of mitochondria, the compartment between the mitochondrial outer and inner membrane, harbors many proteins that perform important functions in cellular processes, such as generation of ATP by oxidative phosphorylation, synthesis of iron sulfur clusters, apoptosis, transport of metabolites and proteins across the IMS and detoxification from reactive oxygen species. Except for a few proteins of the inner membrane and the mitochondrial matrix, all mitochondrial proteins are synthesized in the cytoplasm from nuclear DNA [50-52] and have to cross the outer membrane through a translocation complex of outer membrane (TOM) [50]. Some precursors of IMS proteins contain bipartite targeting information in their N-terminal region: a presequence-type matrix targeting signal followed by a hydrophobic sorting signal, which arrests the translocation at the translocase inner membrane (TIM23) complex. Through a subsequent proteolytic step, the mature protein is released into the IMS. A family of small IMS proteins is synthesized without a cleavable presequence, but contain conserved cysteine residues arranged in a coiled-coil-helix-coiled-coil-helix (CHCH) folding domain [50]. They contain twin -CX_nC- cysteine motifs and are imported through the TOM channel as reduced, unfolded polypeptides. In the IMS, these proteins interact with Mia40, which induces folding and oxidation through the formation of intramolecular disulfide bonds. Once oxidized and folded, they are unable to cross the outer membrane back, and therefore they are trapped in the IMS [53]. Among these small IMS proteins are Cox17, which is a copper chaperone essential for delivering Cu(I) through interaction with Sco1, Sco2 and Cox11 to the CuA site of cytochrome *c* oxidase [54-56], and the small Tim proteins, which are chaperones essential for import and maturation of mitochondrial integral membrane proteins [57, 58].

Mia40 is a 16 kDa protein with a CHCH structure analogous to its substrates and it harbors six conserved cysteines forming three disulfide bonds. Two of them are formed across the twin -CX_9C- motifs between Cys64-Cys97 and Cys74-Cys87 and have a structural role. In the mature form of Mia40, they hold together two alpha-helices forming an antiparallel α-hairpin (α2 residues 65-77 and α3 residues 88-100) [59, 60].

The third disulfide bond is formed by Cys53-Cys55 and constitutes the catalytic CPC redox site (Fig. 2). This is crucial for the oxidation of the reduced substrates. Thus in mitochondria, Mia40 cycles between two redox states: the fully oxidized one ($Mia40^{3S-S}$) containing three intramolecular disulfide bonds, and the partially reduced form where the CPC is reduced ($Mia40^{2S-S}$).

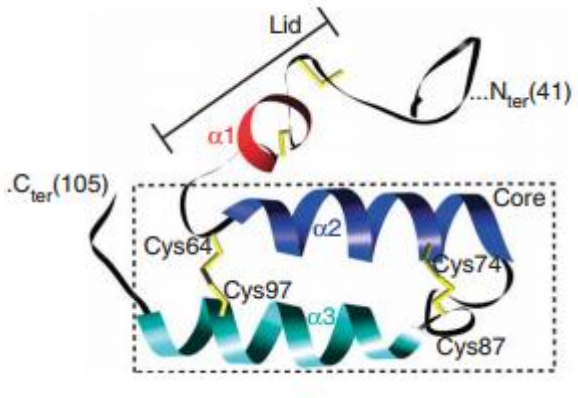


Figure 2 Solution structure of the folded $Mia40^{2S-S}$. Ribbon diagram of the lowest-energy conformer of $Mia40^{2S-S}$. Helix α1 of the N-terminal lid is shown in red, and helices a2 and a3 composing the a-hairpin core are shown in blue and cyan. Disulfide pairings (or free thiols) are shown in yellow. From [59].

Mia40 is synthesized in the cytoplasm and needs to be in a fully reduced and unfolded state in order to translocate to the IMS through the TOM channel. Therefore, Mia40 must remain unfolded and completely reduced during its entire journey through the cytoplasm, until it is imported into mitochondria. It has been demonstrated that Mia40 is imported into the IMS through the same pathway of its substrates, and in the IMS it is oxidized by pre-existing Mia40, in a self-propagating manner [61-63].

Mia40 interacts with its substrates, after they enter the IMS in the unfolded reduced state, through hydrophobic interaction between the hydrophobic cleft on Mia40 and an internal targeting signal (ITS) present on the substrate [64]. This interaction orients the substrate in a way to expose a cysteine side chain to the active CPC site of Mia40, where a mixed disulfide bond is formed between the substrate cysteine and the docking Cys55 of Mia40. It has been demonstrated by NMR experiments that covalent binding to Mia40 induces formation of first α-helix in the substrate, thus demonstrating the chaperone role of Mia40 [60]. Upon isomerization of the S-S intermolecular bond, the first disulfide bond is formed in the substrate inducing the folding of the second α-helix and the release from Mia40, which is now CPC-reduced (Fig. 3). The reduced active

site of Mia40 is re-oxidized by reaction with the Augmenter of Liver Regeneration protein (ALR) a FAD-linked thiol oxidase, which in turn shuttles electrons to cytochrome oxidase (and eventually to O_2) via reaction with cytochrome *c* [65, 66]. Together form the disulfide relay system of the IMS of mitochondria [67]. The generation of the second disulfide bond in the substrate could follow different routes depending on the substrate and on different cellular condition. This disulfide can be formed upon oxidation by oxygen and other small oxidant or upon the formation of a ternary complex among substrate, Mia40 and ALR [68].

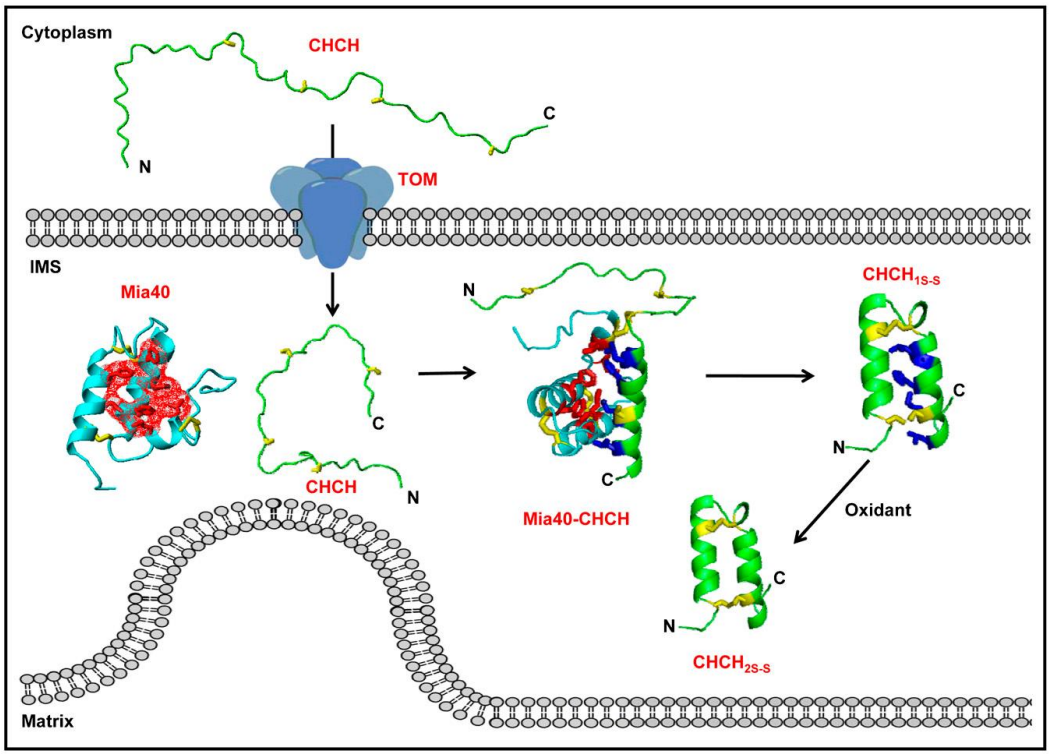


Figure 3 Steps of the oxidative folding mechanism of CHCH domain-containing proteins in the IMS. Substrates of the oxidative folding pathway are imported in the IMS through the TOM channel in the unfolded state. Upon interaction with Mia40, an intermolecular disulfide bond is formed, inducing the formation of the first α-helix. The oxidized substrate is then released from Mia40. From [60].

1.4 Non structural protein 5A

The non structural protein 5A (NS5A) is a phosphoprotein without any enzymatic activity that plays key roles in both genome replication and virion assembly/release of Hepatitis C virus (HCV). HCV virus chronically infects approximately 130-150 million people worldwide (http://www.who.int/mediacentre/factsheets/fs164/en/). Acute infections are often asymptomatic and are spontaneously cleared within several weeks.

Persistent, or chronic, infections are defined by detection of HCV RNA in the blood for 6 months or more after infection. It is estimated that up to 85% of all HCV infections progress to chronicity, causing fibrosis which ultimately leads to cirrhosis, end-stage liver disease, and/or hepatocellular carcinoma [69, 70]. Six closely related viral genotypes, classified by sequence homology, comprise the HCV genus [71]. Genotype 1 is the most abundant worldwide, and includes subtypes 1a and 1b [72]. The traditional standard of care (SOC) for HCV-infected patients has been limited to a regimen of pegylated-interferon alpha (pegIFN) and ribavirin; displaying low cure rates in a majority of patients and severe side effects [73, 74]. However, in 2011 the first direct-acting antivirals (DAA) were licensed to treat HCV-infected patients in combination with SOC, which served to elevate treatment response rates [75]. The HCV drug development pipeline is currently populated with many additional and improved DAAs; primarily molecules that target the virus-encoded protease, polymerase enzymes or non structural proteins (http://www.uptodate.com/contents/direct-acting-antivirals-for-the-treatment-of-hepatitis-c-virus-infection) but their mode of action is still not well understood. Therefore, a better understanding of the molecular mechanisms involved in the HCV life cycle is required to develop new antiviral therapies.

HCV is an enveloped virus, belonging to *Flaviviridae* family, containing a positive-sense, single-stranded RNA genome of about 9.6-kb [76]. The RNA genome codes for an ~3,000 amino acids polyprotein that is cleaved co- and posttraslationally by both host and viral proteases to produce 10 mature viral proteins: the structural proteins, Core, E1 and E2, which make up the virus particle; the viroporin p7, which has roles in virus assembly and release, and the nonstructural (NS) proteins NS2, NS3, NS4A, NS4B, NS5A, NS5B, which are involved in both genome replication and virion assembly [77, 78].

NS5A plays a critical role in RNA replication and particle assembly, however the precise nature of these two functions of the protein and how they are regulated remain to be elucidated [79]. NS5A consists of three domains linked by two low complexity sequences (LCS-1 and LCS-2) and an amino-terminal amphipathic helix essential for its membrane association (Fig. 4) [80, 81].

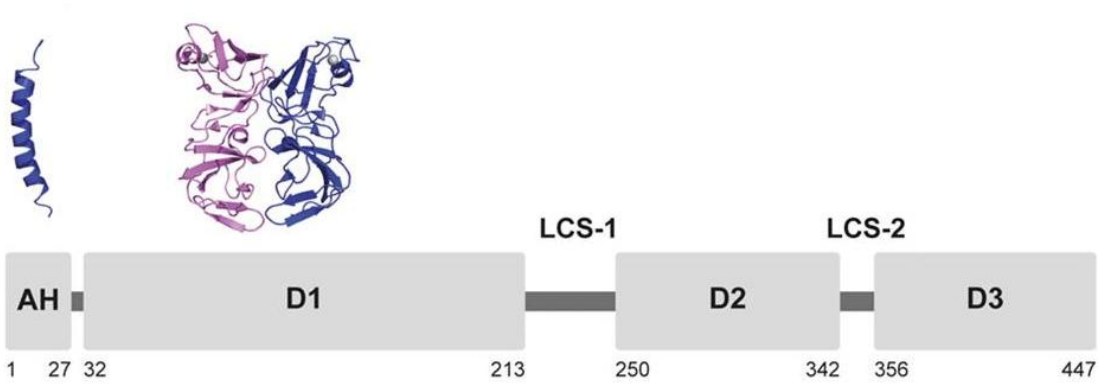


Figure 4 Schematic representation of the domain organization of NS5A. Ribbon representations of the N-terminal amphipathic helix (AH, blue, PDB entry 1R7E) and dimeric globular domain 1 (D1, pink and blue, PDB entry 3FQM) are shown. Reprinted from [82].

Domain 1 (D1) comprises a zinc-binding motif, which is essential for HCV RNA replication [83] and forms a homodimer with a large RNA binding groove located at the interface of the monomer [84, 85]. It is also required for the interaction with lipid droplets, the organelles where the assembly of infectious particles starts [86]. Domains 2 and 3 (D2, D3) have been shown to be both intrinsically disordered with a propensity to form secondary structure elements [87-89]. D2 is essential for RNA replication, while D3 is thought to be dispensable for that role [90, 91]. It has been shown, however, that D3 is critical in the virus assembly process, as it is required for the interaction between NS5A and the core protein [92].

NS5A exists in two forms with slightly different mobility on an SDS-PAGE gel: a basally phosphorylated form (apparent molecular mass of 56 kDa) and a hyperphosphorylated form (58 kDa) [93, 94]. Pulse-chase and mutational analysis suggest that both phosphorylated forms are produced after completion of the proteolytic cleavages at the N- and C-termini of NS5A, and that p58 is the product of p56 hyperphosphorylation. Hyperphosphorylation sites have been identified in a serine-rich region in LCS-1 and in D2, and have been shown to act as negative regulators of the RNA replication process [94-97]. It has been demonstrated that hyperphosphorylation of NS5A requires NS3, NS4A and NS5B [98].The cellular kinases that are mainly involved in the generation of the p58 form are the α isoform of casein kinase 1 (CK1-α) [99] and polo-like kinase 1 (Plk1) [100]. Suzuki *et al.* suggested that CK1-α phosphorylates NS5A on S225 and S232 [101]. It has been demonstrated that also S222 is involved in hyperphosphorylation form of NS5A but actually the responsible kinase has not been identified [97, 102]. The basally phosphorylated sites are mainly serine

residues located in both D2 and D3 [94-97]. The basal phosphorylation of D3 plays a critical role in the virion assembly, as serine to alanine mutations and/or deletions of parts of D3 have been shown to disrupt the interaction with core protein and impair the production of infectious virions [92, 103]. It has been shown that casein kinase 2 (CK2) is responsible for the basal phosphorylation of D3 in multiple sites [104, 105]. Potential CK2 phosphorylation sites were identified but only one residue (Serine 457 in 2a JFH-1 genotype) was confirmed [103, 106].

Mechanisms regulating NS5A phosphorylation and its exact function in the HCV life cycle have not been clearly defined. Hyperphosphorylated NS5A form was predominantly localized in low-density membrane structures around lipid droplets, in which NS5A interacts with the core for virion assembly, while reduction of NS5A hyperphosphorylation led to a decrease in NS5A abundance in the low-density membrane structures. This suggests that NS5A-associated kinases and NS5A phosphorylation have a regulatory role in the viral life cycle especially as a molecular switch governing the transition between viral replication and virion assembly [101].

2. Results

2.1 Visualization of Redox-Controlled Protein Fold in Living Cells

The folding and oxidation states of Mia40 were investigated as a function of the sample conditions in the human cell cytoplasm by in-cell NMR. In this cellular compartment, Mia40 need to be in the reduced, unfolded state. However, when it is overexpressed in HEK293T cells cytoplasm, it is present in a folded and in disulfide-oxidized state as monitored through 2D $^{1}H,^{15}N$-SOFAST-HMQC and ^{1}H NMR spectra. This suggested that some thiol-disulfide regulation mechanism is necessary to keep Mia40 reduced in the cytoplasm until it reaches the outer membrane.

Co-expression with glutaredoxin 1 (Grx1) and thioredoxin 1 (Trx1), two cytoplasmic oxidoreductases involved in the regulation of the oxidation state of protein thiol groups influenced the ratio between folded and unfolded species of Mia40. When Mia40 was co-expressed with various amounts of Grx1, a large fraction of Mia40 was present in the reduced, unfolded state. This effect was confirmed by a decrease of the methyl signal at -0.7 ppm, used as a marker of the folded state of Mia40, and the total amount of protein was determined by Westerm Blot. In presence of increasing amounts of Grx1, oxidized Mia40 ($Mia40^{2S\text{-}S}$) in the cytoplasm decreased to ~25% of the total Mia40.

The effect of Trx1 on Mia40 state was also investigated, by inducing co-expression of Trx1 with Mia40. Co-expression of Trx1 affected the folding state of Mia40 to a lesser extent than Grx1 when expressed at similar levels, as ~50% of cytoplasmic Mia40 was still in the folded, oxidized state. Therefore, the two proteins have different efficacy in keeping Mia40 reduced, despite being reported to have overall similar functions in the cytoplasm.

Chemistry & Biology
Brief Communication

Visualization of Redox-Controlled Protein Fold in Living Cells

Lucia Banci,[1,2,*] Letizia Barbieri,[1] Enrico Luchinat,[1,3] and Erica Secci[1]
[1]Magnetic Resonance Center - CERM, University of Florence, Via Luigi Sacconi 6, 50019 Sesto Fiorentino, Florence, Italy
[2]Department of Chemistry, University of Florence, Via della Lastruccia 3, 50019 Sesto Fiorentino, Florence, Italy
[3]Department of Biomedical, Clinical and Experimental Sciences, University of Florence, Viale Morgagni 50, 50134 Florence, Italy
*Correspondence: banci@cerm.unifi.it
http://dx.doi.org/10.1016/j.chembiol.2013.05.007

SUMMARY

Most mitochondrial proteins are encoded by nuclear DNA, synthesized in the cytoplasm, and imported into mitochondria. Several proteins of the intermembrane space (IMS) are imported and localized through an oxidative process, being folded through the formation of structural disulfide bonds catalyzed by Mia40, and trapped in the IMS. To be imported, these proteins need to be reduced and unfolded; however, no structural information in situ exists on these proteins in the cytoplasm. In humans, Mia40 undergoes the same mechanism, although its folding state in the cytoplasm is unknown. We provide atomic-level details on the Mia40 folding state in the human cell cytoplasm through in-cell nuclear magnetic resonance. Overexpressed cytoplasmic Mia40 is folded, and its folding state depends on the glutaredoxin 1 (Grx1) and thioredoxin 1 (Trx1) systems. Specifically, increased Grx1 levels keep most Mia40 unfolded, while Trx1 is less effective.

INTRODUCTION

The majority of the human proteins are produced by nuclear DNA and released in the cytoplasm and/or in the endoplasmic reticulum (ER). These proteins then need to complete their folding and maturation process, which could involve several steps, from cofactor binding to cysteine oxidation to other posttranslational modifications. Furthermore, if the protein destiny is a cellular compartment other than the cytoplasm, some or all maturation and folding steps may take place in the final cellular localization. This is particularly true for a large share of proteins present in the intermembrane space (IMS) of mitochondria, but which do not feature any target sequence for this organelle (Banci et al., 2009b; Longen et al., 2009; Neupert and Herrmann, 2007). They were proposed to be in an unfolded state in the cytoplasm and thus can enter mitochondria, thanks to their conformational flexibility that allows them to go through the translocase of the outer membrane (TOM) channel (Neupert and Herrmann, 2007). Once they have entered the IMS, they fold to their native form, thus being blocked in a defined, more rigid conformation, which prevents them from crossing back to the outer membrane (Chacinska et al., 2004; Lu et al., 2004; Mesecke et al., 2005). It is therefore evident that the folding state of a protein depends on the cellular compartment where the protein is located and on its properties. It has been shown that the import of some of these IMS proteins is in fact modulated by cytosolic thiol-disulfide regulation systems (Durigon et al., 2012). However no detailed, atomic resolution study has confirmed such findings.

In this work we show, by exploiting in-cell nuclear magnetic resonance (in-cell NMR; Banci et al., 2013; Inomata et al., 2009; Ogino et al., 2009; Reckel et al., 2007; Selenko et al., 2008), that indeed the properties of the cell compartment in terms of redox-regulating components influence the folding state of one of such proteins, whose oxidation state is compartment-dependent. We have specifically characterized Mia40, a hub protein for the mitochondrial protein import process. Mia40 is an oxidoreductase that catalyzes in the IMS the formation of internal disulfide bonds on its protein substrates through the intermediate formation of a mixed disulfide bond between the substrate and its catalytic CPC motif (Banci et al., 2009a; Chacinska et al., 2004; Grumbt et al., 2007; Naoé et al., 2004). Together, these proteins constitute the disulfide relay system of the IMS (Mesecke et al., 2005; Tokatlidis, 2005). Upon interaction with Mia40 and subsequent formation of their disulfide bonds, the substrates of Mia40 become folded and are trapped in the IMS (Banci et al., 2010; Gabriel et al., 2007). Interestingly, Mia40 itself obtains its final structure in the IMS upon the formation of two internal disulfide bonds, likely by acting as a substrate of itself (Chacinska et al., 2008; Grumbt et al., 2007). Like the other substrates of the disulfide relay system, Mia40 has to cross the outer mitochondrial membrane in an unfolded, reduced state. Glutaredoxin 1 (Grx1) and thioredoxin 1 (Trx1) are cytoplasmic oxidoreductases involved in the regulation of protein thiol groups and in the cellular defense against oxidative stress (Holmgren, 1989; Meyer et al., 2009). Trx1 was recently shown to be responsible for facilitating the mitochondrial import of the small Tim proteins (Durigon et al., 2012). It can be hypothesized that other small proteins of the IMS sharing the same import mechanism, including Mia40, are regulated by such thiol-regulating proteins.

We show here that Mia40, even in the reducing environment of the cytoplasm, is largely in the oxidized, folded state when it is overexpressed, thus indicating that the high cytoplasmic level of reduced glutathione is not sufficient alone to maintain the reduced state of Mia40. Co-expression of Grx1 (Meyer et al., 2009; Sagemark et al., 2007) keeps Mia40 mostly in the unfolded, reduced state, while co-expression of Trx1, which has a similar role in keeping protein thiols reduced in the

cytoplasm (Holmgren, 1979; Meyer et al., 2009), affects the oxidation state of Mia40 to a lesser extent. Additionally, Grx1 does not catalyze the reduction of Mia40 in the presence of reducing agents in vitro, implying some effect of the cytoplasmic environment that is not reproduced in vitro. Overall, these results indicate that Mia40, and likely other mitochondrial proteins that share the same structural motif and import pathway, tend to reach the oxidized, folded conformation in the absence of specific proteins, even in the reducing environment of the cytoplasm, and therefore require intervention of the cytoplasmic thiol-disulfide regulation mechanisms, especially those of Grx1, to reach the outer mitochondrial membrane in the reduced, import-competent state.

RESULTS

Transient expression of human Mia40, either alone or together with Grx1 or Trx1, was induced in human embryonic kidney (HEK293T) cells, and the relative expression levels were modulated by transfecting different amounts of DNA for each gene. The expression levels of each protein were determined with western blot analysis of the cell extracts (Figure 1). The concentration of overexpressed Mia40 remained constant in most of the samples, around 60 ± 10 μM. Grx1 was co-expressed in various amounts, ranging from 20 ± 10 μM to 100 ± 30 μM. Trx1 was co-expressed in two amounts: 35 ± 4 μM and 70 ± 15 μM. We assessed the cellular distribution of the overexpressed proteins in isolated cytosolic and mitochondrial fractions. Irrespective of the total protein abundance, >96% of Mia40 and >99% of either Grx1 or Trx1 were present in the cytosolic fraction (Figure S1A available online). Endogenous Mia40 remained undetected (<20 nM, data not shown), while endogenous levels of Grx1 were 60 ± 20 nM, ~10% of which was found in mitochondria, and 5 ± 2 μM of Trx1 was localized in the cytoplasm only, as measured on extracts of untransfected cells (Figure S1B).

Mia40, overexpressed alone in the cytoplasm of human cells, took a folded conformation, as monitored through ^{1}H-^{15}N and ^{1}H NMR spectra, which corresponds to the functional state of Mia40 that is normally found in the IMS (Figure 2A), and is the same conformation as that of oxidized Mia40 in vitro ($Mia40^{2S\text{-}S}$; Figure S2A). The central region of the protein is stably folded in two α helices while the N- and C-terminal regions are intrinsically unstructured. The cross-peaks of the folded part were well dispersed in the spectrum, while the peaks of the unstructured regions appeared as intense overlapped signals, which fell in the central part of the spectrum. Analysis of the methyl region of the ^{1}H NMR spectrum, which provides more sensitivity to assess the relative amount of folded Mia40 (Figure 2B), indicated that the folded protein was ~85% of the total protein amount (Figure 4). When it was co-expressed with various amounts of Grx1, a large fraction of Mia40 was present in the reduced, unfolded state. This effect was confirmed by a decrease of the methyl signal and also reflected in the ^{1}H-^{15}N NMR spectra, in which the cross-peaks of the folded part of Mia40 were barely detected as expected, while the cross-peaks of the unfolded parts were still visible, both in intact cells (Figures 2C and 2D) and in cell extracts (Figures 3C and 3D). In the presence of increasing amounts of Grx1, oxidized Mia40 ($Mia40^{2S\text{-}S}$) in the cytoplasm decreased from ~85% to ~25% of the total

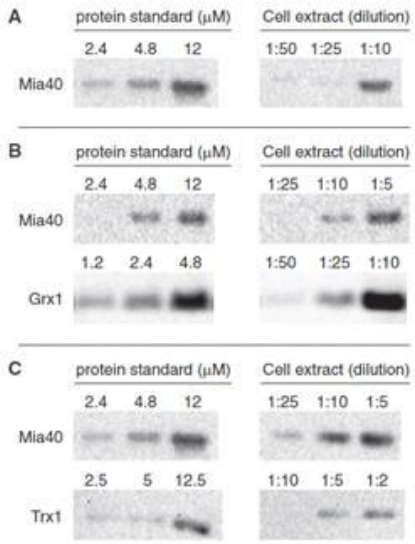


Figure 1. The Total Amount of Mia40, Grx1, and Trx1 in the Cell Extracts Is Measured with Western Blot Analysis

Samples of cell extracts were blotted at increasing dilutions together with samples of pure Mia40, Grx1 and Trx1 at known concentrations.

(A) Extract from cells expressing Mia40 alone.

(B) Extract from cells co-expressing Mia40 and Grx1 (1:0.25:2 *hMia40:hGrx1*:PEI ratio).

(C) Extract from cells co-expressing Mia40 and Trx1 (1:1:2 *hMia40:hTrx1*:PEI ratio) in selenium-supplemented medium. A calibration curve was obtained for each blot from the intensities of the standard Mia40 samples.

The concentration of Mia40 in the cell extracts was calculated for each dilution from the calibration curve and averaged.

See also Figure S1.

Mia40 (Figure 4). Co-expressed Grx1 was in the fully reduced and active state (Figure S1C), consistent with its reported redox potential (E_0 = −230 mV, compared to −290 mV for the GSH/GSSG couple in the cytoplasm; Sagemark et al., 2007). Grx1 is invisible in the in-cell NMR spectra because its signals are broadened beyond detection as a consequence of its slow tumbling rate, which is likely due to interactions with the cellular environment (a similar effect has been reported for several proteins overexpressed in *E. coli* cytoplasm, including wild-type ubiquitin [Sakai et al., 2006], cytochrome c [Crowley et al., 2011], *Escherichia coli* thioredoxin, and FKBP [Reckel et al., 2012]). This behavior was confirmed by the empty ^{1}H-^{15}N in-cell NMR spectra when Grx1 alone was expressed (Figure S3A), whereas Grx1 signals became visible upon cell lysis (Figure S3B). The NMR properties of cytoplasmic Grx1 therefore allowed us to obtain ^{1}H-^{15}N NMR spectra of Mia40 free of interference from Grx1 signals.

The effect of Trx1 on the Mia40 state was also investigated by inducing co-expression of Trx1 with Mia40. Similar to Grx1, Trx1 in the human cell cytoplasm is not detectable with NMR, consistent with what was reported in *E. coli* (Reckel et al., 2012; Figure S3C). The intracellular activity of Trx1 depends on the

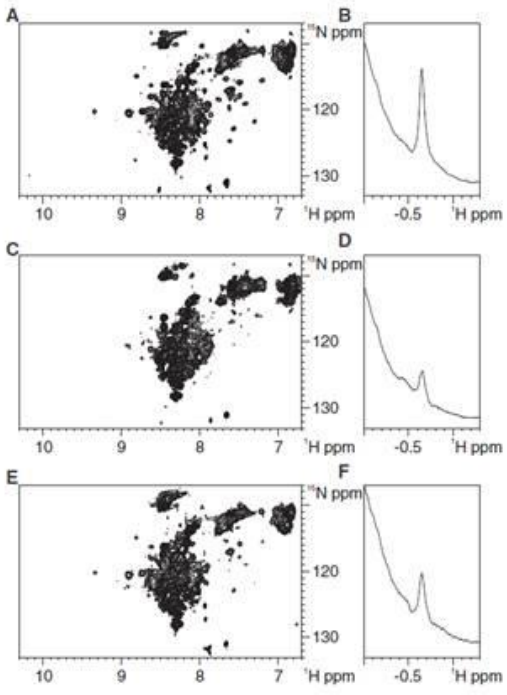


Figure 2. The Folding State of Mia40 in the Cytoplasm Depends on the Presence of Different Redox-Regulating Proteins

NMR spectra were acquired on human cells expressing uniformly ^{15}N-labeled Mia40 in the cytoplasm.

(A and B) Spectra of cells expressing Mia40.

(C and D) Spectra of cells co-expressing Mia40 and Grx1.

(E and F) Spectra of cells co-expressing Mia40 and Trx1 in selenium-supplemented medium.

(A, C, and E) ^{1}H-^{15}N SOFAST HMQC spectra. The strong, overlapped cross-peaks between 8.0 and 8.5 ppm (^{1}H) correspond to the unfolded regions, while the weaker, dispersed cross-peaks belong to the residues of the folded region of Mia40 (see also Figure S2).

(B, D, and F) Aliphatic region of ^{1}H spectra showing the ^{1}Hγ peak of Ile 53, which is a marker of the folded conformation of Mia40 and falls in a cellular background-free region.

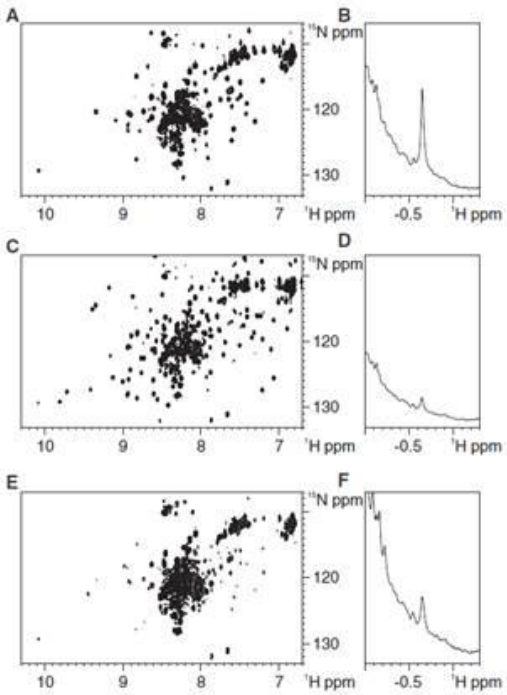


Figure 3. NMR Spectra Acquired on the Cell Extracts Corresponding to the Samples in Figure 2

NMR spectra were acquired on the cell extract of samples expressing uniformly ^{15}N-labeled Mia40. The relative amounts of the two Mia40 folding states remained unchanged upon cell lysis.

(A and B) Spectra of the cell extract containing Mia40.

(C and D) Spectra of the cell extract containing Mia40 and Grx1.

(E and F) Spectra of the cell extract containing Mia40 and Trx1 in selenium-supplemented medium.

Upon cell lysis, both Grx1 and Trx1 became visible in the ^{1}H-^{15}N spectra (C and E; see also Figure S3). Glutathionylation of unfolded Mia40 was excluded with mass spectrometry (see also Figure S4).

availability of selenium, which is required for the activation of thioredoxin reductase (TrxR; Ueno et al., 2007). Indeed, Trx1 co-expression in cells grown with basal amounts of selenium did not affect the folding state of Mia40 (data not shown). To ensure the complete activation of TrxR, and consequently active Trx1, sodium selenite was supplemented to cells co-expressing Trx1 and Mia40. In these conditions, co-expressed Trx1 was fully reduced (Figure S1C) and active, consistent with the reported Trx1 redox state at endogenous levels (Watson et al., 2003). Co-expression of active Trx1 affected the folding state of Mia40 to a lesser extent than Grx1 when expressed at similar levels because ~50% of cytoplasmic Mia40 was still in the folded, oxidized state (Figures 2E, 2F, 3E, and 3F). Therefore, the two thiol-regulating proteins have different efficacy in keeping Mia40 reduced, despite being reported to have overall similar functions in the cytoplasm (Figure 4).

In contrast to what was observed in the cells, fully reduced Grx1 had no effect on the Mia40 redox state in vitro. U-^{15}N Mia40$^{2S\text{-}S}$ was incubated in reduction conditions, either in the presence of 20 mM dithiothreitol (DTT) or 20 mM reduced glutathione (GSH), with increasing concentrations of unlabeled, fully reduced Grx1, and each step was monitored with NMR. No change in the ^{1}H-^{15}N NMR spectrum occurred upon addition of up to 2 eq of Grx1 to Mia40$^{2S\text{-}S}$ and 48 hr incubation, thus excluding a direct mechanism of reduction of Mia40 by Grx1. In a control experiment, U-^{15}N Mia40$^{2S\text{-}S}$ (Figure S2A) was completely reduced by heat denaturation at 95°C in buffer containing either 20 mM DTT or 20 mM GSH. No protein degradation occurred. The protein remained reduced in both reductants when cooled down at 25°C (Figure S2C). In these conditions,

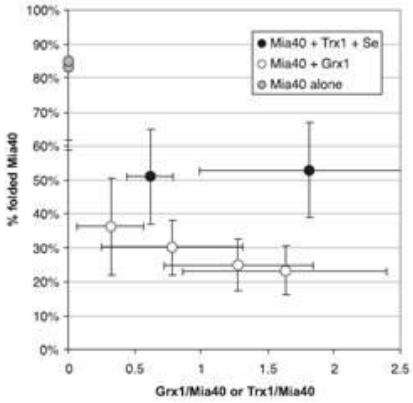


Figure 4. Effect of Grx1 and Trx1 on the Mia40 Folding State in the Cytoplasm

The ratio of folded Mia40 over total Mia40 was determined in cell samples with varying relative amounts of co-expressed Grx1 or Trx1. The amount of each protein was measured on cell extracts with western blot analysis. The amount of folded Mia40 was measured on the cell extracts by NMR. When only Mia40 was expressed, it was largely found in the folded state in the cytoplasm (gray circles). Co-expression of Grx1 caused a remarkable decrease of folded Mia40, which also occurred at substoichiometric amounts relative to Mia40 (white circles). Co-expression of Trx1 in the presence of selenium affected the folding state of Mia40 to a lesser extent (black circles).
Error bars represent SDs; n = 3.

the ^{1}H-^{15}N cross-peaks of the α-helical region disappeared, analogous to what was observed in the cytoplasm in the presence of Grx1. The alkylation reaction with 4-acetamido-4′-maleimidylstilbene-2,2′-disulfonic acid (AMS) depicted with SDS-PAGE confirmed the complete reduction of all cysteines (Figures S2B and S2D). Upon removal of the reducing agent and exposure to air, Mia40 rapidly reverted back to the folded, oxidized conformation. To test whether Grx1 could bind fully reduced Mia40 and prevent its oxidation, 2 eq of Grx1 was added to a sample of U-^{15}N labeled, unfolded Mia40 in presence of 20 mM GSH. Upon exposure to air, Mia40 rapidly folded, thus indicating that Grx1 did not protect reduced Mia40 from oxidation. Grx1-catalyzed glutathionylation of reduced Mia40, either in vitro in presence of GSH or in the cytoplasm, was excluded by mass spectrometry analysis of the in vitro Mia40 samples and of Mia40 isolated from cell extracts (Figure S4).

DISCUSSION

The current model of Mia40 maturation pathway requires that the protein, which is natively expressed in the cytoplasm from nuclear mRNA, crosses the outer mitochondrial membrane through the TOM channel in a reduced and unfolded conformation. No Mia40 has been reported to reside in the cytoplasm, except for the time required to translocate to the mitochondria (Hofmann et al., 2005). However, when Mia40 was overexpressed, it did not translocate quantitatively to the mitochondria, and close to all of the Mia40 remained in the cytoplasm. This effect could be explained with the limited availability of the import machinery on the outer mitochondrial membrane (i.e., the TOM channel), which determines an upper limit to the import level of Mia40. Consequently, unimported Mia40 accumulates in the cytoplasm. Our results show that in these conditions, cytoplasmic Mia40 reached the folded state, which is thus a thermodynamically favored conformation in the cytoplasm. Therefore, in physiologic conditions there is the need for a mechanism to keep Mia40 reduced until it reaches the outer mitochondrial membrane because the amount and ratio of cytoplasmic GSH are apparently not sufficient for that purpose.

Overexpression of Grx1 allowed most Mia40 to remain in the reduced, unfolded state. This effect also occurred with substoichiometric amounts of Grx1, consistent with a catalytic role of Grx1. Conversely, when oxidized Mia40 was incubated in vitro with Grx1 in the presence of GSH, it did not change its redox state and the reduction of the structural disulfide bonds was only possible upon heat denaturation in reducing conditions. Moreover, Grx1 did not prevent oxidation of heat-denatured Mia40 when added in the presence of GSH and molecular oxygen, and neither partially oxidized Mia40 nor glutathionylated Mia40 were formed. The lack of a direct interaction of Grx1 with Mia40 suggests that the effect of Grx1 on the oxidation state of intracellular Mia40 is mediated by some other component of the cytoplasm.

Overexpression of Trx1 resulted in a weaker effect because a sizable fraction of Mia40 reached the folded state. As with Grx1, the effect occurred also with a substoichiometric amount of Trx1, therefore suggesting that the effect of Grx1, and to a lesser extent that of Trx1, is not the consequence of a generic increase in reducing power in the cytoplasm, which would be concentration-dependent.

Overall, the data show that cytoplasmic redox-regulation systems, Grx1 in particular, have a specific catalytic—although likely indirect—role in keeping a sizable fraction of Mia40 unfolded in the cytoplasm, implying a link between these systems and the Mia40 maturation pathway.

SIGNIFICANCE

We characterized the folding state of Mia40 in the cytoplasm, obtaining atomic-level information in living human cells with NMR. We found that the folded form of Mia40, which cannot be imported into mitochondria, is thermodynamically favored in the cytoplasm. We also showed that the folding of Mia40 is controlled by the cytoplasmic Grx1 and Trx1 redox-regulation systems, Grx1 having a stronger effect. These results show the general relevance of atomic resolution studies performed in living cells, which provide a way to describe cellular physiologic processes such as the redox-controlled protein folding, and to understand how other involved pathways can affect and regulate such processes.

EXPERIMENTAL PROCEDURES

Overexpression of Mia40, Grx1, and Trx1 in human cells was performed by following a previously established protocol (Aricescu et al., 2006; Banci et al., 2013). Briefly, the cDNA sequences encoding Mia40 (amino acids 1–142, GenBank accession number: NP_001091972.1), Grx1 (amino acids

1–106, GenBank accession number: NP_001112362.1), and Trx1 (amino acids 1–105, GenBank accession number: NP_003320.2) were amplified with PCR and subcloned into the pHLsec vector (Aricescu et al., 2006; Banci et al., 2013) between EcoRI and XhoI restriction enzyme sites. The clones were verified by gene sequencing. Transient transfection was obtained by treating the cells with a DNA:polyethylenimine mixture. Cells were collected after 48 hr for in-cell NMR and western blot analysis. Intracellular distribution was assessed by separating the cytoplasmic and the mitochondrial fractions from cell extracts using a mitochondria isolation kit for cultured cells (Thermo Scientific).

Grx1 was co-expressed with Mia40, either unlabeled or with uniform ^{15}N labeling. Different ratios of DNA were chosen to obtain different protein ratios: 1:0.75:2; 1:0.5:2; 1:0.25:2; and 1:0.125:2 *hMia40:hGrx1*:PEI. Trx1 was co-expressed with Mia40 with the following ratios of DNA: 1:1:2 and 0.5:1:2 *hMia40:hTrx1*:PEI. To ensure complete activation of Trx1, sodium selenite (kindly provided by Prof. A. Arcangeli lab, University of Florence) was supplemented to the cell culture media to a final concentration of 100 nM, starting 24 hr before transfection.

The amount of folded Mia40 in the cell extracts was measured with NMR through the standard addition of pure, folded Mia40 at known concentration. The ^1H resonance at −0.7 ppm of Ile 53 Hγ was used as a marker of the folded conformation of Mia40. Total Mia40 was determined on the same cell extracts by western blot analysis, by using the same pure Mia40 sample at increasing dilutions as a reference; Grx1 and Trx1 concentrations in the cell extracts were determined as above by using samples of purified Grx1 and Trx1 as references. Mia40 was stained with a rabbit polyclonal anti-Mia40 antibody (Abcam: ab87033, diluted to 0.5 μg/ml); Grx1 with a rabbit polyclonal anti-glutaredoxin 1 antibody (Abcam: ab45953, 1.0 μg/ml); Trx1 with a rabbit polyclonal anti-thioredoxin/TRX antibody (Abcam: ab26320, 0.4 μg/ml). Goat anti-rabbit IgG (whole molecule)-peroxidase secondary antibody (Sigma: A0545) was used for detection, diluted at 1:160,000 (for Mia40) or 1:80,000 (for Grx1 and Trx1). For checking the purity of the cytoplasmic and mitochondrial fractions, antibodies against a cytoplasmic marker (rabbit polyclonal anti-GAPDH antibody, Abcam: ab9485) and a mitochondrial marker (rabbit polyclonal anti-COX IV, Abcam: ab16056) were used.

U-^{15}N-labeled Mia40 for in vitro experiments and unlabeled Mia40 for NMR quantifications were produced as previously described (Banci et al., 2009b). Glutaredoxin 1 for in vitro interaction with Mia40 was produced as follows: a pTH34 vector containing the human Grx1 gene (N-term fused with His-tag and TEV recognition site) was transformed in *E. coli* BL21(DE3) gold competent cells. Cells were grown at 37°C in minimal medium to an optical density 0.6 and then induced with 0.5 mM IPTG for 16 hr at 25°C. Glutaredoxin 1 was purified by affinity chromatography using a nickel-chelating HisTrap (GE Healthcare) column. After digestion with AcTEV protease (Invitrogen) O/N at 25°C, the protein was separated from the affinity tag in a HisTrap column. The sample buffer was then exchanged with 50 mM potassium phosphate, 0.5 mM EDTA, pH = 7. Pure thioredoxin 1 for western blot analysis was prepared from HEK293T cells overexpressing Trx1. Cells were lysed in 20 mM potassium phosphate buffer (pH 7) and the cleared extract was incubated at 70°C for 4 min. After centrifugation (30 min at 16,000 × *g*), Trx1 was purified by anionic exchange chromatography using a HiTrap DEAE FF (Amersham Biosciences) column applying a linear gradient of potassium phosphate buffer (10–100 mM, pH 7).

Mia40 for mass spectrometry analysis was isolated from cell extracts by ionic exchange chromatography using a HiTrap DEAE FF (Amersham Biosciences) column applying a linear gradient of NaCl (0–500 mM NaCl in 20 mM TRIS buffer, pH 8). For both in vitro and cell extracts samples, the buffer was exchanged with 0.1% trifluoroacetic acid/ 50% acetonitrile by Zip Tip_{C18} tips (MILLIPORE). Mass spectrometry analysis was performed at a Bruker Ultraflex III MALDI TOF/TOF instrument, using an α-cyano-4-hydroxycinnamic acid matrix.

For thiol alkylation reaction with 4-acetamido-4′-maleimidylstilbene-2,2′-disulfonic acid (AMS), either cell or protein samples were precipitated with 10% trichloroacetic acid, washed with acetone, and resuspended in 100 mM Tris pH 7 + 2% SDS. The mixtures were then incubated 1 hr at 37°C with 20 mM AMS, and run on a nonreducing SDS-PAGE.

NMR experiments were acquired with a 950 MHz Bruker Avance III spectrometer equipped with a CP TCI CryoProbe. 1D ^1H and 2D $^1H,^{15}N$-SOFAST-HMQC (Schanda and Brutscher, 2005) spectra were acquired at 305K. The total acquisition time for each cell sample ranged from 1 to 2 hr. The supernatant of each cell sample was checked for protein leakage in the same experimental conditions. The same NMR spectra were also acquired on the cell extracts. Cell viability before and after NMR experiments was assessed with trypan blue staining (Freshney, 1987). Cell viability remained above 90% because damaged cells ranged from 3% before the experiments to 8% after the experiments. 2D $^1H,^{15}N$-SOFAST-HMQC NMR spectra of cell samples and extracts were processed with Bruker Topspin 3.1 software by subtracting a spectrum of untransfected cells/cell extract acquired within the same experimental conditions, thereby eliminating the interference of signals arising from unspecific ^{15}N labeling.

SUPPLEMENTAL INFORMATION

Supplemental Information includes four figures and can be found with this article online at http://dx.doi.org/10.1016/j.chembiol.2013.05.007.

ACKNOWLEDGMENTS

This work was supported by the Programmi di Ricerca di Rilevante Interesse Nazionale (PRIN; 2009FAKHZT_001 "Biologia strutturale meccanicistica: avanzamenti metodologici e biologici"), the Italian "Rete Nazionale per lo studio della Proteomica Umana (Italian Human ProteomeNet)" FIRB PROTEOMICA MIUR -RBRN07BMCT, and Ente Cassa di Risparmio di Firenze ("Biologia Strutturale Integrata").

Received: January 5, 2013
Revised: May 8, 2013
Accepted: May 17, 2013
Published: June 20, 2013

REFERENCES

Aricescu, A.R., Lu, W., and Jones, E.Y. (2006). A time- and cost-efficient system for high-level protein production in mammalian cells. Acta Crystallogr. D Biol. Crystallogr. *62*, 1243–1250.

Banci, L., Bertini, I., Cefaro, C., Ciofi-Baffoni, S., Gallo, A., Martinelli, M., Sideris, D.P., Katrakili, N., and Tokatlidis, K. (2009a). MIA40 is an oxidoreductase that catalyzes oxidative protein folding in mitochondria. Nat. Struct. Mol. Biol. *16*, 198–206.

Banci, L., Bertini, I., Ciofi-Baffoni, S., and Tokatlidis, K. (2009b). The coiled coil-helix-coiled coil-helix proteins may be redox proteins. FEBS Lett. *583*, 1699–1702.

Banci, L., Bertini, I., Cefaro, C., Cenacchi, L., Ciofi-Baffoni, S., Felli, I.C., Gallo, A., Gonnelli, L., Luchinat, E., Sideris, D.P., and Tokatlidis, K. (2010). Molecular chaperone function of Mia40 triggers consecutive induced folding steps of the substrate in mitochondrial protein import. Proc. Natl. Acad. Sci. USA *107*, 20190–20195.

Banci, L., Barbieri, L., Bertini, I., Luchinat, E., Secci, E., Zhao, Y., and Aricescu, A.R. (2013). Atomic-resolution monitoring of protein maturation in live human cells by NMR. Nat. Chem. Biol. *9*, 297–299. http://dx.doi.org/10.1038/nchembio.1202.

Chacinska, A., Pfannschmidt, S., Wiedemann, N., Kozjak, V., Sanjuan Szklarz, L.K., Schulze-Specking, A., Truscott, K.N., Guiard, B., Meisinger, C., and Pfanner, N. (2004). Essential role of Mia40 in import and assembly of mitochondrial intermembrane space proteins. EMBO J. *23*, 3735–3746.

Chacinska, A., Guiard, B., Müller, J.M., Schulze-Specking, A., Gabriel, K., Kutik, S., and Pfanner, N. (2008). Mitochondrial biogenesis, switching the sorting pathway of the intermembrane space receptor Mia40. J. Biol. Chem. *283*, 29723–29729.

Crowley, P.B., Chow, E., and Papkovskaia, T. (2011). Protein interactions in the Escherichia coli cytosol: an impediment to in-cell NMR spectroscopy. ChemBioChem *12*, 1043–1048.

Durigon, R., Wang, Q., Ceh Pavia, E., Grant, C.M., and Lu, H. (2012). Cytosolic thioredoxin system facilitates the import of mitochondrial small Tim proteins. EMBO Rep. *13*, 916–922.

Freshney, R. (1987). Culture of Animal Cells: A Manual of Basic Technique (New York: Alan R. Liss).

Gabriel, K., Milenkovic, D., Chacinska, A., Müller, J., Guiard, B., Pfanner, N., and Meisinger, C. (2007). Novel mitochondrial intermembrane space proteins as substrates of the MIA import pathway. J. Mol. Biol. *365*, 612–620.

Grumbt, B., Stroobant, V., Terziyska, N., Israel, L., and Hell, K. (2007). Functional characterization of Mia40p, the central component of the disulfide relay system of the mitochondrial intermembrane space. J. Biol. Chem. *282*, 37461–37470.

Hofmann, S., Rothbauer, U., Mühlenbein, N., Baiker, K., Hell, K., and Bauer, M.F. (2005). Functional and mutational characterization of human MIA40 acting during import into the mitochondrial intermembrane space. J. Mol. Biol. *353*, 517–528.

Holmgren, A. (1979). Reduction of disulfides by thioredoxin. Exceptional reactivity of insulin and suggested functions of thioredoxin in mechanism of hormone action. J. Biol. Chem. *254*, 9113–9119.

Holmgren, A. (1989). Thioredoxin and glutaredoxin systems. J. Biol. Chem. *264*, 13963–13966.

Inomata, K., Ohno, A., Tochio, H., Isogai, S., Tenno, T., Nakase, I., Takeuchi, T., Futaki, S., Ito, Y., Hiroaki, H., and Shirakawa, M. (2009). High-resolution multi-dimensional NMR spectroscopy of proteins in human cells. Nature *458*, 106–109.

Longen, S., Bien, M., Bihlmaier, K., Kloeppel, C., Kauff, F., Hammermeister, M., Westermann, B., Herrmann, J.M., and Riemer, J. (2009). Systematic analysis of the twin cx(9)c protein family. J. Mol. Biol. *393*, 356–368.

Lu, H., Allen, S., Wardleworth, L., Savory, P., and Tokatlidis, K. (2004). Functional TIM10 chaperone assembly is redox-regulated in vivo. J. Biol. Chem. *279*, 18952–18958.

Mesecke, N., Terziyska, N., Kozany, C., Baumann, F., Neupert, W., Hell, K., and Herrmann, J.M. (2005). A disulfide relay system in the intermembrane space of mitochondria that mediates protein import. Cell *121*, 1059–1069.

Meyer, Y., Buchanan, B.B., Vignols, F., and Reichheld, J.P. (2009). Thioredoxins and glutaredoxins: unifying elements in redox biology. Annu. Rev. Genet. *43*, 335–367.

Naoé, M., Ohwa, Y., Ishikawa, D., Ohshima, C., Nishikawa, S.I., Yamamoto, H., and Endo, T. (2004). Identification of Tim40 that mediates protein sorting to the mitochondrial intermembrane space. J. Biol. Chem. *279*, 47815–47821.

Neupert, W., and Herrmann, J.M. (2007). Translocation of proteins into mitochondria. Annu. Rev. Biochem. *76*, 723–749.

Ogino, S., Kubo, S., Umemoto, R., Huang, S., Nishida, N., and Shimada, I. (2009). Observation of NMR signals from proteins introduced into living mammalian cells by reversible membrane permeabilization using a pore-forming toxin, streptolysin O. J. Am. Chem. Soc. *131*, 10834–10835.

Reckel, S., Hänsel, R., Löhr, F., and Dötsch, V. (2007). In-cell NMR spectroscopy. Prog. Nucl. Magn. Reson. Spectrosc. *51*, 91–101.

Reckel, S., Lopez, J.J., Löhr, F., Glaubitz, C., and Dötsch, V. (2012). In-cell solid-state NMR as a tool to study proteins in large complexes. ChemBioChem *13*, 534–537.

Sagemark, J., Elgán, T.H., Bürglin, T.R., Johansson, C., Holmgren, A., and Berndt, K.D. (2007). Redox properties and evolution of human glutaredoxins. Proteins *68*, 879–892.

Sakai, T., Tochio, H., Tenno, T., Ito, Y., Kokubo, T., Hiroaki, H., and Shirakawa, M. (2006). In-cell NMR spectroscopy of proteins inside Xenopus laevis oocytes. J. Biomol. NMR *36*, 179–188.

Schanda, P., and Brutscher, B. (2005). Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds. J. Am. Chem. Soc. *127*, 8014–8015.

Selenko, P., Frueh, D.P., Elsaesser, S.J., Haas, W., Gygi, S.P., and Wagner, G. (2008). In situ observation of protein phosphorylation by high-resolution NMR spectroscopy. Nat. Struct. Mol. Biol. *15*, 321–329.

Tokatlidis, K. (2005). A disulfide relay system in mitochondria. Cell *121*, 965–967.

Ueno, H., Kajihara, H., Nakamura, H., Yodoi, J., and Nakamuro, K. (2007). Contribution of thioredoxin reductase to T-cell mitogenesis and NF-kappaB DNA-binding promoted by selenite. Antioxid. Redox Signal. *9*, 115–121.

Watson, W.H., Pohl, J., Montfort, W.R., Stuchlik, O., Reed, M.S., Powis, G., and Jones, D.P. (2003). Redox potential of human thioredoxin 1 and identification of a second dithiol/disulfide motif. J. Biol. Chem. *278*, 33408–33415.

Chemistry & Biology, Volume *20*

Supplemental Information

Visualization of Redox-Controlled Protein Fold in Living Cells

Lucia Banci, Letizia Barbieri, Enrico Luchinat, and Erica Secci

Inventory of Supplemental Information

Figure S1: WB analysis of cytoplasmic and mitochondrial fractions of cell samples overexpressing Mia40, Grx1 or Trx1, which allowed their subcellular localization to be assessed; dot blot analysis of endogenous levels of Mia40, Grx1 and Trx1 (related to Figure 1).

Figure S2: NMR spectra showing the folding and redox states of Mia40 *in vitro*, which are used as a reference to analyze the folding state of cytoplasmic Mia40 (related to Figure 2).

Figure S3: NMR spectra showing that intracellular Grx1 and Trx1 are not detectable by NMR, and only become visible in the cell extracts (related to Figure 3).

Figure S4: MALDI mass spectrometry analysis of Mia40 isolated from cell extracts and *in vitro* Mia40 samples, in which is shown the absence of glutathionylated forms of Mia40 (related to Figure 3).

Supplementary Figures:

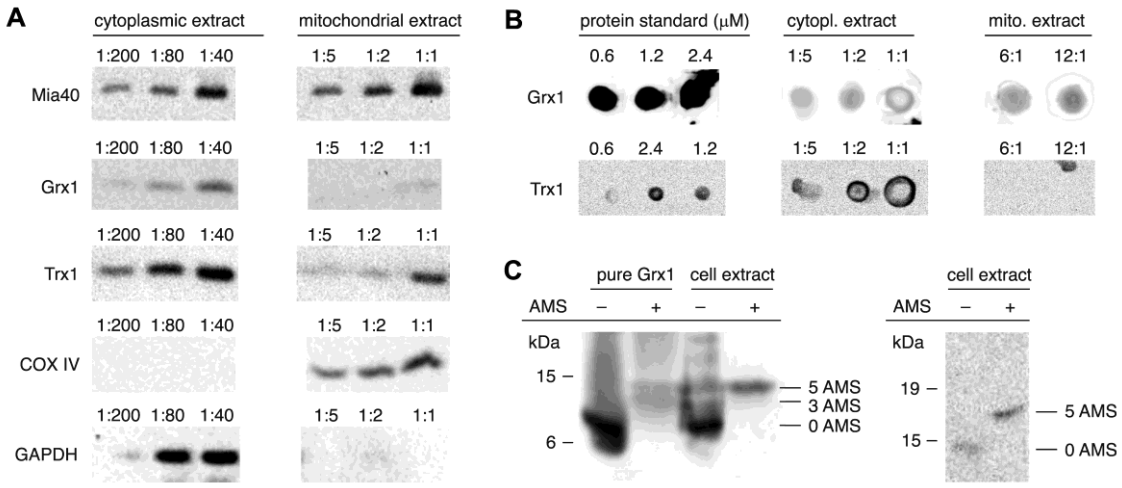


Figure S1, related to Figure 1: Overexpressed Mia40, Grx1 and Trx1 are mostly localized in the cytoplasm; Grx1 and Trx1 are fully reduced. (A) Cytoplasmic and mitochondrial fractions of cells were blotted at increasing dilutions to assess the intracellular distribution of overexpressed Mia40, Grx1 and Trx1. Dilutions are expressed relative to the most concentrated mitochondrial extract. The subcellular fractions were obtained using a mitochondria isolation kit for cultured cells (Thermo Scientific). The fraction purity was checked with a mitochondrial and a cytoplasmic marker (COX IV and GAPDH, respectively). (B) Endogenous levels and subcellular localization of Grx1 and Trx1 were measured by dot blot analysis of cytoplasmic and mitochondrial extracts. Pure samples at known concentration were blotted as a reference. (C) AMS thiol alkylation reaction on a cell extract containing overexpressed Grx1 was detected by Western Blot, showing that all Grx1 cysteines are reduced. Purified Grx1 was run as a reference. The same reaction was performed on a cell extract containing Trx1 overexpressed in presence of fully active TrxR, showing that all Trx1 cysteines are reduced.

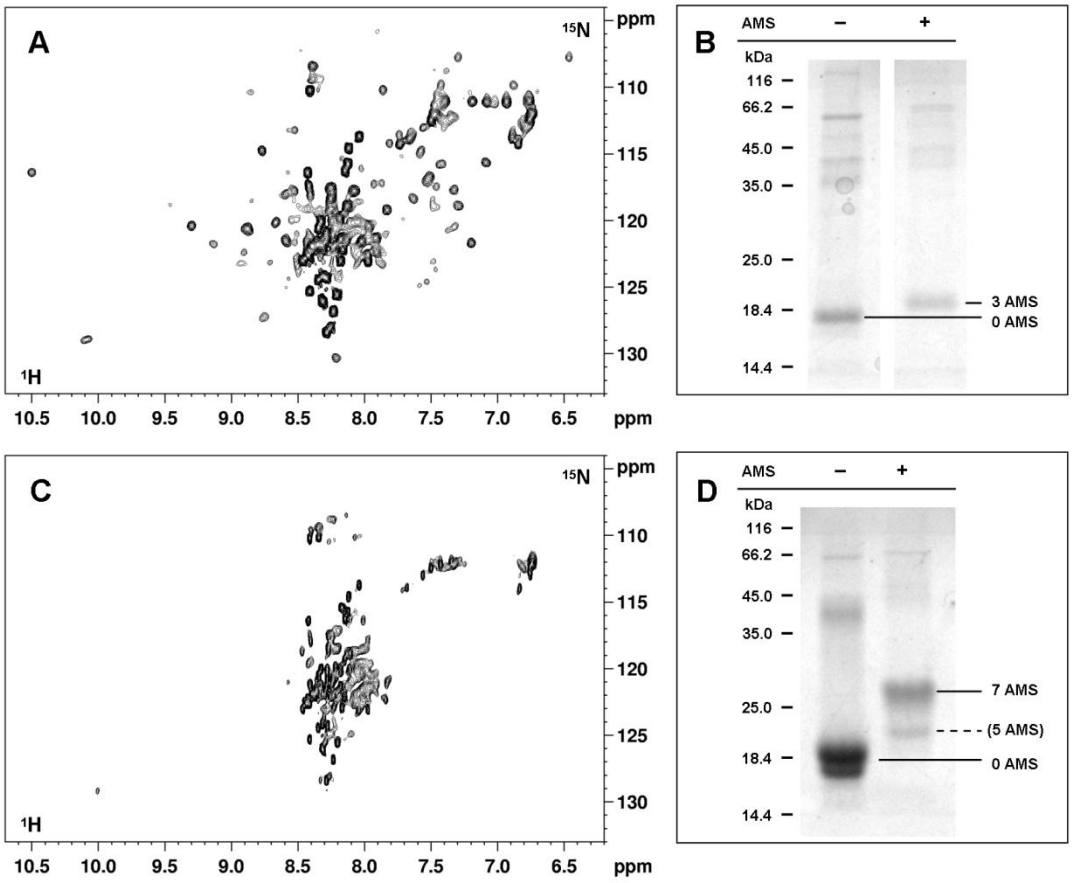


Figure S2, related to Figure 2: Oxidized, folded Mia40 is reduced *in vitro* upon heat denaturation in reducing conditions. ^{1}H-^{15}N SOFAST HMQC spectra were acquired on pure U-^{15}N Mia40$^{2S\text{-}S}$ (A) before and (C) after denaturation at 95°C in presence of 10 mM GSH. After denaturation, the crosspeaks of the N- and C-terminal unfolded segments of Mia40 are still visible (with sharper lines), while those corresponding to the core segment of oxidized Mia40 disappear. AMS reaction was performed on Mia40 samples before and after denaturation, which were run on Coomassie-stained SDS-PAGE (B, D), confirming the reduction of the two structural disulfide bonds.

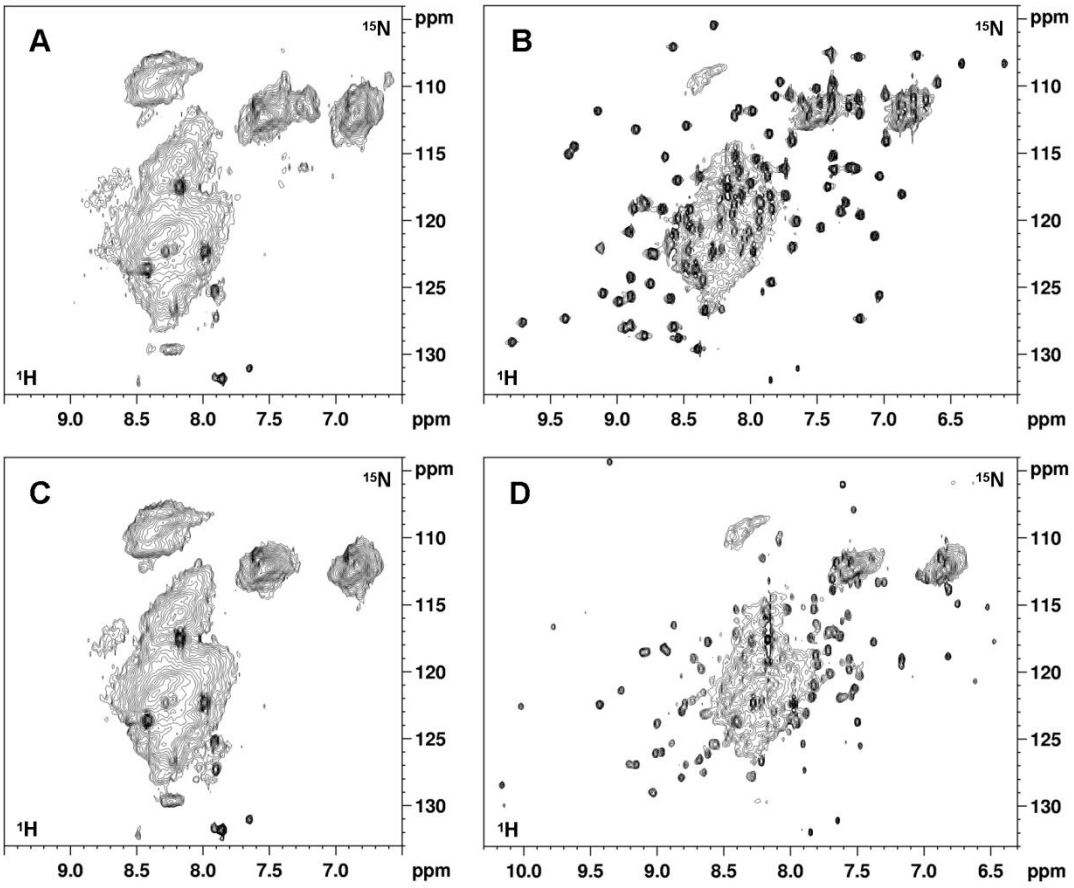


Figure S3, related to Figure 3: The amide signals of cytoplasmic Grx1 and Trx1 are broadened beyond detection. ^{1}H-^{15}N SOFAST-HMQC spectra were acquired on samples of human cells expressing (A) Grx1 and (C**)** Trx1, and on the corresponding cell extracts (B, D). Only the cellular background signals were visible in the spectra of intact cells, where the amide crosspeaks of Grx1 and Trx1 are broadened beyond detection. Upon cell lysis, the crosspeaks of both proteins were detected.

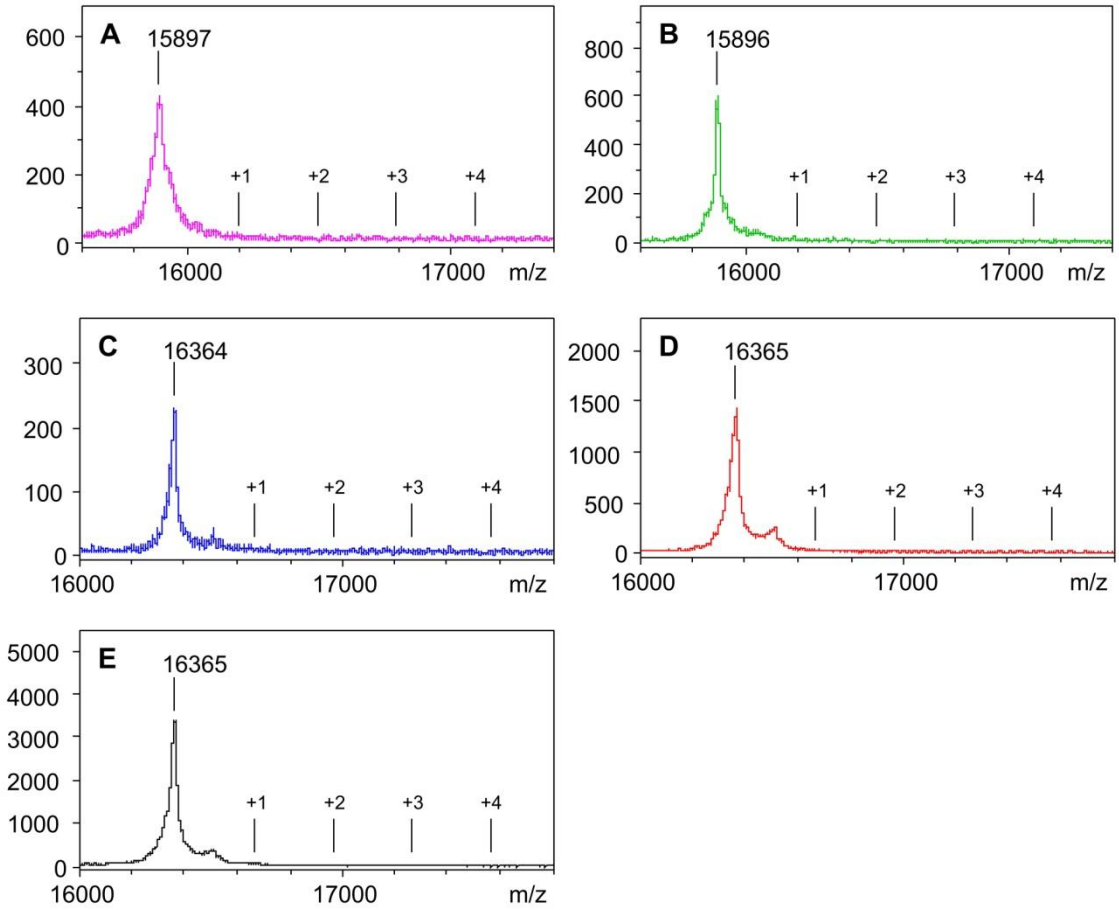


Figure S4, related to Figure 3: Mass spectrometry analysis shows absence of glutathionylated Mia40. MALDI TOF/TOF analysis of Mia40 isolated from cell extracts (native Mia40) and Mia40 *in vitro* samples (GSFT-Mia40). (A) Extract of cells overexpressing Mia40 alone; (B) extract of cells co-expressing Mia40 and Grx1; (C) *in vitro* Mia40 heat denatured in 20 mM GSH (phosphate buffer, pH 7), after removal of GSH; (D) same sample as (C), incubated with 2 eq. of Grx1 in 20 mM GSH, after removal of GSH. Points in the spectra corresponding to increasingly glutathionylated forms of Mia40 are indicated (+305.3 Da for each GSH molecule).

2.2 Sequential protein expression and selective labeling for in-cell NMR in human cells

When the genes encoding two proteins are transiently co-transfected, both proteins are overexpressed and labeled at the same time. In order to effectively characterize protein-protein interactions in mammalian cells, the in-cell NMR signals of only one partner need to be detected. This requires the selective labeling of only one protein. A way to reach this goal is the overexpression of a protein in a transient way while the other is constitutively overexpressed by the cells. Stable HEK293T cell lines overexpressing HAH1, a cytosolic copper chaperone protein were established. The HAH1 expression levels were checked for each cell line by Western Blot analysis, and the one with the highest levels was chosen. To test the protocol, the stable cell line was grown in unlabeled medium and was transiently transfected with either SOD1 or Trx1 (whose signals are not detectable due to interactions with the cellular environment) in [U-^{15}N]-labeled medium. In either case, the NMR signals of ^{15}N-HAH1 were still detected, due to HAH1 being continuously expressed together with the second protein and therefore becoming labeled to some extent. A possible way to increase the labeling selectivity is to decrease the expression rate of HAH1 during the transient transfection, by exploiting the RNA interference (RNAi) machinery of the cell to silence the HAH1 gene, and by performing the medium switch after the silencing has taken place. We used a transcription plasmid (pSilencer vector) that it is able to produce a shRNA specific against the mRNA of HAH1. We prepared three pSilencer vectors, each encoding a different shRNA sequence, and we tested them by checking the HAH1 expression level in stably transfected cells. The relative levels of HAH1 mRNA were assessed by RT-PCR analysis, and we verified that after 24 hours of silencing a combination of the three shRNA sequences had higher efficiency compared to each sequence alone. The HAH1 mRNA and protein levels were checked respectively by RT-PCR and Western Blot at different times after the transfection, ranging from 6 to 72 hours. A useful time window to perform the medium switch and the labeling of the transiently expressed protein was found, from 14 to 30 hours after the transfection, in which the HAH1 gene was silenced while the protein levels did not decrease much due to protein turnover. NMR experiments were performed on cells where the stably expressed HAH1 was silenced and SOD1 was transiently expressed, keeping the cells in unlabeled medium for 14 hr and 16 hr in ^{15}N-labeled medium. In the ^{1}H-^{15}N SOFAST-

HMQC spectra, only the signals of SOD1 were detected in both cells and lysate samples. In the 1D ^{1}H NMR spectra, we verified that the total final levels of HAH1 were ~70% of the control sample with no silencing. This approach can be a useful tool for the characterization of protein-protein interactions, allowing the detection of the NMR signals of only one intracellular partner.

Sequential protein expression and selective labeling for in-cell NMR in human cells

Enrico Luchinat[a,b,*], Erica Secci[a], Francesca Cencetti[b], Paola Bruni[b]

Affiliations:

[a]Magnetic Resonance Center - CERM, University of Florence, Via Luigi Sacconi 6, 50019 Sesto Fiorentino, Florence, Italy.

[b]Department of Biomedical, Experimental and Clinical Sciences, "Mario Serio", University of Florence, Viale Morgagni 50, 50134 Florence, Italy.

* To whom correspondence should be addressed E.L. (eluchinat@cerm.unifi.it).

SUBMITTED

Abstract

Background: In-cell NMR is a powerful technique to investigate proteins in living cells at atomic resolution, especially when applied to human cells. Ideally, when studying functional processes involving protein-protein interactions by NMR, only one partner should be isotopically labeled. Here we show that constitutive and transient protein expression can be combined together with protein silencing to obtain selective protein labeling in human cells.

Methods: We established a human cell line stably overexpressing the copper binding protein HAH1. A second protein (human superoxide dismutase 1, SOD1) was overexpressed by transient transfection and isotopically labeled. A silencing vector containing shRNA sequences against the HAH1 gene was used to decrease the rate of HAH1 synthesis during the expression of SOD1. The levels of HAH1 mRNA and protein were measured as a function of time following transfection by RT-PCR and Western Blot, and the final cell samples were analyzed by in-cell NMR.

Results: SOD1 was ectopically expressed and labeled in a time window during which HAH1 biosynthesis was strongly decreased by shRNA, thus preventing its labeling. In-cell NMR spectra confirmed that, while both proteins were present in the sample, only SOD1 was selectively labeled and could be detected by ^{1}H-^{15}N heteronuclear NMR.

Conclusions and General Significance: We showed that controlling protein expression by specifically silencing a stably expressed protein is a useful strategy to obtain selective labeling of only one protein. This approach relies on established techniques thus permitting the investigation of protein-protein interactions by NMR in human cells.

Keywords: in-cell NMR; nuclear magnetic resonance; protein-protein interactions; isotopic labeling; mammalian cells.

Introduction

In-cell NMR spectroscopy has been proven to be an ideal technique to obtain atomic-level structural and functional information on biological macromolecules directly in living cells, i.e. in their native environment (1-3). The data obtained by in-cell NMR are

likely closest to the physiological conditions than any other atomic-resolution technique. In-cell NMR allows physiological processes to be investigated such as protein folding and misfolding (4–6), interactions with other proteins or nucleic acids (7–10), binding of metal ions (11–13), and ligand screening (14). Most of the initial works focused on proteins overexpressed in *E. coli* cells, due to the high protein levels attainable and to the ease of growing and handling the cells. Although bacterial cells are considered a good model to investigate cytoplasmic proteins, they lack many cellular machineries for protein folding and post-translational modifications, which are crucial for the correct maturation of many eukaryotic proteins. Due to such limitations of the bacterial models, in-cell NMR in eukaryotic cells is a necessary step forward for studying proteins in their native environment. Several approaches have been developed to deliver a purified labeled protein into the cytoplasm of eukaryotic cells, including microinjection in *X. laevis* oocytes (15), fusion with cell-penetrating peptides (16), and cell permeabilization either through pore-forming toxins (17, 18) or by cell electroporation (19, 20). Alternatively, methods have been developed to overexpress the protein of interest directly into eukaryotic cells, and have been applied to yeast and insect cells (21, 22) and to cultured human cells (13).

Protein overexpression in human cells has been used to follow the various steps of protein maturation right after protein biosynthesis, including folding and metal binding (5, 13, 23), changes in redox state (24), and targeting towards cellular compartments such as mitochondria (25). With the same approach, ectopic co-expression of two or more proteins has permitted to observe the effects of functional interactions with specific partners on the protein of interest (5, 13, 24). The employed strategy for protein co-expression causes the two (or more) proteins to be expressed simultaneously by the cells, and does not allow any labeling selectivity between them. Therefore, as soon as the isotopically enriched medium is provided, all the expressed proteins are equally labeled, and are detected in the in-cell NMR spectra. Ideally, when studying biomolecular interactions by NMR, only one partner should be labeled in order to allow the correct interpretation of any spectral changes upon interaction. In bacterial cells, methods have been developed for this purpose, in which orthogonal induction systems are used for two or more proteins, allowing them to be expressed at different times and with different labeling (7, 8). To date, similar approaches that permit sequential protein expression and labeling in human cells are lacking.

Here we show that constitutive and transient protein expression can be combined with protein knock-down by gene silencing to allow two human proteins to be expressed at different times in human cells, permitting selective isotopic labeling of one protein. This novel methodological approach was tested and optimized on two human proteins that have been extensively characterized by NMR both in-cell and *in vitro*: HAH1 (26, 27) and SOD1 (5, 11, 13, 28, 29). We are confident that this novel experimental procedure will enable the study of protein-protein interactions in mammalian cells by NMR.

Materials and Methods

Gene cloning

The gene encoding full-length human HAH1 (amino acids 1-68, GenBank accession number: NP_004036.1) was amplified from cDNA by PCR and cloned in the pURD vector (kindly provided by Y. Zhao, University of Oxford) (30) between HindIII and XhoI restriction enzymes. Three sequences of small hairpin RNA (shRNA) were designed to perform silencing of HAH1, in the form of sense-loop-antisense sequences (31, 32). One sequence, shRNA-1, targets the HAH1 open reading frame (5'-G**ACGAGTTCTCTGTGGACATGA**CGAA**TCATGTCCACAGAGAACTCGT**TTTTTTTGGAA-3'; sense and antisense sequences are shown in bold) while other two target unique UTR elements present in the pURD vector: shRNA-2 targets the woodchuck hepatitis virus post-transcriptional regulatory element (WPRE: 5'-**GCAATCAACCTCTGGATTACA**CGAA**TGTAATCCAGAGGTTGATTGC**TTTTTTTGGAA-3'); shRNA-3 targets the bovine growth hormone polyadenylation signal (bGHpA: 5'-**GGAAGACAATAGCAGGCATGC**CGAA**GCATGCCTGCTATTGTCTTCC**TTTTTTTGGAA-3'). The corresponding synthetic double stranded DNA oligonucleotides – flanked by compatible ends for BamHI and HindIII restriction sites – were individually cloned in the pSilencer 2.1-U6 Hygro vector (Life Technologies) following the vector manual. pHLsec-derived vectors containing the cDNA encoding the full-length human SOD1 (pHLsec-SOD1) and the full-length human thioredoxin 1 (pHLsec-TRX1) were used for transient expression as previously described (13).

Generation of stable cell lines

HEK293T-derived cell lines stably overexpressing HAH1 were obtained following a reported protocol (33). Parental 293T cells (ATCC CRL-3216) were transfected in a 6-well dish with the pURD vector containing the HAH1 cDNA, together with the PhiC31 integrase gene (pgk-phiC31/pCB92) in a 1:3 (HAH1:phiC31) ratio, to facilitate integration into the host cell genome. The following day the cells were seeded into a 15 cm dish, and let grown for the next 5 days. Once they reached ~100% confluence, stable cell line selection was performed by treating the cells with 2 µg/ml puromycin for 10 days, replacing the growth medium every 3-4 days to remove dead cells, until cell clumps (single colonies) were visible. Single colonies were manually picked from the dish and individually seeded 96-well plate wells. During the following days, 36 growing colonies (H1 – H36) were selected by manual inspection and seeded to 12-well plates. 12 lines out of these were selected based on the growth rate, and were expanded to 6-well plates and further to T25 flasks. The HAH1 expression levels were measured by Western Blot analysis, and the cell line (H22) with the highest expression levels was chosen for the NMR experiments.

Cell culture and transient transfection

HEK293T and HAH1 stably-transfected HEK293T cells were maintained in Dulbecco's Modified Eagle's medium (DMEM; high glucose, Gibco) supplemented with L-glutamine (Gibco), antibiotics (penicillin and streptomycin) (Gibco) and 10% FBS (Gibco) in uncoated 75-cm^2 plastic flasks and incubated at 37 °C, 5% CO_2 in a humidified atmosphere. Stably-transfected cells were additionally supplemented with MEM NEAA (Gibco). Transient transfection was performed by DNA:polyethylenimine complex as previously described (13). Cells expressing SOD1 with silenced HAH1 were obtained by transfection with pHLsec-SOD1 and the three silencing vectors (3:1:1:1 ratio); unsilenced cells expressing SOD1 were obtained by transfection with pHLsec-SOD1 and empty pHLsec (1:1 ratio); control cells not expressing SOD1 were transfected as above, by replacing pHLsec-SOD1 with empty pHLsec. The total DNA was kept constant at 25 µg / 75 cm^2 flask in all experiments. [U-^{15}N]-BioExpress 6000 medium (Cambridge Isotope Laboratories) was used for ^{15}N labeling. Cell samples

expressing SOD1 were supplemented with 10 μM $ZnSO_4$ starting at the time of transfection.

Western Blot and protein quantification

Protein content was measured by Western Blot analysis. HAH1 was stained with a rabbit monoclonal anti-HAH1 antibody (Abcam: ab154179, diluted to 53.2 ng/mL). GAPDH was used as reference (rabbit polyclonal anti-GAPDH antibody, Abcam: ab9485, diluted at 1:2,000). Goat anti-rabbit IgG (whole molecule)-peroxidase secondary antibody (Sigma:A0545) was used, diluted at 1:80,000. For detection, LiteAblot EXTEND chemiluminescent substrate (EuroClone) was used. Purified recombinant HAH1 was used as standard for the quantification. HAH1 (GenBank: NP_004036.1) was expressed with a 6-His tag at C-terminal part in *E. coli* BL21(DE3) gold competent cells for 4 h at 37 °C. After cell lysis, a first purification step was performed using a HiTrap chelating (GE Healthcare) column charged with Zn(II). The protein was eluted in 20 mM Na_2HPO_4, 0.5 M NaCl, 5 mM Imidazole, 20 mM EDTA, pH=8, buffer. After buffer exchange and digestion with Factor Xa protease (New England Biolabs) for 24 hr at 25 °C, the protein was separated from the affinity tag in the HiTrap column charged with Zn(II).

Real-time PCR

Total RNA (1 μg) was extracted from cells using TRI Reagent® (Sigma-Aldrich) and reverse-transcribed by the employment of high-capacity cDNA reverse transcription kit (Applied Biosystems), according to the manufacturer's instructions. Quantification of the mRNA level of target genes was performed by Real-Time PCR using TaqMan® gene expression assays with the automated ABI Prism 7700 sequence detector system (Applied Biosystems) as previously described (34). Samples were run in triplicate in Micro-Amp optical 96-well plates (Applied Biosystems) with a TaqMan Universal PCR Master Mix (Applied Biosystems). Simultaneous amplification of the target sequence of HAH1 (Hs01076125_m1, FAM™-MGB dye) together with the housekeeping gene, human ACTB (beta-actin) (Applied Biosystems Endogenous Control VIC®/TAMRA™ Probe) was carried out with the following profile: initial denaturation for 10 minutes at

95 °C, followed by denaturation for 15 seconds at 95 °C and primer annealing and elongation at 60 °C for 1 minute for 40 cycles. The results were analyzed by ABI Prism Sequence Detection System software (version 1.7; Applied Biosystems). The 2^(-ΔΔCT) method was used as a comparative method of quantification (35).

The data are reported as mean ± SEM of 2^(-ΔΔCT) of samples run in triplicate; statistical analysis was performed by One-way ANOVA followed by Bonferroni post-hoc test (*$p<0.05$).

NMR experiments

NMR experiments on cells and lysates were acquired at a 950 MHz Bruker Avance spectrometer equipped with a TCI CryoProbe. 1D ^{1}H (zgesgp Bruker pulse sequence) and 2D ^{1}H-^{15}N SOFAST HMQC spectra (36) were acquired at 308 K. The total acquisition time for each cell sample was about 1 h (64 scans, 128 increments) or 1.5 h (96 scans, 128 increments). The supernatant of each cell sample was checked in the same experimental conditions to exclude protein leakage. The same NMR spectra were also acquired on the cell lysates. All the spectra were processed with Bruker Topspin software. To eliminate the signals arising from unspecific ^{15}N labeling of the cells, the 2D ^{1}H-^{15}N SOFAST HMQC spectra of the cell lysates were further processed by subtracting the spectrum of a lysate from untransfected cells, acquired in the same experimental conditions and identically processed.

Results

Stable HAH1 cell line characterization

HEK293T cell lines stably expressing human HAH1 were obtained from single colonies of cells transfected with the HAH1 and the integrase genes, after antibiotic selection, and the relative expression levels of HAH1 in each cell line were assessed by Western Blot analysis and SDS-PAGE (Fig. 1a). The cell line in which HAH1 reached the expression levels suitable for in-cell NMR experiments was selected (HEK293T-HAH1-22, H22 hereafter, Fig. 1b). The effective concentration of HAH1 in a NMR sample of H22 cells was ~45 μM, as estimated by 1D ^{1}H NMR by comparing the signal

intensity between 0.1 and −0.2 ppm in the cell lysate spectrum with an *in vitro* sample of recombinantly expressed HAH1 at known concentration (Fig. 2a). The amide crosspeaks arising from [U-^{15}N]-labeled HAH1 in a sample of H22 cells grown in [U-^{15}N]-labeled medium were detected in the ^{1}H-^{15}N correlation NMR spectra of both cells and lysate, showing that stably transfected cells can reach constitutive proteins levels sufficient for detection by heteronuclear in-cell NMR (Fig. 2b,c and Fig. S1a,b). The chemical shifts of the detected crosspeaks corresponded to those previously assigned on [U-^{15}N]-apo-HAH1 *in vitro* (Fig. S1c) (26).

Transient transfection of the second protein gives poor labeling selectivity

A sequential expression strategy was tested, in which H22 cells were transiently transfected with the gene of a second protein, and the unlabeled medium was replaced with [U-^{15}N]-labeled medium at the time of transfection. SOD1 was chosen as a test case for globular proteins that are easily detected by in-cell NMR (5, 11, 13); thioredoxin 1 (TRX1) was chosen to represent globular proteins that cannot be detected by in-cell solution NMR without recurring to deuterium enrichment, due to extensive weak interactions with the cellular environment (20, 24, 37). In either case, transient transfection did not decrease the levels of stably expressed HAH1, and the overexpressed protein reached the same levels as previously observed in the parental HEK cells (13, 24). However, signals arising from ^{15}N-labeled HAH1 were still being detected in the resulting in-cell NMR spectra (Fig. 3a,b and Fig. S2a,c) and in the corresponding lysates (Fig. 3c,d and Fig. S2b,d), due to HAH1 being still synthesized in the [U-^{15}N]-labeled medium, together with the second protein.

Silencing of HAH1 reduces the rate of protein synthesis

To avoid the unwanted labeling of HAH1 during the expression of the second protein, the rate of HAH1 synthesis needs to be reduced after transfection. Gene silencing by RNA interference (RNAi) is a widely used technique to knock-down gene expression in mammalian cells. With that purpose, a silencing vector was used to transcribe sequences of short hairpin RNA (shRNA) complementary to different regions of the HAH1 mRNA. Three different sequences were individually cloned in the silencing

vector (Fig. 4a). The silencing vector was co-transfected with the gene of the second protein. The relative levels of HAH1 mRNA and protein were measured at different times after transfection by RT-PCR and Western Blot, respectively. Efficient silencing was achieved by mixing together the three shRNA-encoding sequences (Fig. 4b). When comparing the levels of mRNA with those of HAH1 protein as a function of time after transfection, it was observed that the protein levels decreased much slower than those of mRNA, indicating that the protein turnover rate is low enough to permit the expression of the second protein in the time-frame following transfection, without compromising the level of HAH1 (Fig. 4c and Fig. S3).

Silencing combined with timed medium switch improves the labeling selectivity

A temporal window was identified, during which the levels of HAH1 mRNA had reached a minimum, while the HAH1 protein levels did not decrease excessively (Fig. 4c). Specific labeling of the second protein could be performed during that window with minimal unwanted labeling of HAH1. To test this procedure, H22 cells were transfected with either SOD1 or the empty vector together with the vectors encoding the shRNAs against HAH1, and were kept in unlabeled medium for 14 h followed by 16 h in [U-^{15}N]-medium (Fig. 5 and Fig. S4). Corresponding control samples were produced, in which HAH1 was not silenced. The resulting in-cell NMR spectra confirmed our expectations: HAH1 signals were almost not detected in the ^{1}H-^{15}N correlation spectra acquired on both cells (Fig. 5b,d) and lysates (Fig. 5f,h), while the ^{1}H spectra revealed the presence of unlabeled HAH1 (~80% of the levels of unsilenced HAH1, Fig. 5i-l). The relative amounts of total and labeled HAH1 in each lysate were estimated from the ratio between the aliphatic ^{1}H signal intensities (arising from the total HAH1, Fig. 5i-l) and the amide ^{1}H-^{15}N crosspeak intensities (arising from ^{15}N-labeled HAH1, Fig. 5e-h), normalized by the same ratio obtained from a sample of purified ^{15}N-labeled HAH1. In the "fully" ^{15}N-labeled H22 cell sample (Fig. 2c), ~93% of the total HAH1 was labeled, close to the nominal isotopic enrichment of the growth medium (U-^{15}N, 98%). The "14+16" labeling strategy alone, without HAH1 silencing, caused ~32% of the total HAH1 to be ^{15}N-labeled, while the silencing further increased the labeling selectivity as only ~15% of the total HAH1 was ^{15}N-labeled, and was barely detected by ^{1}H-^{15}N correlation NMR.

Discussion

HEK293T cells stably expressing HAH1 were co-transfected with two vectors: one encoding SOD1, the other containing a shRNA which knocked-down the expression of HAH1 by degrading its mRNA via RNAi. ^{15}N-labeling of SOD1 was achieved by replacing the unlabeled growth medium with ^{15}N-labeled medium at a proper time after transfection. The optimal timing for medium switch and sample collection after SOD1 expression was chosen by measuring HAH1 protein and mRNA levels at different time points following cell transfection. Since HAH1 has a sufficiently long intracellular half-life, a time window was identified (between 14 h and 30 h following transient transfection) during which the shRNA had degraded the HAH1 mRNA, so that the rate of HAH1 synthesis was greatly decreased, while its cellular protein content was not yet decreased below the concentration required for in-cell NMR (Fig. 4c). During this time window, SOD1 was transiently overexpressed and selectively ^{15}N-labeled with respect to HAH1.

The strategy described above combines stable and transient protein expression to allow two (and in principle more) proteins to be overexpressed in human cells at different times. Without any protein expression control mechanism, the stably expressed protein would still be produced by the cells after the expression of the second protein has started, resulting in poor labeling selectivity, which is critical for proper interpretation of the NMR spectra. Our strategy exploits the RNAi machinery of the cells to control the rate of protein expression as a function of time (31, 32). For the same purpose, other methods may be used, such as the employment of synthetic small interference RNA (siRNA) molecules instead of shRNAs (38), however these approaches would require the protocol to be adapted to combine DNA transfection with siRNA treatment. Alternatively, different strategies could be pursued to control protein expression. Inducible gene expression systems, for example the tetracycline-controlled Tet-ON/Tet-OFF systems, are commonly used for that purpose (39–41). Stable cell lines could then be created that can overexpress the proteins of interest separately, by combining alternative inducible promoters, such as the ecdysone-controlled system (42), to the tetracycline-controlled one. However, such strategies would require additional vector designing, and may not reach the expression levels required by the relatively low sensitivity of NMR spectroscopy.

This work shows that controlled protein expression is a useful approach for studying protein interactions in human cells by NMR. Although our test proteins were not supposed to interact directly, they were useful to assess the labeling specificity and to demonstrate the feasibility of our strategy. This method represents an innovative approach for studying *bona fide* interactions between proteins at the atomic level, and in principle it can be adapted to different labeling schemes or combined with the existing protein-delivery approaches. Hopefully, this work will provide a starting point for developing advanced methods that allow finer control of protein synthesis in mammalian cells for in-cell NMR applications.

Acknowledgments

We would like to thank Prof. Lucia Banci (Univ. of Florence) for critically reading the manuscript and providing advice; Prof. A. Radu Aricescu and Dr. Yuguang Zhao (Univ. of Oxford) for providing support for the stable cell line selection. This work has been supported by "MEDINTECH: Tecnologie convergenti per aumentare la sicurezza e l'efficacia di farmaci e vaccini" and by Instruct, part of the European Strategy Forum on Research Infrastructures (ESFRI) and supported by national member subscriptions. Specifically, we thank the EU ESFRI Instruct Core Centres CERM-Italy. The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under BioStruct-X (grant agreement N°283570).

References

1. Freedberg, D. I. & Selenko, P. Live cell NMR. *Annu. Rev. Biophys.* **43,** 171–192 (2014).

2. Hänsel, R., Luh, L. M., Corbeski, I., Trantirek, L. & Dötsch, V. In-cell NMR and EPR spectroscopy of biomacromolecules. *Angew. Chem. Int. Ed Engl.* **53,** 10300–10314 (2014).

3. Smith, A. E., Zhang, Z., Pielak, G. J. & Li, C. NMR studies of protein folding and binding in cells and cell-like environments. *Curr. Opin. Struct. Biol.* **30,** 7–16 (2015).

4. Schlesinger, A. P., Wang, Y., Tadeo, X., Millet, O. & Pielak, G. J. Macromolecular crowding fails to fold a globular protein in cells. *J. Am. Chem. Soc.* **133,** 8082–8085 (2011).

5. Luchinat, E. *et al.* In-cell NMR reveals potential precursor of toxic species from SOD1 fALS mutants. *Nat. Commun.* **5,** 5502 (2014).

6. Monteith, W. B., Cohen, R. D., Smith, A. E., Guzman-Cisneros, E. & Pielak, G. J. Quinary structure modulates protein stability in cells. *Proc. Natl. Acad. Sci. U. S. A.* **112,** 1739–1742 (2015).

7. Burz, D. S., Dutta, K., Cowburn, D. & Shekhtman, A. Mapping structural interactions using in-cell NMR spectroscopy (STINT-NMR). *Nat. Methods* **3,** 91–93 (2006).

8. Burz, D. S. & Shekhtman, A. In-cell biochemistry using NMR spectroscopy. *PloS One* **3,** e2571 (2008).

9. Augustus, A. M., Reardon, P. N. & Spicer, L. D. MetJ repressor interactions with DNA probed by in-cell NMR. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 5065–5069 (2009).

10. Luh, L. M. *et al.* Molecular crowding drives active Pin1 into nonspecific complexes with endogenous proteins prior to substrate recognition. *J. Am. Chem. Soc.* **135,** 13796–13803 (2013).

11. Banci, L., Barbieri, L., Bertini, I., Cantini, F. & Luchinat, E. In-cell NMR in E. coli to monitor maturation steps of hSOD1. *PloS One* **6,** e23561 (2011).

12. Hembram, D. S. S. *et al.* An in-cell NMR study of monitoring stress-induced increase of cytosolic Ca2+ concentration in HeLa cells. *Biochem. Biophys. Res. Commun.* **438,** 653–659 (2013).

13. Banci, L. *et al.* Atomic-resolution monitoring of protein maturation in live human cells by NMR. *Nat. Chem. Biol.* **9,** 297–299 (2013).

14. Xie, J., Thapa, R., Reverdatto, S., Burz, D. S. & Shekhtman, A. Screening of small molecule interactor library by using in-cell NMR spectroscopy (SMILI-NMR). *J. Med. Chem.* **52,** 3516–3522 (2009).

15. Thongwichian, R. & Selenko, P. In-cell NMR in Xenopus laevis oocytes. *Methods Mol. Biol. Clifton NJ* **895,** 33–41 (2012).

16. Inomata, K. *et al.* High-resolution multi-dimensional NMR spectroscopy of proteins in human cells. *Nature* **458,** 106–109 (2009).

17. Ogino, S. *et al.* Observation of NMR signals from proteins introduced into living mammalian cells by reversible membrane permeabilization using a pore-forming toxin, streptolysin O. *J. Am. Chem. Soc.* **131,** 10834–10835 (2009).

18. Kubo, S. *et al.* A Gel-Encapsulated Bioreactor System for NMR Studies of Protein-Protein Interactions in Living Mammalian Cells. *Angew. Chem. Int. Ed Engl.* (2012). doi:10.1002/anie.201207243

19. Bekei, B. In-cell NMR Spectroscopy in Mammalian Cells. (Freie Universität Berlin, Germany, 2013). at <http://www.diss.fu-berlin.de/diss/receive/FUDISS_thesis_000000095311?lang=en>

20. Majumder, S. *et al.* Probing protein quinary interactions by in-cell nuclear magnetic resonance spectroscopy. *Biochemistry (Mosc.)* **54,** 2727–2738 (2015).

21. Bertrand, K., Reverdatto, S., Burz, D. S., Zitomer, R. & Shekhtman, A. Structure of proteins in eukaryotic compartments. *J. Am. Chem. Soc.* **134,** 12798–12806 (2012).

22. Hamatsu, J. *et al.* High-resolution heteronuclear multidimensional NMR of proteins in living insect cells using a baculovirus protein expression system. *J. Am. Chem. Soc.* **135,** 1688–1691 (2013).

23. Luchinat, E., Gianoncelli, A., Mello, T., Galli, A. & Banci, L. Combining in-cell NMR and X-ray fluorescence microscopy to reveal the intracellular maturation states of human superoxide dismutase 1. *Chem. Commun. Camb. Engl.* **51,** 584–587 (2015).

24. Banci, L., Barbieri, L., Luchinat, E. & Secci, E. Visualization of redox-controlled protein fold in living cells. *Chem. Biol.* **20,** 747–752 (2013).

25. Barbieri, L., Luchinat, E. & Banci, L. Structural insights of proteins in sub-cellular compartments: In-mitochondria NMR. *Biochim. Biophys. Acta* **1843,** 2492–2496 (2014).

26. Anastassopoulou, I. *et al.* Solution structure of the apo and copper(I)-loaded human metallochaperone HAH1. *Biochemistry (Mosc.)* **43,** 13046–13053 (2004).

27. Arnesano, F. *et al.* Probing the interaction of cisplatin with the human copper chaperone Atox1 by solution and in-cell NMR spectroscopy. *J. Am. Chem. Soc.* **133,** 18361–18369 (2011).

28. Banci, L., Bertini, I., Cantini, F., D'Amelio, N. & Gaggelli, E. Human SOD1 before harboring the catalytic metal: solution structure of copper-depleted, disulfide-reduced form. *J. Biol. Chem.* **281,** 2333–2337 (2006).

29. Banci, L. *et al.* Human superoxide dismutase 1 (hSOD1) maturation through interaction with human copper chaperone for SOD1 (hCCS). *Proc. Natl. Acad. Sci. U. S. A.* **109,** 13555–13560 (2012).

30. Zhao, Y., Malinauskas, T., Harlos, K. & Jones, E. Y. Structural insights into the inhibition of Wnt signaling by cancer antigen 5T4/Wnt-activated inhibitory factor 1. *Struct. Lond. Engl. 1993* **22,** 612–620 (2014).

31. Paddison, P. J., Caudy, A. A. & Hannon, G. J. Stable suppression of gene expression by RNAi in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 1443–1448 (2002).

32. Sui, G. *et al.* A DNA vector-based RNAi technology to suppress gene expression in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 5515–5520 (2002).

33. Seiradake, E., Zhao, Y., Lu, W., Aricescu, A. R. & Jones, E. Y. Production of cell surface and secreted glycoproteins in mammalian cells. *Methods Mol. Biol. Clifton NJ* **1261,** 115–127 (2015).

34. Donati, C. *et al.* Sphingosine 1-phosphate mediates proliferation and survival of mesoangioblasts. *Stem Cells Dayt. Ohio* **25,** 1713–1719 (2007).

35. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods San Diego Calif* **25,** 402–408 (2001).

36. Schanda, P. & Brutscher, B. Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds. *J. Am. Chem. Soc.* **127,** 8014–8015 (2005).

37. Reckel, S., Lopez, J. J., Löhr, F., Glaubitz, C. & Dötsch, V. In-cell solid-state NMR as a tool to study proteins in large complexes. *Chembiochem Eur. J. Chem. Biol.* **13,** 534–537 (2012).

38. Amarzguioui, M., Rossi, J. J. & Kim, D. Approaches for chemically synthesized siRNA and vector-mediated RNAi. *FEBS Lett.* **579,** 5974–5981 (2005).

39. Gossen, M. & Bujard, H. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc. Natl. Acad. Sci. U. S. A.* **89,** 5547–5551 (1992).

40. Freundlieb, S., Schirra-Müller, C. & Bujard, H. A tetracycline controlled activation/repression system with increased potential for gene transfer into mammalian cells. *J. Gene Med.* **1,** 4–12 (1999).

41. Nishijima, H., Yasunari, T., Nakayama, T., Adachi, N. & Shibahara, K. Improved applications of the tetracycline-regulated gene depletion system. *Biosci. Trends* **3,** 161–167 (2009).

42. No, D., Yao, T. P. & Evans, R. M. Ecdysone-inducible gene expression in mammalian cells and transgenic mice. *Proc. Natl. Acad. Sci. U. S. A.* **93,** 3346–3351 (1996).

Figures

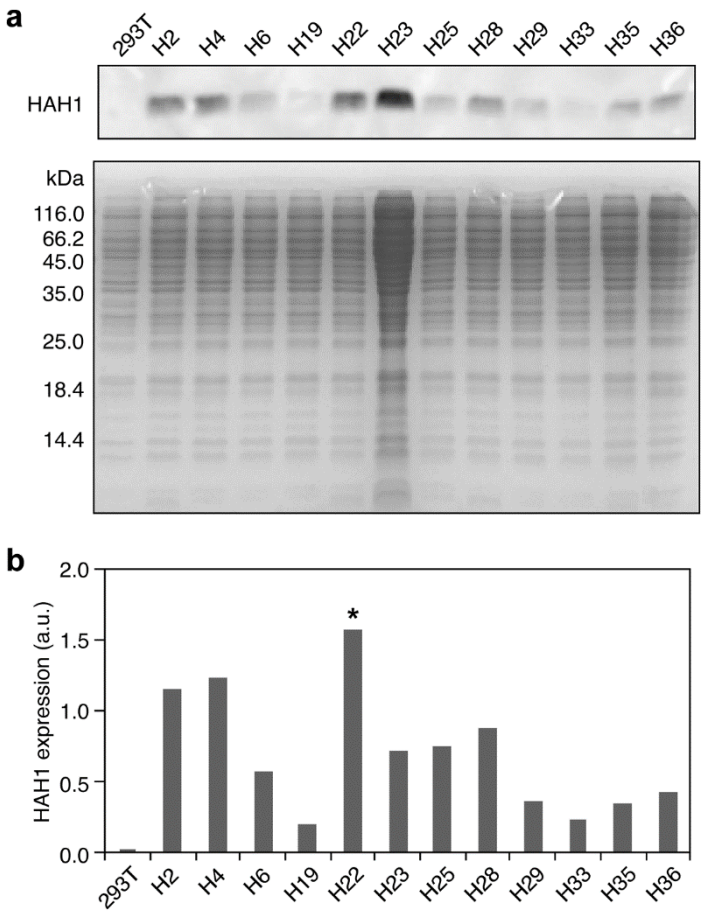


Fig. 1. Stable cell line characterization. (a) Western Blot and SDS-PAGE analysis of 12 HEK293T-HAH1 cell lines stably expressing HAH1 selected from 36 single colonies (H1 – H36) obtained after antibiotic selection; the parental cell line (293T) is shown as comparison. (b) Relative HAH1expression levels obtained by normalizing the Western Blot intensity by the SDS-PAGE lane intensity; the cell line with the highest expression was chosen (H22, marked with an asterisk).

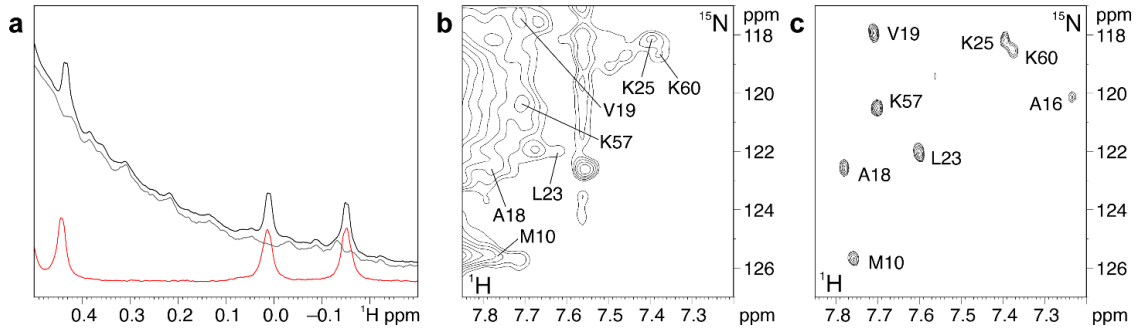


Fig. 2. HAH1 stably expressed in human cells can be observed by in-cell NMR. (a) Aliphatic region of the 1D ^{1}H NMR spectrum of lysed H22 cells (black), showing signals arising from stably expressed HAH1. The spectrum of lysed HEK293T cells (gray) and the spectrum of purified HAH1 (red) are shown for reference. (b) Selected region of the ^{1}H-^{15}N SOFAST-HMQC spectrum recorded on a sample of [^{15}N]-labeled H22 cells; (c) Same region as (b) of the ^{1}H-^{15}N SOFAST-HMQC spectrum recorded on the corresponding cell lysate. The amide crosspeaks arising from HAH1 are labeled according to the resonance assignment. The spectrum in (c) was further processed to subtract the cellular background. See also Fig. S1.

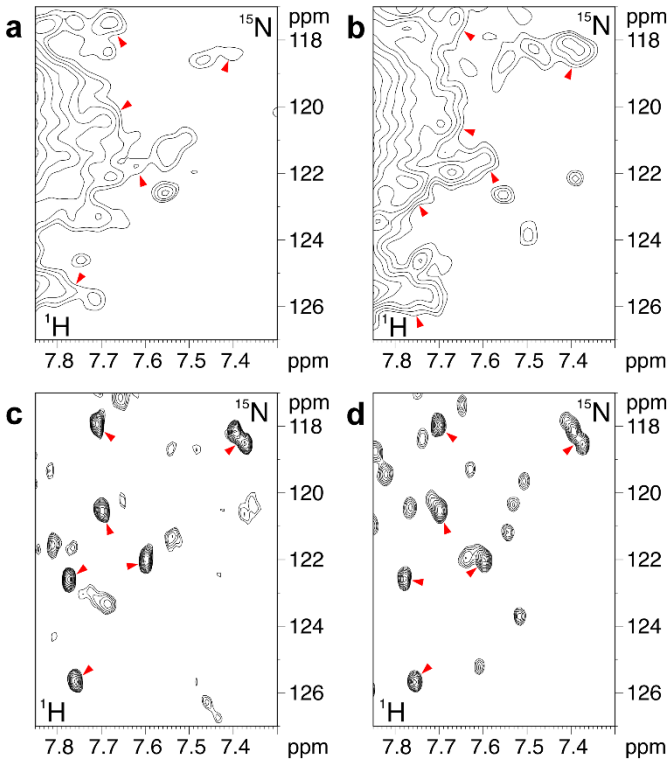


Fig. 3. Transient expression and labeling of the second protein causes partial labeling of the stably expressed HAH1. Selected region of ^{1}H-^{15}N SOFAST-HMQC spectra of H22 cells (a,b) and the corresponding lysates (c,d) transiently transfected either with SOD1 (a,c) or with TRX1 (b,d). The cells were collected either 24 h (a,c) or 48 h (b,d) following transfection. The amide crosspeaks arising from [^{15}N]-labeled HAH1 are indicated by red arrowheads. See also Fig. S2.

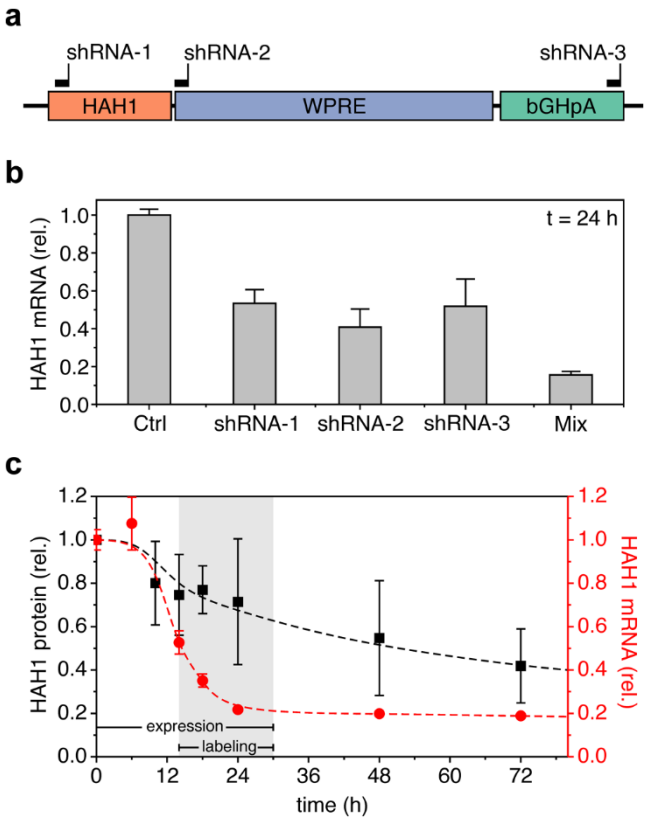


Fig. 4. The expression of HAH1 can be efficiently silenced with shRNA-encoding vectors. (a) Schematic representation of the HAH1 gene integrated in the H22 genome; from left to right: open reading frame from HAH1 cDNA (salmon), woodchuck hepatitis virus post-transcriptional regulatory element (WPRE, light blue), bovine growth hormone polyadenylation signal (bGHpA, light green). The recognition sites of each shRNA sequence are shown (black bars). (b) Silencing efficiency of each shRNA-encoding vector, assessed from the relative levels of HAH1 mRNA 24 h after transfection. The combination of shRNA vectors (Mix) gives the highest transfection efficiency (error bars: SEM) (c) Relative levels of HAH1 mRNA (red circles) and HAH1 protein (black squares) as a function of time after transfection, measured by Real Time RT-PCR (error bars: SEM) and Western Blot (error bars: SD, n = 3), respectively. Trends are qualitatively shown as dashed lines. The timing chosen for expression and labeling of the second protein is shown (expression in the white + gray area; labeling in the gray area). See also Fig. S3.

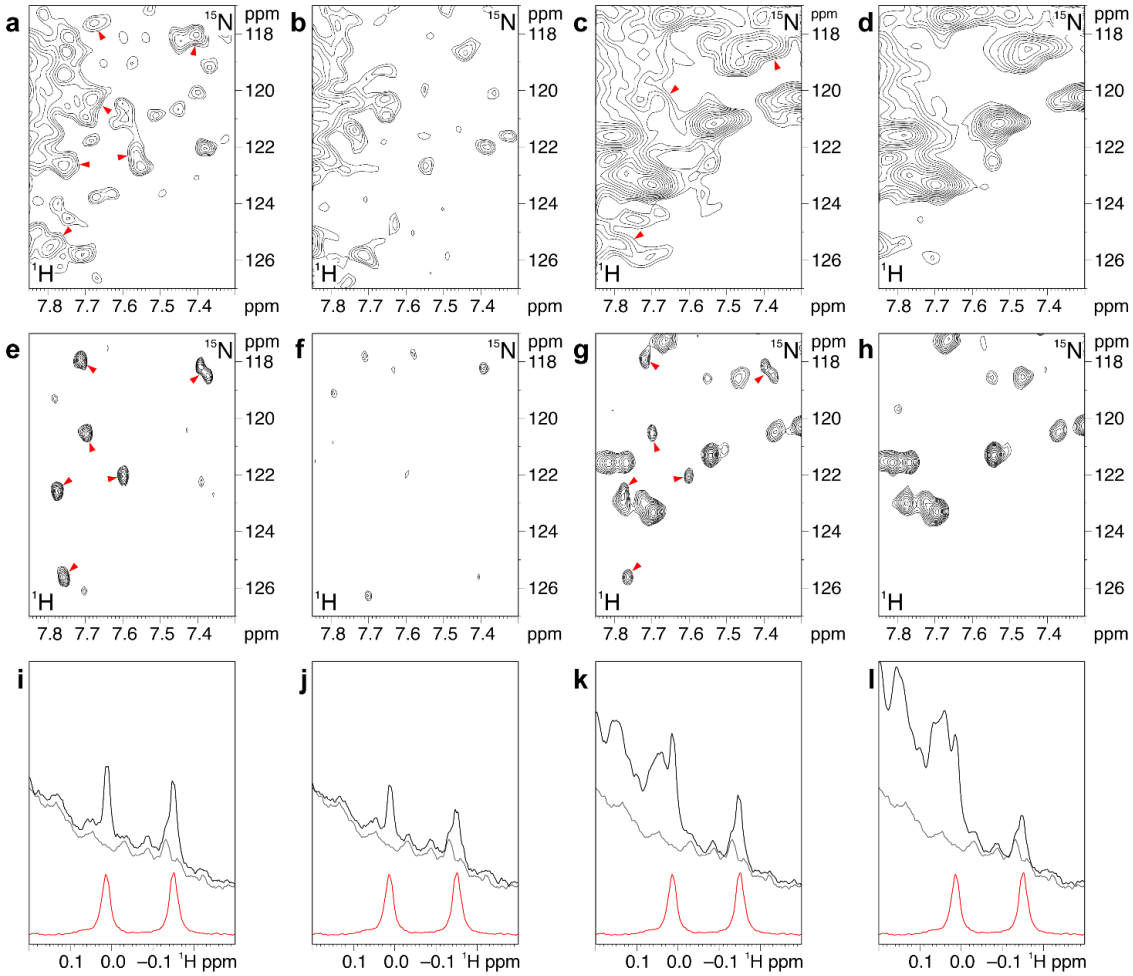
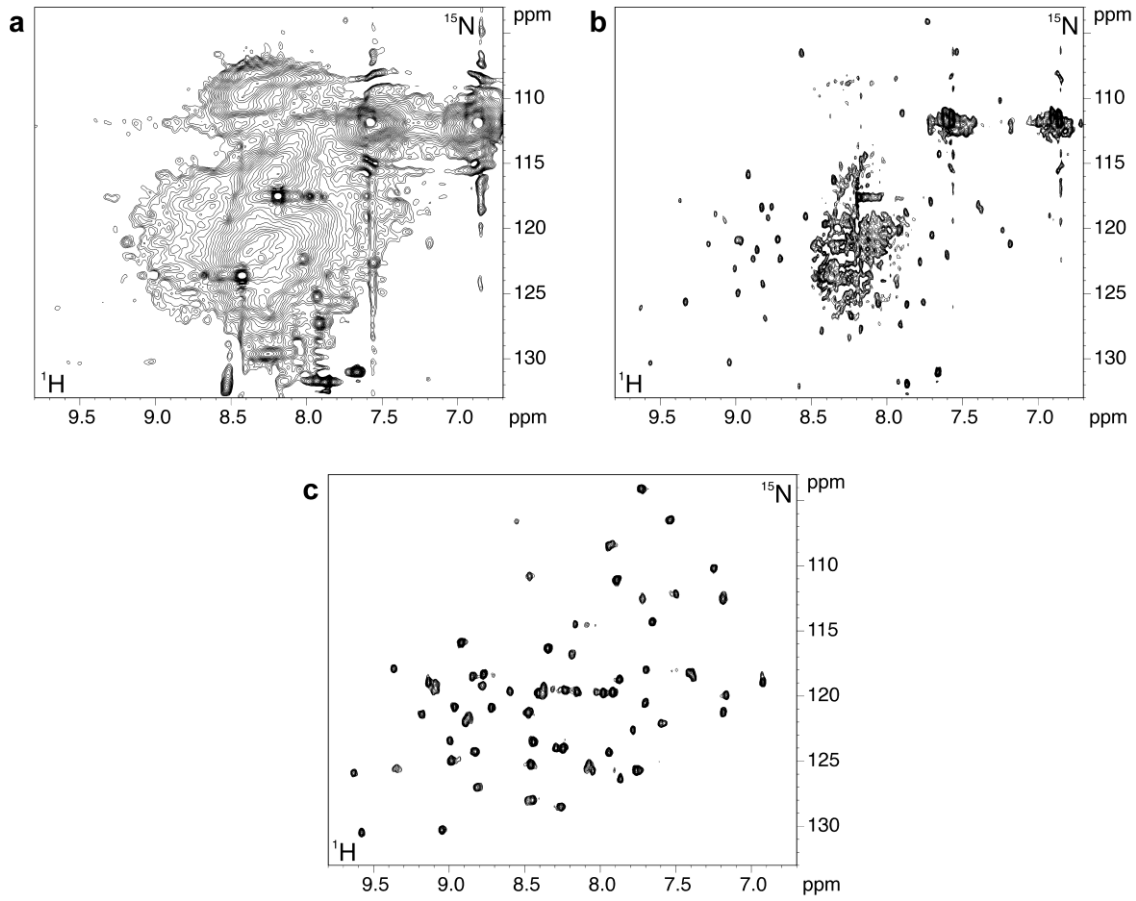


Fig. 5. HAH1 silencing combined with the appropriate labeling strategy improves the labeling selectivity. (a-d) Selected region of ^{1}H-^{15}N SOFAST-HMQC spectra of four samples of H22 cells. Cells were transiently transfected with empty vector (a), shRNA mix (b), SOD1 (c) or shRNA mix + SOD1 (d). The [^{15}N]-medium was provided after 14 h, and the cells were collected 30 h following transfection. (e-h) Selected region of ^{1}H-^{15}N SOFAST-HMQC spectra of the corresponding lysates. The amide crosspeaks arising from [^{15}N]-labeled HAH1 are indicated by red arrowheads. (i-l) Aliphatic region of the 1D ^{1}H spectra of the same samples (black); HEK293T lysate (gray) and purified HAH1 (red) are shown for reference. See also Fig. S4.

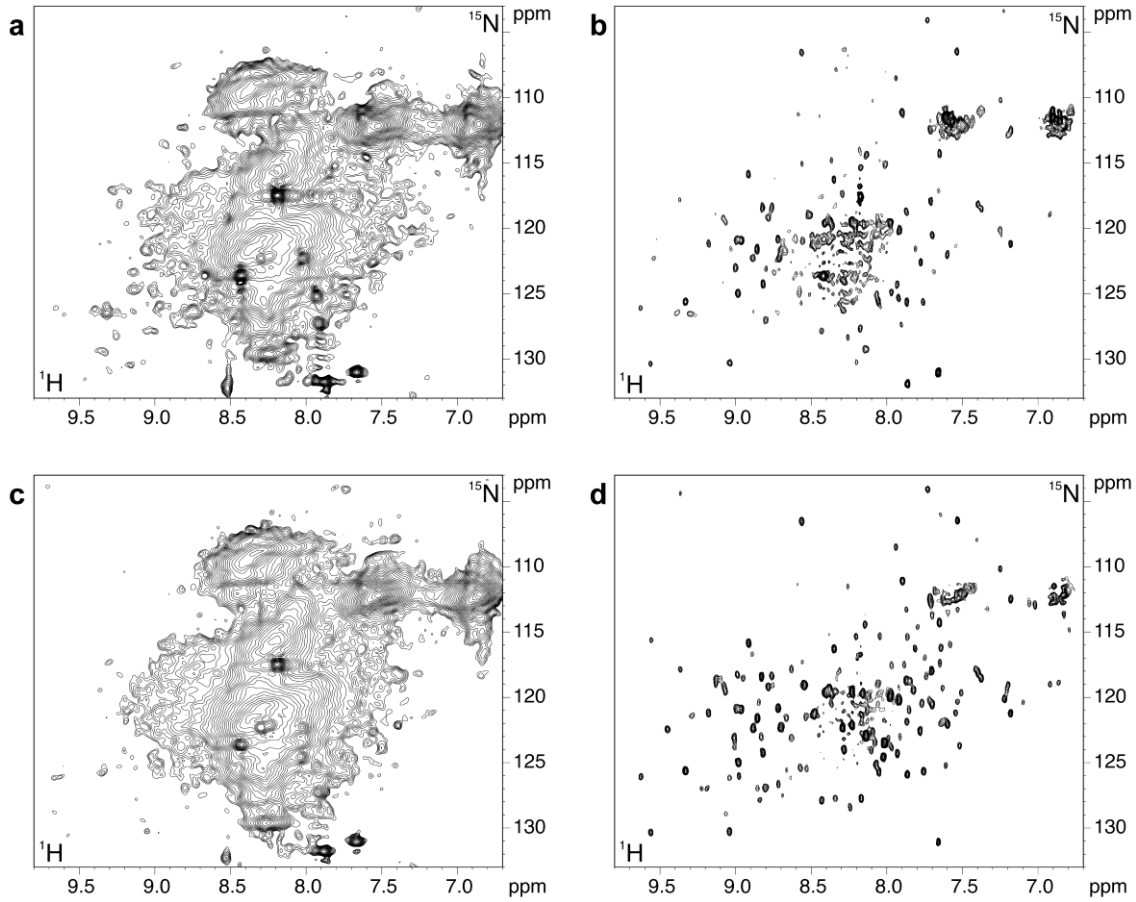
Supporting Information:

Sequential protein expression and selective labeling for in-cell NMR in human cells

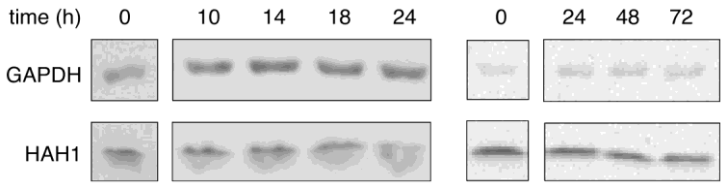
Supplementary Figures



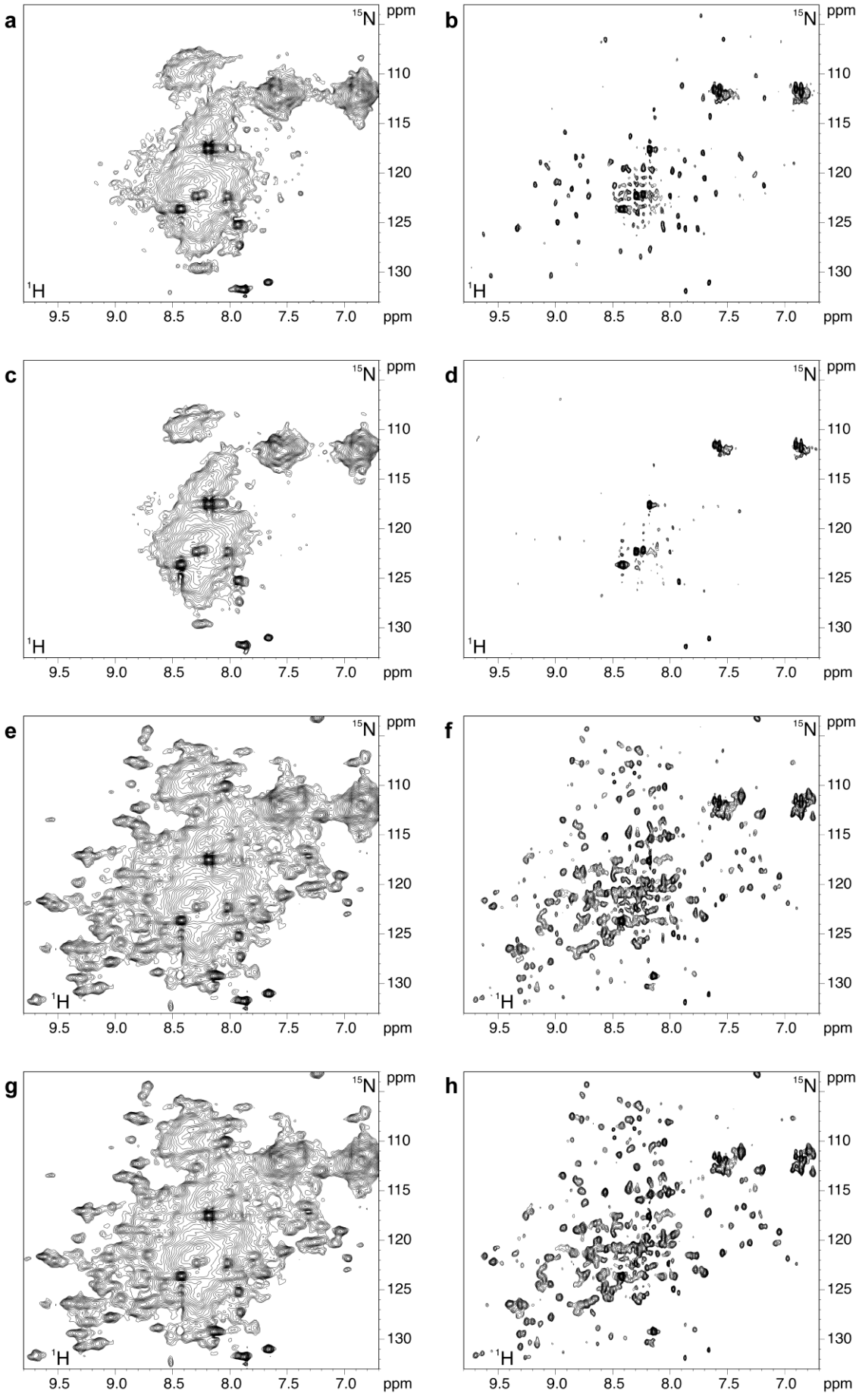
Supplementary Figure S1. (a) ^{1}H-^{15}N SOFAST-HMQC spectrum recorded on a sample of [^{15}N]-labeled H22 cells; (b) ^{1}H-^{15}N SOFAST-HMQC spectrum recorded on the corresponding cell lysate. The spectrum in (b) was further processed to subtract the cellular background; (c) ^{1}H-^{15}N SOFAST-HMQC spectrum recorded on purified [U-^{15}N]- HAH1.



Supplementary Figure S2. (a,b) ^{1}H-^{15}N SOFAST-HMQC spectra of H22 cells (a) and the corresponding lysate (b) transiently transfected with SOD1 and collected 24 h following transfection; (c,d) ^{1}H-^{15}N SOFAST-HMQC spectra of H22 cells (c) and the corresponding lysate (d) transiently transfected with TRX1 and collected either 48 h following transfection.



Supplementary Figure S3. Levels of HAH1 protein measured by Western Blot at different times after transfection of the shRNA vectors. GAPDH was used as reference.


a
b
c
d
e
f
g
h
^{15}N
ppm
110
115
120
125
130
^{1}H
9.5
9.0
8.5
8.0
7.5
7.0
ppm

Supplementary Figure S4. Selected region of ^{1}H-^{15}N SOFAST-HMQC spectra of four samples of H22 cells and the corresponding lysates transiently transfected with empty vector (a,b), shRNA mix (c,d), SOD1 (e,f) or shRNA mix + SOD1 (g,h). The [^{15}N]-medium was provided after 14 h, and the cells were collected 30 h following transfection.

2.3 The casein kinase 2-dependent phosphorylation of NS5A domain 3 from hepatitis C virus followed by time-resolved NMR

In order to identify CK2-dependent phosphorylation sites in the sequence of NS5A domain 3 (from the HCV-1b-J4 genotype), the reaction between the protein and CK2 was monitored *in vitro* by acquiring a series of ^{1}H-^{15}N SOFAST-HMQC spectra during the course of the reaction, which was carried on for 24 h at 303 K. The crosspeaks arising from the phosphorylated species were assigned *de novo* from sets of 3D triple resonance NMR experiments acquired on U-^{13}C,^{15}N-labeled NS5A-D3 samples obtained by inactivating the kinase at different times. Upon reaction with CK2, four serine residues were unambiguously identified as phosphorylation sites (Ser408, Ser401, Ser429 and Ser434). Ser408 was phosphorylated with a higher rate compared to the other residues.

Phosphorylation of S429 and S434 gave rise to two amide crosspeaks for each residue, with different buildup curves. As the reaction proceeded, the singly phosphorylated species were depleted, as reflected by the decreasing intensity of the corresponding crosspeaks. The curves of the intermediate and final species of S434 are compatible with a consecutive reaction model. This suggests that other reactions may occur at dependence of S434 phosphorylation. A suitable model was not found to explain the formation of the intermediate and final species of S429.

Ser434 is not a canonical phosphorylation site of CK2 because there is not an acidic amino acid (E/D) in position +1, +2 or +3 but becomes a secondary phosphorylation site if a serine in position +3 is phosphorylated. Interestingly, both S434 and S437 are highly conserved among the HCV genotypes and their phosphorylation has been found to be critical for the interaction between NS5A and the core protein. Although phosphorylation of S437 was not observed in our experimental conditions, a decrease in the crosspeak intensity of the residues 435-437 was observed, although no new crosspeaks could be identified. This behaviour suggests that, during the reaction with CK2, either one species is formed that undergoes conformational exchange phenomena or multiple species are formed with different phosphorylation patterns.

These results provide a starting point to elucidate the role of phosphorylation in the modulation of the viral activity, and in the mechanisms of viral assembly, at the molecular level.

The casein kinase 2-dependent phosphorylation of NS5A domain 3 from hepatitis C virus followed by time-resolved NMR

Erica Secci[a], Enrico Luchinat[a,b], Lucia Banci[a,c,*]

Affiliations:

[a]Magnetic Resonance Center - CERM, University of Florence, Via Luigi Sacconi 6, 50019 Sesto Fiorentino, Florence, Italy.

[b]Department of Biomedical, Experimental and Clinical Sciences, "Mario Serio", University of Florence, Viale Morgagni 50, 50134 Florence, Italy.

[c]Department of Chemistry, University of Florence, Via della Lastruccia 3, 50019, Sesto Fiorentino, Florence, Italy.

* To whom correspondence should be addressed L.B. (banci@cerm.unifi.it).

SUBMITTED

Background: The Hepatitis C virus (HCV) protein NS5A is involved in viral replication and assembly.

Results: CK2-dependent phosphorylation of NS5A Domain 3 is monitored *in vitro* by time-resolved NMR.

Conclusions: We unambiguously identified four serines, which are phosphorylated with different rates.

Significance: Phosphorylation of NS5A plays a critical role in the viral assembly efficiency and infectivity of HCV.

Keywords: nuclear magnetic resonance (NMR); phosphorylation; Hepatitis C virus (HCV); viral protein; virus assembly; enzyme kinetics; time-resolved NMR; intrinsically disordered protein (IDP); post-translational modification (PTM).

Abstract

Hepatitis C virus (HCV) chronically affects millions of individuals worldwide, and an effective vaccine is not available to date. The HCV non-structural protein 5A (NS5A) is involved in the formation of the viral RNA replication complex, and plays a critical role in the viral assembly pathway. Domain 3 (D3) of NS5A is an unstructured polypeptide responsible for the interaction between the replication complex and the core particle assembly structure. Casein kinase 2 (CK2) is known to phosphorylate NS5A-D3 at multiple sites, which have been largely predicted and only indirectly observed. In order to identify the CK2-dependent phosphorylation sites, we monitored *in vitro* the reaction between NS5A-D3 and CK2 by time-resolved NMR. By performing *de novo* resonance assignments on the phosphorylated protein, we could identify unambiguously four serine residues as substrates of CK2. The time-dependent phosphorylation curves obtained for each site and for the neighboring residues allowed to determine the apparent rate constants. One serine (S408) was quickly phosphorylated, while the other three (S401, S429, S434) reacted noticeably slower. Additional reactions might occur close to the protein C-terminus, which caused a decrease in the NMR signal intensities. These results provide a starting point to elucidate the role of phosphorylation in the

mechanisms of viral assembly, and in the modulation of the viral activity, at the molecular level.

Introduction

Worldwide, an estimated 130–150 million individuals are infected by the Hepatitis C virus (HCV), a member of the *Flaviviridae* family[1]. In 85% of the cases, the virus is able to establish chronic infections, leading to liver cirrhosis, hepatocellular carcinoma and liver failure (1,2). Six closely related viral genotypes, classified by sequence homology, comprise the HCV genus (3). Genotype 1 is the most abundant worldwide, and includes subtypes 1a and 1b (4). There is still no effective vaccine available, and the antiviral treatments currently applied show some limitation (5,6). Therefore, a better understanding of the molecular mechanisms involved in the HCV life cycle is required to develop new antiviral therapies. For this reason, the biology of HCV remains a topical research subject (7).

HCV is an enveloped virus containing a positive-sense, single-stranded RNA genome of about 9.6 kb (8). The RNA genome of HCV contains a single large open reading frame (ORF), which encodes a polyprotein approximately 3,000 amino acids long. This precursor polyprotein undergoes a complex series of co- and post-translational proteolytic events, catalyzed by both host and viral proteases, to produce 10 mature viral proteins: the virion structural proteins E1, E2 (envelope glycoproteins) and the core protein; the viroporin p7 involved in virion assembly/release and the nonstructural (NS) proteins NS2, NS3, NS4A, NS4B, NS5A, NS5B, which are involved in both genome replication and virion assembly (9,10).

NS5A is a zinc-binding phosphoprotein lacking enzymatic activity, which binds the viral RNA and simultaneously interacts with other HCV proteins such as NS5B and with host cell proteins such as the human vesicle-associated membrane protein-associated protein A (VAPA), thus contributing to the replication complex formation (11). In addition, NS5A has been shown to be involved in a variety of cellular regulation processes, such as interferon resistance and apoptotic regulation, although the precise role of these functions in the HCV life cycle has yet to be defined (12,13).

[1]http://www.who.int/mediacentre/factsheets/fs164/en/

NS5A consists of three domains separated by two low complexity sequences (LCS1 and LCS2) (14) and an N-terminal amphipathic helix essential for its membrane association (15). Domain 1 (D1) comprises a zinc-binding motif, which is essential for RNA replication (16) and forms a homodimer with a claw-like shape that can accommodate a single strand RNA molecule (17,18). It is also required for the interaction with lipid droplets, the organelles where the assembly of infectious particles starts (19). Domains 2 and 3 (D2, D3) have been shown to be both intrinsically disordered with a propensity to form secondary structure elements (20–22). D2 is essential for RNA replication, while D3 is thought to be dispensable for that role (23,24). It has been shown, however, that D3 is critical in the virus assembly process, as it is required for the interaction between NS5A and the core protein (24,25).

NS5A exists in two forms with slightly different mobility on an SDS-PAGE gel: a basally phosphorylated form (apparent molecular mass of 56 kDa) and a hyperphosphorylated form (58 kDa) (26,27). Pulse-chase and mutational analysis suggested that both phosphorylated forms are produced after completion of the proteolytic cleavages at the N- and C-termini of NS5A, and that p58 is the product of p56 hyperphosphorylation (26,27). The degree of phosphorylation does not affect NS5A localization to the ER membrane (28). Hyperphosphorylation sites have been identified in a serine-rich region in LCS1 and in D2, and have been shown to act as negative regulators of the RNA replication process (27–30). The cellular kinases that are mainly involved in the generation of the p58 form are the α isoform of casein kinase 1 (31) and polo-like kinase 1 (32). The basally phosphorylated sites are mainly serine residues located in both D2 and D3 (27–30). The basal phosphorylation of D3 plays a critical role in the virion assembly, as deletions of part of D3 and/or serine to alanine mutations have been shown to impair the production of infectious virions, likely by disrupting the interaction with core protein (25,33). It has been shown that casein kinase 2 (CK2) is responsible for the basal phosphorylation of D3 in multiple sites, and several CK2 phosphorylation sites have been identified (34,35). However, with few exceptions, the positions of such phosphorylation sites, obtained with combined approaches such as mass spectrometry and predictive analysis (30,33,36), have not been unambiguously determined.

In this work, we applied time-resolved NMR spectroscopy to characterize *in vitro* the CK2 phosphorylation pattern of HCV-1b NS5A-D3. High-resolution NMR is an ideal

technique to characterize chemical modifications of proteins at the residue level, and has been shown to be particularly suited to follow protein phosphorylation events (37–40). The amide chemical shifts of an intrinsically disordered protein are extremely sensitive to post-translational modifications, such as phosphorylation, that change the chemical environment in proximity of the substrate residue (41,42). Moreover, fast NMR experiments acquired with sensitive equipment allow protein phosphorylation events to be monitored in real time, also in intact cells and extracts (43–45). By time-resolved *in vitro* NMR, we unambiguously identified four serine phosphorylation sites in NS5A-D3, which react with CK2 at different rates. Interestingly, some sites do not have a 'canonical' CK2 recognition motif, and had not been predicted previously. Finally, our data suggest that additional phosphorylation reactions occur close to the C-terminus of NS5A-D3, which could cause either conformational exchange phenomena, or the formation of multiple phosphorylated products.

Experimental Procedures

Expression and purification of isotope enriched NS5A-D3

The synthetic gene (Eurofins Genomics) encoding NS5A-D3, corresponding to amino acids 356–447 of full length NS5A of HCV genotype 1b (strain HC-J4) was cloned into the expression vector pET-16b+ between NdeI and BamHI restriction sites and was verified by DNA sequencing. The plasmid was transformed in *E. coli* BL21(DE3) Gold competent cells (Agilent Technologies). The cells were grown at 37 °C in minimal medium with ^{15}N-ammonium sulfate and either ^{13}C-glucose or ^{12}C-glucose as the sole nitrogen and carbon sources to an optical density 0.6. Protein expression was induced with 0.5 mM IPTG for 16 hr at 18 °C. NS5A-D3 was purified by affinity chromatography using a nickel-chelating HisTrap (GE Healthcare) column; the protein was recovered in elution buffer containing 300 mM imidazole. After buffer exchange and digestion with Factor Xa protease (New England Biolabs) for 24 hr at 25 °C, the protein was separated from the affinity tag in a HisTrap column. The buffer was exchanged with phosphate buffer saline (PBS, Abcam), 0.5 mM EDTA, 1 mM DTT, pH=7, and the protein was concentrated for NMR analysis.

Phosphorylation of NS5A-D3 by CK2

The phosphorylation reaction was performed on a sample containing 100 µM NS5A-D3 (either ^{15}N-labeled or ^{13}C,^{15}N-labeled) with 1000 U of CK2 (New England Biolabs) at 303 K in kinase reaction buffer (20 mM Tris-HCl, 50 mM KCl, 10 mM $MgCl_2$, 1 mM DTT, 2 mM ATP, pH=7.3). The samples of ^{13}C,^{15}N-labeled phosphorylated NS5A-D3 at different reaction times were prepared by stopping the reaction by inactivating CK2 at 70 °C for 20 min.

NMR experiments

NMR experiments for the sequence-specific resonance assignment of NS5A-D3 were acquired at a 900 MHz Bruker Avance spectrometer equipped with a TCI CryoProbe. A ^{1}H-^{15}N 2D SOFAST-HMQC (46) and a set of BEST triple-resonance 3D experiments (HNCO, HN(CA)CO, HNCA, HN(CO)CA, HNCACB, HN(CO)CACB) (47) were acquired at 278 K on a sample of unphosphorylated [U-^{13}C,^{15}N]-NS5A-D3 in PBS buffer with 0.5 mM EDTA, 1 mM DTT, pH=7. For the assignment of the phosphorylated NS5A-D3, the same 3D experiments were acquired at a 950 MHz Bruker Avance spectrometer equipped with a TCI CryoProbe at 278 K on samples of [U-^{13}C,^{15}N]-NS5A-D3, which had been previously incubated with CK2 at different times in kinase reaction buffer at 303 K. The resonance assignments at 278 K were transferred on the spectra acquired at 303 K with the help of SOFAST-HMQC spectra acquired at increasing temperatures. The CK2-dependent phosphorylation of NS5A-D3 was monitored real-time by NMR at 950 MHz. A ^{1}H-^{15}N 2D SOFAST-HMQC spectrum was acquired at t = 0 on a sample of unphosphorylated [U-^{15}N]-NS5A-D3 in kinase reaction buffer (pH=7.3). 1000 U of CK2 were then added to the sample, which was quickly placed back in the instrument. The reaction was monitored starting from t = 7.5 min., with a series of 15 min.-long SOFAST-HMQC spectra for the first 6 hours, followed by a series of 60 min.-long SOFAST-HMQC spectra for the next 18 h (until t = 24 h). All the spectra were processed with Bruker Topspin software.

NMR data analysis

The NMR spectra were analyzed with the software CARA (Computer Aided Resonance Assignment) (48). Crosspeak intensities were integrated at each time point. The curves obtained at 15 min. intervals over 6 h were used to analyze the faster events, while the curves obtained at 1 h intervals over 24 h were used to analyze the slower events. For the latter, the first 6 points were calculated each by summing the intensity over four consecutive 15 min. spectra. The reaction curves were fitted in Origin Pro with the Nonlinear Curve Fitting tool as single exponential decays/buildups, approximating the enzymatic reaction as irreversible first order reactions with the equation:

$$y = y_0 + S_0 e^{-k_{app} t} \qquad 1$$

$$y = y_0 + S_0\left(1 - e^{-k_{app} t}\right) \qquad 2$$

For each phosphorylation site, k_{app} (the apparent rate constant) was calculated by globally fitting the curves of the neighboring residues. The buildup curve obtained from the overlapped crosspeaks of pS401 and pS408 was fitted afterwards as the sum of two exponential buildups, by imposing k_{app}^{S401} = 0.127 h^{-1} and k_{app}^{S408} = 1.11 h^{-1}. For S429, the intermediate species was independently fitted with the equation:

$$y = \frac{k_1 S_0}{k_2 - k_1}(e^{-k_1 t} - e^{-k_2 t}) \qquad 3$$

where k_1 and k_2 are the apparent rate constants of two consecutive reactions. The final phosphorylated species pS429 was independently fitted with equation 2. For S434, the intermediate and the final species were fitted together as consecutive first order reactions, respectively with equation 3 and the equation:

$$y = S_0 \frac{1}{k_1 - k_2}(k_2 e^{-k_1 t} - k_1 e^{-k_2 t}) \qquad 4$$

Chemical shift perturbations between unphosphorylated and phosphorylated [U-^{15}N]-NS5A-D3 were calculated as Combined Chemical Shift Difference (CCSD) from ^{1}H and ^{15}N amide chemical shifts measured at 278 K using the formula:

$$CCSD = \sqrt{\frac{1}{2}(\Delta\delta^{1}H)^2 + \frac{1}{2}(\Delta\delta^{15}N/5)^2} \qquad 5$$

Neighbor-corrected Structural Propensity (ncSP) plots of unphosphorylated and phosphorylated [U-^{13}C,^{15}N]-NS5A-D3 were calculated from the ^{15}N, ^{13}C', ^{13}Cα, ^{13}Cβ secondary chemical shifts measured at 278 K using the ncSPC web tool (http://nmr.chem.rug.nl/ncSPC; Tamiola et al., 2010 algorithm) (49).

Results

NS5A-D3 is an IDP with low secondary structure propensity

NS5A-D3 (residues 356-447, Fig. 1a) from the HCV-1b HC-J4 strain, was recombinantly expressed in *E. coli* cells, either U-^{15}N- or U-^{13}C,^{15}N-labeled, and purified for NMR characterization. The ^{1}H-^{15}N SOFAST-HMQC spectrum shows poor dispersion of the amide chemical shifts, typical of an intrinsically disordered protein (IDP, Fig. 1b). The backbone chemical shifts were assigned from a set of 3D triple resonance NMR experiments on the U-^{13}C,^{15}N-labeled protein. The neighbor-corrected secondary structure propensity plot was calculated from the secondary chemical shifts, and showed that NS5A-D3 has a region with very low α-helix propensity close to the N-terminus, whereas the rest of the protein lacks any significant secondary structure propensity, as previously reported (Fig. 1c) (50).

CK2 phosphorylates NS5A-D3 on four serine residues with different rates

The phosphorylation of U-^{15}N NS5A-D3 by CK2 was monitored *in vitro* by acquiring a series of ^{1}H-^{15}N NMR spectra during the course of the reaction, which was carried on for 24 h at 303 K. During this time, a number of crosspeaks appeared in a region of the spectrum typical of phosphorylated serines/threonines, accompanied by the decrease in intensity of some crosspeaks of the unphosphorylated D3 (Fig. 2a). Additionally, crosspeaks arising from residues close to the phosphorylated sites shifted during the reaction. The crosspeaks arising from the phosphorylated species were assigned *de novo* from sets of 3D triple resonance NMR experiments acquired on U-^{13}C,^{15}N-labeled NS5A-D3 samples at different reaction times. Upon reaction with CK2, four serine residues were unambiguously identified as phosphorylation sites: S401, S408, S429, S434 (Fig. 2a and Fig. 3). The crosspeaks arising from pS401 and pS408 were

overlapped in the ^{1}H-^{15}N NMR spectra acquired during the reaction (at 303 K), but were well separated at 278 K. Time-dependent phosphorylation curves were obtained for each site by integrating the crosspeak intensities of the phosphorylated residue and of the neighboring residues at each time point (Fig. 2b-e and Supplementary Fig. 1-4). To determine the apparent rate constant for each site ($k_{app} \cong k_{cat} \times [CK2] / K_M$), the curves were fitted with a single exponential decay/buildup approximating the enzymatic reactions as irreversible first order reactions. S408 reacted with the highest rate constant ($k_{app}^{S408} = 1.11 \pm 0.15\ h^{-1}$), and was completely phosphorylated in about 4 h (Fig. 2b and Supplementary Fig. 1), while the other serines reacted significantly slower ($k_{app}^{S401} = 0.127 \pm 0.005\ h^{-1}$; $k_{app}^{S429} = 0.178 \pm 0.005\ h^{-1}$; $k_{app}^{S434} = 0.191 \pm 0.007\ h^{-1}$, Fig. 2c-e and Supplementary Fig. 2-4), and were not completely phosphorylated in 24 hours (possibly also due to decreased kinase activity and/or ATP depletion). Phosphorylation of S429 and S434 gave rise to two amide crosspeaks for each residue, with different buildup curves (Fig. 2d,e and Fig. 3). As the reaction proceeds, the singly phosphorylated species are depleted, as reflected by the decreasing intensity of the corresponding crosspeaks. The curves of the intermediate and final species of S434 are compatible with a consecutive reaction model (Fig. 2e, $k_1^{S434} = 0.17 \pm 0.01\ h^{-1}$; $k_2^{S434} = 0.092 \pm 0.003\ h^{-1}$), suggesting that other reactions may occur at dependence of S434 phosphorylation. A suitable model was not found to explain the formation of the intermediate and final species of S429 (Fig. 2d) that seem to occur with different rate constants ($k_1^{S429} = 0.71 \pm 0.03\ h^{-1}$; $k_2^{S429} = 0.011 \pm 0.001\ h^{-1}$ for the blue curve; $k_{app}^{S429} = 0.111 \pm 0.004\ h^{-1}$ for the red curve).

Phosphorylation of NS5A-D3 does not influence the secondary structure

The chemical shifts of phosphorylated NS5A-D3 at 278 K were compared to those of the non-phosphorylated protein. Phosphorylation only caused small perturbations of the amide chemical shifts of the residues close to the phosphorylated serines, while the other residues remained unaffected, indicating that no changes in the protein conformational features have occurred (Fig. 4a). The neighbor-corrected secondary structure propensity plot confirmed this result, as phosphorylation caused no significant changes in the calculated structure propensity (Fig. 4b).

Discussion

Previous studies have identified several potential CK2-dependent phosphorylation sites in the sequence of NS5A-D3 (29,30,33,36). However, it was not always possible to pinpoint the exact position of each site, due to the intrinsically ambiguous nature of the prediction-based approaches. Indeed, the type and number of predicted phosphorylation sites depends on how the consensus sequence for CK2-dependent phosphorylation is defined. Usually, the consensus sequence used to predict CK2 sites requires an acidic residue (D/E) at the +3 position (i.e. X-pS/pT-X-X-D/E) (36,51). However, other motifs are known to be recognized by CK2 which allow D/E at +1,+2 (i.e. X-pS/pT-X-D/E and X-pS/pT-D/E, respectively) and, rarely, at even different positions (52), thereby making the predictions less reliable. Moreover, the phosphorylation sites were investigated on the NS5A protein from different HCV genotypes. Consequently, different sites were predicted to be recognized by CK2 for each genotype.

The four phosphorylation events monitored here by NMR occur on serine residues. Among them, S408 is phosphorylated at the highest rate, has a 'canonical' (+3) CK2 recognition sequence and its phosphorylation is consistent with previously reported data on other 1b HCV strains (36). S401 is phosphorylated at a much lower rate and, unlike S408, has a 'non-canonical' recognition sequence (+1,+2) in the strain analyzed here. S401 has been identified before as a CK2 phosphorylation site in other NS5A sequences from 1b and 2a HCV strains, in which a +3 recognition sequence was present (30,36). S429 has a +1 recognition sequence in this strain and, to our knowledge, has been identified here for the first time as a phosphorylation site of NS5A-D3.

The observed phosphorylation of S434 was somewhat unexpected, as it lacks a proper recognition sequence for CK2: the closest following acidic residue is E438 (+4), which is rarely the case in CK2 recognition motifs. S434 may be recognized as a secondary phosphorylation site of CK2, dependent on the phosphorylation of S437 (X-pS/pT-X-X-pS/pT). Interestingly, both S434 and S437 are highly conserved among the HCV genotypes and their phosphorylation has been found to be critical for the interaction between NS5A and the core protein, which serves as a scaffold during the assembly of the viral particle (25). Moreover, phosphorylation of S437 has been reported in the 2a strain (S457 in 2a JFH-1 genotype) (33,36). While phosphorylation of S437 was not observed in our experimental conditions, a decrease in the crosspeak intensity of the

residues 435-437 was observed, although no new crosspeaks (shifted due to the formation of pS437) could be identified. This behavior suggests that, during the reaction with CK2, either one species is formed that undergoes conformational exchange phenomena, which would hinder the identification of phosphorylated residues in that region, or multiple species are formed with different phosphorylation patterns.

The basal phosphorylation of NS5A plays a critical role in the virion assembly pathway of HCV. Although the exact mechanism is still largely uncharacterized, NS5A is known to interact with the core protein at the surface of lipid droplets through phosphorylated serines in the D3. Despite the available data, the exact position of several of the reported phosphorylation sites is still unknown. Here, we have shown by NMR that CK2 reacts *in vitro* with NS5A-D3 and we have identified unambiguously four sites of phosphorylation. The reactions occurred independently of each other and with very different speeds, with almost 10-fold difference between the formation of pS408 and the others. While the biological significance of such difference is obscure, it can be speculated that pS408 acts as a primer for other events, such as interaction with other cellular kinases, which may be necessary for the correct protein maturation and assembly of the virion. Finally, our data validates the current hypothesis of the interaction between NS5A and the core protein, which requires the phosphorylation of serines close to the C-terminus of D3, such as S434, in a region conserved among all HCV genotypes. Further investigation should focus on the possible concerted action of CK2 with other kinases, such as the recently reported c-Abl tyrosine kinase (53), which may require NS5A-D3 to first react with CK2 in order to be recognized as a substrate of c-Abl.

Acknowledgments

This work has been supported by "MEDINTECH: Tecnologie convergenti per aumentare la sicurezza e l'efficacia di farmaci e vaccini" and by Instruct, part of the European Strategy Forum on Research Infrastructures (ESFRI) and supported by national member subscriptions. Specifically, we thank the EU ESFRI Instruct Core Centre CERM-Italy.

Conflict of interest

The authors declare that they have no conflicts of interest with the contents of this article.

Author contributions

LB, EL and ES conceived the study and wrote the manuscript. EL and ES designed the experiments and analyzed the data. ES performed the gene cloning, protein expression and purification, and protein phosphorylation reactions. EL performed the NMR experiments, the resonance assignments and the fittings of the reaction curves. All authors reviewed the results and approved the final version of the manuscript.

References

1. Marukian, S., Andrus, L., Sheahan, T. P., Jones, C. T., Charles, E. D., Ploss, A., Rice, C. M., and Dustin, L. B. (2011) Hepatitis C virus induces interferon-λ and interferon-stimulated genes in primary liver cultures. *Hepatol. Baltim. Md.* **54**, 1913–1923

2. Maasoumy, B., and Wedemeyer, H. (2012) Natural history of acute and chronic hepatitis C. *Best Pract. Res. Clin. Gastroenterol.* **26**, 401–412

3. Robertson, B., Myers, G., Howard, C., Brettin, T., Bukh, J., Gaschen, B., Gojobori, T., Maertens, G., Mizokami, M., Nainan, O., Netesov, S., Nishioka, K., Shin i, T., Simmonds, P., Smith, D., Stuyver, L., and Weiner, A. (1998) Classification, nomenclature, and database development for hepatitis C virus (HCV) and related viruses: proposals for standardization. International Committee on Virus Taxonomy. *Arch. Virol.* **143**, 2493–2503

4. Davis, G. L. (1999) Hepatitis C virus genotypes and quasispecies. *Am. J. Med.* **107**, 21S–26S

5. Torresi, J., Johnson, D., and Wedemeyer, H. (2011) Progress in the development of preventive and therapeutic vaccines for hepatitis C virus. *J. Hepatol.* **54**, 1273–1285

6. Pawlotsky, J.-M. (2013) Treatment of chronic hepatitis C: current and future. *Curr. Top. Microbiol. Immunol.* **369**, 321–342

7. Dubuisson, J., and Cosset, F.-L. (2014) Virology and cell biology of the hepatitis C virus life cycle--an update. *J. Hepatol.* **61**, S3–S13

8. Choo, Q. L., Richman, K. H., Han, J. H., Berger, K., Lee, C., Dong, C., Gallegos, C., Coit, D., Medina-Selby, R., and Barr, P. J. (1991) Genetic organization and diversity of the hepatitis C virus. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 2451–2455

9. Bartenschlager, R., Cosset, F.-L., and Lohmann, V. (2010) Hepatitis C virus replication cycle. *J. Hepatol.* **53**, 583–585

10. Moradpour, D., Penin, F., and Rice, C. M. (2007) Replication of hepatitis C virus. *Nat. Rev. Microbiol.* **5**, 453–463

11. Evans, M. J., Rice, C. M., and Goff, S. P. (2004) Phosphorylation of hepatitis C virus nonstructural protein 5A modulates its protein interactions and viral RNA replication. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 13038–13043

12. Macdonald, A., and Harris, M. (2004) Hepatitis C virus NS5A: tales of a promiscuous protein. *J. Gen. Virol.* **85**, 2485–2502

13. Pawlotsky, J. M., and Germanidis, G. (1999) The non-structural 5A protein of hepatitis C virus. *J. Viral Hepat.* **6**, 343–356

14. Tellinghuisen, T. L., Marcotrigiano, J., Gorbalenya, A. E., and Rice, C. M. (2004) The NS5A protein of hepatitis C virus is a zinc metalloprotein. *J. Biol. Chem.* **279**, 48576–48587

15. Penin, F., Brass, V., Appel, N., Ramboarina, S., Montserret, R., Ficheux, D., Blum, H. E., Bartenschlager, R., and Moradpour, D. (2004) Structure and function of the membrane anchor domain of hepatitis C virus nonstructural protein 5A. *J. Biol. Chem.* **279**, 40835–40843

16. Huang, L., Hwang, J., Sharma, S. D., Hargittai, M. R. S., Chen, Y., Arnold, J. J., Raney, K. D., and Cameron, C. E. (2005) Hepatitis C virus nonstructural protein 5A (NS5A) is an RNA-binding protein. *J. Biol. Chem.* **280**, 36417–36428

17. Love, R. A., Brodsky, O., Hickey, M. J., Wells, P. A., and Cronin, C. N. (2009) Crystal structure of a novel dimeric form of NS5A domain I protein from hepatitis C virus. *J. Virol.* **83**, 4395–4403

18. Tellinghuisen, T. L., Marcotrigiano, J., and Rice, C. M. (2005) Structure of the zinc-binding domain of an essential component of the hepatitis C virus replicase. *Nature*. **435**, 374–379

19. Miyanari, Y., Atsuzawa, K., Usuda, N., Watashi, K., Hishiki, T., Zayas, M., Bartenschlager, R., Wakita, T., Hijikata, M., and Shimotohno, K. (2007) The lipid droplet is an important organelle for hepatitis C virus production. *Nat. Cell Biol.* **9**, 1089–1097

20. Feuerstein, S., Solyom, Z., Aladag, A., Favier, A., Schwarten, M., Hoffmann, S., Willbold, D., and Brutscher, B. (2012) Transient structure and SH3 interaction sites in an intrinsically disordered fragment of the hepatitis C virus protein NS5A. *J. Mol. Biol.* **420**, 310–323

21. Hanoulle, X., Verdegem, D., Badillo, A., Wieruszeski, J.-M., Penin, F., and Lippens, G. (2009) Domain 3 of non-structural protein 5A from hepatitis C virus is natively unfolded. *Biochem. Biophys. Res. Commun.* **381**, 634–638

22. Liang, Y., Ye, H., Kang, C. B., and Yoon, H. S. (2007) Domain 2 of nonstructural protein 5A (NS5A) of hepatitis C virus is natively unfolded. *Biochemistry (Mosc.)*. **46**, 11550–11558

23. Tellinghuisen, T. L., Foss, K. L., Treadaway, J. C., and Rice, C. M. (2008) Identification of residues required for RNA replication in domains II and III of the hepatitis C virus NS5A protein. *J. Virol.* **82**, 1073–1083

24. Appel, N., Zayas, M., Miller, S., Krijnse-Locker, J., Schaller, T., Friebe, P., Kallis, S., Engel, U., and Bartenschlager, R. (2008) Essential role of domain III of nonstructural protein 5A for hepatitis C virus infectious particle assembly. *PLoS Pathog.* **4**, e1000035

25. Masaki, T., Suzuki, R., Murakami, K., Aizaki, H., Ishii, K., Murayama, A., Date, T., Matsuura, Y., Miyamura, T., Wakita, T., and Suzuki, T. (2008) Interaction of hepatitis C virus nonstructural protein 5A with core protein is critical for the production of infectious virus particles. *J. Virol.* **82**, 7964–7976

26. Kaneko, T., Tanji, Y., Satoh, S., Hijikata, M., Asabe, S., Kimura, K., and Shimotohno, K. (1994) Production of two phosphoproteins from the NS5A region of the hepatitis C viral genome. *Biochem. Biophys. Res. Commun.* **205**, 320–326

27. Tanji, Y., Kaneko, T., Satoh, S., and Shimotohno, K. (1995) Phosphorylation of hepatitis C virus-encoded nonstructural protein NS5A. *J. Virol.* **69**, 3980–3986

28. Huang, Y., Staschke, K., De Francesco, R., and Tan, S.-L. (2007) Phosphorylation of hepatitis C virus NS5A nonstructural protein: a new paradigm for phosphorylation-dependent viral RNA replication? *Virology*. **364**, 1–9

29. Appel, N., Pietschmann, T., and Bartenschlager, R. (2005) Mutational analysis of hepatitis C virus nonstructural protein 5A: potential role of differential phosphorylation in RNA replication and identification of a genetically flexible domain. *J. Virol.* **79**, 3187–3194

30. Ross-Thriepland, D., and Harris, M. (2014) Insights into the complexity and functionality of hepatitis C virus NS5A phosphorylation. *J. Virol.* **88**, 1421–1432

31. Quintavalle, M., Sambucini, S., Di Pietro, C., De Francesco, R., and Neddermann, P. (2006) The alpha isoform of protein kinase CKI is responsible for hepatitis C virus NS5A hyperphosphorylation. *J. Virol.* **80**, 11305–11312

32. Chen, Y.-C., Su, W.-C., Huang, J.-Y., Chao, T.-C., Jeng, K.-S., Machida, K., and Lai, M. M. C. (2010) Polo-like kinase 1 is involved in hepatitis C virus replication by hyperphosphorylating NS5A. *J. Virol.* **84**, 7983–7993

33. Tellinghuisen, T. L., Foss, K. L., and Treadaway, J. (2008) Regulation of hepatitis C virion production via phosphorylation of the NS5A protein. *PLoS Pathog.* **4**, e1000032

34. Kim, J., Lee, D., and Choe, J. (1999) Hepatitis C virus NS5A protein is phosphorylated by casein kinase II. *Biochem. Biophys. Res. Commun.* **257**, 777–781

35. Huang, L., Sineva, E. V., Hargittai, M. R. S., Sharma, S. D., Suthar, M., Raney, K. D., and Cameron, C. E. (2004) Purification and characterization of hepatitis C virus non-structural protein 5A expressed in Escherichia coli. *Protein Expr. Purif.* **37**, 144–153

36. Dal Pero, F., Di Maira, G., Marin, O., Bortoletto, G., Pinna, L. A., Alberti, A., Ruzzene, M., and Gerotto, M. (2007) Heterogeneity of CK2 phosphorylation sites in the NS5A protein of different hepatitis C virus genotypes. *J. Hepatol.* **47**, 768–776

37. Landrieu, I., Lacosse, L., Leroy, A., Wieruszeski, J.-M., Trivelli, X., Sillen, A., Sibille, N., Schwalbe, H., Saxena, K., Langer, T., and Lippens, G. (2006) NMR analysis of a Tau phosphorylation pattern. *J. Am. Chem. Soc.* **128**, 3575–3583

38. Liokatis, S., Dose, A., Schwarzer, D., and Selenko, P. (2010) Simultaneous detection of protein phosphorylation and acetylation by high-resolution NMR spectroscopy. *J. Am. Chem. Soc.* **132**, 14704–14705

39. Cordier, F., Chaffotte, A., Terrien, E., Préhaud, C., Theillet, F.-X., Delepierre, M., Lafon, M., Buc, H., and Wolff, N. (2012) Ordered phosphorylation events in two independent cascades of the PTEN C-tail revealed by NMR. *J. Am. Chem. Soc.* **134**, 20533–20543

40. Liokatis, S., Stützer, A., Elsässer, S. J., Theillet, F.-X., Klingberg, R., van Rossum, B., Schwarzer, D., Allis, C. D., Fischle, W., and Selenko, P. (2012) Phosphorylation of histone H3 Ser10 establishes a hierarchy for subsequent intramolecular modification events. *Nat. Struct. Mol. Biol.* **19**, 819–823

41. Dose, A., Liokatis, S., Theillet, F.-X., Selenko, P., and Schwarzer, D. (2011) NMR profiling of histone deacetylase and acetyl-transferase activities in real time. *ACS Chem. Biol.* **6**, 419–424

42. Theillet, F.-X., Smet-Nocca, C., Liokatis, S., Thongwichian, R., Kosten, J., Yoon, M.-K., Kriwacki, R. W., Landrieu, I., Lippens, G., and Selenko, P. (2012)

Cell signaling, post-translational protein modifications and NMR spectroscopy. *J. Biomol. NMR*. **54**, 217–236

43. Selenko, P., Frueh, D. P., Elsaesser, S. J., Haas, W., Gygi, S. P., and Wagner, G. (2008) In situ observation of protein phosphorylation by high-resolution NMR spectroscopy. *Nat. Struct. Mol. Biol.* **15**, 321–329

44. Theillet, F.-X., Rose, H. M., Liokatis, S., Binolfi, A., Thongwichian, R., Stuiver, M., and Selenko, P. (2013) Site-specific NMR mapping and time-resolved monitoring of serine and threonine phosphorylation in reconstituted kinase reactions and mammalian cell extracts. *Nat. Protoc.* **8**, 1416–1432

45. Thongwichian, R., Kosten, J., Benary, U., Rose, H. M., Stuiver, M., Theillet, F.-X., Dose, A., Koch, B., Yokoyama, H., Schwarzer, D., Wolf, J., and Selenko, P. (2015) A Multiplexed NMR-Reporter Approach to Measure Cellular Kinase and Phosphatase Activities in Real-Time. *J. Am. Chem. Soc.* **137**, 6468–6471

46. Schanda, P., and Brutscher, B. (2005) Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds. *J. Am. Chem. Soc.* **127**, 8014–8015

47. Lescop, E., Schanda, P., and Brutscher, B. (2007) A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. *J. Magn. Reson. San Diego Calif 1997*. **187**, 163–169

48. Keller, R. L. J. (2004) *Diss. ETH No. 15947*. Ph.D. thesis, Swiss Federal Institute Of Technology Zurich

49. Tamiola, K., Acar, B., and Mulder, F. A. A. (2010) Sequence-specific random coil chemical shifts of intrinsically disordered proteins. *J. Am. Chem. Soc.* **132**, 18000–18003

50. Verdegem, D., Badillo, A., Wieruszeski, J.-M., Landrieu, I., Leroy, A., Bartenschlager, R., Penin, F., Lippens, G., and Hanoulle, X. (2011) Domain 3 of NS5A protein from the hepatitis C virus has intrinsic alpha-helical propensity and is a substrate of cyclophilin A. *J. Biol. Chem.* **286**, 20441–20454

51. Pinna, L. A. (2002) Protein kinase CK2: a challenge to canons. *J. Cell Sci.* **115**, 3873–3878

52. Meggio, F., and Pinna, L. A. (2003) One-thousand-and-one substrates of protein kinase CK2? *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* **17**, 349–368

53. Yamauchi, S., Takeuchi, K., Chihara, K., Sun, X., Honjoh, C., Yoshiki, H., Hotta, H., and Sada, K. (2015) Hepatitis C Virus Particle Assembly Involves Phosphorylation of NS5A by the c-Abl Tyrosine Kinase. *J. Biol. Chem.* 10.1074/jbc.M115.666859

Figures

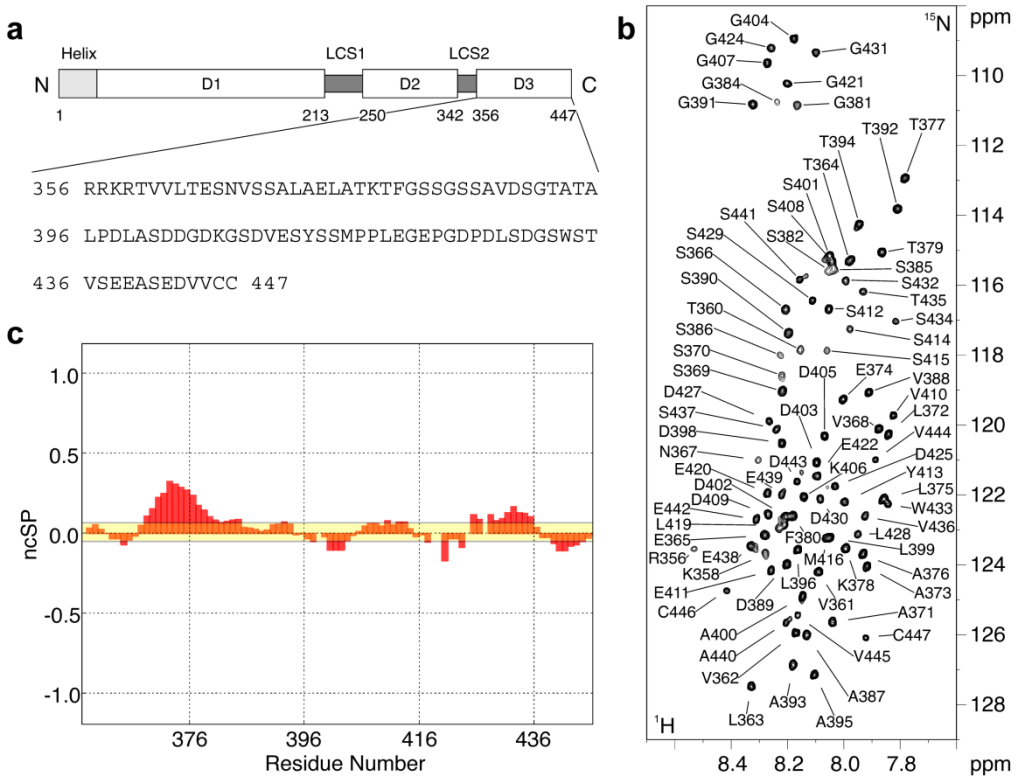


Figure 1. HCV NS5A-D3 has low structural propensity. (a) Domain organization of NS5A. Top, from N- to C-terminus: amphipathic α-helix (light gray); domain 1; low-complexity sequence (LCS) 1; domain 2; LCS2; domain 3. Bottom, amino acid sequence of NS5A-D3. (b) ^{1}H-^{15}N SOFAST-HMQC spectrum of [U-^{15}N]-NS5A-D3 at 278 K. The sequence-specific resonance assignment is shown. (c) Neighbor-corrected Structural Propensity (ncSP) of NS5A-D3 calculated from the ^{15}N, ^{13}C', ^{13}Cα, ^{13}Cβ secondary chemical shifts measured at 278 K; positive and negative values indicate propensities for α-helix and β-strand, respectively.

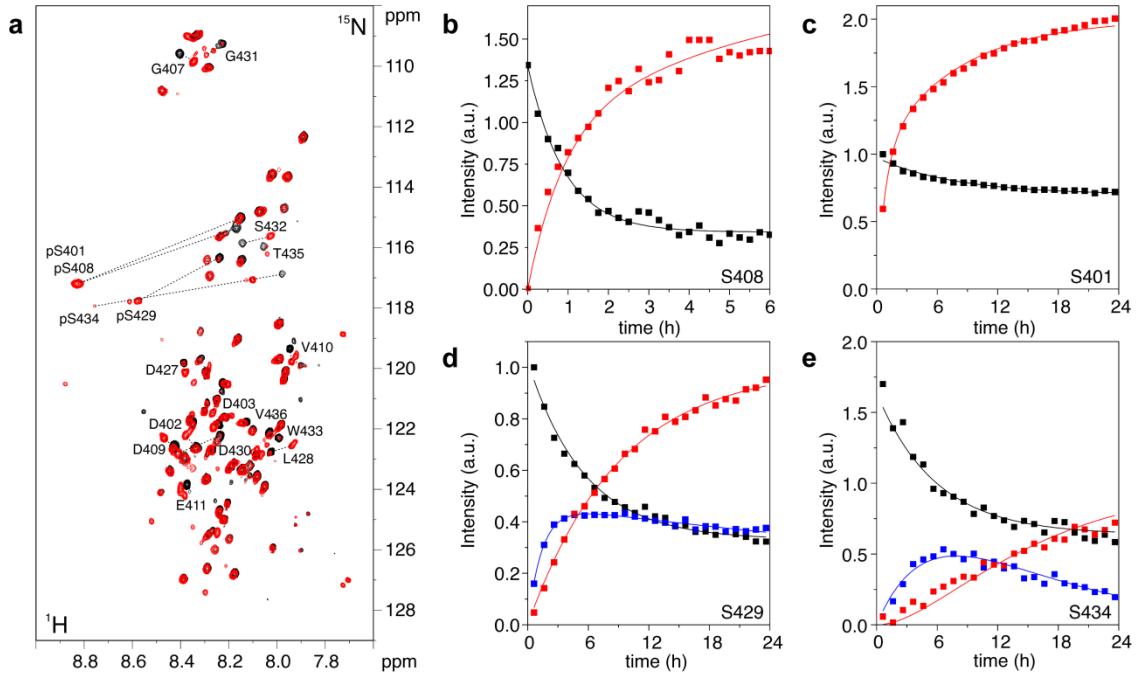


Figure 2. CK2-dependent phosphorylation of NS5A-D3. (a) ^{1}H-^{15}N SOFAST-HMQC spectrum (black) of [U-^{15}N]-NS5A-D3 at 303 K; ^{1}H-^{15}N SOFAST-HMQC spectrum (red) of the same protein sample after CK2-dependent phosphorylation for 24 h at 303 K. The crosspeaks shifted during the reaction are labeled according to the assigned residues. (b-e) Time-dependent phosphorylation curves of each phosphorylated residue, obtained by integrating the crosspeak intensity at each time point. The values are normalized on the highest value of the fitting curves. Unreacted species (black squares) are fitted with a single exponential decay function (black lines); phosphorylated species (red squares) with a single exponential buildup function (red lines); intermediate species (blue squares) with a two-exponential function (see Methods).

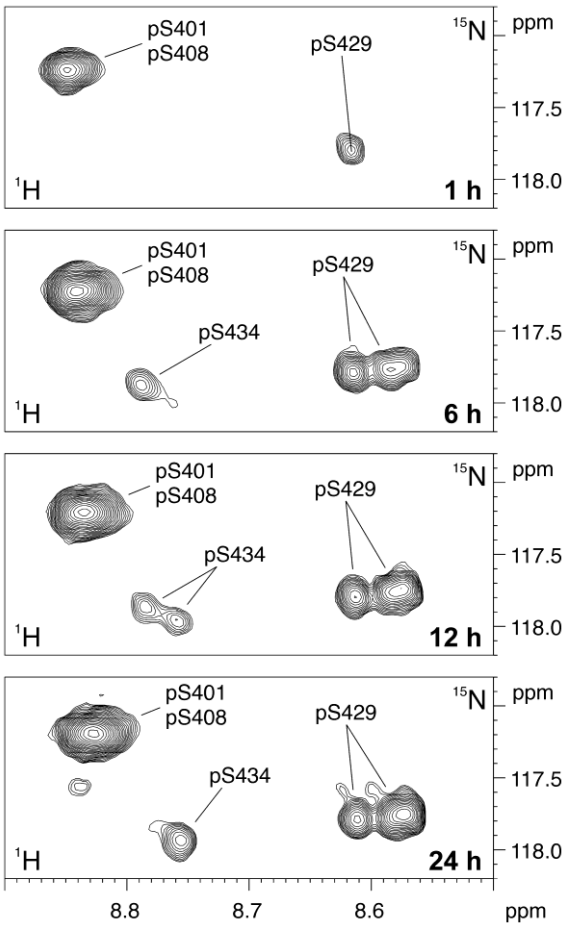


Figure 3. Phosphorylated residues at different reaction times. Detailed view of ^{1}H-^{15}N SOFAST-HMQC spectra acquired on [U-^{15}N]-NS5A-D3 at increasing times of CK2-dependent phosphorylation at 303 K. The crosspeaks arising from phosphorylated serines are labeled according to the residue number.

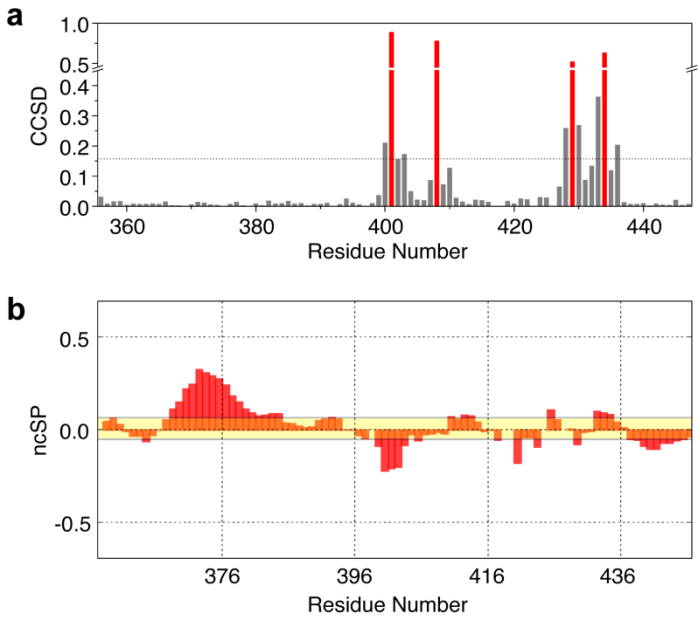
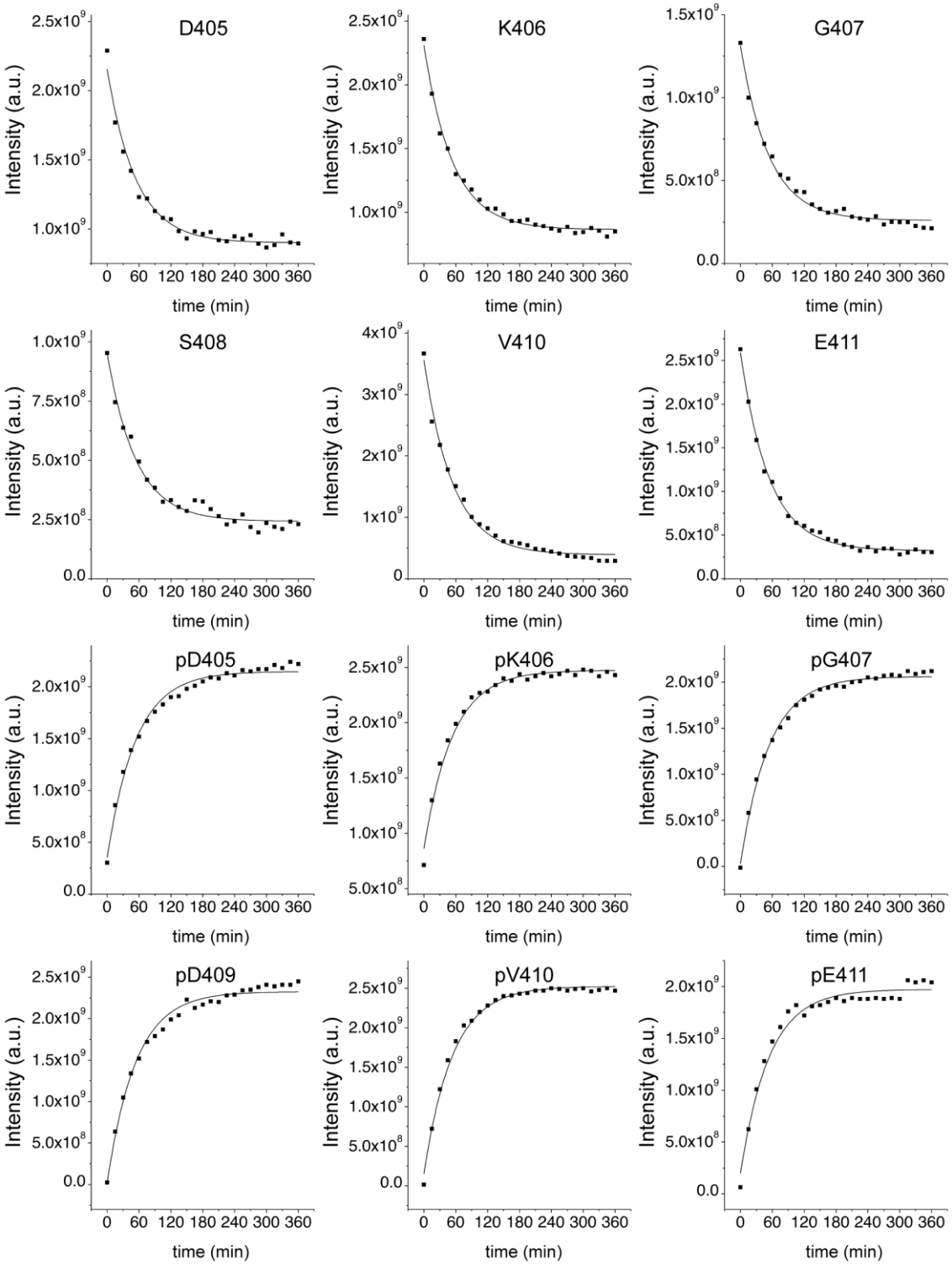


Figure 4. Effect of phosphorylation on the conformational properties of NS5A-D3. (a) Combined Chemical Shift Difference (CCSD) plot of phosphorylated vs. non-phosphorylated NS5A-D3 calculated from the ^{1}H and ^{15}N chemical shifts at 278 K. The phosphorylated residues are shown in red; CCSD standard deviation is shown as dashed line. (b) Neighbor-corrected Structural Propensity (ncSP) of phosphorylated NS5A-D3 calculated from the ^{15}N, $^{13}C'$, $^{13}C\alpha$, $^{13}C\beta$ secondary chemical shifts at 278 K. Due to the shift induced by the phosphate moieties, $^{13}C\alpha$ and $^{13}C\beta$ of the phosphorylated residues were not included in the calculation.

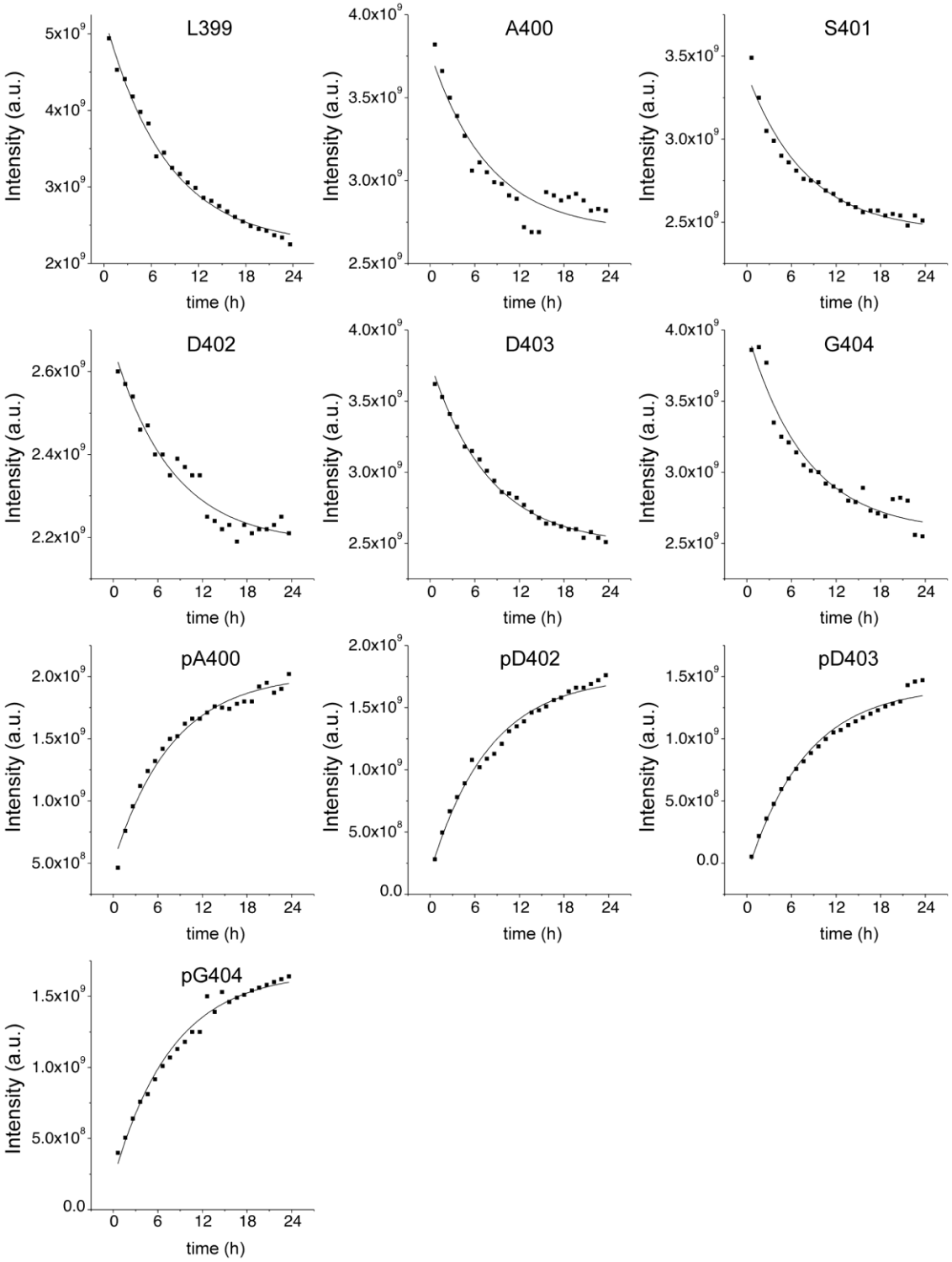
Supporting Information:

The casein kinase 2-dependent phosphorylation of NS5A domain 3 from hepatitis C virus followed by time-resolved NMR

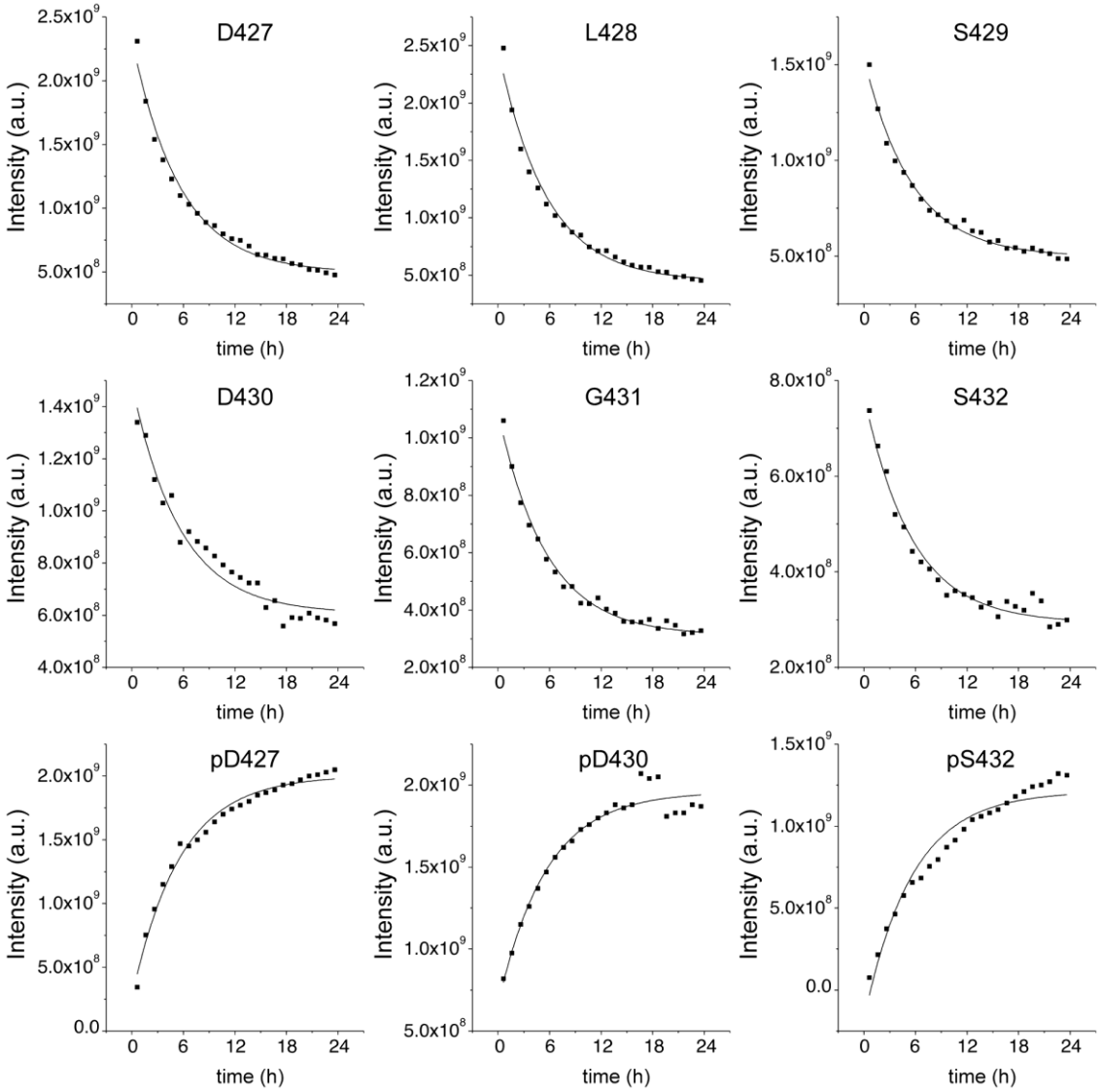
Supplementary Figures


D405
K406
G407
S408
V410
E411
pD405
pK406
pG407
pD409
pV410
pE411
Intensity (a.u.)
time (min)
0 60 120 180 240 300 360

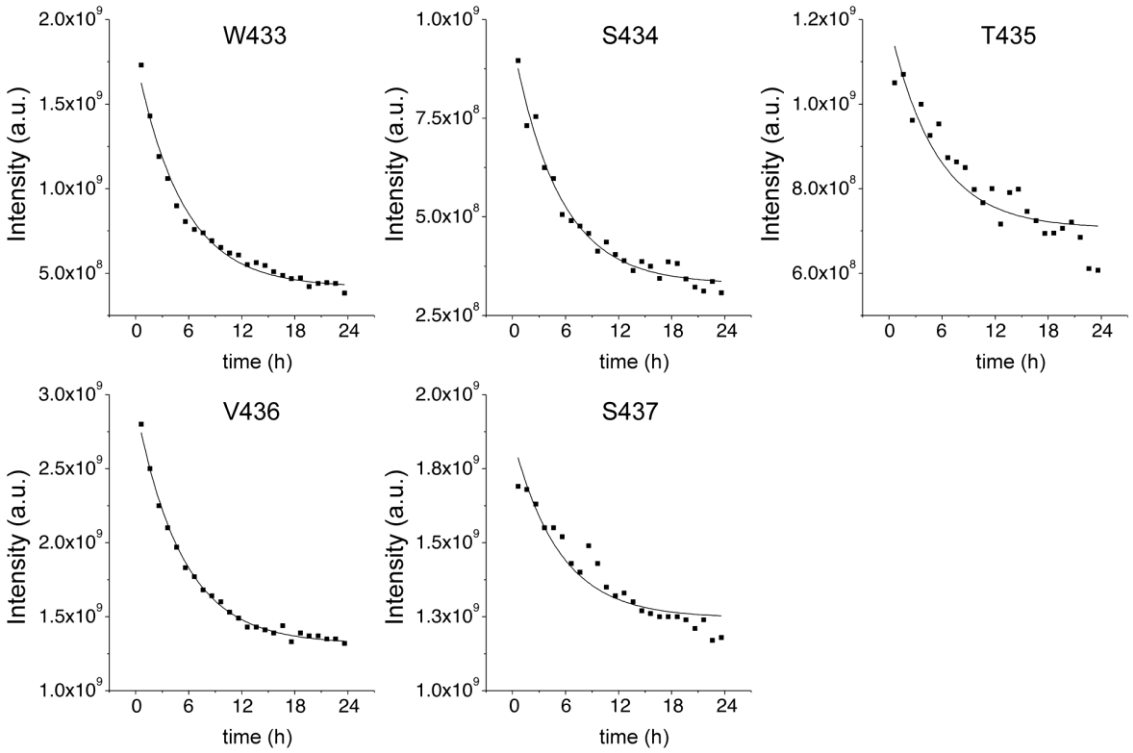
Supplementary Figure 1. Time-dependent reaction curves of the residues in proximity of (and including) S408, obtained by integrating the NMR crosspeak intensity at each time point. Decreasing signals (arising from the residues of unphosphorylated NS5A-D3) are fitted with a single exponential decay function; increasing signals (arising from NS5A-D3 phosphorylated at S408) are fitted with a single exponential buildup function. The curves were globally fitted to calculate the apparent rate constant of S408 phosphorylation (k_{app}^{S408} = 1.11 ± 0.15 h^{-1}). Weak or overlapping crosspeaks (not shown) were not included in the analysis.


L399
A400
S401
D402
D403
G404
pA400
pD402
pD403
pG404
Intensity (a.u.)
time (h)

Supplementary Figure 2. Time-dependent reaction curves of the residues in proximity of (and including) S401, obtained by integrating the NMR crosspeak intensity at each time point. Decreasing signals (arising from the residues of unphosphorylated NS5A-D3) are fitted with a single exponential decay function; increasing signals (arising from NS5A-D3 phosphorylated at S401) are fitted with a single exponential buildup function. The curves were globally fitted to calculate the apparent rate constant of S401 phosphorylation ($k_{app}^{S401} = 0.127 \pm 0.005\ h^{-1}$). Weak or overlapping crosspeaks (not shown) were not included in the analysis.



Supplementary Figure 3. Time-dependent reaction curves of the residues in proximity of (and including) S429, obtained by integrating the NMR crosspeak intensity at each time point. Decreasing signals (arising from the residues of unphosphorylated NS5A-D3) are fitted with a single exponential decay function; increasing signals (arising from NS5A-D3 phosphorylated at S429) are fitted with a single exponential buildup function. The curves were globally fitted to calculate the apparent rate constant of S429 phosphorylation (k_{app}^{S429} = 0.178 ± 0.005 h^{-1}). Weak or overlapping crosspeaks (not shown) were not included in the analysis.



Supplementary Figure 4. Time-dependent reaction curves of the residues in proximity of (and including) S434, obtained by integrating the NMR crosspeak intensity at each time point. Decreasing signals (arising from the residues of unphosphorylated NS5A-D3) are fitted with a single exponential decay function; increasing signals (arising from NS5A-D3 phosphorylated at S434) are fitted with a single exponential buildup function. The curves were globally fitted to calculate the apparent rate constant of S434 phosphorylation ($k_{app}^{S434} = 0.191 \pm 0.007$ h^{-1}). Weak or overlapping crosspeaks (not shown) were not included in the analysis.

3. Methodological Aspects

3.1 Cell line

Cultured human embryonic kidney 293T (HEK293T) cells were continuously propagated in culture flasks for adherent cells, by incubation under controlled 5% CO_2 atmosphere at 37 °C. HEK293T cells have a doubling time of ~24 h, and are propagated every 3-4 days by 1:10 dilution and plating on a new flask. This cell line has some advantages like to be easy-handling, excellent transfectability and high capacity for recombinant protein expression. They maintain unaltered these properties for ~20 propagation steps, after which a new cell culture is started from a frozen cell stock. During cell growth, Dulbecco's modified Eagle medium (DMEM) is used, supplemented with L-glutamine, antibiotics (penicillin/streptomycin) and 10% fetal bovine serum (FBS).

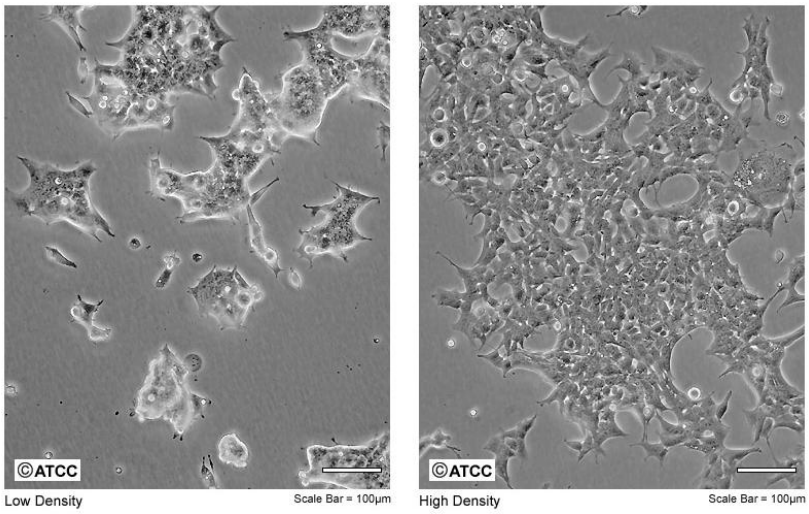


Figure 5 HEK293T cells. Reprinted from ATCC website.

3.2 Gene cloning

The genes of hMia40, hTrx1 and hGrx1 were cloned into the pHLsec vector (Fig. 6) [107].

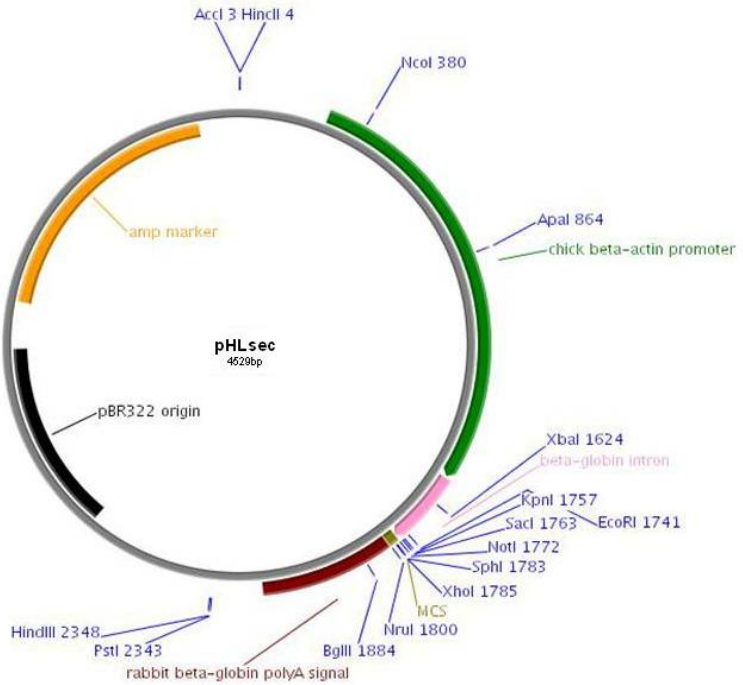


Figure 6 pHLsec vector map [107].

This plasmid is optimized for the expression in mammalian cells and contains the cytomegalovirus (CMV) enhancer, a chicken beta actin promoter, a Kozak sequence, an optimized chicken protein receptor tyrosine phosphatase (RPTP) sigma secretion signal, cloning sites, and a rabbit betaglobin polyadenylation (PolyA) sequence. Using the EcoR I and Xho I restriction enzymes, the secretion signal sequence was removed allowing the native protein to be expressed in the cytoplasm.

The gene of hHAH1 was cloned in the pURD plasmid between Hind III and Xho I restriction enzymes (Fig. 7) [108].

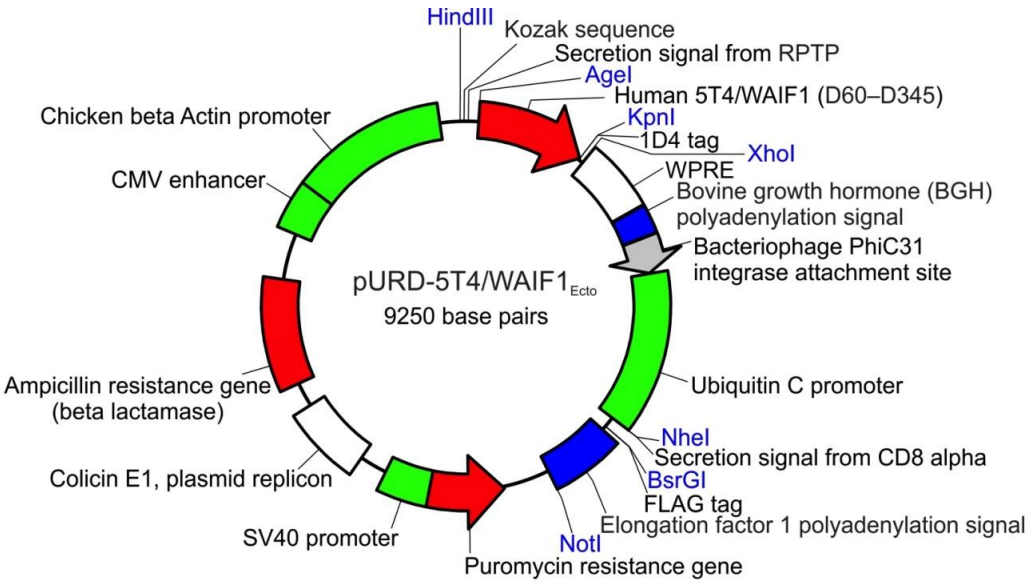


Figure 7 pURD vector map [108].

pURD vector is optimized for generating stable cell lines. It contains a phiC31 attachment site (AttB) for site-specific recombination with pseudo AttP sites in the mammalian cell genome when the PhiC31 integrase gene (pgk-phiC31/pCB92) is cotransfected, and a mammalian cell selection cassette (SV40 promoter, puromycin resistance gene, PolyA). The expression cassette contains the same promoter as pHLSec, a Woodchuck Hepatitis Virus posttranscriptional regulatory element (WPRE), and a bovine growth hormone (BGH) PolyA.

All the DNA clones were verified by DNA sequencing. DNA was purified using the Endotoxin-Free Plasmid Maxi Kit (Macherey-Nagel) in order to ensure successful transfection of human cells and protein expression.

The synthetic gene (Eurofins Genomics) encoding NS5A-D3, corresponding to amino acids 356–447 of full length NS5A of HCV genotype 1b (strain HC-J4), was cloned into the expression vector pET-16b+ between Nde I and BamH I restriction sites. The pET-16b+ vector contains a T7 promoter for expression in bacterial cells and carries an N-terminal His-Tag sequence followed by a Factor Xa site (Fig. 8).

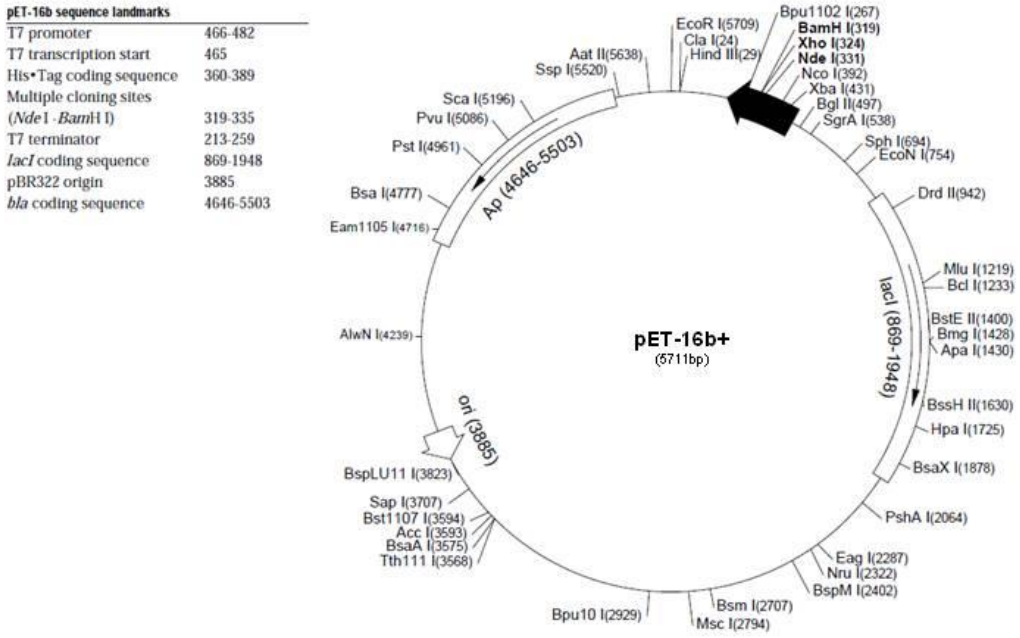


Figure 8 pET1-6b+ vector map. Reprinted from http://www.helmholtz-muenchen.de/fileadmin/PEPF/pET_vectors/pET-16b_map.pdf

3.3 DNA transfection in mammalian cells

Transfection is the process of introducing nucleic acids (such as supercoiled plasmid DNA or siRNA constructs) into the cells. There are various methods of introducing foreign DNA into a eukaryotic cell: some rely on physical treatment (electroporation, cell squeezing, nanoparticles), other on chemical materials or biological particles (viruses) that are used as carriers. A very affordable and highly efficient transfection reagent is polyethylenimine (PEI) [109, 110]. A large selection of PEI forms is available, varying in molecular weight and branching. The employed PEI is '25 kDa branched', which was found to be most effective in transfecting various cells lines with minimal cytotoxic effects [111]. Stock solutions are made in water, first at 100 mg ml^{-1} (PEI is an extremely viscous liquid). Once the solution is homogeneous, it is further diluted to 1 mg ml^{-1}, the pH adjusted to 7 with HCl, filter sterilized and aliquoted.

PEI is a polycationic reagent that forms large complexes with the negatively charged DNA and promotes the binding of the complex to anionic proteoglycans that are present on the cell surface. Through endocytosis, the DNA/PEI complex enters into the cells, where the DNA is released and it migrates to the nucleus (Fig. 9). The DNA transfection can be transient or stable. In the case of stable transfection, the exogenous

DNA is integrated into cell genome and it remains also in daughter cells. In transient transfection, the transfected genetic material is only transiently expressed. Since the DNA introduced in the transient transfection process is not integrated into the nuclear genome, the foreign DNA will be diluted through mitosis or degraded.

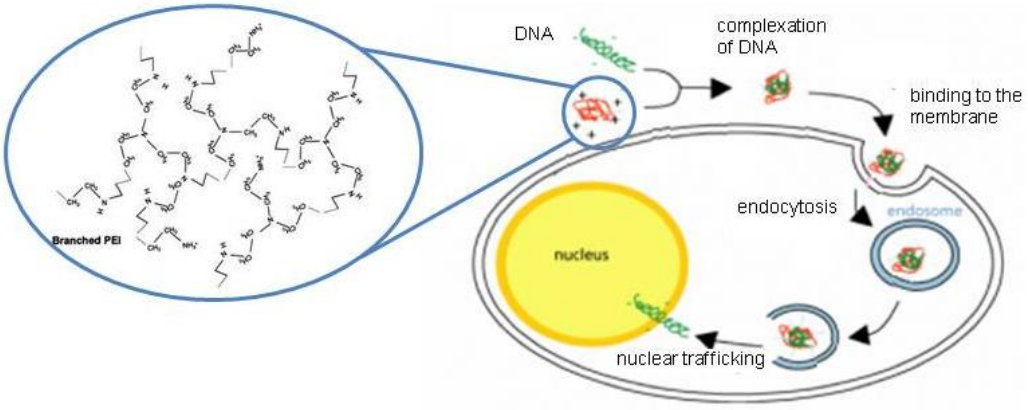


Figure 9 Schematic representation of PEI-mediated DNA transfection of mammalian cells. Adapted from http://www.nano-lifescience.com.

- **DNA transient transfection**

To perform transient transfection, optimal conditions are when adherent cells reach about 90% confluency and a mixture of DNA and PEI in 1:2 ratio (w/w) is used [107]. For a standard 75 cm^2 flask, 25 µg of plasmid DNA are required. The DNA is added to 2.5 ml serum-free medium and is followed by the addition of 50 µg PEI and gently mixed. During complex formation, media from the plates to be transfected should be changed, lowering the serum concentration to 2%. Finally, the DNA–PEI complex is added to the dish, which is then briefly rotated to allow mixing and the cells are placed in the incubator. For co-expression of two proteins, the amount of DNAs can be varied to obtain different expression levels for each protein (the expression level is not linear with the DNA amount).

- **DNA stable transfection**

To produce stably transfected cell lines, parental 293T cells were transfected with the pURD vector containing the HAH1 cDNA together with the PhiC31 integrase gene [112]. Once they reached ~100% confluence, stable cell line selection was performed by

treating the cells with 2 µg/ml puromycin for 10 days, replacing the growth medium every 3-4 days to remove dead cells, until cell clumps (single colonies) were visible. Single colonies were manually picked from the dish and individually seeded 96-well plate wells. During the following days, growing colonies were selected by manual inspection and seeded to 12-well plates. Lines were selected based on the growth rate, and were expanded to 6-well plates and further to T25 flasks. The HAH1 expression levels were measured by Western Blot analysis.

3.4 Gene silencing in mammalian cells

RNA interference (RNAi) is a biological process of silencing gene expression in a range of organisms (reviewed in [113]). The silencing of a gene is a consequence of degradation of double-stranded RNA into short RNAs that activate ribonucleases to target homologous mRNA. The resulting phenotypes either are identical to those of genetic null mutants or resemble an allelic series of mutants. Extensive genetic and biochemical analysis revealed a two-step mechanism of RNAi-induced gene silencing. The first step involves degradation of dsRNA into small interfering RNAs (siRNAs), 21 to 25 nucleotides long, by Dicer protein. In the second step, the siRNAs join an RNase complex, RISC (RNA-induced silencing complex), which acts on the cognate mRNA and degrades it. RNAi has been associated with regulatory processes such as antiviral defense mechanisms and gene regulation.

RNAi could be induced in cultured cells with exogenous RNA by directly transfection with siRNA or with a plasmid that encodes for a short hairpin RNA (shRNA). Because siRNA is composed of RNA and it is inherently fragile, it is 99% degraded after 48 hours while shRNA is expressed for up 96 hours. This means also that a higher dose of siRNA is required respect to the plasmid that generates shRNA. For these reasons, pSilencer 2.1-U6 hygro siRNA Expression vector (Life Technologies) was used (Fig. 10).

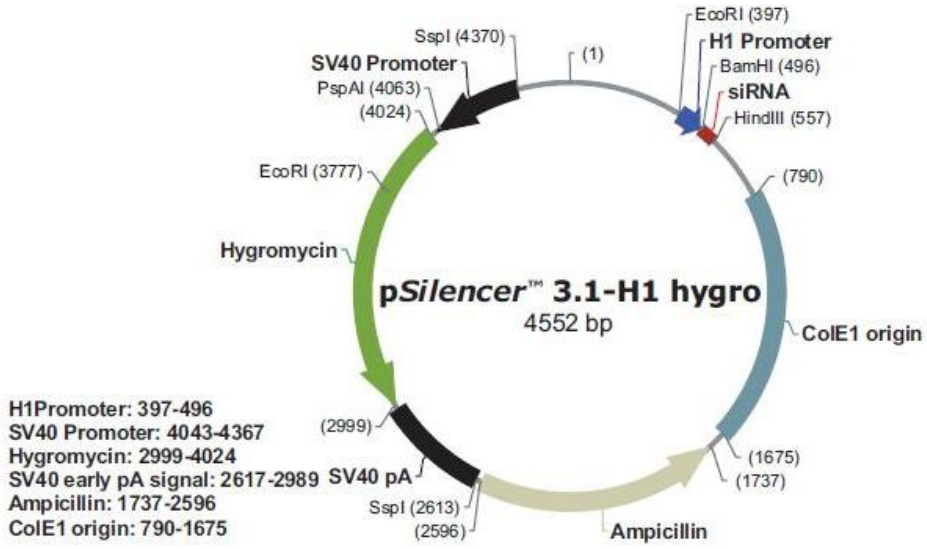


Figure 10 pSilecer 3.1-H1 hygro vector map.

The pSilencer 2.1-U6 hygro vector employs a human U6 RNA polymerase III (pol III) promoter that generates large amounts of small RNA across a variety of cell types [114, 115]. The terminator consists of a short stretch of four uridines; this is compatible with the original siRNA design that terminates with a two uridine 3’ overhang. An efficient approach is to express a single RNA that is a 21-mer hairpin sequences specific to the mRNA target, a loop sequence separating the two complementary domains, and a polyuridine (Fig. 11).

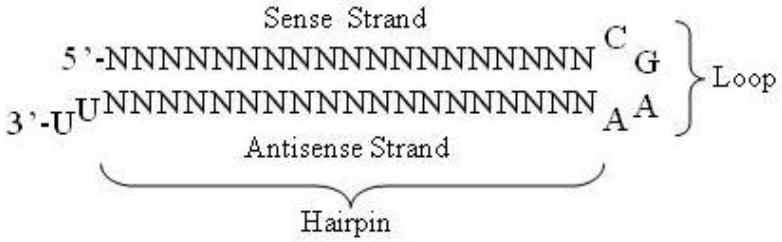


Figure 11 Hairpin siRNA

The 21-mer template oligonucleotide is calculated entering siRNA target sequence into the web-based insert design tool (www.ambion.com/techlib/misc/psilencer_converter.html) and the loop sequence is 5’-CGAA-3’. Near the end of the hairpin siRNA template is a 6 nucleotide poly(T)-tract recognized as a termination signal by RNA pol III that will terminate RNA synthesis. For each target gene, two complementary oligonucleotides are designed with 5'

noncomplementary ends that form the BamH I and Hind III restriction site overhangs upon annealing. The pSilencer siRNA Expression Vectors was linearized with both BamH I and Hind III to facilitate directional cloning.

Three sequences of shRNA were designed to perform silencing of HAH1: one targets the HAH1 open reading frame (sense strand 5'-ACGAGTTCTCTGTGGACATGA-3'), while other two target the unique UTR elements WPRE (5'-GCAATCAACCTCTGGATTACA-3') and BGH (5'-GGAAGACAATAGCAGGCATGC-3') that are present in the pURD vector. This three shRNA were produced into the cells by transient transfection of pSilencer vectors. Transcription and expression of the HAH1 gene were monitored respectively by RT-PCR and Western Blot.

3.5 Protein expression in mammalian cells

Protein expression in HEK293T cells was carried out in 48 hr and the cells were incubated under controlled 5% CO_2 atmosphere at 37 °C. For unlabeled cell samples conventional DMEM medium was used; BioExpress® 6000 (CIL) was used for uniform ^{15}N labeling.

To perform sequential and selective labeling, the expression was carried out in 30 hr: 14 hr in unlabeled medium and further 16 hr in U-^{15}N medium.

Metal ions can be added to expression medium when they are necessary for maturation or activation of overexpressed protein. When hSOD1 was expressed, Zn(II) was supplemented as $ZnSO_4$ in the medium to a final concentration of 10 µM immediately after the transfection. To ensure complete activation of Trx1, sodium selenite was supplemented to a final concentration of 100 nM both in culture medium starting 24 hr before transfection and in expression medium.

3.6 In-cell NMR sample preparation

To prepare the NMR sample, HEK293T cells from a 75 cm^2 culture flask were detached by incubation with trypsin-EDTA 0.05%, and washed in phosphate buffered saline (PBS) through gentle centrifugation (800 x g). Cells were re-suspended in 180 µl of NMR buffer (DMEM supplemented with 70 mM HEPES, 65 mM Glucose, 20% (v/v) D_2O) and pipetted in a 3 mm Shigemi NMR tube. Cells were sedimented to the bottom with the help of a manual centrifugation device.

After the NMR experiments, cells were re-suspended in their supernatant and pelleted again out of the tube. The supernatant was checked by NMR for protein leakage. Cells were lysed by the freeze-thaw method in 150 µl of PBS buffer supplemented with 0.5 mM EDTA and Pefabloc. After centrifugation at 16,000 x g for 60 min at 4 °C, the cleared cell extract was recovered for NMR and SDS-PAGE analysis.

3.7 NMR experiments

During in-cell NMR experiments, it is fundamental to limit cell death and disruption. A way to overcome this problem is to reduce acquisition time using fast recycling NMR experiments.

In-cell NMR experiments (1D ^1H and 2D ^1H-^{15}N SOFAST-HMQC) were acquired at 305 K and 308 K using a 950 MHz Bruker Avance III spectrometer equipped with a CP TCI CryoProbe. The use of SOFAST sequence allows having a total acquisition time ranged from 1 to 2 hr.

- 1D ^1H NMR spectra

^1H NMR spectra were used to check protein overexpression over other cellular components. In fact, a set of protein signals falls in the aliphatic region from -1 to 1 ppm of 1D spectrum that is free from cellular background. Spectrum of cellular background was obtained transfecting the cells with empty vector. In the case of Mia40, the $^1H\gamma$ peak of Ile 53 at -0.7 ppm is a marker of the folded conformation of Mia40. The

effective concentration of HAH1 in a NMR sample of stable transfected cells was estimated by 1D ^{1}H NMR by comparing the signal intensity between 0.1 and -0.2 ppm in the cell lysate spectrum with an *in vitro* sample of recombinantly expressed HAH1 at known concentration.

- 2D ^{1}H-^{15}N SOFAST-HMQC

Fast data acquisition in SOFAST-HMQC is realized by using very short inter-scan delays. SOFAST-HMQC combines the advantages of a small number of radio-frequency pulses, Ernst-angle excitation, and longitudinal relaxation optimization to obtain an increased signal to noise ratio for high repetition rates of the experiment.

The main features of SOFAST-HMQC are the following: (i) the HMQC-type H–X transfer steps require only few radio-frequency pulses, which limit signal loss due to B1-field inhomogeneities, and pulse imperfections. A reduced number of radio-frequency pulses become important if the experiment is performed on a cryogenic probe, where B1-field inhomogeneities are more pronounced. (ii) The band-selective ^{1}H pulses reduce the effective spin-lattice relaxation times (T1) of the observed ^{1}H spins. The presence of a large number of non-perturbed ^{1}H spins, interacting with the observed ^{1}H via dipolar interactions (NOE effect), significantly reduces longitudinal relaxation times and the equilibrium spin polarization is more quickly restored. (iii) The adjustable flip angle of the ^{1}H excitation pulse allows further enhancement of the available steady-state magnetization for a given recycle delay The performance of SOFAST-HMQC critically depends on the choice of the pulse shapes for the band-selective excitation and refocusing pulses on the ^{1}H channel [116].

3.8 Assessment of cell viability

Cell viability before and after each in-cell NMR experiment was estimated by staining with trypan blue. The trypan blue dye selectively stains damaged or dead cells, as is able to diffuse through damaged cell membranes, and accumulates in the cytoplasm. Conversely, the plasma membrane of live cells is not permeable to the dye. The aliquots of cells are diluted and incubated with trypan blue. They are laid on a Burker chamber

and counted under phase-contrast microscope. On average, cell viability always remained above 90% after NMR experiment.

3.9 Western blotting

The Western Blot is an analytical technique used to identify specific proteins in a sample of cells or extract. The mixture of proteins was separated through gel electrophoresis. The proteins were then immobilized on a nitrocellulose membrane following electrophoretic transfer from the gel. Non-protein binding areas on the membrane were blocked with non-fat milk or BSA to prevent any non-specific binding of antibodies. The membrane was first incubated with a primary antibody that specifically binds to the target protein and then with a secondary antibody conjugated with a reporter used for the detection. Chemiluminescent detection method was used. It is based on secondary antibody conjugated with horseradish poroxidase (HRP) enzyme. On the addition of a peroxide-luminol mixture, the HRP enzyme catalyses the oxidation of luminol, which is accompanied by blue-green luminescence. The light was detected by CCD cameras that capture a digital image of the western blot. The detected signal is proportional to the amount of protein and a quantitative analysis was performed by densitometry using ImageJ program.

3.10 Proteins purification

- hMia40

Unlabeled and U-^{15}N labeled sample of hMia40 for *in vitro* interaction with hMia40 and for Western Blot analysis was produced as follows: a cell culture of *E. coli* BL21(DE3) Gold competent cells transformed with pDEST-His-MBP plasmid containing the hMia40 gene, was grown overnight at 37 °C in 40 ml LB, harvested and re-suspended in 2 L of LB or ^{15}N-labeled M9 medium. Protein expression was induced with 0.7 mM IPTG for 16 hr at 25 °C. hMia40 was purified by affinity chromatography using a nickel chelating HisTrap (GE Healthcare) column; the protein was eluted in 500 mM imidazole. hMia40 was then cut from MBP by incubation with AcTEV protease O/N at

25 °C, and separated from His-tagged MBP with a second purification step using Ni-chelating resin. Finally, two size-exclusion chromatography steps were performed to obtain pure hMia40. The sample buffer was then exchanged with 50 mM potassium phosphate, 0.5 mM EDTA, 1 mM DTT, pH=7 to perform NMR experiment. Reduction of the structural disulfide bonds of hMia40 was performed by incubating hMia40 at 95 °C 10 min with 10 mM DTT in phosphate buffer.

- hGlutaredoxin 1

hGlutaredoxin 1 for *in vitro* interaction with hMia40 and for Western Blot analysis was produced as follows: a pTH34 vector containing the human Grx1 gene (N-term fused with His-tag and TEV recognition site) was transformed in *E. coli* BL21(DE3) Gold competent cells. A cell culture was grown overnight at 37 °C in 10 ml of LB, harvested and re-suspended in 1 L of minimal medium M9. Cells were grown at 37 °C until O.D. 0.6 and protein expression was induced with 0.5 mM IPTG for 16 h at 25 °C. hGrx1 was purified by affinity chromatography using a nickel chelating HisTrap (GE Healthcare) column. The fractions containing pure hGrx1 (checked by SDS-PAGE) were digested with AcTEV protease (Invitrogen) O/N at 25 °C. The protein was separated from the affinity tag in a HisTrap column. The sample buffer was then exchanged with 50 mM potassium phosphate, 0.5 mM EDTA, pH=7.

- hThioredoxin 1

Pure hThioredoxin 1 for Western Blot analysis was prepared from HEK293T cells overexpressing Trx1. Cells were lysed by the freeze-thaw method in 20 mM potassium phosphate buffer (pH=7). After centrifugation at 16,000 x g for 60 min at 4 °C, the cleared cell extract was incubated at 70 °C for 4 min to remove contaminating proteins. After centrifugation (30 min at 16,000 g), Trx1 was purified by anionic exchange chromatography using a HiTrap™ DEAE FF (Amersham Biosciences) column applying a linear gradient of potassium phosphate buffer (10 to 100 mM, pH=7). The fractions containing pure hTrx1 were identified by SDS-PAGE.

- hHAH1

Pure hHAH1 was produced as follows: a pET21a vector containing the human HAH1 gene (C-term fused with Factor Xa recognition site and His-tag) was transformed in *E. coli* BL21(DE3) Gold competent cells. A cell culture was grown overnight at 37 °C in 20 ml LB, harvested and re-suspended in 1 L of LB. Protein expression was induced using 1 mM IPTG when O.D. 0.6 and the cells were harvested 4 hr thereafter. After cell lysis, a first purification step was performed using a HiTrap chelating (GE Healthcare) column previously charged with Zn(II). The protein was eluted with the 20 mM Na_2HPO_4, 0.5 M NaCl, 5 mM Imidazole, 20 mM EDTA, pH=8. After buffer exchange and digestion with Factor Xa protease (New England Biolabs) for 24 hr at 25 °C, the protein was separated from the affinity tag in a HisTrap column charged with Zn(II). The fractions containing pure hHAH1 were identified by SDS-PAGE.

- NS5A-D3

Pure NS5A-D3 for *in vitro* phosphorylation by CK2 was prepared as follows: a pET16b+ containing the gene of NS5A-D3 was transformed in *E. coli* BL21(DE3) Gold competent cells. A cell culture was grown overnight at 37 °C in 40 ml LB, harvested and re-suspended in 2 L of ^{15}N or $^{13}C/^{15}N$ M9 medium. The cells were grown at 37 °C to an optical density 0.6. Protein expression was induced with 0.5 mM IPTG for 16 hr at 18 °C. NS5A-D3 was purified by affinity chromatography using a nickel-chelating HisTrap (GE Healthcare) column; the protein was recovered in elution buffer containing 300 mM imidazole. After buffer exchange and digestion with Factor Xa protease (New England Biolabs) for 24 hr at 25 °C, the protein was separated from the affinity tag in a HisTrap column. The buffer was exchanged with phosphate buffer saline (PBS, Abcam), 0.5 mM EDTA, 1 mM DTT, pH=7, and the protein was concentrated for NMR analysis.

3.11 *In vitro* phosphorylation of NS5A-D3 by CK2

The *in vitro* phosphorylation reaction of NS5A-D3 was performed on a sample containing 100 µM of the protein (either ^{15}N-labeled or ^{13}C,^{15}N-labeled) with 1000 U of CK2 (New England Biolabs) at 30 °C in 20 mM Tris-HCl, 50 mM KCl, 10 mM $MgCl_2$, 1 mM DTT, 2 mM ATP, pH=7.3. The samples of ^{13}C,^{15}N-labeled phosphorylated NS5A-D3 at different reaction times were prepared by stopping the reaction by inactivating CK2 at 70 °C for 20 min.

4. Conclusions

During my PhD studies I applied the in-cell NMR approach to investigate redox-dependent protein folding processes in living cells. In particular, I characterized the folding and redox state of the mitochondrial protein Mia40 in the cytoplasm of living human cells. We found that the folded form of Mia40, which cannot be imported into mitochondria, is accumulated in the cytoplasm, suggesting that there is the need for a mechanism to keep cytoplasmic Mia40 reduced until it reaches the outer mitochondrial membrane. By increasing the expression levels of two cytoplasmic thiol-disulfide oxidoreductases, Grx1 and Trx1, we showed that Mia40 was mostly reduced, indicating that the folding of Mia40 is controlled by these cytoplasmic redox-regulation systems, Grx1 having a stronger effect than Trx1. Conversely, this effect was not observed in *in vitro* experiments suggesting that the effect of Grx1 on oxidation state of intracellular Mia40 is mediated by other components of the cytoplasm. These results show the relevance of NMR studies performed in living cells, which provide a way to describe cellular physiological processes at atomic resolution, and to understand how the proteins involved are affected and regulated by such processes.

In-cell NMR is an ideal approach for the characterization of protein-protein interactions. In order to extract meaningful information from heteronuclear in-cell NMR spectra, only one protein needs to be detectable and/or labeled. Currently, methods that allow expressing two or more proteins at different times and with different labeling have been developed only in *E. coli* cells. To extend this approach to mammalian cells, we designed and optimized a new protocol combining stable and transient transfection with gene silencing, to allow the sequential expression of two proteins, with only one being selectively labeled. With this method, we provide a useful tool for studying interactions between two proteins directly inside human cells. In principle, this approach could be adapted for the characterization of more than two proteins in combination with existing protein-delivery approaches.

Finally, I contributed to characterize *in vitro* the CK2-dependent phosphorylation of the domain 3 of the viral protein NS5A by time-dependent NMR. We identified unambiguously four serines as phosphorylation sites, some of which had not been predicted previously. The phosphorylation reactions occurred independently of each

other and with different rates. Domain 3 of NS5A is fundamental for the interaction with core protein, which is the first step of virion assembly, therefore a better understanding of the molecular mechanisms underlying such process is critical for future development of new antiviral therapies. Further investigation should focus on characterizing the phosphorylation pattern of NS5A-D3 directly in hepatic cells or in cellular lysates. Such studies could reveal how the interplay between CK2 and other cytoplasmic kinases affect NS5A-D3 phosphorylation and ultimately regulate virion assembly.

5. Reference list

[1] Ellis R.J., Macromolecular crowding: obvious but underappreciated, *Trends Biochem. Sci.*, **2001**, *26*, 597-604.

[2] Gershenson A., Gierasch L.M., Protein folding in the cell: challenges and progress. *Curr. Opin. Struct. Biol.*, **2011**, *21*, 32-41.

[3] Elcock A.H., Models of macromolecular crowding effects and the need for quantitative comparisons with experiment. *Curr. Opin. Struct. Biol*., **2010**, *20*, 196-206.

[4] Zhou H.X., Rivas G., Minton A.P., Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences. *Annu. Rev. Biophys*., **2008**, *37*, 375-397.

[5] Serber Z., Dötsch V., In-cell NMR spectroscopy. *Biochemistry*, **2001**, *40*, 14317-14323.

[6] Serber Z., Ledwidge R., Miller S.M., Dötsch V., Evaluation of parameters critical to observing proteins inside living Escherichia coli by in-cell NMR spectroscopy. *J. Am. Chem. Soc.*, **2001**, *123*, 8895-8901.

[7] Selenko P., Serber Z., Gadea B., Ruderman J., Wagner G., Quantitative NMR analysis of the protein G B1 domain in *Xenopus laevis* egg extracts and intact oocytes. *Proc. Natl. Acad. Sci. USA*, **2006**, *103*, 11904-11909.

[8] Ogino S., Kubo S., Umemoto R., Huang S., Nishida N., Shimada I., Observation of NMR signals from proteins introduced into living mammalian cells by reversible membrane permeabilization using a pore-forming toxin, streptolysin O. *J. Am. Chem. Soc.*, **2009**, *131*, 10834-10835.

[9] Inomata K., Ohno A., Tochio H., Isogai S., Tenno T., Nakase I., Takeuchi T., Futaki S., Ito Y., Hiroaki H., Shirakawa M., High-resolution multi-dimensional NMR spectroscopy of proteins in human cells. *Nature*, **2009**, *458*, 106-109.

[10] Hamatsu J., O'Donovan D., Tanaka T., Shirai T., Hourai Y., Mikawa T., Ikeya T., Mishima M., Boucher W., Smith B.O., Laue E.D., Shirakawa M., Ito Y.,

High-resolution heteronuclear multidimensional NMR of proteins in living insect cells using a baculovirus protein expression system. *J. Am. Chem. Soc*., **2013**, *135*, 1688-1691.

[11] Bertrand K., Reverdatto S., Burz D.S., Zitomer R., Shekhtman A., Structure of proteins in eukaryotic compartments. *J. Am. Chem. Soc.*, **2012**, *134*, 12798-12806.

[12] Mullineaux C.W., Nenninger A., Ray N., Robinson C. Diffusion of green fluorescent protein in three cell environments in Escherichia coli. *J Bacteriol*., **2006**, *188*, 3442-8.

[13] Serber Z., Corsini L., Durst F., Dötsch V., In-cell NMR spectroscopy. *Methods Enzymol*., **2005**, *394*, 17-41.

[14] Ye Y., Liu X., Zhang Z., Wu Q., Jiang B., Jiang L., Zhang X., Liu M., Pielak G.J., Li C., (19) F NMR spectroscopy as a probe of cytoplasmic viscosity and weak protein interactions in living cells. *Chem. Eur. J*., **2013**, *19*, 12705-12710.

[15] Wang Q., Zhuravleva A., Gierasch L.M., Exploring weak, transient protein--protein interactions in crowded in vivo environments by in-cell nuclear magnetic resonance spectroscopy. *Biochemistry*, **2011**, *50*, 9225-9236.

[16] Luh L. M., Hänsel R., Löhr F., Kirchner D.K., Krauskopf K., Pitzius S., Schäfer B., Tufar P., Corbeski I., Güntert P., Dötsch V., Molecular crowding drives active Pin1 into nonspecific complexes with endogenous proteins prior to substrate recognition. *J. Am. Chem. Soc*., **2013**, *135*, 13796-13803.

[17] Kosol S., Contreras-Martos S., Cedeño C., Tompa P., Structural characterization of intrinsically disordered proteins by NMR spectroscopy. *Molecules*, **2013**, *18*, 10802-10828.

[18] Amata I., Maffei M., Igea A., Gay M., Vilaseca M., Nebreda A.R., Pons M., Multi-phosphorylation of the intrinsically disordered unique domain of c-Src studied by in-cell and real-time NMR spectroscopy. *ChemBioChem*, **2013**, *14*, 1820-1827.

[19] Cino E.A., Karttunen M., Choy W.Y., Effects of molecular crowding on the dynamics of intrinsically disordered proteins. *PLoS One*, **2012**, *7*, e49876.

[20] Li C., Charlton L.M., Lakkavaram A., Seagle C., Wang G., Young G.B., Macdonald J.M., Pielak G.J., Differential dynamical effects of macromolecular crowding on an intrinsically disordered protein and a globular protein: Implications for in-cell NMR spectroscopy. *J. Am. Chem. Soc.*, **2008**, *130*, 6310-6311.

[21] Wang Y., Li C., Pielak G.J., Effects of proteins on protein diffusion. *J. Am. Chem. Soc.*, **2010**, *132*, 9392-9397.

[22] Sarkar M., Smith A.E., Pielak G.J., Impact of reconstituted cytosol on protein stability. *Proc. Natl. Acad. Sci. USA*, **2013**, *110*, 19342-19347.

[23] Crowley P.B., Chow E., Papkovskaia T., Protein interactions in the Escherichia coli cytosol: an impediment to in-cell NMR spectroscopy. *ChemBioChem*, **2011**, *12*, 1043-1048.

[24] Serber Z., Straub W., Corsini L., Nomura A.M., Shimba N., Craik C.S., Ortiz de Montellano C.S., Dötsch V., Methyl groups as probes for proteins and complexes in in-cell NMR experiments. *J. Am. Chem. Soc.*, **2004**, *126*, 7119-7125.

[25] Hänsel R., Foldynová-Trantírková S., Löhr F., Buck J., Bongartz E., Bamberg E., Schwalbe H., Dötsch V., Trantírek L., Evaluation of parameters critical for observing nucleic acids inside living Xenopus laevis oocytes by in-cell NMR spectroscopy. *J. Am. Chem. Soc.*, **2009**, *131*, 15761-15768.

[26] Hänsel R., Foldynová-Trantírková S., Dötsch V., Trantírek L., Investigation of quadruplex structure under physiological conditions using in-cell NMR. *Top. Curr. Chem.*, **2013**, *330*, 47-65.

[27] Dedmon M.M., Patel C.N., Young G.B., Pielak G.J., FlgM gains structure in living cells. *PNAS*, **2002**, *99*, 12681-12684.

[28] Egan W., Barile M., Rottem S., 31P-NMR studies of Mycoplasma gallisepticum cells using a continuous perfusion technique. *FEBS Lett.*, **1986**, *204*, 373-376.

[29] Barnes C.O., Monteith, W.B., Pielak G.J., Internal and global protein motion assessed with a fusion construct and in-cell NMR spectroscopy. *Chembiochem*., **2011**, *12*, 390-391.

[30] McNulty B.C., Young G.B., Pielak G.J., Macromolecular crowding in the Escherichia coli periplasm maintains alpha-synuclein disorder. *J. Mol. Biol*., **2006**, *355*, 893-897.

[31] Sakakibara D., Sasaki A., Ikeya T., Hamatsu J., Hanashima T., Mishima M., Yoshimasu M., Hayashi N., Mikawa T., Wälchli M., Smith B.O., Shirakawa M., Güntert P., Ito Y., Protein structure determination in living cells by in-cell NMR spectroscopy. *Nature*, **2009**, *458*, 102-105.

[32] Schlesinger A.P., Wang Y., Tadeo X., Millet O., Pielak G.J., Macromolecular crowding fails to fold a globular protein in cells. *J. Am. Chem. Soc*., **2011**, *133*, 8082-8085.

[33] Binolfi A., Theillet F.X., Selenko P., Bacterial in-cell NMR of human alpha-synuclein: a disordered monomer by nature? *Biochem. Soc. Trans*., **2012**, *40*, 950-954.

[34] Banci L., Barbieri L., Bertini I., Cantini F., Luchinat E., In-cell NMR in *E. coli* to monitor maturation steps of hSOD1. *PLoS One*, **2011**, *6(8)*.

[35] Burz D.S., Dutta K., Cowburn D., Shekhtman A., Mapping structural interactions using in-cell NMR spectroscopy (STINT-NMR). *Nat. Methods*, **2006**, *3*, 91-93.

[36] Burz D.S., Shekhtman A., In-cell biochemistry using NMR spectroscopy. *PLoS One*, **2008**, *3*, e2571.

[37] Xie J., Thapa R., Reverdatto S., Burz D.S., Shekhtman A., Screening of small molecule interactor library by using in-cell NMR spectroscopy (SMILI-NMR). *J. Med. Chem*., **2009**, *52*, 3516-3522.

[38] Augustus A.M., Reardon P.N., Spicer L.D., MetJ repressor interactions with DNA probed by in-cell NMR. *Proc. Natl. Acad. Sci. USA*, **2009**, *106*, 5065-5069.

[39] Selenko P., Frueh D.P., Elsaesser S.J., Haas W., Gygi S.P., Wagner G., In situ observation of protein phosphorylation by high-resolution NMR spectroscopy. *Nat. Struct. Mol. Biol*., **2008**, *15*, 321-329.

[40] Bodart J.F., Wieruszeski J.M., Amniai L., Leroy A., Landrieu I., Rousseau-Lescuyer A., Vilain J.P., Lippens G., NMR observation of Tau in Xenopus oocytes. *J. Magn. Reson*., **2008**, *192*, 252-257.

[41] Hänsel R., Löhr F., Foldynová-Trantírková S., Bamberg E., Trantírek L., Dötsch V., The parallel G-quadruplex structure of vertebrate telomeric repeat sequences is not the preferred folding topology under physiological conditions. *Nucleic Acids Res*., **2011**, *39*, 5768-5775.

[42] Danielsson J., Inomata K., Murayama S., Tochio H., Lang L., Shirakawa M., Oliveberg M., Pruning the ALS-associated protein SOD1 for in-cell NMR. *J. Am. Chem. Soc*., **2013**, *135*, 10266–69.

[43] Deshayes S., Morris M., Heitz F., Divita G., Delivery of proteins and nucleic acids using a non-covalent peptide-based strategy. *Adv Drug Deliv Rev.*, **2008**, *60*, 537-47.

[44] Bekei B., In-cell NMR Spectroscopy in Mammalian Cells. Ph.D. Dissertation, Freie Universität Berlin, Berlin, **2013**.

[45] Majumder S., Xue J., DeMott C.M., Reverdatto S., Burz D.S., Shekhtman A., Probing Protein Quinary Interactions by In-Cell Nuclear Magnetic Resonance Spectroscopy. *Biochemistry*, **2015**, *54*, 2727-2738.

[46] Banci L., Barbieri L., Bertini I., Luchinat E., Secci E., Zhao Y., Aricescu A.R., Atomic-resolution monitoring of protein maturation in live human cells by NMR. *Nat. Chem. Biol*., **2013**, *9*, 297-299.

[47] Freedberg D.I., Selenko P., Live cell NMR. *Annu Rev Biophys*., **2014**, *43*, 171-92.

[48] Hofmann S., Rothbauer U., Mühlenbein N., Baiker K., Hell K., Bauer M.F., Functional and Mutational Characterization of Human MIA40 Acting During

Import into the Mitochondrial Intermembrane Space. *J. Mol. Biol.*, **2005**, *353*, 517-528.

[49] Chacinska A., Pfannschmidt S., Wiedemann N., Kozjak V., Sanjua'n Szklarz L.K., Schulze-Specking A., Truscott K.N., Guiard B., Meisinger C., Pfanner N., Essential role of Mia40 in import and assembly of mitochondrial intermembrane space proteins. *EMBO J.*, **2004**, *23*, 3735-3746.

[50] Herrmann J.M., Neupert W., Protein transport into mitochondria. *Curr Opin Microbiol*, **2000**, *3*, 210-214.

[51] Jensen R.E., Johnson A.E., Opening the door to mitochondrial protein import. *Nat Struct Biol*, **2001**, *8*, 1008–1010.

[52] Sickmann A., Reinders J., Wagner Y., Joppich C., Zahedi R., Meyer H.E., Schönfisch B., Perschil I., Chacinska A., Guiard B., Rehling P., Pfanner N., Meisinger C., The proteome of Saccharomyces cerevisiae mitochondria. *Proc Natl Acad Sci USA*, **2003**, *100*, 13207-13212.

[53] Stojanovski D., Müller J.M., Milenkovic D., Guiard B., Pfanner N., Chacinska A., The MIA system for protein import into the mitochondrial intermembrane space. *Biochim. Biophys. Acta*, **2008**, *1783*, 610-617.

[54] Cobine P.A., Pierrel F., Winge D.R., Copper trafficking to the mitochondrion and assembly of copper metalloenzymes. *Biochim. Biophys. Acta*, **2006**, *1763*, 759-772.

[55] Banci L., Bertini I., Ciofi-Baffoni S., Hadjiloi T., Martinelli M., Palumaa P., Mitochondrial copper(I) transfer from Cox17 to Sco1 is coupled to electron transfer. *Proc. Natl. Acad. Sci. USA*, **2008**, *105*, 6803-6808.

[56] Banci L., Bertini I., Cavallaro G., Rosato A., The functions of Sco proteins from genomebased analysis. *J. Proteome Res.*, **2007**, *6*, 1568-1579.

[57] Bauer M.F., Hofmann S., Neupert W., Brunner M., Protein translocation into mitochondria: the role of TIM complexes. *Trends Cell Biol.*, **2000**, *10*, 25-31.

[58] Vial S., Lu H., Allen S., Savory P., Thornton D., Sheehan J., Tokatlidis K., Assembly of Tim9 and Tim10 into a functional chaperone. *J. Biol. Chem.*, **2002**, 277, 36100-36108.

[59] Banci L., Bertini I., Cefaro C., Ciofi-Baffoni S., Gallo A., Martinelli M., Sideris D.P., Katrakili N., Tokatlidis K., MIA40 is an oxidoreductase that catalyzes oxidative protein folding in mitochondria. *Nat Struct Mol Biol.*, **2009**, *16*, 198-206.

[60] Banci L., Bertini I., Cefaro C., Cenacchi L., Ciofi-Baffoni S., Felli I.C., Gallo A., Gonnelli L., Luchinat E., Sideris D., Tokatlidis K., Molecular chaperone function of Mia40 triggers consecutive induced folding steps of the substrate in mitochondrial protein import. *Proc Natl Acad Sci USA.*, **2010**, *107*, 20190-5.

[61] Grumbt B., Stroobant V., Terziyska N., Israel L., Hell K., Functional Characterization of Mia40p, the Central Component of the Disulfide Relay System of the Mitochondrial Intermembrane Space. *J. Biol. Chem.*, **2007**, *282*, 37461-37470.

[62] Chacinska A., Guiard B., Müller J.M., Schulze-Specking A., Gabriel K., Kutik S., Pfanner N., Mitochondrial Biogenesis, Switching the Sorting Pathway of the Intermembrane Space Receptor Mia40. *J. Biol. Chem.*, **2008**, *283*, 29723-29729.

[63] Chatzi A., Sideris D.P., Katrakili N., Pozidis C., Tokatlidis K., Biogenesis of yeast Mia40-uncoupling folding from import and atypical recognition features, *FEBS Journal*, **2013**, *280*, 4960-4969.

[64] Sideris D.P., Petrakis N., Katrakili N., Mikropoulou D., Angelo Gallo A., Ciofi-Baffoni S., Banci L., Bertini I., Tokatlidis K., A novel intermembrane space-targeting signal docks cysteines onto Mia40 during mitochondrial oxidative folding. *J. Cell. Biol.*, **2009**, *187*, 1007-1022.

[65] Banci L., Bertini I., Calderone V., Cefaro C., Ciofi-Baffoni S., Gallo A., Kallergi E., Lionaki E., Pozidis C., Tokatlidis K., Molecular recognition and substrate mimicry drive the electron-transfer process between MIA40 and ALR. *Proc. Natl. Acad. Sci. U.S.A.*, **2011**, *108*, 4811.

[66] Banci L., Bertini I., Calderone V., Cefaro C., Ciofi-Baffoni S., Gallo A., Tokatlidis K., An electron-transfer path through an extended disulfide relay system: the case of the redox protein ALR. *J. Am. Chem. Soc*., **2012**, *134*, 1442-1445.

[67] Mesecke N., Terziyska N., Kozany C., Baumann F., Neupert W., Hell K., Herrmann J.M., A disulfide relay system in the intermembrane space of mitochondria that mediates protein import. *Cell*, **2005**, *121*, 1059-1069.

[68] Stojanovski D., Milenkovic D., Müller J.M., Gabriel K., Schulze-Specking A., Baker M.J., Ryan M.T., Guiard B., Pfanner N., Chacinska A., Mitochondrial protein import: precursor oxidation in a ternary complex with disulfide carrier and sulfhydryl oxidase. *J Cell Biol.*, **2008**, *183*, 195-202.

[69] Marukian S., Andrus L., Sheahan T.P., Jones C.T., Charles E.D., Ploss A., Rice C.M., Dustin L.B., Hepatitis C virus induces interferon-λ and interferon-stimulated genes in primary liver cultures. *Hepatology*, **2011**, *54*, 1913-23.

[70] Maasoumy B., Wedemeyer H., Natural history of acute and chronic hepatitis C. *Best Pract. Res. Clin. Gastroenterol*., **2012**, *26*, 401-412.

[71] Robertson B., Myers G., Howard C., Brettin T., Bukh J., Gaschen B., Gojobori T., Maertens G., Mizokami M., Nainan O., Netesov S., Nishioka K., Shin i T., Simmonds P., Smith D., Stuyver L., Weiner A., Classification, nomenclature, and database development for hepatitis C virus (HCV) and related viruses: proposals for standardization. *International Committee on Virus Taxonomy. Arch. Virol*, **1998**, *143*, 2493-2503.

[72] Davis G.L. Hepatitis C virus genotypes and quasispecies. *Am. J. Med*., **1999**, *107*, 21S-26S.

[73] Ghany M.G., Strader D.B., Thomas D.L., Seeff L.B., Diagnosis, management, and treatment of hepatitis C: an update. *Hepatology*, **2009**, *49*, 1335-1374.

[74] Tan S.L., Pause A., Shi Y., Sonenberg N., Hepatitis C therapeutics: current status and emerging strategies. *Nature reviews. Drug Discovery*, **2002**, *1*, 867-881.

[75] Ghany M.G., Nelson D.R., Strader D.B., Thomas D.L., Seeff L.B., An update on treatment of genotype 1 chronic hepatitis C virus infection: 2011 practice guideline by the American Association for the Study of Liver Diseases. *Hepatology*, **2011**, *54*, 1433-1444.

[76] Choo Q.L., Richman K.H., Han J.H., Berger K., Lee C., Dong C., Gallegos C., Coit D., Medina-Selby R., Barr P.J., et al., Genetic organization and diversity of the hepatitis C virus. *Proc. Natl. Acad. Sci. USA*, **1991**, *88*, 2451-2455.

[77] Bartenschlager R., Cosset F.L., Lohmann V., Hepatitis C virus replication cycle. *J. Hepatol*., **2010**, *53*, 583-585.

[78] Moradpour D., Penin F., Rice C.M., Replication of hepatitis C virus. *Nat. Rev. Microbiol*., **2007**, *5*, 453-463.

[79] Ross-Thriepland D., Harris M., Hepatitis C virus NS5A: enigmatic but still promiscuous 10 years on! *J. Gen. Virol*., **2015**, *96*, 727-738.

[80] Tellinghuisen T.L., Marcotrigiano J., Gorbalenya A.E., Rice C.M., The NS5A protein of hepatitis C virus is a zinc metalloprotein. *J. Biol. Chem*., **2004**, *279*, 48576-48587.

[81] Penin F., Brass V., Appel N., Ramboarina S., Montserret R., Ficheux D., Blum H.E., Bartenschlager R., Moradpour D., Structure and function of the membrane anchor domain of hepatitis C virus nonstructural protein 5A. *J. Biol. Chem*., **2004**, *279*, 40835-40843.

[82] Schwarten M., Sólyom Z., Feuerstein S., Aladağ A., Hoffmann S., Willbold D., Brutscher B., Interaction of nonstructural protein 5A of the hepatitis C virus with Src homology 3 domains using noncanonical binding sites. *Biochemistry*., **2013**, *52*, 6160-8.

[83] Huang L., Hwang J., Sharma S.D., Hargittai M.R., Chen Y., Arnold J.J., Raney K.D., Cameron C.E., Hepatitis C virus nonstructural protein 5A (NS5A) is an RNA-binding protein. *J. Biol. Chem*., **2005**, *280*, 36417-36428.

[84] Love R.A., Brodsky O., Hickey M.J., Wells P.A., Cronin C.N., Crystal structure of a novel dimeric form of NS5A domain I protein from hepatitis C virus. *J. Virol*., **2009**, *83*, 4395-4403.

[85] Tellinghuisen T.L., Marcotrigiano J., Rice C.M., Structure of the zinc-binding domain of an essential component of the hepatitis C virus replicase. *Nature*, **2005**, *435*, 374-379.

[86] Miyanari Y., Atsuzawa K., Usuda N., Watashi K., Hishiki T., Zayas M., Bartenschlager R., Wakita T., Hijikata M., Shimotohno K., The lipid droplet is an important organelle for hepatitis C virus production. *Nat. Cell Biol*., **2007**, *9*, 1089-1097.

[87] Feuerstein S., Sólyom Z., Aladag A., Favier A., Schwarten M., Hoffmann S., Willbold D., Brutscher B., Transient structure and SH3 interaction sites in an intrinsically disordered fragment of the hepatitis C virus protein NS5A. *J. Mol. Biol*., **2012**, *420*, 310-323.

[88] Hanoulle X., Verdegem D., Badillo A., Wieruszeski J.M., Penin F., Lippens G., Domain 3 of non-structural protein 5A from hepatitis C virus is natively unfolded. *Biochem. Biophys. Res. Commun*., **2009**, *381*, 634-638.

[89] Liang Y., Ye H., Kang C.B., Yoon H.S., Domain 2 of nonstructural protein 5A (NS5A) of hepatitis C virus is natively unfolded. *Biochemistry*, **2007**, *46*, 11550-11558.

[90] Tellinghuisen T.L., Foss K.L., Treadaway J.C., Rice C.M., Identification of residues required for RNA replication in domains II and III of the hepatitis C virus NS5A protein. *J. Virol*., **2008**, *82*, 1073-1083.

[91] Appel N., Zayas M., Miller S., Krijnse-Locker J., Schaller T., Friebe P., Kallis S., Engel U., Bartenschlager R., Essential role of domain III of nonstructural protein 5A for hepatitis C virus infectious particle assembly. *PLoS Pathog*., **2008**, *4*, e1000035.

[92] Masaki T., Suzuki R., Murakami K., Aizaki H., Ishii K., Murayama A., Date T., Matsuura Y., Miyamura T., Wakita T., Suzuki T., Interaction of hepatitis C virus

nonstructural protein 5A with core protein is critical for the production of infectious virus particles. *J. Virol.*, **2008**, *82*, 7964-7976.

[93] Kaneko T., Tanji Y., Satoh S., Hijikata M., Asabe S., Kimura K., Shimotohno K., Production of two phosphoproteins from the NS5A region of the hepatitis C viral genome. *Biochem. Biophys. Res. Commun.*, **1994**, *205*, 320-326.

[94] Tanji Y., Kaneko T., Satoh S., Shimotohno K., Phosphorylation of hepatitis C virus-encoded nonstructural protein NS5A. *J. Virol.*, **1995**, *69*, 3980-3986.

[95] Huang Y., Staschke K., De Francesco R., Tan S.L., Phosphorylation of hepatitis C virus NS5A nonstructural protein: a new paradigm for phosphorylation-dependent viral RNA replication? *Virology*, **2007**, *364*, 1-9.

[96] Appel N., Pietschmann T., Bartenschlager R., Mutational analysis of hepatitis C virus nonstructural protein 5A: potential role of differential phosphorylation in RNA replication and identification of a genetically flexible domain. *J. Virol.*, **2005**, *79*, 3187-3194.

[97] Ross-Thriepland D., Harris M., Insights into the complexity and functionality of hepatitis C virus NS5A phosphorylation. *J. Virol.*, **2014**, *88*, 1421-1432.

[98] Koch J.O., Bartenschlager R.., Modulation of hepatitis C virus NS5A hyperphosphorylation by nonstructural proteins NS3, NS4A, and NS4B. *J. Virol.*, **1999**, *73*, 7138-46.

[99] Quintavalle M., Sambucini S., Di Pietro C., De Francesco R. Neddermann P., The alpha isoform of protein kinase CKI is responsible for hepatitis C virus NS5A hyperphosphorylation. *J. Virol.*, **2006**, *80*, 11305-11312.

[100] Chen Y.C., Su W.C., Huang J.Y., Chao T.C., Jeng K.S., Machida K., Lai M.M., Polo-like kinase 1 is involved in hepatitis C virus replication by hyperphosphorylating NS5A. *J. Virol.*, **2010**, *84*, 7983-7993.

[101] Masaki T., Matsunaga S., Takahashi H., Nakashima K., Kimura Y., Ito M., Matsuda M., Murayama A., Kato T., Hirano H., Endo Y., Lemon S.M., Wakita T., Sawasaki T., Suzuki T., Involvement of hepatitis C virus NS5A

hyperphosphorylation mediated by casein kinase I-α in infectious virus production. *J Virol*., **2014**, *88*, 7541-55.

[102] Lemay K.L., Treadaway J., Angulo I., Tellinghuisen T.L., A Hepatitis C Virus NS5A Phosphorylation Site That Regulates RNA Replication. *J Virol*. **2013**, *87*, 1255-60.

[103] Tellinghuisen T.L., Foss K.L., Treadaway J., Regulation of hepatitis C virion production via phosphorylation of the NS5A protein. *PLoS Pathog*., **2008**, *4*, e1000032.

[104] Kim J., Lee D., Choe J., Hepatitis C virus NS5A protein is phosphorylated by casein kinase II. *Biochem. Biophys. Res. Commun*., **1999**, *257*, 777-781.

[105] Huang L., Sineva E.V., Hargittai M.R., Sharma S.D., Suthar M., Raney K.D., Cameron C.E., Purification and characterization of hepatitis C virus non-structural protein 5A expressed in Escherichia coli. *Protein Expr. Purif*., **2004**, *37*, 144-153.

[106] Dal Pero F., Di Maira G., Marin O., Bortoletto G., Pinna L.A., Alberti A., Ruzzene M., Gerotto M., Heterogeneity of CK2 phosphorylation sites in the NS5A protein of different hepatitis C virus genotypes. *J. Hepatol*., **2007**, *47*, 768-776.

[107] Aricescu A.R., Lu W., Jones E.Y., A time- and cost-efficient system for high level protein production in mammalian cells. *Acta Crystallogr. D. Biol. Crystallogr*., **2006**, *62*, 1243-1250.

[108] Zhao Y., Malinauskas T., Harlos K., Jones E.Y., Structural insights into the inhibition of Wnt signaling by cancer antigen 5T4/Wntactivated inhibitory factor 1. *Structure*, **2014**, *22*, 612-620.

[109] Kichler A., Gene transfer with modified polyethylenimines. *J. Gene Med*., **2004**, *6*, S3-S10.

[110] Demeneix B., Behr J.P., Polyethylenimine (PEI) *Adv Genet*. **2005**, *53PA*, 215-230.

[111] Durocher Y., Perret S., Kamen A., High-level and high-throughput recombinant protein production by transient transfection of suspension-growing human 293-EBNA1 cells. *Nucleic Acids Res*., **2002**, *30*, E9.

[112] Seiradake E., Zhao Y., Lu W., Aricescu A.R., Jones E.Y., Production of cell surface and secreted glycoproteins in mammalian cells. *Methods Mol. Biol. Clifton NJ*, **2015**, *1261*, 115-127.

[113] Agrawal N., Dasaradhi P.V., Mohmmed A., Malhotra P., Bhatnagar R.K., Mukherjee S.K., RNA interference: biology, mechanism, and applications. *Microbiol Mol Biol Rev*., **2003**, *67*, 657-85.

[114] Myslinski E., Ame J.C., Krol A., Carbon P., An unusually compact external promoter for RNA polymerase III transcription of the human H1RNA gene. *Nucleic Acids Res*., **2001**, *29*, 2502-9.

[115] Kunkel G.R. Pederson T., Transcription of a human U6 small nuclear RNA gene in vivo withstands deletion of intragenic sequences but not of an upstream TATATA box. *Nucleic Acids Res*., **1989**, *17*, 7371-7379.

[116] Schanda P., Kupce E., Brutscher B., SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds. *J. Biomol NMR*, **2005**, *33*, 199-211.