UNIVERSITÀ
DEGLI STUDI
FIRENZE

DOTTORATO DI RICERCA
INTERNATIONAL DOCTORATE IN STRUCTURAL BIOLOGY

CICLO XXXIII

COORDINATOR Prof. Lucia Banci

Segmental labelling approaches applied to NMR studies of multidomain proteins in vitro and in-cell

Settore Scientifico Disciplinare CHIM/03

PhD student
Dott. Matteo Cremonini

Tutor
Prof. Lucia Banci

Coordinator
Prof. Lucia Banci

November 2017 – 2020

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 Structural biology

"Unfortunately, we cannot accurately describe at the chemical level how a molecule functions unless we first know its structure". These few words, pronounced in 1964 by James Watson, summarize in a very immediate and simple way the relevance that structural biology hold among all the scientific fields. This discipline focuses on providing a unique insight about the architecture, the dynamics and the interactions of biological macromolecules, such as proteins and nucleic acids. Thanks to the combination of biophysical, biochemical and molecular biology techniques it is possible to unveil, at atomic resolution, the three-dimensional structure of biomolecules and obtain fundamental mechanistic information about the biological processes in which they are involved. In particular we can reconstruct the dynamic complexity of cellular machinery, which often relies on intricate networks of competing interactions that form and dissolve in a time-depended way. The structure of enzymes and other categories of proteins is strictly connected to their functions and it is crucial for the correct interaction with other molecular partners. Furthermore, to highlight the importance of the structure, it is well known that many pathological states are caused by protein misfolding and, often, in viral or bacterial infections, the pathogen adhesion process to the host cell is driven by the recognition of precise structural motifs on the cell membrane. In this regard, it is recognized that understanding the exact structures of proteins can accelerate the progress of the drug discovery pipeline, by providing a detailed source of information such as target shape, macromolecules hydrophobic or hydrophilic behaviors and substrate interactions (Blundell, Jhoti, and Abell 2002). The aforementioned collected data can be extremely useful in designing ligands with complementary binding properties thus to optimize their potency and specificity (Bassetto et al. 2016). Various techniques with multiple features can be exploited for structurally characterizing macromolecules. They are Nuclear Magnetic Resonance (NMR), X-ray crystallography and cryo-Electron Microscopy (cryo-EM), regarding the high-resolution techniques, and Small Angle X-ray Scattering (SAXS), Electron Paramagnetic Resonance (EPR), Circular Dichroism (CD), Fluorescence Resonance Energy Transfer (FRET), just to name a few, regarding the low-resolution techniques. However, to adequately investigate and describe in detail all the aspects of a biological system, the application of a single methodology

is often not sufficient. In fact, individual structural data results incomplete in purpose and accuracy due to the intrinsic limitations of each technique. For this reason, more complete structural information can be achieved only through the use of an integrated approach combining together two or more methodologies. In the last years, the interest on the correlation between structure of macromolecules and functional processes they are involved in has grown considerably. As a consequence, obtaining structural information at atomic resolution directly in a cellular context can be crucial to connect structural biology data with that obtained from other disciplines as cell biology and biochemistry. To do that NMR has proven to be the most suitable technique (Luchinat and Banci 2018).

1.2 In vitro and in-cell NMR

Nuclear magnetic resonance spectroscopy has established as one of the most powerful methods within the whole landscape of structural biology techniques. It plays a crucial role, not only in determining high-resolution structures of macromolecules, but also in analyzing dynamic systems. In fact, it provides information about protein conformational equilibria, conformational changes upon binding to ligands, kinetic and thermodynamic aspects and functional dynamics of biomolecules, helping us to unveil the complex mechanisms by which biomolecules perform their multiple functions. However, as previously mentioned, the information obtained from these studies needs to be strictly correlated to that obtained from cell biology and biochemistry. For this reason, it is crucial that the data about a biomolecule are collected under conditions as close as possible to the physiological ones. The relatively recent progresses of NMR, in term of sensitivity, higher resolution, data acquisition and processing hardware and methodologies, make this structural technique one of the most efficient to achieve such requirements (Luchinat and Banci 2018). In 2001, for the first time, a small isotope-labelled globular protein was observed by NMR spectroscopy directly inside the cytoplasm of *E.coli* living cells (Serber and Dötsch 2001). That experiment was the milestone that gave birth to that branch of NMR spectroscopy called in-cell NMR. However, the use of bacteria to study human proteins presents some limitations. In the first place, bacterial machinery is not able to perform all those post-translational modifications that many human proteins require for a correct folding and to exerts

their natural function; secondly, the bacterial cytoplasm, in which human proteins are over-expressed and analyzed, is not entirely comparable to that of eukaryotic cells; for example it lacks possible functional partners of the protein in exam (Luchinat and Banci 2017). Consequently, to overcome these problems, researchers have developed different methodologies that employed eukaryotic cells for in-cell NMR investigation. Initially, easily manageable eukaryotic cells, like *Xenopus laevis* oocytes, were used. In this approach, human isotope labelled proteins were first expressed and purified from bacteria, then inserted in the oocytes by the use of microinjections (Selenko et al. 2006; Sakai et al. 2006). Subsequently, other scientists started to deliver proteins directly inside human cells cytoplasm increasing the membrane permeability by cell-penetrating peptides, electroporation or streptolysin O (Theillet et al. 2016; Binolfi et al. 2016; Ogino et al. 2009; Inomata et al. 2009). Alongside the aforementioned methodologies, an approach that has proven its undoubted efficacy is the intracellular expression of proteins. The challenges of this method were many, starting from the selection of appropriate eukaryotic cell lines, the design of an expression system capable of producing proteins at the high levels required by in-cell NMR, up the identification of the most efficient transfection technique. All of these challenges were overcome by several research groups who developed intracellular expression techniques using either yeast (Bertrand et al. 2012) or insect cells (Hamatsu et al. 2013) but more significantly human cells (Banci et al. 2013; Barbieri, Luchinat, and Banci 2016). Thanks to the last improvements, in-cell NMR became the unique tool able to study macromolecules at atomic resolution within their native cellular environment. As the cells used in this approach are intact and alive, physiochemical parameters like pH, viscosity, redox potential, presence of molecular partners, are much closer to the physiological ones, and for this reason, the information obtained from the investigation about proteins structure and their dynamics acquire a greater relevance in the biological field (Luchinat and Banci 2018; Barbieri, Luchinat, and Banci 2016, 2015; C. Li et al. 2017).

1.3 Protein-protein and protein-ligand interactions: strategies for their characterization and methodological advancements

The majority of the biological processes as well as the metabolic reactions that take place inside the cells are the result of innumerable molecular recognition mechanisms that involve both protein-protein and protein-ligand interactions. These interactions not only underlie cellular life but also play a crucial role in the development of infections and pathological states. For this reason, their functional and structural characterization can be of vital importance to deepen their knowledge and thus to open the way to new therapeutic strategies. As mentioned above, solution NMR spectroscopy can play a crucial role in this type of investigation. However, one of its major limitations is the molecular size of the investigated proteins. In fact, increase in molecular weight determines faster transverse relaxation and therefore broader lines and faster magnetization decay. Furthermore, the higher number of nuclei contributes to increase the spectrum complexity. To overcome the latter problem, one of the mostly used approach is the segmental isotopic labelling that allows to isotopic label only a portion of the polypeptide sequence within the whole native and functional protein. This strategy is particularly suitable for the study of multi-domain proteins and large complexes because it permits to label only the protein fragment of interest with NMR active nuclei, thus simplifying the interpretation of crowded and overlapped NMR spectral regions (Xu et al. 1999). To date different segmental labelling techniques have been developed. The most commonly used are the expressed protein ligation (EPL), the sortase-mediated protein ligation, and the protein trans-splicing (PTS). EPL is a variant of the native chemical ligation (NCL) method (Dawson et al. 1994) and allows to ligate a synthetic peptide with a recombinant expressed protein (Berrade and Camarero 2009). Sortase-mediated protein ligation instead, is a technique that use the ability of a prokaryotic traspeptidase, the sortaseA. This enzyme recognizes a specific aminoacidic sequences (LPXTG) on the donor polypeptide and, through the cleavage between T and G residues is able to ligate carboxyl group of threonine to an amino group of a peptide with a N-terminal glycine (Debelouchina and Muir 2018). The last segmental isotope labelling strategy, the PTS, exploits the catalytic properties of a polypeptide, known as intein, to self-excise from a host protein and ligate with a peptide bond the flanking polypeptide (Vila-Perelló and Muir 2010).

Among all these techniques, the latter is one of the most used in the NMR field. In fact, the PTS method can count on a wide range of inteins with different properties, from those that increase the expression yield to those capable of reacting in different temperature conditions and without too many constraints in the choice of extein junction site. In particular, the latter property can be extremely useful for NMR studies due to the possibility of obtaining a segmental labelled protein as close as possible to the native one. For this reason, PTS technique constitutes one of the most interesting strategies for studying large multi-domain protein and their interactions.

NMR has been also proven to be extremely valuable in the characterization of protein-ligand interactions. Indeed, nowadays it is routinely used within another fundamental scientific area, i.e. the pharmaceutical sciences. Since the 1970s the number of articles in which the terms NMR coexist with "drug" or "inhibitor" has increased exponentially (fig. 1) (Sugiki et al. 2018).

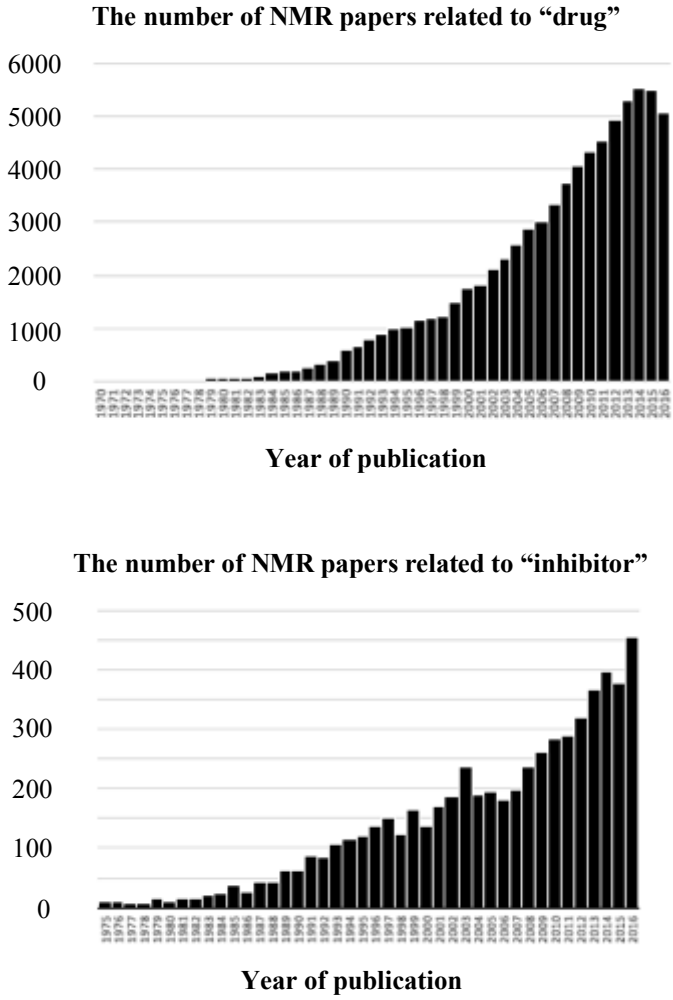


Fig. 1: Growth of the number of NMR papers related to the word "drug" and "inhibitor" (Sugiki et al. 2018)

Ligand-binding validation or identification of potential inhibitors can be obtained just using basic NMR parameters and approaches (fig. 2). For example, the high sensitivity of the chemical shifts to perturbations around the observed atoms can provide useful information about whether a small molecule binds to a target protein or nucleic acids and also which residues or functional groups are involved in this interaction (Pellecchia et al. 2008). The overall mobility of the ligands differs significantly from the free to the bound state; this is easily detectable through the saturation transfer difference analysis (Mayer and Meyer 1999) or simple ^{1}H NMR relaxation experiments (Hajduk, Olejniczak, and Fesik 1997).

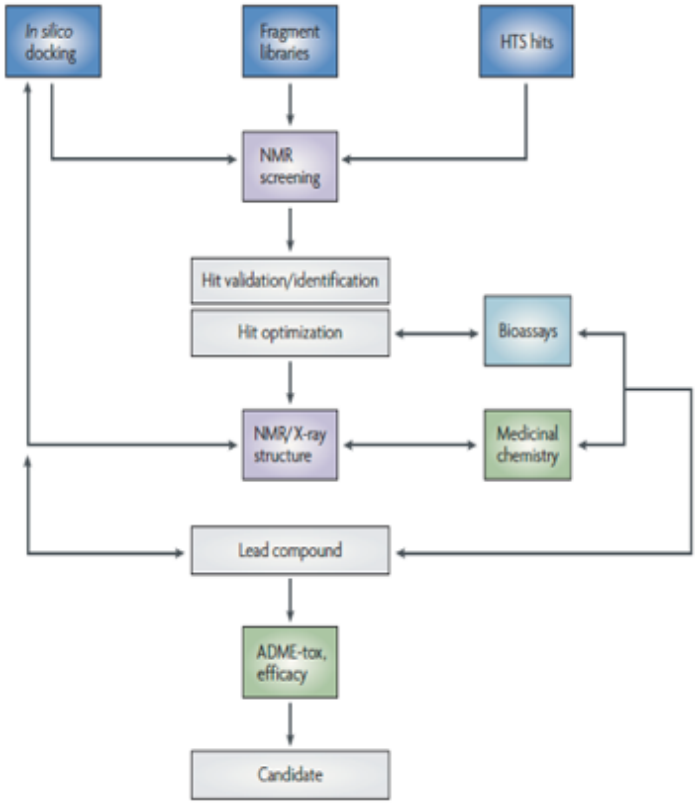


Fig. 2: Overview of applications of NMR in drug screening (Pellecchia et al. 2008)

However, further improvement in the development and identification of new drugs can be achieved through the use of in-cell NMR. In this way, by testing a potential drug directly in the cell, information on its behavior in physiological conditions and also on essential parameters to improve its efficacy and potency can be obtained.

2. PROJECTS

2.1 Segmental isotope labelling approach for the study of multidomain proteins: its application to the characterization for the Copper Chaperone for SOD1 (CCS)

2.1.1 SOD1-CCS overview

Superoxide dismutases (SOD) are well-characterized metalloproteins which, by exerting an antioxidant role, protect cells from the possible damages caused by superoxide anion. This function was first described in 1969 by McCord and Fridovich that discovered how these enzymes are able to dismutate superoxide anion to oxygen and hydrogen peroxide (McCord and Fridovich 1969):

$$2O_2^- + 2H^+ \rightarrow O_2 + H_2O_2$$

Superoxide dismutases family members can be roughly divided in groups according to the folding and to the metal cofactor: Cu/Zn-SOD, Mn-SOD, Fe-SOD and Ni-SOD. The human isoforms are three: SOD1 (Zn/Cu) that has been analyzed in this thesis, SOD2 (Mn) which is located in the mitochondrial matrix and the extracellular form SOD3 (Zn/Cu). SOD1 is largely present inside mammalian cells cytoplasm but also in the mitochondrial inter-membrane space, where it can directly remove the superoxide that is formed as a byproduct of the respiratory chain (Okado-Matsumoto and Fridovich 2001). The relevance of SOD1 has progressively increased due the discovery of the correlation between some mutations on its sequence and cases of familial Amyotrophic Lateral Sclerosis (fALS) (Luchinat, Barbieri, and Banci 2017). ALS is a fatal neurodegenerative disease that progressively brings to the destruction of the motoneurons of the spinal cord, brain stem and cortex with consequent muscular paralyses (Rowland and Shneider 2001). 90-95% of ALS cases have been shown to be sporadic (sALS), while the remaining 5-10% are linked to a previous family history of the disease (fALS). Of these, 20-25% are attributable to mutations mapped on the SOD1 encoding gene located on chromosome 21 (Valentine, Doucette, and Zittin Potter 2005). The development of fALS is not caused by a loss of functionality of the enzyme, but by a gain of toxicity. Indeed, SOD1 knock-out mice do not develop the disease and moreover, many fALS mutant SOD1 proteins maintain a high level of enzymatic activity (Rabizadeh et al. 1995; Borchelt et al. 1994; Reaume et al. 1996). The presence of proteinaceous aggregates rich in SOD1 in the case of ALS patients or rich in other proteins in the

case of Alzheimer, Parkinson, Huntington, transmittable spongiform encephalopathies, and others, is a common feature of many neurodegenerative diseases (Bruijn, Miller, and Cleveland 2004). This suggests a correlation between this SOD1-rich aggregates and the pathology (Bruijn et al. 1998). However, as for the other pathologies cited above, it is currently believed that these inclusions are not the main toxic species themselves, also because they appear at a relatively advanced stage of the disease (Bitan et al. 2003; Sherman and Goldberg 2001; Lashuel et al. 2002). Some of these fALS mutations are known to negatively affect the nascent superoxide dismutase stability, with consequent difficulties of reaching the mature form and then leading to the formation of aggregates (Lindberg, Tibell, and Oliveberg 2002; Hayward et al. 2002). The process of maturation by which SOD1 achieves its final enzymatic-active form is a concerted pattern of molecular events that have been extensively studied. At first, apo-SOD1, in its monomeric disulfide-reduced form, spontaneously binds zinc to form a stable homodimer, then copper-transfer and disulfide bond formation between the Cys57 and the Cys146 occur, mediated by a metallochaperone termed Copper Chaperone for SOD1 (CCS) (Arnesano et al. 2004; Banci et al. 2006; Schmidt et al. 1999; Casareno, Waggoner, and Gitlin 1998). CCS is widely expressed all over the eukaryotic organisms and shares with SOD1 a comparable distribution in the various cellular compartments (Rae et al. 1999; Rothstein et al. 1999). This chaperone also forms a homodimer, each monomer being constituted by three domains each of them with a specific crucial role in the SOD1 maturation process. The N-terminal domain (D1) is essential for copper delivery *in vivo*, it has a high fold similarity with other metallochaperones and shows the typical MXCXXC copper binding motif (Caruano-yzermans, Bartnikas, and Gitlin 2006). The second domain (D2) possesses a high structural and sequence similarity with SOD1 and, due to its SOD-like conformation, is able to interact with the immature form of the molecule (Schmidt et al. 1999; Lamb et al. 1999, 2000). The C-terminal domain (D3) is a short polypeptide segment (30-40 residues) that natively lacks secondary structure. It harbors invariant conserved CXC (Cys-244 and Cys-246) motif that appears to be essential for the correct SOD1 maturation (Furukawa, Torres, and O'Halloran 2004). Despite many efforts made to elucidate the CCS-dependent SOD1 maturation mechanism, some aspects of this multi-steps molecular process remain still unknown. For example, the mechanism of disulfide bond formation, whether

the copper transfer and disulfide bond formation occur in a single step or in different moments, and the exact conformation adopted by the D3 in the SOD1-CCS heterodimer, have remained elusive until now. In one of the most accredited model, the CCS-mediated SOD1 activation takes place in at least six steps (fig. 3): zinc acquisition, formation of the CCS heterodimer with SOD1, copper delivery from CCS to SOD1 thanks to D1, disulfide bond transfer from CCS (Cys244-Cys246) to SOD1 (Cys57-Cys146) probably via the CCS (Cys244) and SOD1 (Cys57) inter-molecular disulfide bond formation, dimerization of SOD1 (Banci et al. 2012).

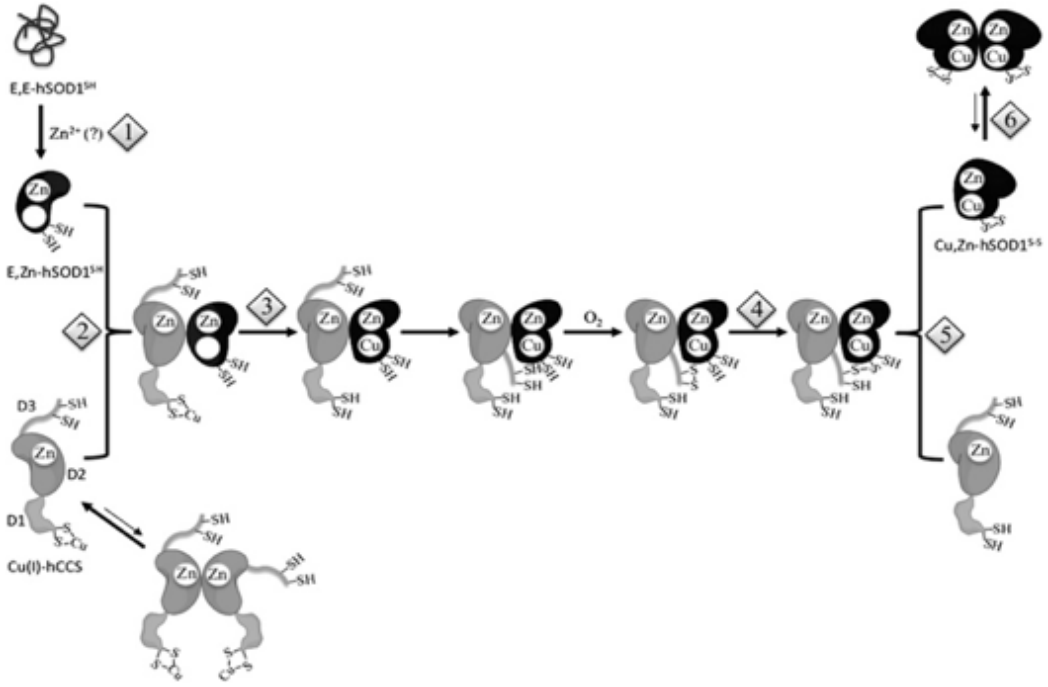


Fig. 3: Schematic mechanism of CCS-dependent SOD1 maturation in vitro. (1) Nascent SOD1 (E,E-SOD^{SH}) acquires zinc producing E,Zn-SOD1^{SH}. (2) The SOD1^{SH} monomer forms a heterodimer with a monomer Cu(I)-CCS. The current study suggests that copper transfer (3) occurs before formation of the intermolecular (4), and then SOD1 intramolecular (5), disulfide bond, resulting in the mature monomer (6) that dimerizes to the active state. Although oxygen in vitro is required in the process before step 4, it is currently unknown precisely at which step this occurs. [Note: E,Zn-SOD1^{SH} can obtain Cu(I) in a CCS-independent manner]. (Banci et al. 2012)

According to the above mechanism, CCS D1 is responsible for copper delivery, D2 is required for the interaction with SOD1 and D3 is involved in the disulfide bond formation. However, other studies revealed conflicting results. In fact, a series of experiments conducted to examine the roles of the various CCS domains at a cellular level reported that in cells transfected with full length CCS containing a mutation on the D3 CXC-binding site neither SOD1 activity nor SOD1 copper acquisition were observed (Caruano-yzermans, Bartnikas, and Gitlin 2006). Moreover, different organisms like *Drosophila melanogaster*, *Anopheles gambiae*

and *Saccharomyces pombe,* lacking the MXCXXC motif in the first domain, show to have the ability to bring SOD to a complete maturation despite the absence of the copper chaperone activity (Kirby et al. 2008; Laliberté et al. 2004).
To underline even more the CCS relevance, it has been recently demonstrated by *in-vitro* and in-cell NMR analysis, that the second domain of CCS co-expressed with a set of SOD1 fALS mutants helps the nascent SOD1 to bind zinc, stabilizes its structure, and addresses it towards the right maturation pathway, therefore acting as a molecular chaperone (Luchinat, Barbieri, and Banci 2017). Overall, CCS is responsible of many molecular events and is a closely related to the pathological states in which SOD1 mutants are involved. Because of this reason, CCS deserves further and in-depth investigations that may shed light on its still obscure aspects.

2.1.2 Aims of the study

The aim of this project consisted in using nuclear magnetic spectroscopy (NMR) to characterize CCS D3 conformation within full-length CCS, in different metalation and redox states, both in the CCS homodimer and in the heterodimer with SOD1. The application of a segmental labelling technique would allow the incorporation of NMR active nuclei only in the CCS D3 making easier the study of all its possible conformations through to the analysis of the backbone chemical shifts, thanks to the reduced spectral crowding.

2.1.3 State of the art

The structural features of the CCS third domain have been the subject of numerous studies over the years. The lack of secondary structure gives this domain a high level of flexibility allowing it to assume radically different conformations in solution. This has been further confirmed by a study of Fetherolf et al. in 2017. The authors reported a hSOD-yCCS heterodimer x-ray structure in which the D3 conformation strikingly differs from that reported for the same domain in the previous ySOD-yCCS structure (fig. 4) (Fetherolf et al. 2017; Lamb et al. 2001).

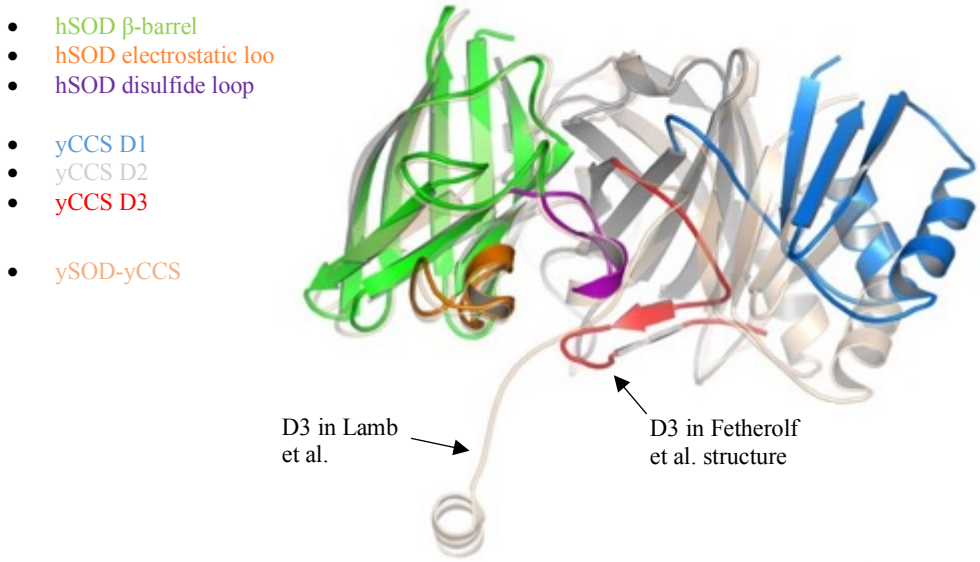


Figure 4. Superposition of ySOD-yCCS (Lamb et al. 2001) and hSOD1-yCCS (Fetherolf et al. 2017)

In this new structure it is evident that CCS D3 takes a β-hairpin shape, which is in contrast with the disordered and/or α-helix conformations observed in earlier works (Lamb et al. 1999, 2001). The β-hairpin is stabilized by several interactions with residues of SOD, of D1, and of D2 that further contribute to strengthen the entire SOD-CCS complex (fig. 5).

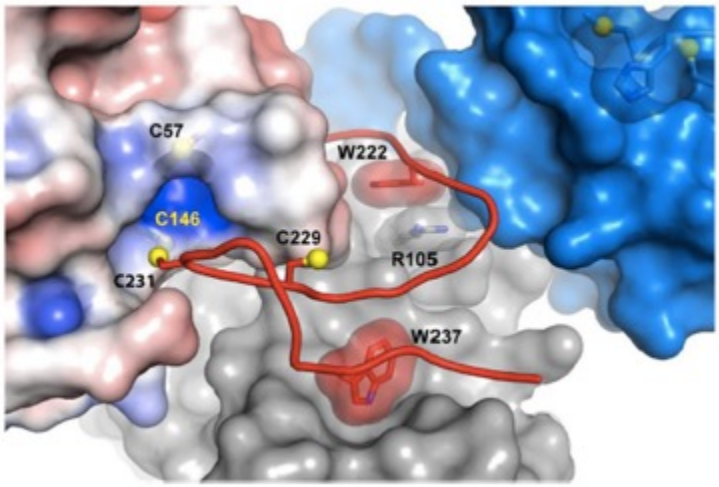


Figure 5. D3 β-hairpin interaction with SOD1 β-barrel and disulfide loop. Cys57–Cys146 Sod1 disulfide bond exposes an electropositive zone near Cys146. CCS D1 (blue) and D2 (gray) are represented as surfaces and the stabilizing positioning of residues Trp222 and Trp237 from CCS D3 are highlighted. (Fetherolf et al. 2017)

2.1.4 Methodologies

Large multi-domain protein complexes investigation has always represented a challenge for solution NMR spectroscopy. In solution, CCS is present as a homodimer, and this means that its size and the numerous interactions between domains make its study via NMR challenging. Therefore, to overcome these problems, we sought to apply the segmental labelling technique known as protein trans-splicing (PTS). This kind of approach basically consists in a post-translational modification catalysed by an intervening polypeptide sequence termed intein. Intein is able to excide itself out from a precursor protein and to ligate the flanking polypeptide sequences, termed exteins, with a native peptide bond. As in this research project, intein can also be split into its N- and C-terminus regions. In this case the two separate intein parts individually have no activity but once combined they associate noncovalently to give a functional protein splicing element followed by the ligation between N- and C- extein sequences constituting the full-length protein (fig. 6). A fundamental requirement for successful ligation is the presence of a cysteine in the splicing junction sequence.

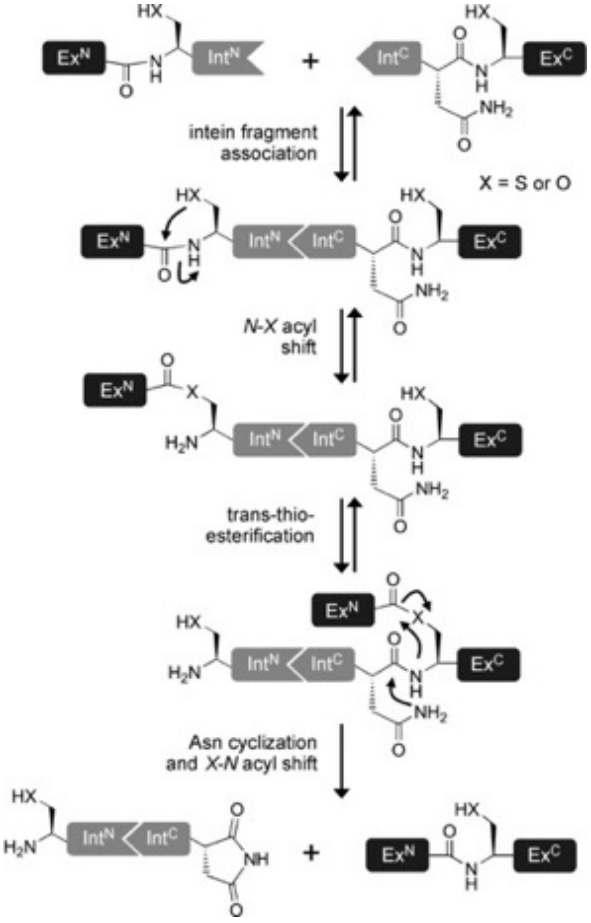


Figure 6. Protein trans-splicing scheme (Mootz 2009). Intein is divided in its N- and C- terminus portion (Int^N and Int^C) and conjugated with extein N- (Ex^N) and extein C- (Ex^C) protein domains. After the non-covalent association between Int^N and Int^C a series of reaction bring to the formation of a native peptide bond between exteins and to the detachment of the intein portions.

The incorporation of isotope labels only in the CCS D3 region will overcome the problem of its non-dispersed signals overlapped with those of the other two domains, thus providing an easier analysis of that particular spectral region. The experimental plan included different steps (fig. 7): the first one consists of the separate expression of unlabelled CCS D1D2 and [^{13}C ^{15}N] labelled CCS D3 ligated respectively with intein N- and C-terminus sequences, a second step consists in reconstituting the full-length CCS through the trans splicing reaction between the previous expressed constructs, while the last step is the NMR investigation of both CCS homodimer and heterodimer with SOD1. In the first part of this work, among all the various types of inteins, we decided to use those of a unicellular microorganism known as *Nostoc punctiforme.* Such inteins showed to have a high efficiency in various conditions combined with a high tolerability of the extein flanking residues (Iwai et al. 2006; Wu, Hu, and Liu 1998).

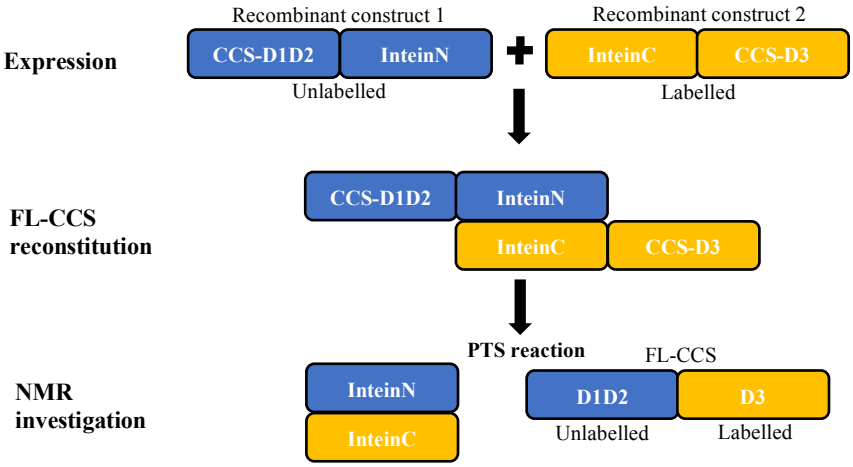


Figure 7. Schematic representation of the experimental workflow.

2.1.5 Results

To achieve the final goal, three different experimental approaches have been adopted:

Cloning and ligation:

N- and C-terminus intein sequences, CCS D1D2 domain and CCS D3 domain were first amplified separately trough PCR, then, the ligase enzyme was used to constitute D1D2-InteinN and InteinC-D3 constructs (fig. 8 a-b). Finally, the obtained fragments were cloned in two PET11 vectors for the expression.

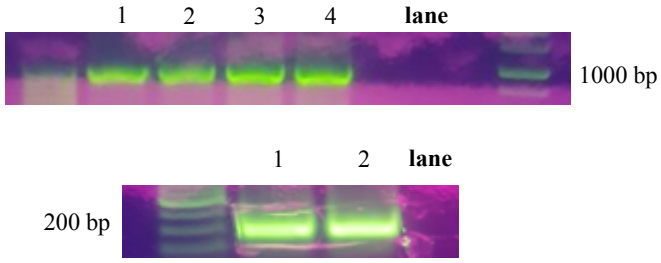


Fig. 8 Electrophoresis gels of fragments **a)** D1D2-InteinN (1023 bp) and **b)** InteinC-D3 (255 bp)

Although the electrophoresis gels reported in fig. 8 showed the presence of bands corresponding to the correct height of the two constructs, the sequencing data revealed that cloning and ligation were not correct. Indeed, both IntN_D1D2 and D3_IntC constructs lacked of some bases in the blunt ligases zone.

Multiple cloning attempts trying to variate annealing temperature and testing different DNA/plasmid ratio were performed with no success. We hypothesized that the problem was due to some intrinsic property of our protein rather than to the cloning technique itself. For this reason, we designed a different construct which was purchased as de-novo synthetized.

Constructs synthesis

D1D2-IntN and IntC-D3 constructs were *de novo* synthesized following the scheme represented in fig. 9 and then cloned in PET11 expression vector between BamH1 and Nde restriction sites. The last CCS-D1D2 fragment residue was mutated from an Arg to a Cys, as a fundamental requirement for the correct happening of the protein trans-splicing reaction.

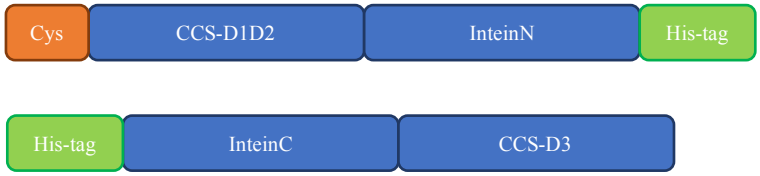


Figure 9. Schematic representation of the two synthetized constructs. We can find the 6-His tail purification tags (green), the two recombinant constructs (blue) and the cysteine residue at the splicing junction (orange).

The purification, performed through a His trap column, showed a poor yield in the soluble fraction and a low level of purity for both the expressed proteins, but, despite these problems, we considered that the amount of products was sufficient to perform a first protein trans-splicing test.

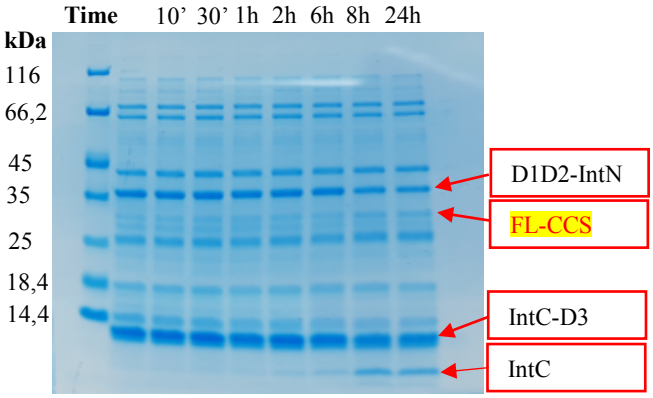


Figure 10. SDS-page gel of protein trans splicing reaction between the two constructs

In the SDS-page of the PTS reaction (fig. 10) we could identify bands corresponding to D1D2-IntN (37 kDa), IntC-D3 (9 kDa), IntC (4,9 kDa) alone and FL-CCS (29 kDa). Instead, IntN (12 kDa) fragment was not clearly visible. The progressive reaction product formation was also visible in the western blot (WB) analysis (fig. 11) using an anti-full-length CCS antibody.

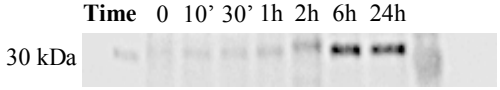


Figure 11. Western Blot analysis of PTS reaction SDS-page gel using an anti-FL-CCS antibody

From both SDS page and WB the formation of the FL-CCS was clear, but resulted in a very poor yield, not sufficient for an NMR experiments. The PTS reaction was repeated in different conditions of temperature (25°C, 30°C and 37°C) and concentration of the two precursors (from ~20 to ~50 µM). However, the results remained unsatisfactory. Furthermore, IntC-D3 fragments seemed to show a progressive hydrolysis during the course of the reaction. To overcome these problems and to be able to recover even the insoluble part of the two constructs, further attempts were made taking in account a previous publication (Zettler, Schütz, and Mootz 2009), in which it was reported that the PTS reaction could take place even in presence of different urea concentrations. According to the aforementioned paper, pellets of expression cultures were resuspended in urea 8 M solution and purified under denaturing conditions. The buffer of the two proteins was then diluted to a 4 M urea concentration and the PTS reaction was performed mixing equimolar amounts of the precursors. However, no formation of the final product took place (fig. 12). A better result was not even achieved by refolding the precursors through a multi-steps dialysis in decreasing urea concentration and then testing PTS reaction with higher precursors concentration (~100 µM).

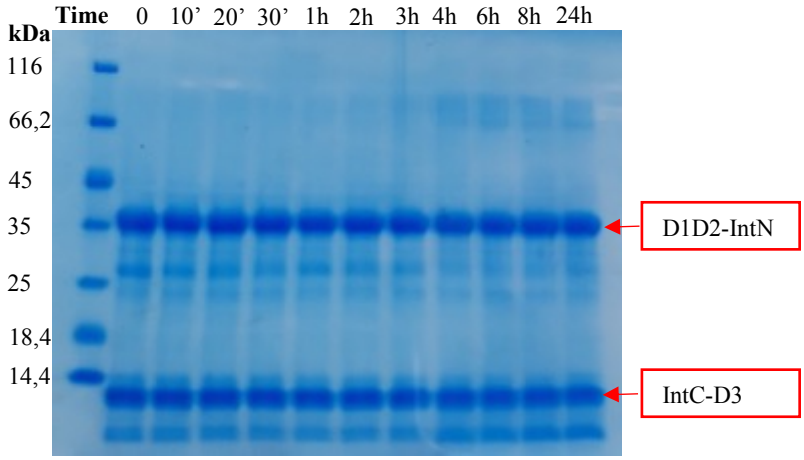


Figure 12. Protein trans splicing reaction SDS-page gel under denaturing condition (UREA 4M). It is possible to identify bands corresponding to D1D2-IntN (37 kDa) and IntC-D3 (9 kDa).

Multiple further attempts testing various solutions and trying to optimize every step of the entire experimental workflow, from the conditions of expression, purification to those of reaction were made. NaCl different concentrations (from 150 mM to 1 M) as well as various reducing agents (TCEP and DTT) were tested both in purification and in PTS reaction buffers. Higher aggregation tendency at low salt

levels was observed. Furthermore, two types of Superdex prep grade columns were used for the gel filtration step, the Superdex 16/60 75 pg and the Superdex 16/60 200 pg, to test different separation resolutions. Despite these numerous attempts the expected results have not been achieved. In fact, we had to face several critical issues including a poor solubility of D1D2-IntN and IntC-D3 expressed proteins, difficulty in purification and an insufficient protein trans splicing reaction yield. In order to overcome these problems, we decided to redesign the constructs.

New constructs design

New constructs were designed following a scheme kindly suggested by Prof. Gaetano Montelione in order to improve both the solubility and the specificity of the reaction. The most important change concerned the type of inteins used. Indeed, the constructs design involved the use of a chimeric split intein system since the two catalytic polypeptides derived from two different organisms. The N-terminal part of the intein came from the *Nostoc punctiforme* DnaE gene while the C-terminus from the *Synechocistis sp.* DnaE gene. The use of this inteine system also required the insertion of a specific recognition sequence composed of SGV aminoacidic residues in the trans-splicing junction (Lockless & Muir, 2009). Both fragments were conjugated with maltose binding protein (MBP) to improve the solubility and a tobacco etch virus cleavage site was introduced to eliminate MBP and StrepII purification tag from the reconstituted FL-CCS. In figure 13 the scheme of the most promising constructs combination is represented.

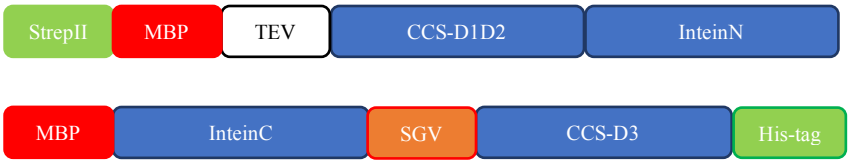


Figure 13. Schematic representation of the two new synthetized constructs. We can find the purification tags (green), the two recombinant constructs (blue), the SGV motif at the splicing junction (orange), the maltose binding protein (red) and the TEV cleavage site (white).

The first tests, unlike what we saw with the previous constructs, confirmed for both D1D2-IntN and Int-D3 a higher expression efficiency with the major part of the expressed protein located in the soluble fraction. PTS reaction was then performed mixing the soluble part of the two expression cultures. From SDS-page gel we

observed a progressive reduction of the precursors and the formation of a ~75 kDa band corresponding to the final expected product, i.e Strep-MBP-FL-CCS-His-tag (fig. 14).

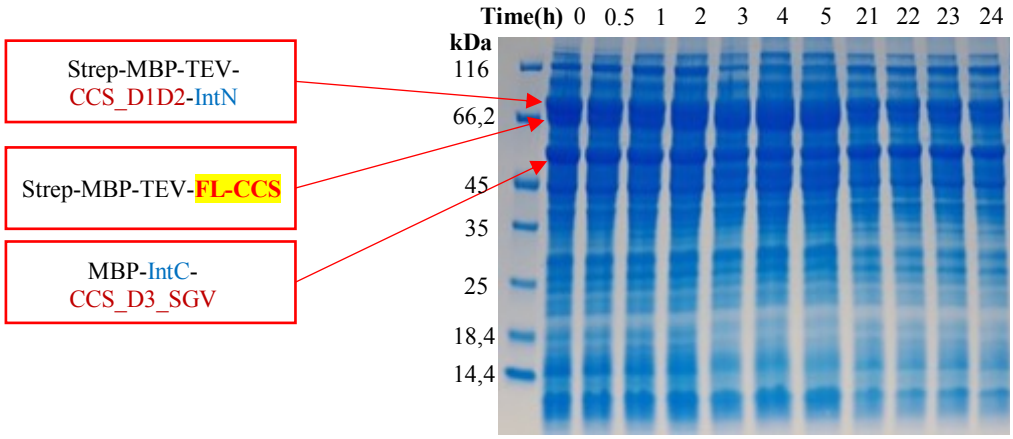


Figure 14. SDS-page gel performed on the PTS mixture with withdrawals at different time. It is possible to identify bands corresponding to Strep-MBP-TEV-CCS_D1D1-IntN (82.66 kDa), Strep-MBP-TEV-FL-CCS (75.43 kDa) and MBP-IntC-CCS_D3_SGV (51.57 kDa).

The reaction solution was purified using a His-trap nickel affinity column, and a small aliquot was incubated over-night (ON) at room temperature with TEV enzyme to obtain a FL-CCS without any additional tags (fig. 15 a-b).

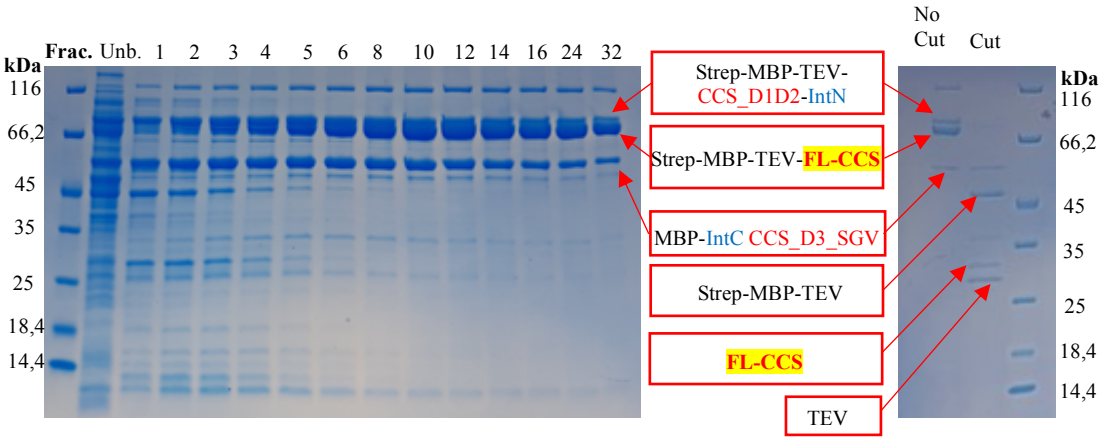


Figure 15. SDS-page gel performed on **a)** fractions collected from the purification through His-trap affinity column of the PTS reaction mixture and on **b)** aliquot collected from the previous purification before and after the cut with TEV enzyme. It is possible to identify bands corresponding to Strep-MBP-TEV-CCS_D1D1-IntN (82.66 kDa), Strep-MBP-TEV-FL-CCS (75.43 kDa), MBP-IntC-CCS_D3_SGV (51.57 kDa), Strep-MBP-TEV (46.20 kDa), FL-CCS (29 kDa) and TEV enzyme (27 kDa).

The success of the PTS reaction was confirmed by the occurrence of the TEV cut with the formation of FL-CCS as shown in figure 14b. However, despite the results of the reaction, its purification constituted a challenging aspect. In fact, as already observed in the previous experimental approach, we observed the formation of aggregates that interfered with the purification procedures, thus making impossible to achieve a sufficient level of purity. To solve this problem we applied different techniques, from gel filtration to anionic exchange assays. Furthermore, we tried different purification buffer compositions increasing NaCl concentration up to 2M in order to break any hydrophilic interactions. However, in none of these attempts it was possible to reach sufficiently high levels of purity to perform NMR experiments.
Further attempts will be made to purify the PTS reaction mixture under denaturing conditions (UREA 8M). As next step, the protein will be refolded and then we will proceed with TEV cut.

2.1.6 Discussion

In the present work, the segmental labelling technique known as protein trans splicing was applied for the purpose of elucidating the conformational features of the CCS third domain in its homodimeric form and in the heterodimer with SOD1. Three different experimental strategies were followed, among which the last one revealed to be the most promising. In fact, the refined constructs optimization, obtained through the insertion of the SGV recognition sequence, the choice of high efficiency inteins as well as the use of the MBP solubility tags, allowed to improve both the expression and the PTS reaction yield with the formation of the final expected product, i.e. FL-CCS. Although these results are encouraging, there are still two crucial problems that we need to overcome. The first one regards the incomplete reaction between the two precursors while the second is the difficulty of obtaining a final product that is sufficiently pure to be analysed by NMR. Multiple issues can negatively affect the expected results. For example, cases of N- and/or C-cleavages (fig. 16) during the occurrence of PTS reaction are reported in literature (Aranko and Volkmann 2011).

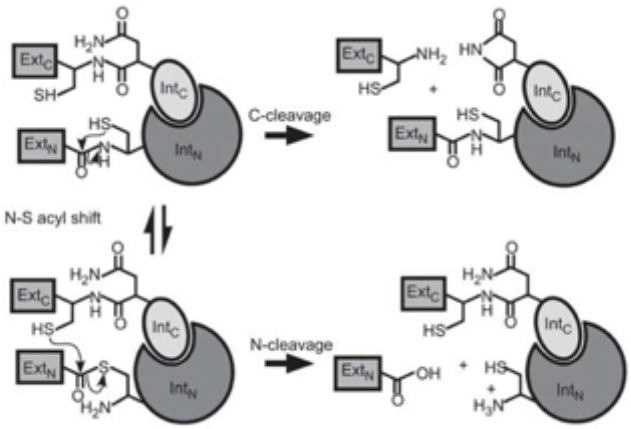


Figure 16. N- and -C possible cleavages during the PTS reaction

The authors explained these off-pathway cleavage reactions as due to a perturbation of the inteins catalytic core by foreign extein sequences, resulting in the loss of N- or C- extein before the splicing reaction is completed. Moreover, it is necessary to take in account the splicing sequence junction nature. In particular, both the residues flanking the N-terminus and C-terminus catalytic sites and the same junction length can be crucial for the effective occurrence of the reaction. However, considering the results obtained until now a possible common cause for the main problems encountered emerges, that is the tendency of the two fragments to be prone to form aggregates. In fact, it is well known that CCS D2 is essential for the formation of the homodimer in solution and therefore, it is possible that these strong interactions can prevent both the splicing reaction and the subsequent purification.

The segmental labelling technique known as protein trans-splicing has never been developed for studying the CCS third domain and its interaction with SOD1. For this reason, all the applied protocols need a refined optimization process in order to solve the difficulties mentioned above and proceed with the NMR investigation.

2.2 Protein based NMR drug screening in human living cells: its application to carbonic anhydrase II inhibitors.

2.2.1 Overview

Nuclear magnetic resonance spectroscopy has been shown to be an extremely powerful method for studying protein-ligand interactions, and, nowadays, with the implementation of in-cell NMR, there has been a further step towards probing these processes in living human cells. The information obtained from the application of this method can be crucial not only to improve the potency and the efficacy of drugs but also to better understand their action mode in the physiological context. In this research project, we showed how in-cell NMR can be applied in a novel manner to overcome some bottlenecks that are typical of the classical drug discovery pipeline in vitro. Indeed, it is very common that many promising drugs that have undergone a long optimization procedure *in vitro* turn out to be totally ineffective *in vivo*. In this work we applied NMR spectroscopy to analyze the binding of carbonic anhydrase inhibitors to the second isoform of the human carbonic anhydrase (CA II) directly in the cytosol of living human cells.

2.2.2 Carbonic anhydrase II

Carbonic anhydrases (CAs) are enzymes that catalyse the reversible hydration of carbon dioxide and bicarbonate:

$$CO_2 + H_2O \leftrightarrow HCO_3^- + H^+$$

These enzymes have been known for over 80 years, and since their discovery in 1933, they have been the subject of numerous studies that embraced different research disciplines, from structural biology, to clinical sciences and drug discovery. There are five different classes in which CAs are divided: α, β, γ, δ and ζ. α-class, the only present in mammals, comprises CAs primarily present in vertebrates. The β family is found in higher plant chloroplasts and some prokaryotes, γ is characteristic of archaebacteria, especially methanogens, whereas δ and ζ are commonly observed in diatoms and some bacteria. These classes are considered to

be a clear example of convergent evolution, in fact, they present differences in primary sequence of amino acids and as consequence in the three-dimensional tertiary structure (α, β, γ). This implies that they have convergently evolved to carry out the same biochemical reaction essential for life (Frost 2014). Among all classes, the α is definitely the most known and deeply studied due to its implications in numerous human pathologies. It consists of sixteen different isoforms which are distinguished by function, localization, tissue-specific expression, kinetic properties and sensitivity to inhibitors (Supuran 2008). Of these, eight are cytosolic proteins (CA I, CA II, CA III, CA VII, CA VIII, CA X, CA XI, CA XIII), two are localized in the mitochondrial matrix (CA VA, CA VB), CA VI is secreted. Finally, five isoforms take contact with cellular membrane, among which CA IX, CA XII and CA XIV are transmembrane proteins, whereas CA IV and CA XV are glycosylphosphatidylinositol (GPI)-anchored proteins (Frost 2014). All these isoforms are involved in a myriad of physiological processes like transport of CO_2 and HCO_3^- across both membrane sides, acid-basic balance, signal transduction, oncogenesis, cellular proliferation, secretion and others that have been well documented along many decades of study (Ditte et al. 2011; Thiry et al. 2006; Henry 1996; Sterling, Reithmeier, and Casey 2001a).

In this project we focused on carbonic anhydrase II (CAII) isoform. CAII is highly expressed in red blood cells (RBCs) and in the kidney intercalated cells, but it can be present at lower levels also in loop of Henle and in proximal tubules (Chegwidden and Carter 2000; Swenson 2000; Brown et al. 1983). The role of this enzyme is to control cytosolic pH balance. This balance is maintained through the transport of CO_2 out of the cell and also through the interaction that this enzyme establishes with different membrane carriers, among which, Cl^-/HCO_3^- and Na^+/H^+ exchangers and the Na^+/HCO_3^- cotransporter (Vince and Reithmeier 1998, 2000; Sterling, Reithmeier, and Casey 2001b; Becker and Deitmer 2007; X. Li et al. 2002). Structurally, CAII presents an active site in which a Zn^{2+} ion is coordinated to three His residues and a molecule of water (fig. 17).

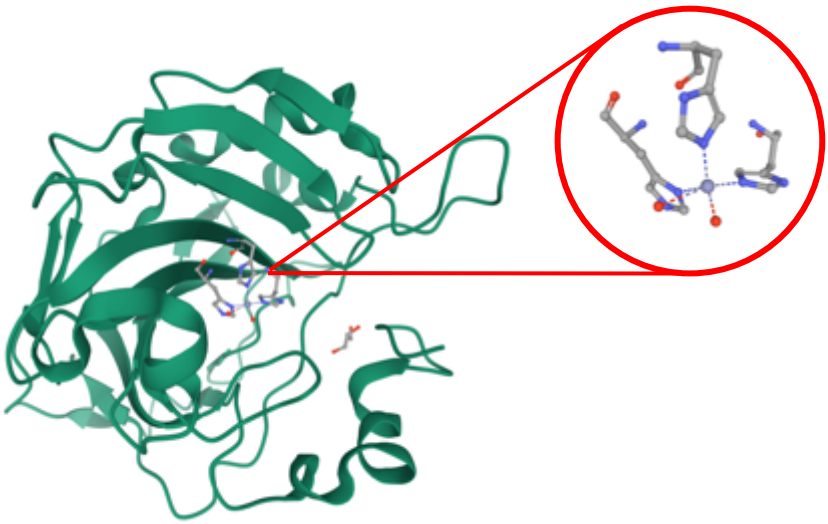


Fig. 17: Structure of human CAII; represented in cartoon the active site of CAII with three His residues that coordinate a Zn^{2+} ion (in grey). PDB: 3KS3.

The overall catalytic pocket is constituted by two different portions, one hydrophilic and the other hydrophobic, with the last one comprising the CO_2 binding site.
CA II in particular is associated with cases of epilepsy and glaucoma and it is still the subject of numerous studies aimed at developing new and increasingly specific drugs. However, all CAs share a high structural similarity of the catalytic pocket thus limiting the isoform specificity of drugs and, as consequence, the possibility of treating pathologies in a selective fashion.
The profound knowledge of this isoform and the numerous drug design studies and already approved and used drugs, as well as the ease of expression in human cells led us to select CAII as a model for our study.

GDCh **Communications** Angewandte International Edition Chemie

NMR Spectroscopy Hot Paper

International Edition: DOI: 10.1002/anie.201913436
German Edition: DOI: 10.1002/ange.201913436

Drug Screening in Human Cells by NMR Spectroscopy Allows the Early Assessment of Drug Potency

*Enrico Luchinat, Letizia Barbieri, Matteo Cremonini, Alessio Nocentini, Claudiu T. Supuran, and Lucia Banci**

Abstract: *Structure-based drug development is often hampered by the lack of in vivo activity of promising compounds screened in vitro, due to low membrane permeability or poor intracellular binding selectivity. Herein, we show that ligand screening can be performed in living human cells by "intracellular protein-observed" NMR spectroscopy, without requiring enzymatic activity measurements or other cellular assays. Quantitative binding information is obtained by fast, inexpensive ^{1}H NMR experiments, providing intracellular dose- and time-dependent ligand binding curves, from which kinetic and thermodynamic parameters linked to cell permeability and binding affinity and selectivity are obtained. The approach was applied to carbonic anhydrase and, in principle, can be extended to any NMR-observable intracellular target. The results obtained are directly related to the potency of candidate drugs, that is, the required dose. The application of this approach at an early stage of the drug design pipeline could greatly increase the low success rate of modern drug development.*

Rational drug design requires ligand-based screenings and protein-based structural studies to characterize the binding to the target protein, followed by lead optimization to increase the binding constant. The ability to reach intracellular targets must then evaluated with ad hoc cell-based assays. Most drug candidates fail here or at later stages, due to lack of activity in cells or to the occurrence of adverse effects in vivo.[1] Such failures are often linked to poor membrane permeability and/or to the lack of binding specificity in the cellular environment. Nuclear magnetic resonance (NMR) spectroscopy applied to living cells[2–6] has the potential to overcome these critical bottlenecks, as it can directly observe macromolecule–ligand interactions at atomic resolution within the cellular environment.[2,7,8] Herein, we report an approach to perform protein-observed ligand screening by NMR directly in the cytosol of living human cells. The approach allows the assessment of membrane permeability and intracellular binding specificity of candidate drugs, which cannot be inferred from in vitro analysis, and can reveal interesting and strikingly different behaviour for compounds that have similar properties in vitro.

We applied the method to the screening of inhibitors of the second isoform of human carbonic anhydrase (CA2). In humans there are 15 CA isoforms,[9] many of which are relevant drug targets involved in several pathologies.[10–14] To date, several CA inhibitors are routinely administered in the treatment of glaucoma, epilepsy, or as diuretics,[15,16] with some of them in clinical development as antitumor agents.[17,18] In order to detect CA2 by NMR, the protein was directly expressed and labelled in the cytosol of human cells (see the Experimental Methods section of the Supporting Information).[19,20] NMR signals arising from [^{15}N]-CA2 were clearly detected in the ^{1}H-^{15}N correlation spectra (Supporting Information, Figure S1), indicating that the intracellular protein is soluble and free from interactions with slow-tumbling cellular components, which would otherwise increase transverse relaxation and cause signal broadening beyond detection.[21] Signals from slow-exchanging histidine side chain amide protons located in the active site were clearly detected in a background-free region of the 1D ^{1}H NMR spectrum between 11 and 16 ppm, allowing protein–ligand interactions to be monitored without the need for isotopic labelling (Supporting Information, Figure S2).

The interaction between intracellular CA2 and two approved drugs, acetazolamide (**AAZ**) and methazolamide (**MZA**), was monitored by ^{1}H-^{15}N NMR. Spectra from cells expressing [^{15}N]-CA2 treated with excess of **AAZ** or **MZA** showed clear differences compared to untreated cells, indicating that both drugs had bound the intracellular protein (Figure 1 a,b). Inhibition of intracellular CA2 was confirmed by total CO_2 hydration activity measured by stopped-flow, which was largely decreased in lysates from cells expressing

[*] Dr. E. Luchinat, Dr. L. Barbieri, M. Cremonini, Prof. L. Banci
CERM—Magnetic Resonance Center,
Università degli Studi di Firenze
via Luigi Sacconi 6, 50019 Sesto Fiorentino (Italy)
E-mail: banci@cerm.unifi.it
Dr. E. Luchinat
Dipartimento di Scienze Biomediche Sperimentali e Cliniche "Mario Serio", Università degli Studi di Firenze
Viale Morgagni 50, 50134 Florence (Italy)
Dr. L. Barbieri
Consorzio Interuniversitario Risonanze Magnetiche di Metalloproteine, Via Luigi Sacconi 6, Sesto Fiorentino (Italy)
Dr. A. Nocentini, Prof. C. T. Supuran
Dipartimento Neurofarba, Sezione di Scienze Farmaceutiche, Università degli Studi di Firenze
Via Ugo Schiff 6, 50019 Sesto Fiorentino (Italy)
Prof. C. T. Supuran, Prof. L. Banci
Dipartimento di Chimica, Università degli Studi di Firenze
Via della Lastruccia 3, 50019 Sesto Fiorentino (Italy)
Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under: https://doi.org/10.1002/anie.201913436.

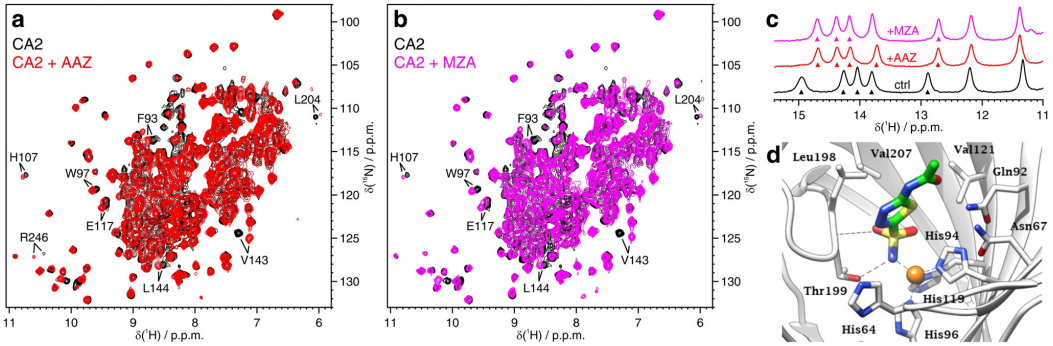


Figure 1. Overlay of ^{1}H-^{15}N NMR spectra of cells expressing [^{15}N]-CA2 in the absence of ligands (black) and treated for 1 hour with a) 100 μM **AAZ** (red); b) 100 μM **MZA** (magenta). Some peaks shifting upon ligand binding are labelled. c) Imino region of the 1D ^{1}H NMR spectra of unlabelled intracellular CA2 in the absence of ligands (black) and bound to **AAZ** (red) or **MZA** (magenta). Peaks used to obtain binding curved are marked with arrows. d) Active site view of CA2/**AAZ** complex (PDB 3HS4): the catalytic zinc ion (orange sphere) is located at the end of a large, conical cavity close to the protein centre and exhibits a tetrahedral coordination with three conserved histidine residues and, in the active form, a water molecule/hydroxide ion as fourth ligand.[9] The latter is replaced by the deprotonated sulfonamide nitrogen in the enzyme/inhibitor adduct. The SO_2NH^- moiety is additionally H-bonded as donor to the OH moiety and as acceptor to the amidic NH of residue Thr199.

CA2 treated with inhibitor with respect to untreated cells (Supporting Information, Figure S3). In vitro, titration of CA2 with one equivalent of **AAZ** or **MZA** resulted in quantitative binding (Supporting Information, Figure S4), consistent with nanomolar dissociation constants ($K_d^{AAZ} = 12.5 \pm 1.0$ nM; $K_d^{MZA} = 14 \pm 0.9$ nM, see the Experimental Methods section in the Supporting Information). The protein–ligand interaction can be mapped at the single-residue level by chemical shift perturbation (CSP) analysis. The high similarity between in-cell and in vitro CSP plots revealed that both ligands bind to intracellular CA2 in an almost identical fashion as in vitro (Supporting Information, Figure S5), that is also consistent with the atomic structure of the protein–drug adducts reported previously (Figure 1d).[22]

Binding of **AAZ** and **MZA** to unlabelled CA2 in the slow exchange regime was clearly observed in the imino region of the 1D ^{1}H NMR spectra, both in cells and in vitro, due to the proximity of the observed nuclei to the ligand binding site (Figure 1c,d and Supporting Information, Figure S6a,b). Despite their known pH-dependence, these nuclei proved to be excellent reporters of the free and bound states of CA2 (Supporting Information, Figure S6c). The surprisingly good signal separation between free and bound states, despite the broader spectral lines of in-cell NMR spectra, allowed the area under each peak to be retrieved by deconvolution analysis (Supporting Information, Figure S7). Simple 1D ^{1}H NMR spectra could therefore provide a quantitative measure of the free and ligand-bound fractions of intracellular CA2, providing a less time- and cost-intensive strategy compared to 2D ^{1}H-^{15}N NMR.

^{1}H in-cell NMR and deconvolution analysis were subsequently applied to screen a larger set of CA inhibitors, selected among recently reported sulfonamide-derivatives, for which the binding affinity to CA2 had been previously characterized in vitro and ranged from low-nanomolar to high-micromolar (Figure 2).[23–25] To establish whether this approach could discriminate ligands based on their cell permeability, a CA inhibitor (**C18**) that is unable to diffuse through the plasma membrane was also included. Strikingly,

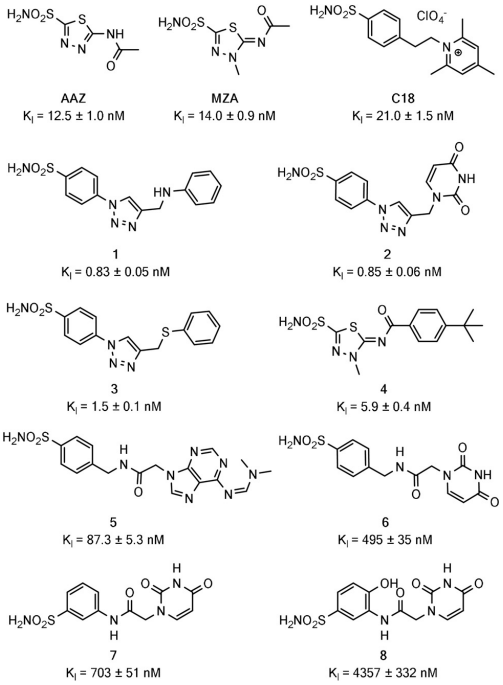


Figure 2. Sulfonamide-derived CA inhibitors analyzed in this study. K_I measured in vitro for CA2 are reported (see the Experimental Methods section of the Supporting Information).

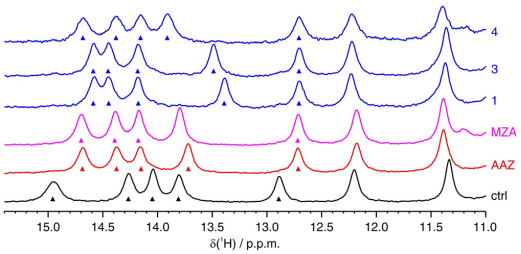
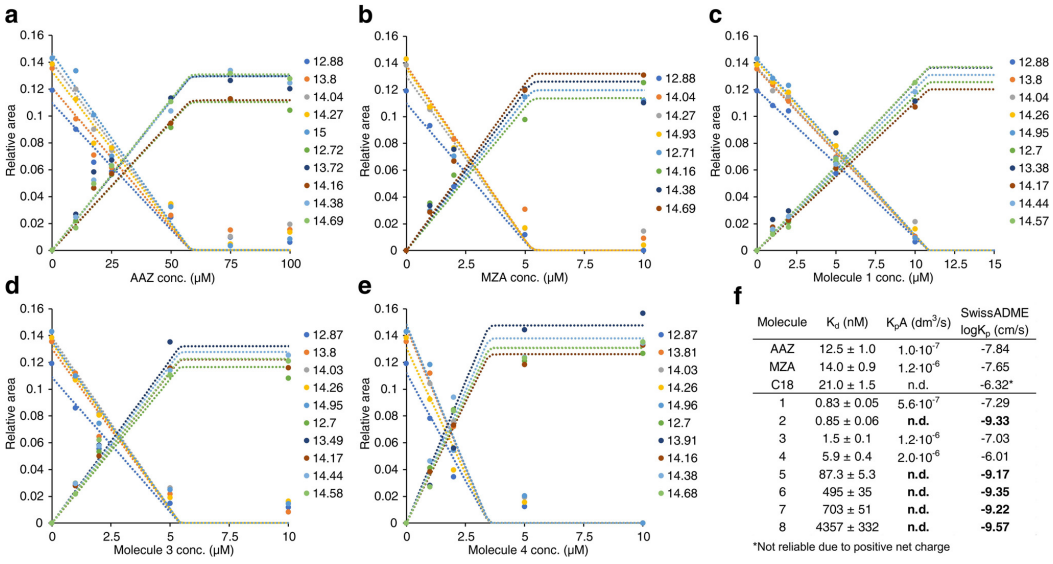


Figure 3. Imino region of the 1D ^{1}H NMR spectra of cells expressing CA2 in the absence of ligands (black) and treated with 100 μM of each ligand for 1 hour, showing complete binding to intracellular CA2: **AAZ** (red), **MZA** (magenta) and ligands **1**, **3**, **4** (blue). Peaks used to obtain binding curves are marked with arrows.

two categories of ligands emerged, that did not correlate with the differences in binding affinity: those that bind intracellular CA2 quantitatively, similar to **AAZ** and **MZA** (**1**, **3**, and **4**, Figure 3), and those for which binding is negligible, similar to **C18** (**2**, **5**, **6**, **7**, and **8**, Supporting Information, Figure S8). In comparison, ligand binding was clearly observed in vitro, both in the ^{1}H-^{15}N and in the ^{1}H NMR spectra (Supporting Information, Figure S9).

Therefore, a simple high-dose ligand screening by ^{1}H in-cell NMR could discriminate between "successful" and "unsuccessful" inhibitors without recurring to any cell-based activity assay, by directly observing intracellular binding. However, no insights were provided on the reasons why some ligands, even those with high in vitro affinity, do not bind intracellular CA2 at all. While several phenomena could prevent a ligand from binding an intracellular target, the most likely reasons are a low permeability of the plasma membrane and a low binding specificity to the target, that is, the presence of other intracellular molecules with comparatively high affinity for the same ligand. To provide a mechanistic hypothesis on the behaviour of the "unsuccessful" ligands, the "successful" ligands were analysed by in-cell NMR in a dose- and in a time-dependent manner. First, cells expressing CA2 and treated for 1 hour with increasing concentrations of each ligand were analysed by ^{1}H in-cell NMR and spectral deconvolution. Except **AAZ**, all the ligands followed a common trend, showing incomplete binding at low doses and reaching complete binding at around 5–15 μM (Figure 4 a–e). At the lowest ligand concentrations, the incomplete binding may be due to a lower-than-one ligand to CA2 molar ratio. Surprisingly, **AAZ** resulted in a much shallower dose-dependent binding curve, showing no binding below 10 μM and complete binding only around approximately 75 μM. Incomplete binding of **AAZ** at higher molar ratios suggests that it either diffuses slowly through the plasma membrane or strongly binds to other intracellular molecules. While competition binding is an equilibrium effect, slow diffusion kinetics should be observed in time-dependent binding experiments at fixed ligand concentration. Indeed, treating the cells with either 27 μM **AAZ** or 2 μM **MZA** for increasing time periods resulted in clear time-dependent



f

Molecule	K_d (nM)	K_pA (dm^3/s)	SwissADME logK_p (cm/s)
AAZ	12.5 ± 1.0	$1.0 \cdot 10^{-7}$	-7.84
MZA	14.0 ± 0.9	$1.2 \cdot 10^{-6}$	-7.65
C18	21.0 ± 1.5	n.d.	-6.32*
1	0.83 ± 0.05	$5.6 \cdot 10^{-7}$	-7.29
2	0.85 ± 0.06	**n.d.**	**-9.33**
3	1.5 ± 0.1	$1.2 \cdot 10^{-6}$	-7.03
4	5.9 ± 0.4	$2.0 \cdot 10^{-6}$	-6.01
5	87.3 ± 5.3	**n.d.**	**-9.17**
6	495 ± 35	**n.d.**	**-9.35**
7	703 ± 51	**n.d.**	**-9.22**
8	4357 ± 332	**n.d.**	**-9.57**

*Not reliable due to positive net charge

Figure 4. a–e) Dose-dependent binding curves observed in cells expressing CA2 treated for 1 hour with different ligands at increasing concentrations, fitted with a time-dependent binding equilibrium (See the Experimental Methods section of the Supporting Information). Each dot represents the area under a single peak in the imino region of the 1D ^{1}H in-cell NMR spectra, normalized to the total spectral area.
f) Dissociation constant measured in vitro (K_d), permeability coefficient × membrane area (K_pA) obtained from curve fitting and predicted skin permeability coefficient (log K_p) for each molecule. "Unsuccessful" molecules correlate with lower skin permeability (shown in bold).

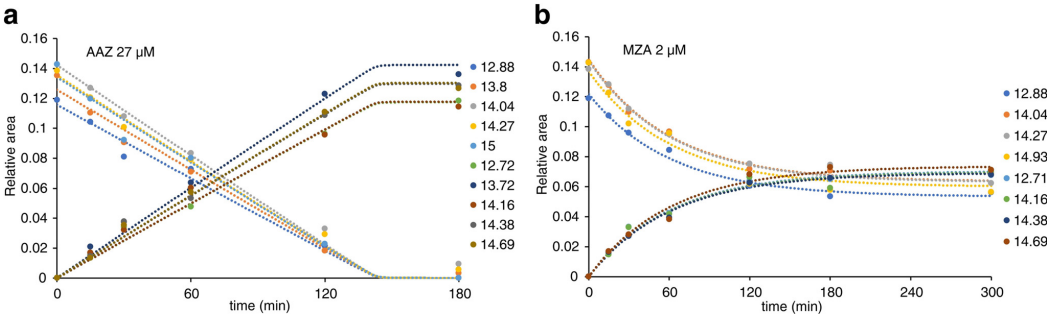


Figure 5. Time-dependent binding curves observed in cells expressing CA2 treated with either 27 µM **AAZ** (a) or 2 µM **MZA** (b) for increasing time periods. Binding of **AAZ** was fitted with a time-dependent binding equilibrium; binding of **MZA** was fitted with a time-dependent diffusion in deficiency of external ligand with respect to the protein (See the Experimental Methods section of the Supporting Information). Each dot represents the area under a single peak in the imino region of the 1D ^{1}H in-cell NMR spectra, normalized to the total spectral area.

binding curves (Figure 5). Dose- and time-dependent curves could then be fitted with a kinetic model to obtain membrane permeability coefficients, revealing that **AAZ** diffuses approximately 12-fold slower than **MZA** (Figure 4 f). As both molecules passively diffuse through the plasma membrane and do not rely on active transport,[26] such strikingly different behaviour should eventually affect drug permeability in human tissues. Indeed, the observed difference is reflected in the pharmacokinetic properties of the two drugs, as the recommended dosage for **AAZ** in the treatment of glaucoma is approximately 10-fold higher than that of **MZA**.[27] Therefore, the kinetics of membrane diffusion can greatly affect the behaviour of different ligands, irrespective of their binding affinity for the intracellular target. Interestingly, all the "unsuccessful" ligands screened here share a nitrogenous base, either a uracil (**2**, **6**, **7**, and **8**) or adenine (**5**) as a common feature, and have a much lower predicted skin permeability than the "successful" ligands (Figure 4 f), thus suggesting that the lack of binding of the former is likely the consequence of an exceedingly slow diffusion through the plasma membrane.

Once the cells are analysed under steady-state conditions, that is, after the ligand had enough time to reach the intracellular target, the in-cell binding curve can be fitted to obtain an apparent K_d. Comparison with the K_d determined in vitro would reveal the extent of competition binding, thereby providing a measure of intracellular binding specificity. To obtain a meaningful apparent K_d, the ligand to protein molar ratio should be higher than one. For this purpose, useful binding curves can be obtained from cells expressing the target protein at lower levels. Indeed, cells expressing an approximately 3-fold lower level of CA2 and treated for 2 hours with concentrations of **MZA** ranging from 0.1 to 2 µM resulted in a steeper binding curve compared to cells with high levels of CA2, giving an apparent K_d similar, within the error, to that measured in vitro (Supporting Information, Figure S10 a,b). Interestingly, treatment with **4** under the same conditions resulted in an apparent $K_d \approx 20$-fold higher than in vitro, thus suggesting that other molecules may compete for binding **4** within the cell (Supporting Information, Figure S10 c,d). Therefore, quantitative binding curves obtained by NMR can provide meaningful information to describe the kinetics and thermodynamics of intracellular ligand binding.

Finally, to assess to which extent the "intracellular protein-observed" approach can be generalized to other intracellular targets, ligand binding was tested on intracellular CA1. While CA1 reached lower expression levels than CA2, histidine amide protons were still detected in the 1D ^{1}H NMR spectra, and free and bound species could be clearly separated (Supporting Information, Figure S11). Therefore, the approach should be applicable to other intracellular CA isoforms, which contain conserved zinc-binding histidines, and in principle to any other protein that gives rise to ^{1}H signals downfield of approximately 11 ppm[19] or in any background-free spectral region. More in general, by recurring to selective isotopic labelling strategies, such as amino acid type-selective [^{13}C]-methyl or [^{15}N]-labelling, this approach can be applied to any soluble intracellular protein, provided that at least one signal is observable by in-cell NMR and is sensitive to ligand binding.

Herein, we show that ligand screening can be performed in human cells towards a specific protein by "intracellular protein-observed" NMR. Since the first proof-of-principle in human cells,[2] in-cell NMR application to protein-observed ligand screening has been limited to bacteria.[28] The approach shown here allows efficient in-depth drug screening in human cells for assessing intracellular target-binding capabilities, and also provides a way to characterize them in more quantitative terms, without recurring to chemical tagging of the protein.[29] The currently low throughput of the approach can be greatly improved by increasing automation (for example, through the use of a temperature-controlled NMR sample changer) and by screening multiple ligands simultaneously (for example, through matrix methods[28]). Importantly, this method does not rely on enzymatic activity measurements. Therefore, it provides a unique novel way to evaluate the effectiveness of drugs also against non-enzymatic targets, each of them otherwise requiring indirect cell-based assays to be established, such as cell proliferation, invasion, and viability/

apoptosis assays. Ultimately, once framed within the modern drug development pipeline, this method could allow the assessment of the potency of a candidate drug, that is, the amount of drug required to exert an effect of given intensity, in a clinically relevant concentration range (0.1–100 μM). Such predictive ability would allow the optimization of potency at an earlier stage of the pipeline compared to cell- and animal-based assays.

Acknowledgements

This work has been supported by iNEXT, grant number 653706, funded by the Horizon 2020 programme of the European Commission, and by Instruct-ULTRA, grant number 731005, an EU H2020 project to further develop the services of Instruct-ERIC. The authors acknowledge the support and the use of resources of Instruct-ERIC, a Landmark ESFRI project, and specifically the CERM/CIRMMP Italy Centre.

Conflict of interest

The authors declare no conflict of interest.

Keywords: drug design · drug screening · in-cell NMR spectroscopy · structural biology · sulfonamide

How to cite: *Angew. Chem. Int. Ed.* **2020**, *59*, 6535–6539
Angew. Chem. **2020**, *132*, 6597–6601

[1] H. Dowden, J. Munro, *Nat. Rev. Drug Discovery* **2019**, *18*, 495–496.
[2] K. Inomata, A. Ohno, H. Tochio, S. Isogai, T. Tenno, I. Nakase, T. Takeuchi, S. Futaki, Y. Ito, H. Hiroaki, et al., *Nature* **2009**, *458*, 106–109.
[3] F.-X. Theillet, A. Binolfi, B. Bekei, A. Martorana, H. M. Rose, M. Stuiver, S. Verzini, D. Lorenz, M. van Rossum, D. Goldfarb, et al., *Nature* **2016**, *530*, 45–50.
[4] E. Luchinat, L. Banci, *Acc. Chem. Res.* **2018**, *51*, 1550–1557.
[5] S. Dzatko, M. Krafcikova, R. Hänsel-Hertsch, T. Fessl, R. Fiala, T. Loja, D. Krafcik, J.-L. Mergny, S. Foldynova-Trantirkova, L. Trantirek, *Angew. Chem. Int. Ed.* **2018**, *57*, 2165–2169; *Angew. Chem.* **2018**, *130*, 2187–2191.
[6] T. Tanaka, T. Ikeya, H. Kamoshida, Y. Suemoto, M. Mishima, M. Shirakawa, P. Güntert, Y. Ito, *Angew. Chem. Int. Ed.* **2019**, *58*, 7284–7288; *Angew. Chem.* **2019**, *131*, 7362–7366.
[7] M. Krafcikova, S. Dzatko, C. Caron, A. Granzhan, R. Fiala, T. Loja, M.-P. Teulade-Fichou, T. Fessl, R. Hänsel-Hertsch, J.-L. Mergny, et al., *J. Am. Chem. Soc.* **2019**, *141*, 13281–13285.
[8] G. Siegal, P. Selenko, *J. Magn. Reson.* **2019**, *306*, 202–212.
[9] V. Alterio, A. Di Fiore, K. D'Ambrosio, C. T. Supuran, G. De Simone, *Chem. Rev.* **2012**, *112*, 4421–4468.
[10] C. T. Supuran, *Nat. Rev. Drug Discovery* **2008**, *7*, 168–181.
[11] C. T. Supuran, *Biochem. J.* **2016**, *473*, 2023–2032.
[12] D. Neri, C. T. Supuran, *Nat. Rev. Drug Discovery* **2011**, *10*, 767–777.
[13] C. T. Supuran, *J. Enzyme Inhib. Med. Chem.* **2016**, *31*, 345–360.
[14] C. T. Supuran, *Expert Opin. Drug Discovery* **2017**, *12*, 61–88.
[15] C. T. Supuran, *Expert Opin. Ther. Pat.* **2018**, *28*, 709–712.
[16] C. T. Supuran, *Expert Opin. Ther. Pat.* **2018**, *28*, 713–721.
[17] C. T. Supuran, *Expert Opin. Invest. Drugs* **2018**, *27*, 963–970.
[18] E. Berrino, C. T. Supuran, *Expert Opin. Drug Discovery* **2019**, *14*, 231–248.
[19] L. Banci, L. Barbieri, I. Bertini, E. Luchinat, E. Secci, Y. Zhao, A. R. Aricescu, *Nat. Chem. Biol.* **2013**, *9*, 297–299.
[20] L. Barbieri, E. Luchinat, L. Banci, *Nat. Protoc.* **2016**, *11*, 1101–1111.
[21] L. Barbieri, E. Luchinat, L. Banci, *Sci. Rep.* **2015**, *5*, 14456.
[22] K. H. Sippel, A. H. Robbins, J. Domsic, C. Genis, M. Agbandje-McKenna, R. McKenna, *Acta Crystallogr. Sect. F* **2009**, *65*, 992–995.
[23] A. Nocentini, S. Bua, C. L. Lomelino, R. McKenna, M. Menicatti, G. Bartolucci, B. Tenci, L. Di Cesare Mannelli, C. Ghelardini, P. Gratteri, et al., *ACS Med. Chem. Lett.* **2017**, *8*, 1314–1319.
[24] A. Nocentini, M. Ferraroni, F. Carta, M. Ceruso, P. Gratteri, C. Lanzi, E. Masini, C. T. Supuran, *J. Med. Chem.* **2016**, *59*, 10692–10704.
[25] A. Rogato, S. Del Prete, A. Nocentini, V. Carginale, C. T. Supuran, C. Capasso, *J. Enzyme Inhib. Med. Chem.* **2019**, *34*, 510–518.
[26] T. H. Maren, *Physiol. Rev.* **1967**, *47*, 595–781.
[27] R. L. Stamper, M. F. Lieberman, M. V. Drake, in *Becker-Shaffer's Diagnosis and Therapy of the Glaucomas*, 8th ed. (Eds.: R. L. Stamper, M. F. Lieberman, M. V. Drake), Mosby, Edinburgh, **2009**, pp. 407–419.
[28] C. M. DeMott, R. Girardin, J. Cobbert, S. Reverdatto, D. S. Burz, K. McDonough, A. Shekhtman, *ACS Chem. Biol.* **2018**, *13*, 733–741.
[29] Y. Takaoka, Y. Kioi, A. Morito, J. Otani, K. Arita, E. Ashihara, M. Ariyoshi, H. Tochio, M. Shirakawa, I. Hamachi, *Chem. Commun.* **2013**, *49*, 2801–2803.

Manuscript received: October 21, 2019
Revised manuscript received: December 9, 2019
Accepted manuscript online: February 5, 2020
Version of record online: February 25, 2020

2.2.4 Supplementary information

Supporting Information

Drug Screening in Human Cells by NMR Spectroscopy Allows the Early Assessment of Drug Potency

*Enrico Luchinat, Letizia Barbieri, Matteo Cremonini, Alessio Nocentini, Claudiu T. Supuran, and Lucia Banci**

anie_201913436_sm_miscellaneous_information.pdf

Contents

1. Experimental Methods

1.1. Gene cloning

To generate the mammalian expression plasmids, the cDNAs encoding full-length human CA2 (amino acids 1–260, GenBank: NP_000058.1) and human CA1 (amino acids 1–261, GenBank: NP_001122303.1) were amplified by PCR and cloned into the pHLsec vector[1] between EcoRI and XhoI restriction enzyme sites (CA2) and between HindIII and XhoI restriction enzyme sites (CA1), following a previously reported cloning strategy.[2] These restriction sites were chosen in order to remove a N-terminal signal peptide and a C-terminal histidine tag present in the original vector, so that the expression vectors obtained with this strategy encode the native protein sequences. Specifically, CA2 was amplified using the following primers:

Forward: 5'-CCGGAATTCGCCACCATGGCCCATCACTGGGGGTACGGC-3';

Reverse: 5'-CCGCTCGAGTTATTTGAAGGAAGCTTTGATTTGCCTGTTCTTC-3';

CA1 was amplified using the following primers:

Forward: 5'-CCCAAGCTTGCCACCATGGCAAGTCCAGACTGGGGATATGATG-3';

Reverse: 5'-CCGCTCGAGTCAAAATGAAGCTCTCACTGTTCTGCCCTTC-3'.

A Kozak sequence was inserted in both forward primers downstream of the EcoRI/HindIII site, while a stop codon was inserted in both reverse primers upstream of the XhoI site. The clones were verified by DNA sequencing.

1.2. Human cell culture and transfection

HEK293T (ATCC CRL-3216) cells were maintained in Dulbecco-modified Eagle medium (DMEM) high glucose (Gibco) supplemented with L-glutamine, antibiotics (penicillin and streptomycin) and 10% fetal bovine serum (FBS, Gibco) in uncoated 75 cm^2 plastic flasks and incubated at 37 °C, 5% CO2 in a humidified atmosphere. Cells were transiently transfected with the pHLsec plasmid containing the gene of CA2 or CA1 using polyethylenimine (PEI), with a DNA:PEI ratio of 1:2 (25 µg/flask DNA, 50 µg/flask PEI). Lower expression levels of CA2 were obtained by transfecting cells with a mixture of 5 µg CA2 DNA and 20 µg empty vector. DMEM medium was used for unlabelled in-cell NMR samples; [U-^{15}N]-BioExpress6000 medium (Cambridge Isotope Laboratories) was used for uniform ^{15}N in-cell NMR samples. Expression media were supplemented with 2% FBS, antibiotics and 10 µM $ZnSO_4$.

1.3. Protein quantification and intracellular distribution

The expression level of CA2 was determined by Coomassie-stained SDS-PAGE. Densitometry analysis was performed with ImageJ. Lysates from cell samples were run at increasing dilutions together with

purified CA2 as a reference. CA2 expression at lower levels was assessed by SDS-PAGE of lysates from cells transfected with lower amounts of CA2 DNA (Figure S12a). CA1 level was estimated by comparison with CA2 (Figure S12b). The values reported in the main text reflect the protein concentrations calculated from cells lysed in 1 cell pellet-volume, therefore corresponding to the effective concentrations in the in-cell NMR samples (mean value ± s.d., n =3). Intracellular distribution was assessed by separating the insoluble fraction (containing the nuclei and other membranous fractions), the cytosolic and the mitochondrial fraction from cell extracts using a mitochondria isolation kit for cultured cells (Thermo Fisher Scientific), followed by analysis via SDS-PAGE of each fraction at serial dilutions. Following the separation, >90% of CA2 was recovered in the cytosolic fraction, whereas <5% was recovered in the insoluble fraction and <2% was recovered in the mitochondrial fraction (Figure S12d).

1.4. Treatment with inhibitors

HEK293T cells overexpressing CA2 or CA1 were treated with each inhibitor 48 hours post-transfection. Dose-dependent binding curves were obtained by treating different cell cultures with increasing amounts of each ligand for a fixed time (1 hour, unless specified differently in the main text). Time-dependent binding curves were obtained by treating different cell cultures with a fixed amount of AAZ or MZA for increasing amounts of time. The cellular response to AAZ and MZA in terms of cell viability was assessed by trypan blue staining of cells overexpressing CA2 and subsequently treated for 1 hour with increasing concentrations of AAZ or MZA (0, 1, 10, 100 μM). Viability remained > 95% in all samples (Figure S12c).

1.5. Stopped-flow measurements

An Applied Photophysics stopped-flow instrument was used for assaying the CA2-catalysed CO_2 hydration activity.[3] CA2 was purchased from Sigma-Aldrich. Phenol red (at a concentration of 0.2 mM) was used as indicator, working at the absorbance maximum of 557 nm, with 20 mM HEPES (pH 7.5) as buffer, and 20 mM Na_2SO_4 (for maintaining constant the ionic strength), following the initial rates of the CA2-catalysed CO_2 hydration reaction for a period of 10–100 s. The CO_2 concentrations ranged from 1.7 to 17 mM for the determination of the kinetic parameters and inhibition constants. For each inhibitor at least six traces of the initial 5–10% of the reaction have been used for determining the initial velocity. The uncatalyzed rates were determined in the same manner and subtracted from the total observed rates. Stock solutions of inhibitor (0.1 mM) were prepared in distilled-deionized water and dilutions up to 0.01 nM were done thereafter with the assay buffer. Inhibitor and enzyme solutions were preincubated together for 15 min at room temperature prior to assay, in order to allow for the formation of the E-I complex. The inhibition constants were obtained by non-linear least-squares methods using PRISM 3 and the Cheng–Prusoff equation and represent the mean (± s.d.) from at least three different determinations.

Cell lysates were prepared by freeze-thaw cycles in phosphate buffered saline (PBS), pH 7.4 followed by centrifugation for 1 hour at 16000 g, 4°C. The supernatant from each sample was collected. Total protein concentration in each sample was measured by BCA Protein Assay (Thermo Fisher), and the sample volumes were adjusted with PBS to equalize total protein concentration, and subsequently diluted 400-fold. Total CO_2 hydration activity in the cell lysates was measured by stopped-flow using the same protocol described above.

1.6. Expression and purification of recombinant CA2

Recombinant CA2 for in vitro experiments was prepared following an existing protocol.[4] Briefly, a 1-liter cell culture of *E. Coli* BL21(DE3) Codon Plus Ripl (Stratagene) transformed with a pCAM plasmid containing the CA2 gene was grown overnight at 37°C in LB, harvested and re-suspended in 1 liter of ^{15}N-labelled M9 medium. $ZnSO_4$ was added in the culture to a final concentration of 500 µM. After 5 h from induction with 1 mM IPTG at 37°C the cells were harvested and re-suspended in 20 mM Tris, pH 8 buffer for lysis. The cleared lysate was loaded onto a nickel chelating HisTrap (GE Healthcare) 5 ml column. The protein was eluted with a linear gradient of 20 mM Tris pH 8, 500 mM imidazole. The fractions containing pure CA2 were collected. Finally, the protein was exchanged in NMR buffer (PBS pH 7.4, Gibco, supplemented with 10% D_2O). The correct metalation of the protein was confirmed by chemical shift comparison against previously reported spectra.

1.7. In vitro NMR experiments

Samples of pure CA2 (170 µM in NMR buffer), both unlabelled and ^{15}N-labelled, were placed in 5 mm NMR tubes and analyzed at 310 K at a 900 MHz Bruker Avance HD spectrometer equipped with a TCI CryoProbe. 1D ^{1}H WATERGATE 3-9-19 (on the unlabelled protein) and 2D ^{1}H-^{15}N SOFAST-HMQC (on the ^{15}N-labelled protein) spectra were acquired both in the absence of ligands and upon addition of 1 equivalent of each ligand. The same CA2 samples were then titrated with up to 2 equivalents of each ligand and analyzed by NMR to exclude the presence of high-affinity secondary binding sites. For some ligands, poor solubility in aqueous buffer caused precipitate formation, without affecting protein solubility. In those circumstances, complete binding was obtained by adding step-wise 2-3 equivalents to the protein, spinning down after each addition to remove the precipitate.

1.8. In-cell NMR experiments

Samples for in-cell NMR were prepared following a reported protocol.[2] Briefly, cells overexpressing CA2/CA1 were detached with trypsin, suspended in DMEM + 10% FBS, washed once with PBS and re-suspended in one pellet volume of DMEM supplemented with 90 mM glucose, 70 mM HEPES and 20% D_2O. The cell suspension was transferred in a 3 mm Shigemi NMR tube, which was gently spun to sediment the cells. Cell viability before and after NMR experiments was assessed by trypan blue staining. In-cell NMR spectra were collected at 310 K at a 900 MHz Bruker Avance HD spectrometer

equipped with a TCI CryoProbe. 2D ^{1}H-^{15}N SOFAST-HMQC spectra were recorded on ^{15}N-labelled cell samples; 1D ^{1}H WATERGATE 3-9-19 spectra were recorded on unlabelled cell samples. The acquisition time for cells expressing CA2 at high levels was ~1 hour for 2D NMR spectra and ~30 minutes for 1D NMR spectra. 3-hours long 1D NMR spectra were recorded on unlabelled cells expressing either CA2 at lower levels or CA1. To preserve cell viability, two identical cell samples for each condition were analyzed for 1.5 hour each and the obtained spectra were accumulated during processing. After the NMR experiments, the cells were collected and the supernatant was checked for protein leakage by NMR.

1.9. NMR data processing and signal deconvolution

The NMR spectra were acquired and processed with Bruker Topspin software. The 2D in-cell NMR spectra of ^{15}N-labelled cells were further processed by subtracting an identical spectrum recorded on cells transfected with empty vector, acquired in the same experimental conditions, to eliminate the signals arising from partial ^{15}N incorporation in other cellular components. Backbone amide resonances were assigned based on available data[5]. Combined chemical shift difference (CCSD) for each amide crosspeak was calculated using the formula:

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

Signals in the 1D ^{1}H in-cell NMR spectra were deconvoluted using Fityk software.[6] The spectral region between 11 and 16 ppm was isolated and globally fitted with a sum of N+1 pseudo-Voigt functions (with gaussian weight fixed at 0.15), where N equals the number of discernible signals in the region. The additional function accounted for baseline distortions caused by strong signals outside the spectral region. Relative signal intensities were obtained as the ratio of the area under each fitted peak and the sum of the N areas. Signals that were not perturbed by ligand binding or could not be separated due to high overlap were discarded. Signal intensities were plotted against ligand concentration or time to obtain dose- and time-dependent binding curves, respectively.

1.10. Binding curve fitting

The binding curves were fitted with either dose- and time-dependent or dose-dependent, time-independent binding equations using OriginPro software. Equations were obtained from a model accounting for passive diffusion of the ligand through the plasma membrane: $L_o \rightleftarrows L_i$, followed by binding to the intracellular protein: $L_i + P \rightleftarrows LP$, where L_o and L_i are the extracellular and intracellular ligand, respectively, and P and LP are the free and the bound protein, respectively. Passive diffusion was considered the rate-determining step, with the rate equation:

$$\frac{d[L_{t.in}]}{dt} = \frac{d([L_i] + [LP])}{dt} = \frac{K_p A}{V_c}([L_o] - [L_i])$$

where $[L_{t.in}]$ is the total intracellular ligand concentration, K_p is the permeability coefficient, A is the total area of the membrane and V_c is the intracellular volume. Free and bound protein were considered to reach the equilibrium instantly at each value of $[L_{t.in}]$, therefore the bound protein [LP] was obtained from the equilibrium binding equation:

$$[LP](t) = \frac{K_d + [P_t] + [L_{t.in}](t) - \sqrt{\left(K_d + [P_t] + [L_{t.in}](t)\right)^2 - 4[P_t][L_{t.in}](t)}}{2}$$

where $[P_t]$ is the total intracellular protein concentration and K_d is the dissociation constant. The above equation was used to fit both the dose-dependent and the time-dependent binding curves (rate-limited by membrane diffusion), to obtain values of $K_p \cdot A$, by expressing $[L_{t.in}](t)$ with the integral diffusion rate equation approximated for $[L_i] << [L_o]$:

$$L_{t.in}(t) = L_t\left(1 - e^{-\frac{K_p A}{V_t}t}\right)$$

where $L_t = L_o + L_i + LP$ are the total moles of ligand. Time-dependent curves in defect of external ligand were fitted with a simplified model where the free intracellular ligand $[L_i]$ is considered in steady-state and at negligible concentration, and LP(t) is directly obtained from the membrane diffusion rate:

$$\frac{d[LP]}{dt} = \frac{K_p A}{V_c}([L_o] - [L_i])\ ;\ LP(t) = L_t\left(1 - e^{-\frac{K_p A}{V_t}t}\right)$$

Finally, to obtain an apparent K_d, dose-dependent curves obtained at longer times were fitted with the time-independent binding equilibrium equation calculated for $[L_o] = [L_i] = [L]$:

$$[LP] = \frac{K_d + [P_t]V_r + [L_t] - \sqrt{(K_d + [P_t]V_r + [L_t])^2 - 4[P_t][L_t]V_r}}{2V_r}$$

where $V_r = V_c/V_t$ is the cell/total volume ratio and $[L_t] = [L] + [LP]V_r$.

References

[1] A. R. Aricescu, W. Lu, E. Y. Jones, *Acta Crystallogr. D Biol. Crystallogr.* **2006**, *62*, 1243–1250.
[2] L. Barbieri, E. Luchinat, L. Banci, *Nat. Protoc.* **2016**, *11*, 1101–1111.
[3] R. G. Khalifah, *J. Biol. Chem.* **1971**, *246*, 2561–2573.
[4] L. Cerofolini, S. Giuntini, A. Louka, E. Ravera, M. Fragai, C. Luchinat, *J. Phys. Chem. B* **2017**, *121*, 8094–8101.
[5] W. L. Nettles, H. Song, E. R. Farquhar, N. C. Fitzkee, J. P. Emerson, *Inorg. Chem.* **2015**, *54*, 5671–5680.
[6] M. Wojdyr, *J. Appl. Crystallogr.* **2010**, *43*, 1126–1128.
[7] H. Shimahara, T. Yoshida, Y. Shibata, M. Shimizu, Y. Kyogoku, F. Sakiyama, T. Nakazawa, S. Tate, S. Ohki, T. Kato, et al., *J. Biol. Chem.* **2007**, *282*, 9646–9656.
[8] S. K. Vasa, H. Singh, K. Grohe, R. Linser, *Angew. Chem. Int. Ed Engl.* **2019**, *58*, 5758–5762.

2. Supplementary Figures

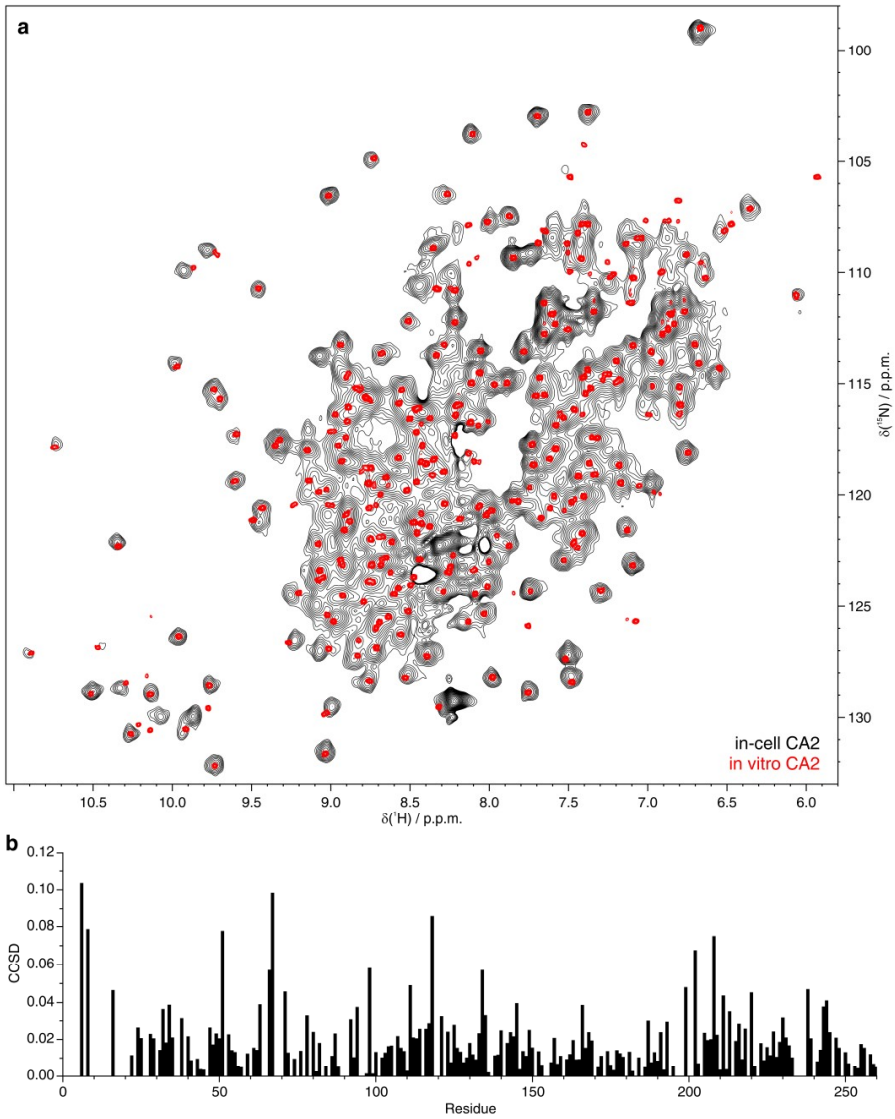


Figure S1. (**a**) Overlay of ^{1}H-^{15}N NMR spectra of cells expressing CA2 (black) and in vitro CA2 (175 µM, red). (**b**) Combined ^{1}H-^{15}N chemical shift difference (CCSD) between in vitro and in-cell CA2.

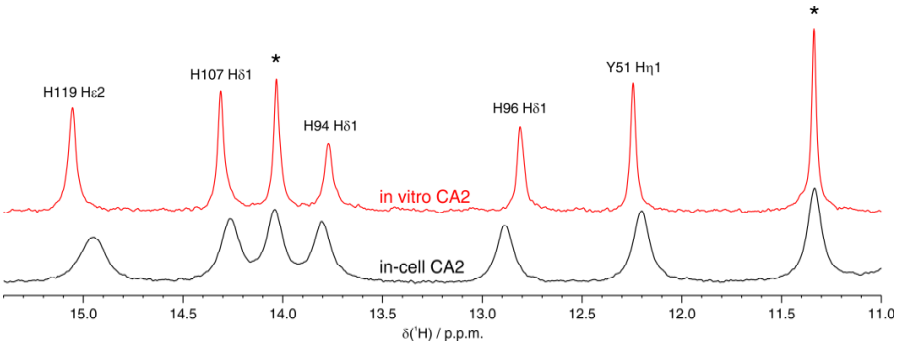


Figure S2. Imino region of the 1D ^{1}H NMR spectrum of CA2 in vitro (red) and in-cell (black). Signals for which the unambiguous assignment has been reported previously are labelled with the corresponding residue number and atom type.[7,8] Signals arising from unassigned protons are labelled with an asterisk. H94, H96 and H119 are the zinc-coordinating histidines in the active site of CA2.

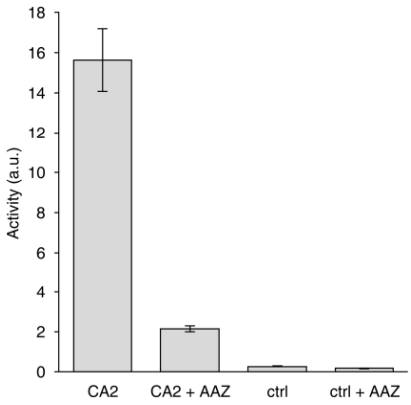


Figure S3. Total CO_2 hydrase activity measured on lysates from HEK293T cells overexpressing CA2 (CA2) and control cells with basal levels of CA2 (ctrl), either untreated or treated with 100 µM AAZ for 1 hour. Error bars represent s.d. (n = 3 independent samples).

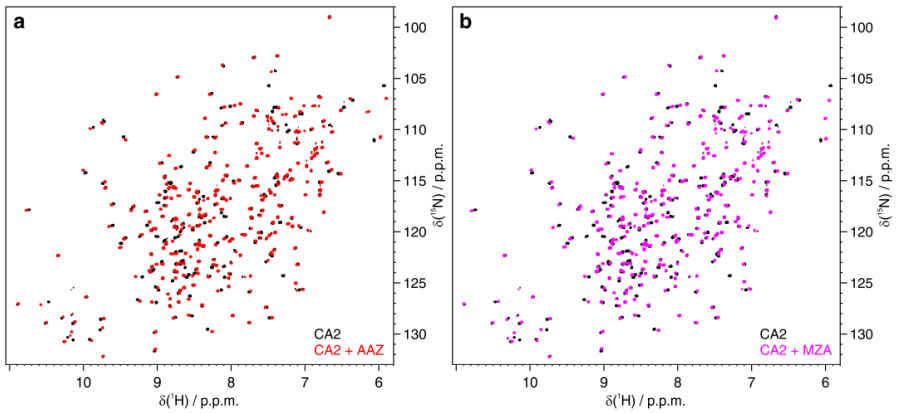


Figure S4. Overlay of ^{1}H-^{15}N NMR spectra of CA2 in vitro (175 μM) in the absence of ligands (black) and in the presence of 1 equivalent of (**a**) AAZ (red) or (**b**) MZA (magenta). The appearance of new sets of peaks in the presence of AAZ/MZA together with the complete disappearance of signals arising from free CA2 indicate quantitative ligand binding.

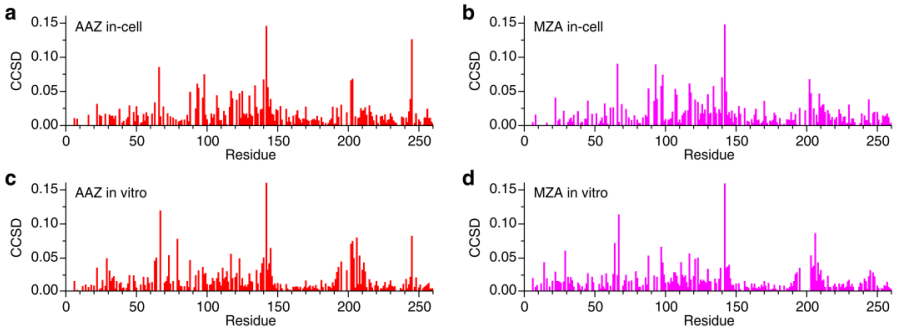


Figure S5. (**a, b**) in-cell ^{1}H-^{15}N CCSD between CA2 in the absence of ligands and bound to (**a**) AAZ; (**b**) MZA. (**c, d**) in vitro ^{1}H-^{15}N CCSD between CA2 in the absence of ligands and bound to (**c**) AAZ; (**d**) MZA.

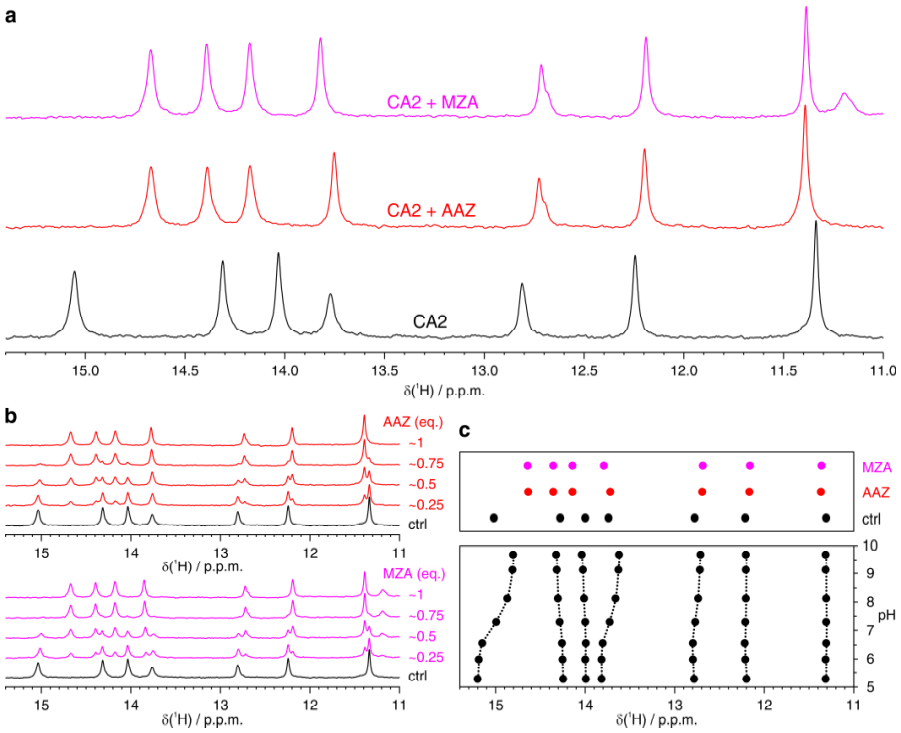
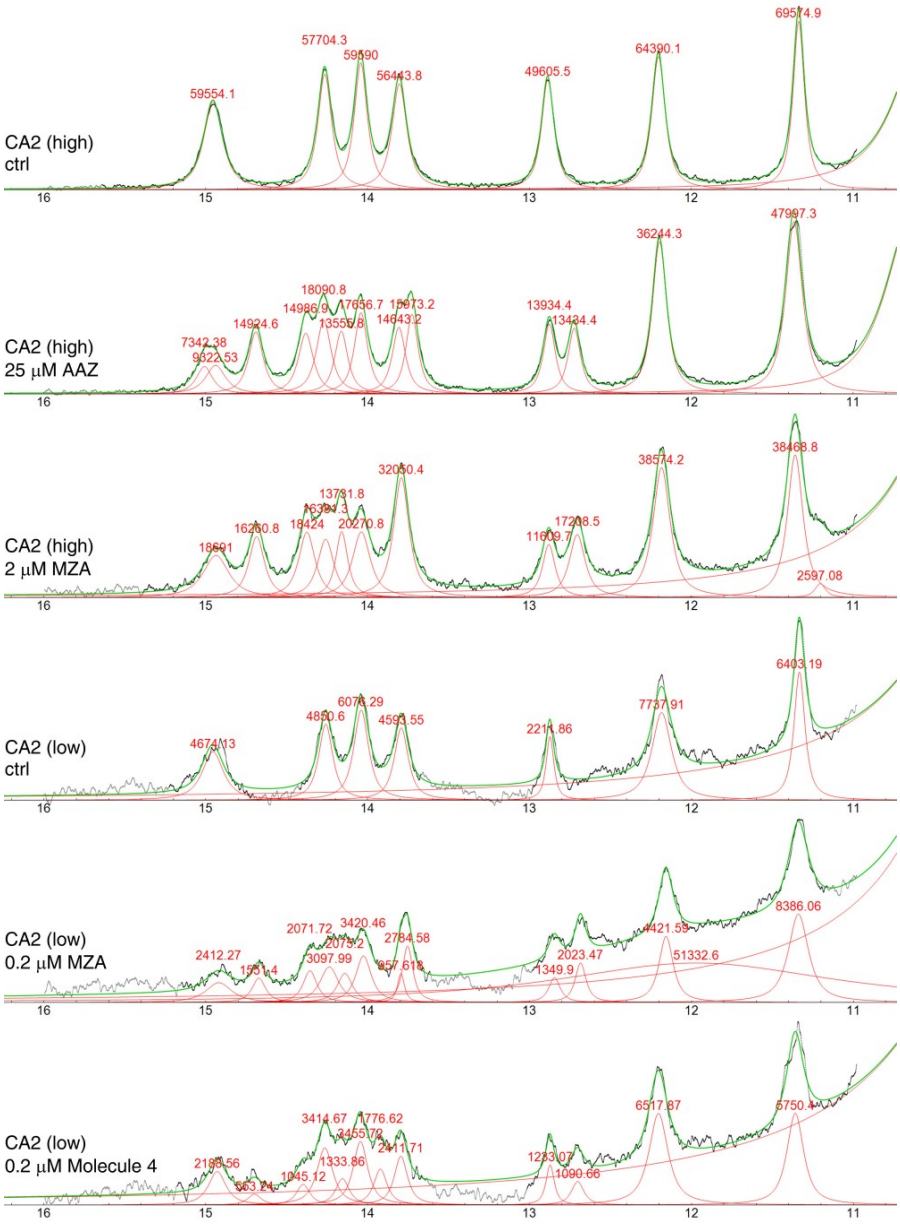


Figure S6. (**a**) Imino region of the 1D ^{1}H NMR spectra of in vitro CA2 in the absence of ligands (black) and bound to 1 equivalent of AAZ (red) or MZA (magenta). (**b**) 1D ^{1}H NMR spectra of in vitro CA2 in the absence of ligands (black) and upon addition of sub-stoichiometric amounts of AAZ (red) or MZA (magenta). Peaks arising from both free and bound CA2 are observed simultaneously. (**c**) Plot of the ^{1}H chemical shifts observed in the same spectral region as in (**a**) and (**b**) in the in vitro NMR spectra of free CA2 (black), AAZ-bound CA2 (red) and MZA-bound CA2 (magenta), measured at pH 7.4, compared to the ^{1}H chemical shift plot of free CA2 as a function of pH.


CA2 (high)
ctrl
59554.1
57704.3
59590
56443.8
49605.5
64390.1
69574.9
CA2 (high)
25 µM AAZ
7342.38
9522.53
14924.6
18090.8
14986.9
17656.7
15973.2
13555.8
14643.2
13934.4
13434.4
36244.3
47997.3
CA2 (high)
2 µM MZA
18691
16260.8
18424
13781.8
20270.8
32050.4
11609.7
17208.5
38574.2
38468.8
2597.08
CA2 (low)
ctrl
4674.13
4850.6
6076.29
2211.86
7737.91
6403.19
CA2 (low)
0.2 µM MZA
2412.27
1551.4
2071.72
3420.46
2075.2
3097.99
2784.58
857.618
1349.9
2023.47
4421.58
51332.6
8386.06
CA2 (low)
0.2 µM Molecule 4
2185.56
563.24
1045.12
3414.67
1776.62
3455.72
1333.86
2411.71
1233.07
1090.66
6517.87
5750.4
16
15
14
13
12
11

12

Figure S7. Spectral deconvolution in the imino region of different in-cell NMR spectra expressing CA2 at high (~150 μM) or low (~45 μM) levels. Both untreated cells (ctrl) and cells treated with concentrations of ligands that result in the incomplete binding are shown. The raw NMR data (black), single peak Pseudo-Voigt functions (red) and the overall sum (green) are shown. Calculated peak areas are indicated in red. In some instances, peaks with non-ideal shapes were fitted as the sum of two Pseudo-Voigt functions. Additional functions were used to correct for baseline distortions, the area of which was not used in the normalization process.

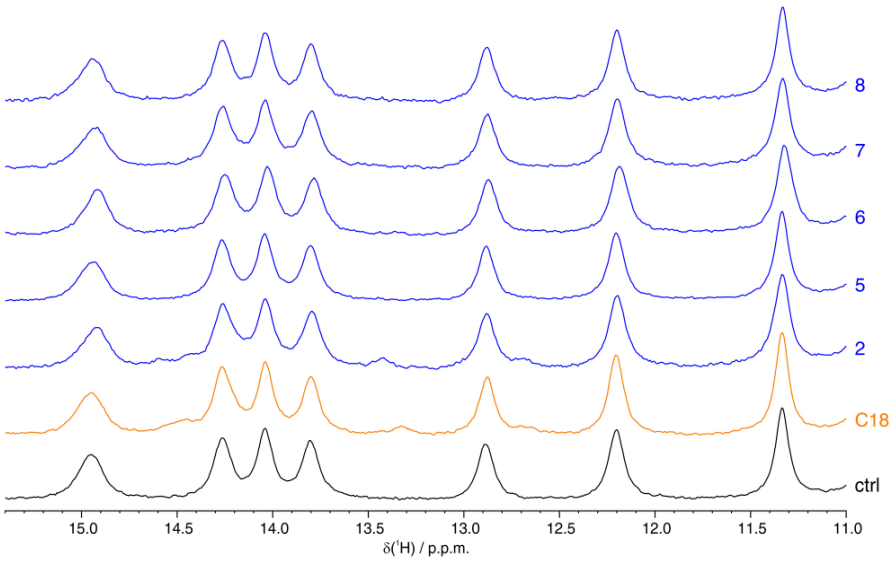
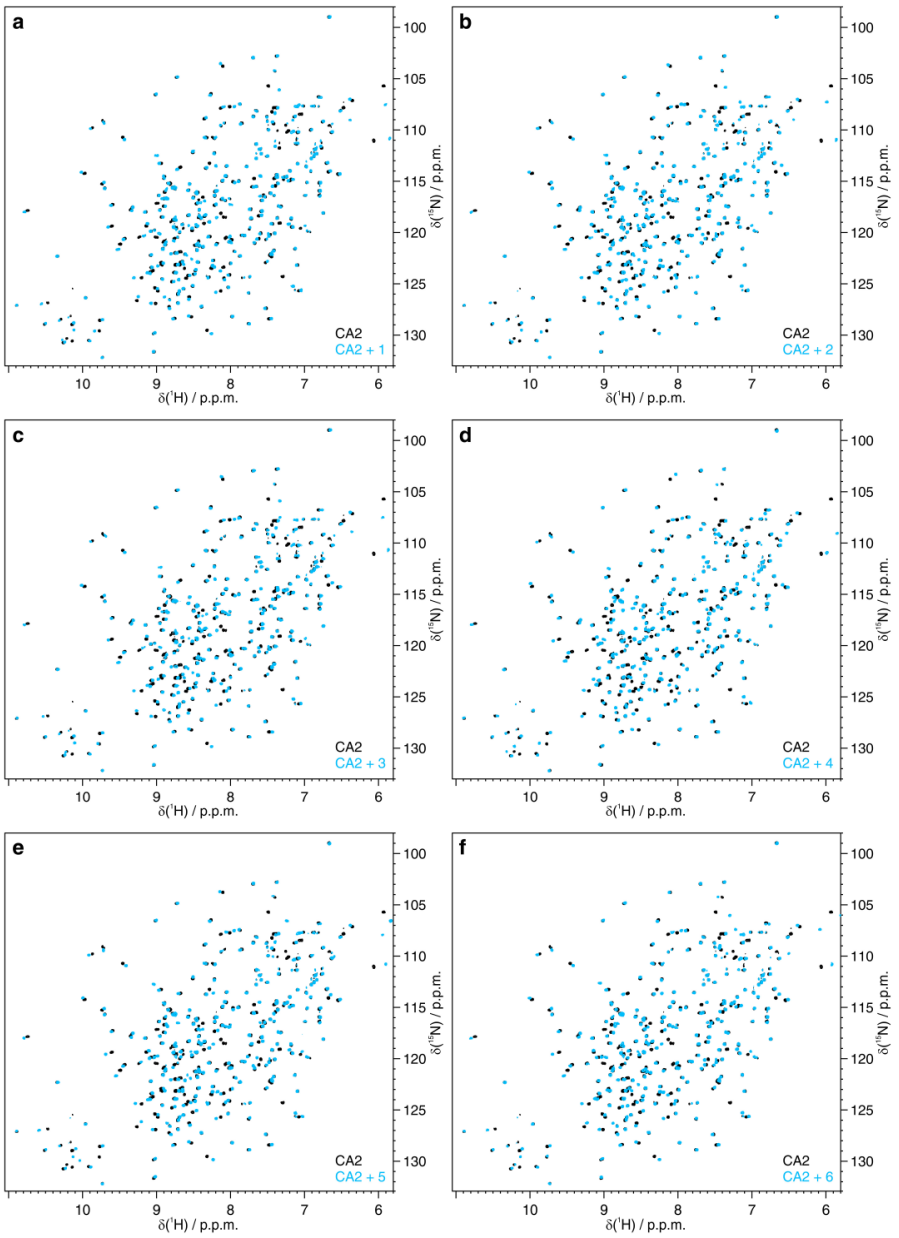


Figure S8. Imino region of the 1D ^{1}H NMR spectra of cells expressing CA2 in the absence of ligands (black) and treated with ligands showing negligible binding to intracellular CA2: C18 (orange), 2, 5, 6, 7 and 8 (blue).


a
CA2
CA2 + 1
b
CA2
CA2 + 2
c
CA2
CA2 + 3
d
CA2
CA2 + 4
e
CA2
CA2 + 5
f
CA2
CA2 + 6
δ(¹H) / p.p.m.
δ(¹⁵N) / p.p.m.

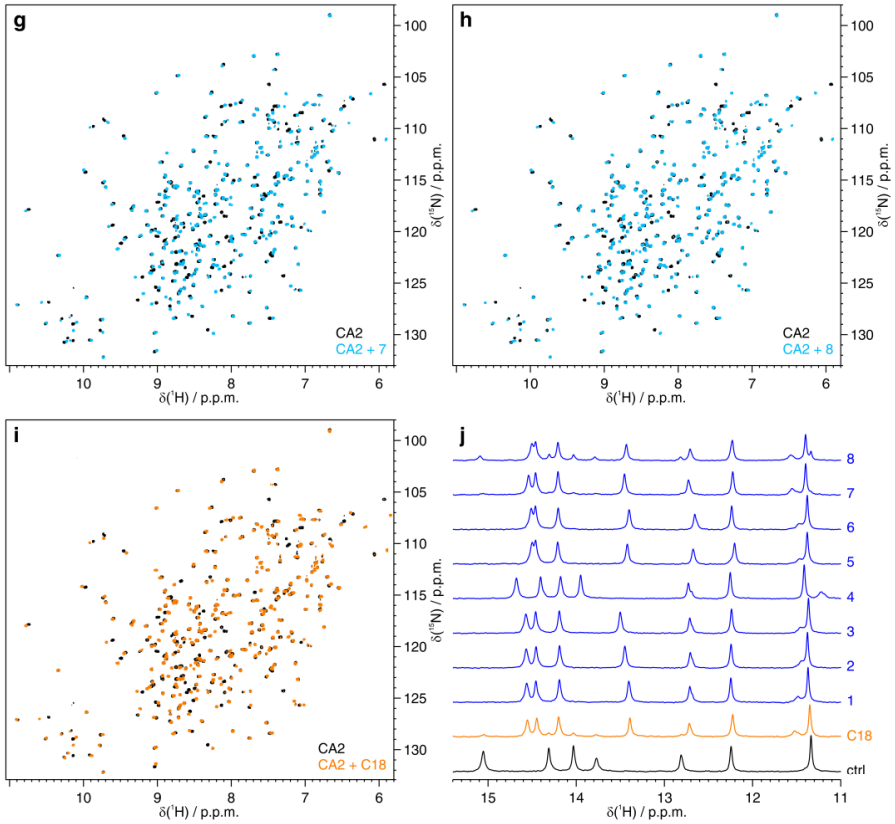


Figure S9. Overlay of ^{1}H-^{15}N NMR spectra of CA2 in vitro (175 µM) in the absence of ligands (black) and in the presence of 1 equivalent of (**a-h**) ligands 1 to 8 (light blue) or (**i**) C18 (orange). (**j**) Imino region of the 1D ^{1}H NMR spectra of CA2 in vitro in the absence of ligands (black) and bound to ligands 1 to 8 (blue) or C18 (orange). Complete binding was not observed for the lowest-affinity ligand, 8, likely due to its precipitation in aqueous buffer at the concentrations required.

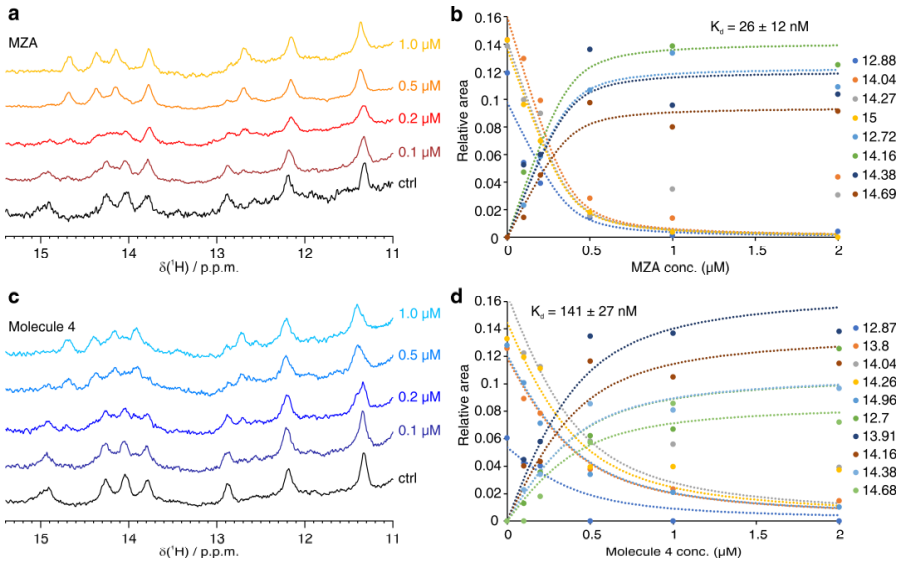


Figure S10. (**a, c**) Imino region of the 1D ^{1}H NMR spectra of cells expressing CA2 at lower levels (~45 µM in the NMR tube) in the absence of ligands (black) and treated for 2 hours with increasing concentrations of MZA (**a**, brown to yellow) or Molecule 4 (**c**, dark blue to cyan) at increasing concentrations. (**b, d**) Dose-dependent binding curves obtained from the NMR spectra shown in (**a**) and (**c**), respectively, fitted with a time-independent binding equation (see Experimental Methods). Binding of MZA reaches 50% at around 0.3 µM, giving an apparent K_d of 26 ± 12 nM (**b**), close to the K_d measured in vitro (14 ± 0.9 nM). Instead, binding of Molecule 4 requires a higher dose, giving an apparent K_d of 141 ± 27 nM (**d**), ~20-fold higher than in vitro (5.9 ± 0.4 nM), which suggests the presence of competing binding sites within the cell.

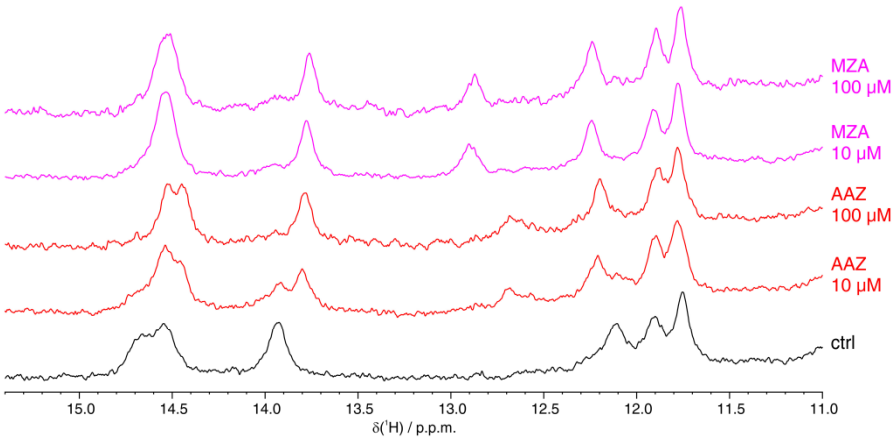


Figure S11. Imino region of the 1D ^{1}H NMR spectra of cells expressing CA1 (~50 µM in the NMR sample) in the absence of ligands (black) and in the presence of AAZ 10 µM and 100 µM (red), or MZA 10 µM and 100 µM (magenta).

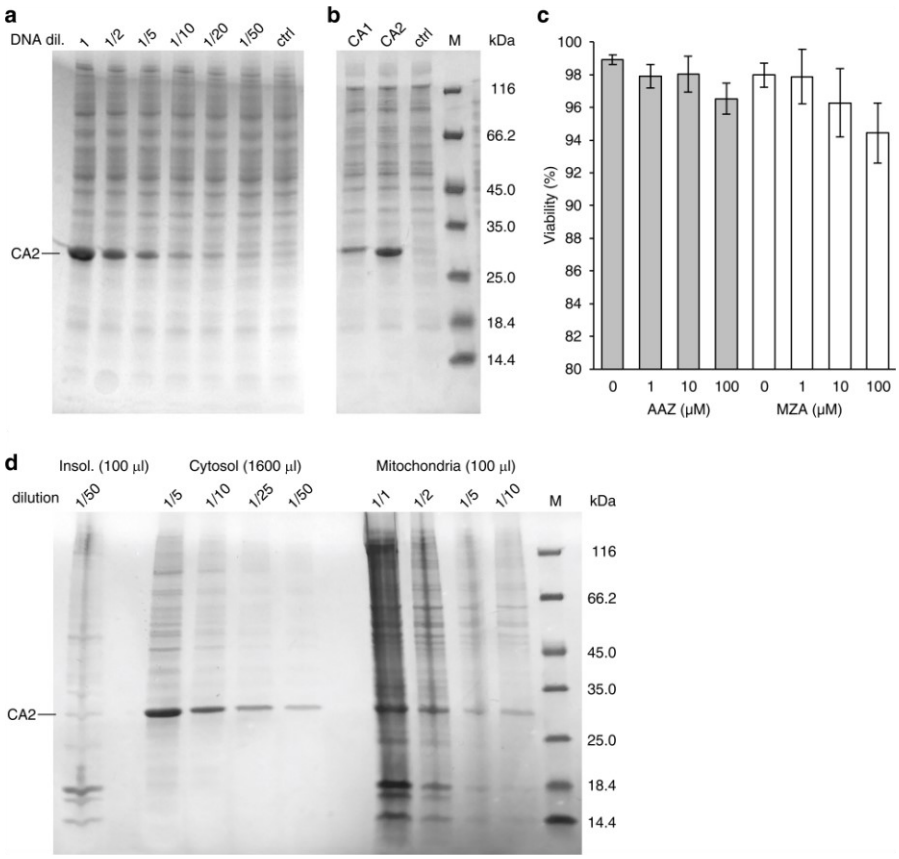


Figure S12. (a) Coomassie-stained SDS-PAGE of cleared lysates from cells transfected with the vector containing the CA2 gene at different dilutions with empty vector. Protein concentration was estimated by densitometry analysis with respect to a sample of recombinant CA2 at known concentration (not shown); **(b)** Coomassie-stained SDS-PAGE comparing lysates from cells expressing CA1 and CA2. M = marker; **(c)** cell viability measured by Trypan blue on cells treated for 1 hour with increasing concentration of AAZ (grey bars) or MZA (white bars). Error bars represent standard deviation (n = 3). **(d)** Intracellular distribution of overexpressed CA2. Insol.: insoluble fraction containing nuclei and cell membranes. For each fraction, the total volume and the loading dilution are indicated. M = marker.

2.3 Intracellular binding/unbinding kinetics of approved drugs to carbonic anhydrase II observed by in-cell NMR

In-cell NMR has been shown to be successful in characterization, at an atomic level, of interactions between a protein and its ligand within the cytoplasm of living cells. The approach that we have previously developed and applied to the study of CA II inhibitors, allowed us to identify and describe the behaviour of various ligands in physiological conditions. In fact, through the analysis of the chemical shift perturbations induced by the binding of different drugs, we were able to obtain pharmacologically relevant information regarding kinetic parameters and properties, such as cell permeability and affinity for an intracellular target.

In this project, the same approach was applied to a set of approved CAII ligands, including some that were originally designed to inhibit different targets. Through in-cell NMR it was possible to determine dose- and time- dependent binding curves from which a range of unpredicted behaviours emerged, in some cases probably correlated to the off-target binding activity of these compounds. This, in addition of being of interest in outlining a preliminary profile of drugs, can be very interesting for testing and consequently accelerating the development of new potential multi-target drugs in that branch of pharmacology, known as polypharmacology.

2.3.1 Attached manuscript (accepted on *ACS chemical biology*)

Intracellular binding/unbinding kinetics of approved drugs to carbonic anhydrase II observed by in-cell NMR

Enrico Luchinat[1,2,*], Letizia Barbieri[1,3], Matteo Cremonini[1], Alessio Nocentini[4], Claudiu T. Supuran[4,5], Lucia Banci[1,5,*]

[1] CERM – Magnetic Resonance Center, Università degli Studi di Firenze, Via Luigi sacconi 6, 50019 Sesto Fiorentino, Italy;
[2] Consorzio per lo Sviluppo dei Sistemi a Grande Interfase – CSGI, Via della Lastruccia 3, 50019 Sesto Fiorentino, Italy;
[3] Consorzio Interuniversitario Risonanze Magnetiche di Metalloproteine, Via Luigi Sacconi 6, Sesto Fiorentino, Italy;
[4] Dipartimento Neurofarba, Sezione di Scienze Farmaceutiche, Università degli Studi di Firenze, Via Ugo Schiff 6, 50019 Sesto Fiorentino, Italy;
[5] Dipartimento di Chimica, Università degli Studi di Firenze, Via della Lastruccia 3, 50019 Sesto Fiorentino, Italy;

[*]Corresponding authors:
Enrico Luchinat: eluchinat@cerm.unifi.it;
Lucia Banci: banci@cerm.unifi.it

Abstract

Candidate drugs rationally designed in vitro often fail due to low efficacy in vivo caused by low tissue availability or because of unwanted side effects. To overcome the limitations of in vitro rational drug design, the binding of candidate drugs to their target needs to be evaluated in the cellular context. Here, we applied in-cell NMR to investigate the binding of a set of approved drugs to the isoform II of carbonic anhydrase (CA) in living human cells. Some compounds were originally developed towards other targets, and were later found to inhibit CAs. We observed strikingly different dose- and time-dependent binding, wherein some drugs exhibited a more complex behavior than others. Specifically, some compounds were shown to gradually unbind from intracellular CA II, even in the presence of free compound in the external medium, therefore preventing the quantitative formation of a stable protein-ligand complex. Such observations could be correlated to the known off-target binding activity of these compounds, suggesting that this approach could provide information on the pharmacokinetic profiles of lead candidates at the early stages of multi-target drug design.

Introduction

Classical rational drug design approaches aim at maximizing the activity towards a specific target in vitro. However, drug efficacy in vivo can be affected by many factors, such as low tissue availability, binding to off-target biomolecules, or unwanted side effects. Therefore, ideally, the binding to an intracellular target should not only be evaluated in vitro, in isolated conditions, but also directly in the cellular context, where poor cell penetrance or the occurrence of off-target binding can negatively affect the activity of potential drugs towards their main target. Cell-based activity assays provide an indirect measure of the effect of a drug, but may not easily discriminate the molecular origin of the cellular response. Ideally, the interaction of a potential drug with the intracellular target should be monitored directly at atomic resolution.

Among the existing atomic-resolution structural techniques, Nuclear Magnetic Resonance (NMR) spectroscopy stands out for its ability to investigate protein-ligand interactions in solution at physiological temperatures.[1–3] Furthermore, in-cell NMR approaches have been developed that allow the atomic-level characterization of proteins and nucleic acids directly in living cells.[4–8] In the context of drug development, in-cell NMR has revealed to be a promising tool, as it can characterize interactions between small molecules and an intracellular target.[9–12] Recently, we have employed in-cell NMR to investigate the interaction between multiple ligands and the second isoform of human carbonic anhydrase (CA, EC 4.2.1.1), CA II, in living human cells.[13] This approach relies on the perturbation of the protein chemical shifts induced by ligand binding and allows an atomic-level description of the ligand binding site and, simultaneously, the measurement of intracellular dose- and time-dependent binding curves. From the latter data, important physicochemical properties of the ligands can be estimated, such as membrane permeability and apparent binding affinity, which are critical to assess the potency of the drugs towards the specific target.[13]

Carbonic anhydrases (CAs) are ubiquitous metalloenzymes that catalyze the conversion of CO_2 and H_2O to HCO_3^- and H^+.[14] In humans, 15 isoforms have been identified, all belonging to the α-class, which contain a catalytic zinc ion in the active site coordinated by three conserved histidine residues and a water molecule/hydroxide anion in a distorted tetrahedral geometry.[15] Although they have a high structural homology, the human isoforms differ in catalytic activity,

structural properties of the binding cavity, subcellular localization and response to exogenous molecules.[14–16] The CO_2 hydration reaction catalyzed by CAs is involved in many physiological processes, such as transport of CO_2 between tissues and lungs during respiration, pH homeostasis, electrolyte transport in various tissues, and several biosynthetic pathways.[14] Importantly, CA isoforms have been implicated in several pathological states, such as epilepsy, glaucoma, cardiovascular diseases and cancer.[17] Given the relevance of CAs as pharmacological targets, CA inhibitors have been developed over the course of several decades, some of which are currently administered in the treatment of glaucoma and epilepsy, and as diuretics.[14] Current efforts are now focused on the development of CA inhibitors with higher isoform selectivity, which would allow the selective inhibition of a single isoform, thereby reducing the required dose for treatment and the insurgence of adverse effects.[16,17] Indeed, a selective inhibitor of CA IX is currently in clinical development as antitumor agent.[18,19]

Here, we applied the above in-cell NMR approach to screen in human cells the binding to CA II of a set of approved drugs that are known to inhibit human CAs. We focused on a selection of sulfonamide compounds representative of different categories of drugs, including anti-inflammatory drugs originally designed to inhibit non-CA targets, diuretics and anticonvulsants which exert their function through inhibition of multiple targets, including CAs, and an anticancer drug currently under clinical trials that was specifically designed to inhibit the tumor-associated isoform CA IX. While the chemical shift perturbation confirmed that all the ligands bound CA II in the active site of the protein, in accordance to the in vitro characterization, we observed strikingly different behaviors from the dose- and time-dependent binding data for each ligand. Specifically, while the binding mode of some ligands could be explained by diffusion-limited, strong binding kinetics, other ligands diverged from such simplistic model and exhibited a more complex pharmacodynamic behavior, which could be correlated to the presence of off-target binding activity.

RESULTS AND DISCUSSION

Properties of the investigated drugs

The structures of the investigated drugs and their properties are summarized in Chart 1 and Table 1. Celecoxib **1** and valdecoxib **2** are first- and second-generation COX-2-selective inhibitors, respectively, employed as nonsteroidal anti-inflammatory drugs (valdecoxib was withdrawn from clinical use); indapamide **3** (racemic mixture), furosemide **4** and hydrochlorothiazide **5** are high-ceiling diuretics, which reduce sodium reabsorption in the kidneys by binding to the electroneutral sodium-chloride cotransporter and are employed treat hypertension, edema and congestive heart failure; zonisamide **6** is a widely used antiepileptic drug which binds to voltage-sensitive sodium and calcium channels; ethoxzolamide **7** is a diuretic, also employed to treat glaucoma, that inhibits CAs in proximal renal tubules; SLC-0111 **8** is a recently developed CA IX inhibitor currently undergoing phase II-b clinical trial as an anticancer/antimetastatic agent; methazolamide **9** is a potent CA inhibitor employed in the treatment of glaucoma. Compounds **1**-**6** were not originally designed as CA inhibitors, but have been shown later to inhibit several pharmacologically relevant CA isoforms with nanomolar affinity.[14,20–25] Conversely, compounds **7** and **8** were designed purposefully to inhibit CA activity. **9** is a well-characterized compound, which was previously observed to bind quantitatively CA II in human cells by NMR[13] and was included as a reference compound in the current study.

Drug binding monitored by ^{1}H NMR

Drug binding to the intracellular protein was investigated by in-cell NMR by analyzing HEK293T cells transiently overexpressing CA II and subsequently treated with each compound at different doses and incubation times. In our experimental conditions, CA II reaches an average concentration of 150 ± 20 μM in ~150 μL of cell pellet (see Methods), as previously reported.[13] Amino ^{1}H signals arising from three histidine side chains located in the CA II active site, as well as ^{1}H from other aromatic side chains, were clearly identified in the region of the 1D ^{1}H NMR spectrum between 11 and 16 ppm (Figure 1). This spectral region is well-resolved and free from cellular background signals, therefore it can be analyzed without recurring to ^{15}N isotopic labeling, that is often necessary to reduce spectral

overlap and avoid interference from non-NH background signals. Ligand binding induces changes in the chemical environment of the active site, giving rise to a new set of signals arising from the bound protein (Figure 2). In aqueous buffer, all the compounds quantitatively bound CA II when added in stoichiometric amount, giving rise to distinct patterns of clearly resolved signals (Figure 2a).

Fixed-dose drug screening

The intracellular binding of each compound was initially assessed by treating CA II-expressing cells with a 100 μM dose of each compound in 20 mL of external medium, therefore in ~100:1 molar ratio with respect to the protein (2×10^{-6} mol of compound vs. ~2×10^{-8} mol of CA II), followed by 1-hour incubation, removal of the external solution, cell detachment and NMR analysis. The bound fraction for each compound was obtained as a time average from a 30 minutes-long in-cell NMR spectrum. Binding to CA II was observed for all compounds, although at varying degrees (Figure 2b). In cells, the chemical shifts of the bound state matched those observed in vitro, indicating that the binding mode of the compounds was essentially unchanged in the cellular setting. Additional line broadening, caused by faster transverse relaxation and magnetic inhomogeneity of the cell sample, resulted in an increased overlap between signals, which was overcome by signal deconvolution (Figure S2, see Methods). The fractions of free and bound intracellular protein could then be quantified by comparing the relative signal intensities of the two species in the in-cell ^{1}H NMR spectra (Figure S2 and Figure 3). The large differences observed in the bound protein fractions at such a high dose treatment could not be explained with the different K_Is reported in vitro (Table 1) and suggested that the incomplete binding could be the consequence of poor drug permeability, as previously observed for other CA inhibitors.[13] Notably, the two enantiomers of **3** bound CA II in equal amounts in vitro, resulting in two distinct sets of signals with similar intensities (Figure 2a). A similar pattern was also observed in cells, suggesting that the two enantiomers bound intracellular CA II with similar affinity, however the broader spectral lines and the presence of signals from free CA II prevented a more precise quantification of each enantiomer (Figure 2b).

Dose- and time-dependent drug binding

The dose- and time-dependency of binding was assessed by analyzing CA II-expressing cells treated with increasing doses of each compound (ranging from 10 to 100 μM in the external medium) and incubated for increasing times (ranging from 30 minutes to 2 hours). The bound fraction in each condition was obtained from a 30 minutes-long NMR experiment as above (Figure 4). Compounds **7-9**, which bound CA II completely at 100 μM for 1 hour, gave essentially the same results both with shorter incubation time (30 minutes), and with a much lower dose (10 μM, still a ~10 fold molar excess with respect to the total CA II), whereas compounds **1-6**, which did not bind CA II completely at 100 μM, showed different dose and time dependencies. Specifically, the fraction of CA II bound to **4** and **5** increased linearly as a function of time, consistent with binding kinetics limited by the plasma membrane permeability, irrespective of the binding affinity for CA II. Instead, CA II binding of compounds **1**-**3** and **6** did not show any dose or time dependency in the 50-100 μM range and reached a plateau at ~70-80% (**1**), ~50-60% (**2**), ~80-90% (**3**) and ~35-40% (**6**), indicating that the observed results in that dose range were not dependent on membrane permeability. For compounds **1** and **8**, treatment at 10 μM dose resulted in a decrease of CA II bound fraction and allowed estimating the rates of permeability through the plasma membrane, in addition to those obtained from the time dependency of **4** and **5** at high dose (Table 2, see Methods), whereas no significant decrease was observed at 10 μM dose for compounds **2**, **3**, **6**, **7**, **9**, indicating a permeability $K_P \cdot A > 6 \cdot 10^{-7}$ dm^3 s^{-1}.

Unbinding kinetics by time-resolved in-cell NMR

To further investigate the origin of the plateau effect observed in the 50-100 μM dose range, cells treated with 100 μM for 1 hour (i.e. the same conditions for which time-averaged data was reported, see Figure 3) were analyzed as a function of time by performing a series of short in-cell NMR experiments over a total experimental time of 1 hour (Figure 5, red and magenta curves). Interestingly, the binding of compounds **1**-**3** and **6** to intracellular CA II exhibited a marked time dependence, decreasing respectively from ~100% to ~66% (**1**), from ~90% to ~45% (**2**) from ~100% to ~75% (**3**) and from ~50% to ~30% (**6**) after 1 hour (values at time 0 were extrapolated), whereas compounds **4** and **5** showed a less pronounced time dependence (Figure 5). These values are fully consistent with the measurements

averaged over 30 minutes, and indicate that these compounds are gradually released during the acquisition of the NMR spectra, thereby explaining the incomplete binding observed in the averaged data (Figure 3). In order to assess whether the same effect could occur in the cell culture, cells treated for 1 hour with compounds **1**-**3** and **6** in the CO_2 incubator were washed to remove the external ligand and further incubated in fresh medium without ligands. Quantitative NMR analysis of lysates from cells collected at increasing times showed that a similar unbinding behavior also occurred in the cell culture, although to a lower extent (Figure S3). The unbinding from CA II was also investigated in samples, initially treated with 100 μM compound for 1 hour, where the compound was re-introduced in the external solution of the cells during the NMR measurement (Figure 5, black and blue curves). Notably, while the unbinding of **1**, **3**, **4** and **5** was only partially mitigated in those conditions, compounds **2** and **6** showed a marked increase of bound protein at time 0. Therefore, the observed unbinding from CA II could be partially explained with the diffusion of ligand molecules from the cytosol back to the external solution (Figure S3). However, the observation of unbinding even in the presence of external ligand (Figure 5) indicates that this phenomenon is not driven solely by passive diffusion.

Drug classification based on binding behavior

The above results show that the interplay between active compound, intracellular target, plasma membrane and cellular milieu can generate complex binding behaviors that are not easily described with a simple diffusion model. Nevertheless, the binding data obtained by in-cell NMR allowed a coarse-grained classification of the screened molecules. Based on the diffusion properties of the compounds (Table 2) and on the presence/absence of unbinding kinetics (Figure 5), the following classes are obtained: I) fast-diffusing, stable binding (**7**, **8**, **9**); II) slow-diffusing, stable binding (**4**, **5**); III) fast-diffusing, unstable binding (**1**, **2**, **3**, **6**). Slow-diffusing, unstable binding compounds were not observed, although it is possible that the slow diffusion of **4** and **5** masks the effects of binding instability. Strikingly, all class-I compounds have a mechanism of action that involves strong CA inhibition, or were even rationally designed for selective CA inhibition (**8**), whereas class-II and III compounds were not primarily intended for CA inhibition and have known activity towards other targets. This correlation suggests that the

intracellular screening performed here could be used as a predictive tool to assess the specificity of a drug towards the desired intracellular target.

Binding of class I and II compounds follow a simple diffusion-limited binding behavior, where at low doses / short times the fraction of bound CA II depends linearly on the permeability of the compounds through the plasma membrane, while at high doses / long times CA II is fully bound to the compounds. In the dose and time ranges investigated, complete binding was only observed for class I. In such regime, ligand binding is likely not affected by changes in the intracellular binding affinity, due to the intracellular concentration of CA II being in the ~100s μM range while the K_Is are in the nM-μM range.

Class-III compounds behave essentially like class-I during the cell incubation step, quickly diffusing through the plasma membrane and binding quantitatively to CA II (Figure 4 and Table 2), but are then partially released once the cells are detached for NMR analysis (Figure 5). Notably, while some release was observed prior to cell detachment after the compounds were removed from the cell growth medium (Figure S3), it also occurred when the ligand molecules in excess were kept in the external solution in the NMR tube (Figure 5). In the latter case, a delay in the onset of the unbinding was introduced together with an increase in the final fraction of bound CA II, while the slope of the curves was not affected. Although several scenarios are possible, such behavior may be partially explained by a model in which an additional species competes against CA II for ligand binding. However, this competitor must be introduced in the system after cells are detached for NMR analysis, otherwise it would have already been saturated with ligand molecules during the incubation step. It is possible that a change in membrane protein composition induced by trypsinization affects the turnover and the cellular localization of membrane proteins, including the known targets of class-III compounds, namely the membrane-bound COX-2 (**1**, **2**), the integral membrane proteins sodium-chloride cotransporter (**3**) and voltage-sensitive channels (**6**). This change would cause the emergence of a pool of competing binding sites that subtracts the ligands from CA II, with a kinetic behavior that can be modulated by the presence of additional ligand molecules in the external solution. However, these results only allow for speculation, as they report specifically on the free and bound fractions of CA II, with no information on the binding of the compounds to other intracellular targets. Furthermore, CA II has to be overexpressed to allow for NMR

detection, making it in excess with respect to the competing targets, thus further complicating the interpretation of the results.

Conclusions

The above findings show that the application of in-cell NMR to investigate drug binding to an intracellular pharmacological target could have a relevant role in the drug discovery process. Indeed, this approach could provide early-stage information on the pharmacokinetic profiles of lead candidates and allow preclinical investigations already at the drug-design and lead identification stages. Furthermore, small-scale screenings such as the one reported here could predict the specificity of a drug towards the desired target or, as in the case for class-III compounds described above, warn against possible multi-target behavior.

Recently, in addition to the classical drug development routes, there has been a growing interest in polypharmacological approaches, which aim at exploiting the off-target activities of drugs in combination with the activity towards their original targets for the treatment of complex disease states.[26] Due to their involvement in different pathological states, human CA isoforms are ideal targets for multi-target drugs.[17] Indeed, drugs such as those investigated here could be employed in novel therapeutic strategies in which CA inhibitors have been shown to be effective, such as against obesity, arthritis, cerebral ischemia and neuropathic pain,[27] by exploiting their multi-targeting effect.

Here, intracellular ligand screening by NMR revealed an unpredicted behavior of some of these hybrid drugs, possibly as a consequence of their activity towards multiple targets. In this context, the in-cell NMR approach could have a marked relevance when designing and characterizing multi-target compounds, driving further optimization in the search of a balanced multi-targeting efficacy and accelerating the development of next-generation polypharmacological drugs, also beyond the field of CAs. From a methodological standpoint, further advancements will be needed to allow observing the binding of a compound to multiple intracellular targets simultaneously, without requiring their overexpression. In this respect, NMR bioreactors have proven valuable tools to increase both sensitivity and time resolution of the methodology.[28,29] In parallel, ^{19}F is being increasingly exploited as a sensitive and background-free probe for cellular NMR studies,[11,30,31] hence a ligand-observe in-cell NMR approach could be envisaged that employs ^{19}F-

labeled compounds to obtain a more complete picture of their fate inside the cells as a function of time.

METHODS

Human cell cultures

HEK293T cells (ATCC CRL-3216) were maintained in Dulbecco-modified Eagle medium (DMEM) high glucose (Gibco) supplemented with L-glutamine, antibiotics (penicillin and streptomycin) and 10% fetal bovine serum (FBS, Gibco) in uncoated 75 cm^2 plastic flasks and incubated at 37 °C, 5% CO_2 in a humidified atmosphere. HEK293T cells were transiently transfected, following a previously reported protocol,[32] with the pHLsec[33] plasmid containing the full-length human CA II gene (amino acids 1–260, GenBank: NP_000058.1)[13] using branched polyethylenimine (PEI). A DNA:PEI ratio of 1:2 (25 μg/flask DNA, 50 μg/flask PEI) was used. Protein expression was carried out in DMEM medium supplemented with 2% FBS, antibiotics and 10 μM $ZnSO_4$. CA II concentration was calculated as previously reported[13] from cells lysed in 1 cell pellet volume, corresponding to the effective concentrations in the in-cell NMR samples (mean value ± s.d., n = 3), and measured by SDS-PAGE by comparing serial dilutions with a sample of purified CA II. Compounds 1-9 were purchased from Sigma-Aldrich or TCI Europe and are ≥ 97% HPLC pure. Cells overexpressing CA II were treated with the compounds 48 hours post-transfection, by dissolving a concentrated DMSO stock solution of each compound (80 mM) directly in 20 mL of growth medium in the cell culture flask. Experiments were performed by treating cells with varying amounts of each compound and incubated for varying amounts of time as specified in the Results section. Control cell samples were incubated for 1 hour with pure DMSO (0.125% final concentration). Cell viability remained > 95%, as assessed by trypan blue exclusion assay, and was not affected by treatment with any the compounds in the dose and time ranges employed in the study (data not shown).

In-cell NMR sample preparation

Samples for in-cell NMR were prepared as previously reported.[32,34] Transfected cells were detached with trypsin, suspended in DMEM + 10% FBS, washed once with PBS and re-suspended in one pellet volume of NMR medium, consisting of

DMEM supplemented with 90 mM glucose, 70 mM HEPES and 20% D_2O. The cell suspension was transferred in a 3 mm Shigemi NMR tube, which was gently spun to sediment the cells. For the unbinding experiments in the presence of compound in the external solution, cells were prepared following the above protocol and resuspended in one pellet volume NMR medium supplemented with 100 μM of compound. Cell viability before and after NMR experiments was assessed by trypan blue exclusion assay. After the NMR experiments, the cells were collected and the supernatant was checked for protein leakage by NMR.

Cell lysate sample preparation

Cultures of HEK293T cells overexpressing CA II were incubated with 100 μM of compound for 1 hour. A control cell culture was immediately detached with trypsin, pelleted and frozen at -20°C. The remaining cell cultures were washed once with PBS and resuspended in fresh DMEM supplemented with 2% FBS, antibiotics and 10 μM of $ZnSO_4$, in the absence of compound. Cells were incubated for varying amounts of time (30 min, 1 and 2 hours) and subsequently collected and frozen at -20°C. Cell lysates were prepared by freeze-thaw cycles in PBS buffer followed by centrifugation to remove the insoluble fraction. The supernatants were supplemented with 10% D_2O, placed in a standard 3-mm NMR tube and analyzed by NMR.

In-cell NMR spectra acquisition and analysis

In-cell NMR spectra were collected at 310 K either at a 900 MHz Bruker Avance NEO or at a 950 MHz Bruker Avance III spectrometer, both equipped with a 5 mm TCI CryoProbe. 1D ^1H NMR spectra were recorded with a WATERGATE experiment using a 3-9-19 binomial pulse train for water suppression (Bruker p3919gp pulse program).[35] Time-averaged NMR spectra were recorded with 1024 scans (total experimental time of 28 minutes). Time-resolved NMR data were collected by recording 16 ^1H NMR spectra, with the same parameters as above, with 128 scans each (total experimental time of 57 minutes). The spectra were processed with Bruker Topspin 4.0 by applying zero filling and exponential line broadening (LB = 20 Hz). To retrieve the area under each signal, the region between 11 and 16 ppm of each spectrum was fitted with Fityk[36] using a sum of N+1 pseudo-Voigt functions (gaussian weight fixed at 0.15), where N equals the number of

discernible signals in the region and the additional function accounted for the baseline distortion. Peak areas for each spectrum were normalized by the sum of the N areas. Relative fractions of free and bound CA II were obtained by Multivariate Curve Resolution – Alternate Least Square (MCR-ALS) analysis,[37] using the MCR-ALS 2.0 GUI[38] in MATLAB according to a previously reported protocol.[29] The contribution of each CA II species to the signal intensities was retrieved from 2D arrays (one array per compound) containing the peak areas (rows = experiments, columns = peaks). Error bars for each fitting were obtained as follows: err = 2 × [lack of fit(%)]/100.

Non-linear curve fitting of the binding data was performed in OriginPro 8 with the previously described dose- and time-dependent binding equation:[13]

$$F(t) = \frac{[LP]}{[P_t]}(t)$$

$$= C \cdot \frac{K_d + [P_t] + [L_{t.in}](t) - \sqrt{\left(K_d + [P_t] + [L_{t.in}](t)\right)^2 - 4[P_t][L_{t.in}](t)}}{2[P_t]}$$

where $[P_t]$ is the total intracellular protein concentration, [LP] is the bound protein concentration, K_d is the dissociation constant and C is the apparent fraction of ligand-bound CA II at plateau, which is < 1 in the presence of unbinding effects. $[L_{t.in}](t)$ is the total intracellular ligand concentration defined as:

$$L_{t.in}(t) = L_t\left(1 - e^{-\frac{K_p A}{V_t}t}\right)$$

where $L_t = L_o + L_i + LP$ are the total moles of ligand (L_o and L_i are the extracellular and intracellular ligand, respectively), V_t is the total external volume, K_p is the permeability coefficient and A is the total area of the membrane.

Expression and purification of CA2

Recombinant CA II was prepared following an existing protocol.[39] Briefly, a 1 l cell culture of *E. Coli* BL21(DE3) Codon Plus Ripl (Stratagene) was transformed with a pCAM plasmid containing the CA II gene, grown overnight at 37 °C in LB, harvested and re-suspended in 1 l M9 medium. $ZnSO_4$ was added in the culture to a final concentration of 500 μM. After 5 h from induction with 1 mM isopropyl β-d-1-thiogalactopyranoside (IPTG) at 37 °C, the cells were harvested and re-suspended in 20 mM Tris buffer, pH 8 for lysis. The cleared lysate was loaded onto a nickel chelating HisTrap (GE Healthcare) 5 ml column. The protein was eluted

with a linear gradient of 20 mM Tris pH 8, 500 mM imidazole. Fractions containing pure CA II were collected. Finally, the protein was exchanged in NMR buffer (HEPES 20 mM pH 7.5, supplemented with 10% D2O). The correct metalation of the protein was confirmed by chemical shift comparison against previously reported spectra.[13,28]

Supporting Information

The Supporting Information is available free of charge via the internet at http://pubs.acs.org

- Supplementary Figures S1, S2 (PDF)

Acknowledgements

This work has been supported by iNEXT-Discovery, grant agreement no. 871037, funded by the Horizon 2020 research and innovation programme of the European Commission, by Timb[3], grant agreement no. 810856, funded by the Horizon 2020 research and innovation programme of the European Commission, and by Ministero dell'Istruzione, dell'Università e della Ricerca PRIN grants 20177XJCHX and 2017XYBP2R. The authors acknowledge the support of Instruct-ERIC, a Landmark ESFRI project and the use of resources of the CERM/CIRMMP Italy Centre.

References

(1) Shuker, S. B., Hajduk, P. J., Meadows, R. P., and Fesik, S. W. (1996) Discovering high-affinity ligands for proteins: SAR by NMR. *Science 274*, 1531–1534.
(2) Pellecchia, M., Bertini, I., Cowburn, D., Dalvit, C., Giralt, E., Jahnke, W., James, T. L., Homans, S. W., Kessler, H., Luchinat, C., Meyer, B., Oschkinat, H., Peng, J., Schwalbe, H., and Siegal, G. (2008) Perspectives on NMR in drug discovery: a technique comes of age. *Nat. Rev. Drug Discov. 7*, 738–745.
(3) Baggio, C., Cerofolini, L., Fragai, M., Luchinat, C., and Pellecchia, M. (2018) HTS by NMR for the Identification of Potent and Selective Inhibitors of Metalloenzymes. *ACS Med. Chem. Lett. 9*, 137–142.
(4) 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.
(5) Luchinat, E., and Banci, L. (2017) In-cell NMR: a topical review. *IUCrJ 4*, 108–118.
(6) Dzatko, S., Krafcikova, M., Hänsel-Hertsch, R., Fessl, T., Fiala, R., Loja, T., Krafcik, D., Mergny, J.-L., Foldynova-Trantirkova, S., and Trantirek, L. (2018) Evaluation of the Stability of DNA i-Motifs in the Nuclei of Living Mammalian Cells. *Angew. Chem. Int. Ed Engl. 57*, 2165–2169.

(7) Luchinat, E., and Banci, L. (2018) In-Cell NMR in Human Cells: Direct Protein Expression Allows Structural Studies of Protein Folding and Maturation. *Acc. Chem. Res. 51*, 1550–1557.
(8) Tanaka, T., Ikeya, T., Kamoshida, H., Suemoto, Y., Mishima, M., Shirakawa, M., Güntert, P., and Ito, Y. (2019) High-Resolution Protein 3D Structure Determination in Living Eukaryotic Cells. *Angew. Chem. Int. Ed Engl. 58*, 7284–7288.
(9) DeMott, C. M., Girardin, R., Cobbert, J., Reverdatto, S., Burz, D. S., McDonough, K., and Shekhtman, A. (2018) Potent Inhibitors of Mycobacterium tuberculosis Growth Identified by Using in-Cell NMR-based Screening. *ACS Chem. Biol. 13*, 733–741.
(10) Krafcikova, M., Dzatko, S., Caron, C., Granzhan, A., Fiala, R., Loja, T., Teulade-Fichou, M.-P., Fessl, T., Hänsel-Hertsch, R., Mergny, J.-L., Foldynova-Trantirkova, S., and Trantirek, L. (2019) Monitoring DNA-Ligand Interactions in Living Human Cells Using NMR Spectroscopy. *J. Am. Chem. Soc. 141*, 13281–13285.
(11) Siegal, G., and Selenko, P. (2019) Cells, drugs and NMR. *J. Magn. Reson. San Diego Calif 1997 306*, 202–212.
(12) Kang, C. (2019) Applications of In-Cell NMR in Structural Biology and Drug Discovery. *Int. J. Mol. Sci. 20*.
(13) Luchinat, E., Barbieri, L., Cremonini, M., Nocentini, A., Supuran, C. T., and Banci, L. (2020) Drug Screening in Human Cells by NMR Spectroscopy Allows the Early Assessment of Drug Potency. *Angew. Chem. Int. Ed Engl. 59*, 6535–6539.
(14) Supuran, C. T. (2008) Carbonic anhydrases: novel therapeutic applications for inhibitors and activators. *Nat. Rev. Drug Discov. 7*, 168–181.
(15) Supuran, C. T. (2016) Structure and function of carbonic anhydrases. *Biochem. J. 473*, 2023–2032.
(16) Alterio, V., Di Fiore, A., D'Ambrosio, K., Supuran, C. T., and De Simone, G. (2012) Multiple binding modes of inhibitors to carbonic anhydrases: how to design specific drugs targeting 15 different isoforms? *Chem. Rev. 112*, 4421–4468.
(17) Nocentini, A., and Supuran, C. T. (2019) Advances in the structural annotation of human carbonic anhydrases and impact on future drug discovery. *Expert Opin. Drug Discov. 14*, 1175–1197.
(18) A Study of SLC-0111 and Gemcitabine for Metastatic Pancreatic Ductal Cancer in Subjects Positive for CAIX - Full Text View - ClinicalTrials.gov.
(19) McDonald, P. C., Chia, S., Bedard, P. L., Chu, Q., Lyle, M., Tang, L., Singh, M., Zhang, Z., Supuran, C. T., Renouf, D. J., and Dedhar, S. (2020) A Phase 1 Study of SLC-0111, a Novel Inhibitor of Carbonic Anhydrase IX, in Patients With Advanced Solid Tumors. *Am. J. Clin. Oncol.*
(20) Weber, A., Casini, A., Heine, A., Kuhn, D., Supuran, C. T., Scozzafava, A., and Klebe, G. (2004) Unexpected nanomolar inhibition of carbonic anhydrase by COX-2-selective celecoxib: new pharmacological opportunities due to related binding site recognition. *J. Med. Chem. 47*, 550–557.
(21) Di Fiore, A., Pedone, C., D'Ambrosio, K., Scozzafava, A., De Simone, G., and Supuran, C. T. (2006) Carbonic anhydrase inhibitors: Valdecoxib binds to a different active site region of the human isoform II as compared to the structurally related cyclooxygenase II "selective" inhibitor celecoxib. *Bioorg. Med. Chem. Lett. 16*, 437–442.
(22) Temperini, C., Cecchi, A., Scozzafava, A., and Supuran, C. T. (2008) Carbonic anhydrase inhibitors. Interaction of indapamide and related diuretics with 12 mammalian isozymes and X-ray crystallographic studies for the indapamide-isozyme II adduct. *Bioorg. Med. Chem. Lett. 18*, 2567–2573.
(23) Temperini, C., Cecchi, A., Scozzafava, A., and Supuran, C. T. (2009) Carbonic anhydrase inhibitors. Comparison of chlorthalidone and indapamide X-ray crystal structures in adducts with isozyme II: when three water molecules and the keto-enol tautomerism make the difference. *J. Med. Chem. 52*, 322–328.
(24) Temperini, C., Cecchi, A., Scozzafava, A., and Supuran, C. T. (2008) Carbonic anhydrase inhibitors. Sulfonamide diuretics revisited--old leads for new applications? *Org. Biomol. Chem. 6*, 2499–2506.

(25) De Simone, G., Di Fiore, A., Menchise, V., Pedone, C., Antel, J., Casini, A., Scozzafava, A., Wurl, M., and Supuran, C. T. (2005) Carbonic anhydrase inhibitors. Zonisamide is an effective inhibitor of the cytosolic isozyme II and mitochondrial isozyme V: solution and X-ray crystallographic studies. *Bioorg. Med. Chem. Lett. 15*, 2315–2320.
(26) Moya-García, A., Adeyelu, T., Kruger, F. A., Dawson, N. L., Lees, J. G., Overington, J. P., Orengo, C., and Ranea, J. A. G. (2017) Structural and Functional View of Polypharmacology. *Sci. Rep. 7*, 10102.
(27) Supuran, C. T. (2020) Exploring the multiple binding modes of inhibitors to carbonic anhydrases for novel drug discovery. *Expert Opin. Drug Discov. 15*, 671–686.
(28) Cerofolini, L., Giuntini, S., Barbieri, L., Pennestri, M., Codina, A., Fragai, M., Banci, L., Luchinat, E., and Ravera, E. (2019) Real-Time Insights into Biological Events: In-Cell Processes and Protein-Ligand Interactions. *Biophys. J. 116*, 239–247.
(29) Luchinat, E., Barbieri, L., Campbell, T. F., and Banci, L. (2020) Real-Time Quantitative In-Cell NMR: Ligand Binding and Protein Oxidation Monitored in Human Cells Using Multivariate Curve Resolution. *Anal. Chem.*
(30) Ye, Y., Liu, X., Zhang, Z., Wu, Q., Jiang, B., Jiang, L., Zhang, X., Liu, M., Pielak, G. J., and Li, C. (2013) (19) F NMR spectroscopy as a probe of cytoplasmic viscosity and weak protein interactions in living cells. *Chem. Weinh. Bergstr. Ger. 19*, 12705–12710.
(31) Veronesi, M., Giacomina, F., Romeo, E., Castellani, B., Ottonello, G., Lambruschini, C., Garau, G., Scarpelli, R., Bandiera, T., Piomelli, D., and Dalvit, C. (2016) Fluorine nuclear magnetic resonance-based assay in living mammalian cells. *Anal. Biochem. 495*, 52–59.
(32) Barbieri, L., Luchinat, E., and Banci, L. (2016) Characterization of proteins by in-cell NMR spectroscopy in cultured mammalian cells. *Nat. Protoc. 11*, 1101–1111.
(33) 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.
(34) 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.
(35) Piotto, M., Saudek, V., and Sklenár, V. (1992) Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR 2*, 661–665.
(36) Wojdyr, M. (2010) Fityk: a general-purpose peak fitting program. *J. Appl. Crystallogr. 43*, 1126–1128.
(37) Juan, A. de, and Tauler, R. (2006) Multivariate Curve Resolution (MCR) from 2000: Progress in Concepts and Applications. *Crit. Rev. Anal. Chem. 36*, 163–176.
(38) Jaumot, J., de Juan, A., and Tauler, R. (2015) MCR-ALS GUI 2.0: New features and applications. *Chemom. Intell. Lab. Syst. 140*, 1–12.
(39) Cerofolini, L., Giuntini, S., Louka, A., Ravera, E., Fragai, M., and Luchinat, C. (2017) High-Resolution Solid-State NMR Characterization of Ligand Binding to a Protein Immobilized in a Silica Matrix. *J. Phys. Chem. B 121*, 8094–8101.
(40) Shimahara, H., Yoshida, T., Shibata, Y., Shimizu, M., Kyogoku, Y., Sakiyama, F., Nakazawa, T., Tate, S., Ohki, S., Kato, T., Moriyama, H., Kishida, K., Tano, Y., Ohkubo, T., and Kobayashi, Y. (2007) Tautomerism of histidine 64 associated with proton transfer in catalysis of carbonic anhydrase. *J. Biol. Chem. 282*, 9646–9656.
(41) Vasa, S. K., Singh, H., Grohe, K., and Linser, R. (2019) Assessment of a Large Enzyme-Drug Complex by Proton-Detected Solid-State NMR Spectroscopy without Deuteration. *Angew. Chem. Int. Ed Engl. 58*, 5758–5762.

Figures

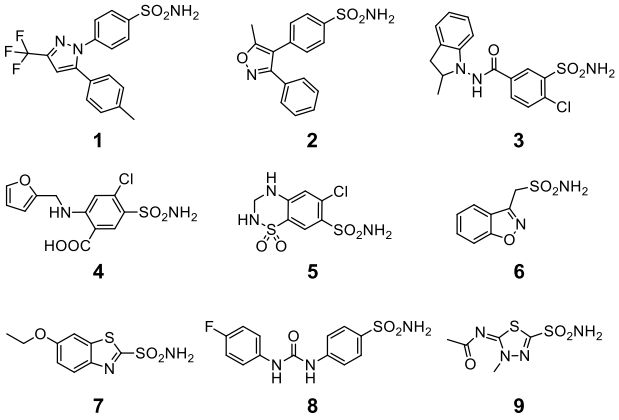


Chart 1. Chemical structures of the compounds analyzed in this study.

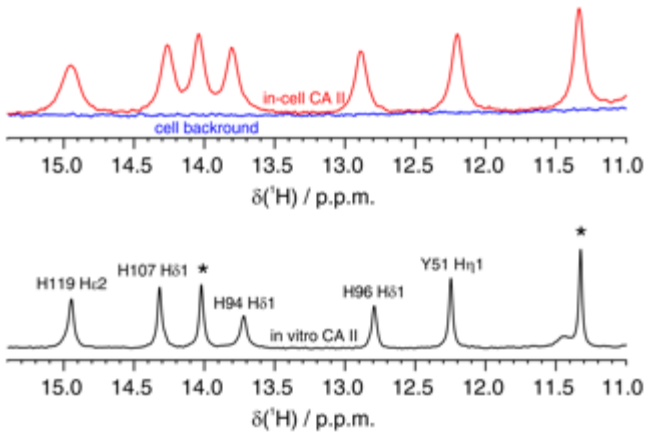


Figure 1. Top: imino region of the 1D ^{1}H NMR spectrum of CA II in human cells (red) overlaid to the spectrum of the cellular background obtained from cells transfected with empty vector (blue). Bottom: 1D ^{1}H NMR spectrum of CA II in aqueous buffer (black). Residues for which the unambiguous assignment has been reported previously are labeled with the corresponding residue number and atom type.[40,41] H94, H96 and H119 coordinate the zinc ion in the active site. Signals arising from unassigned protons are labeled with an asterisk.

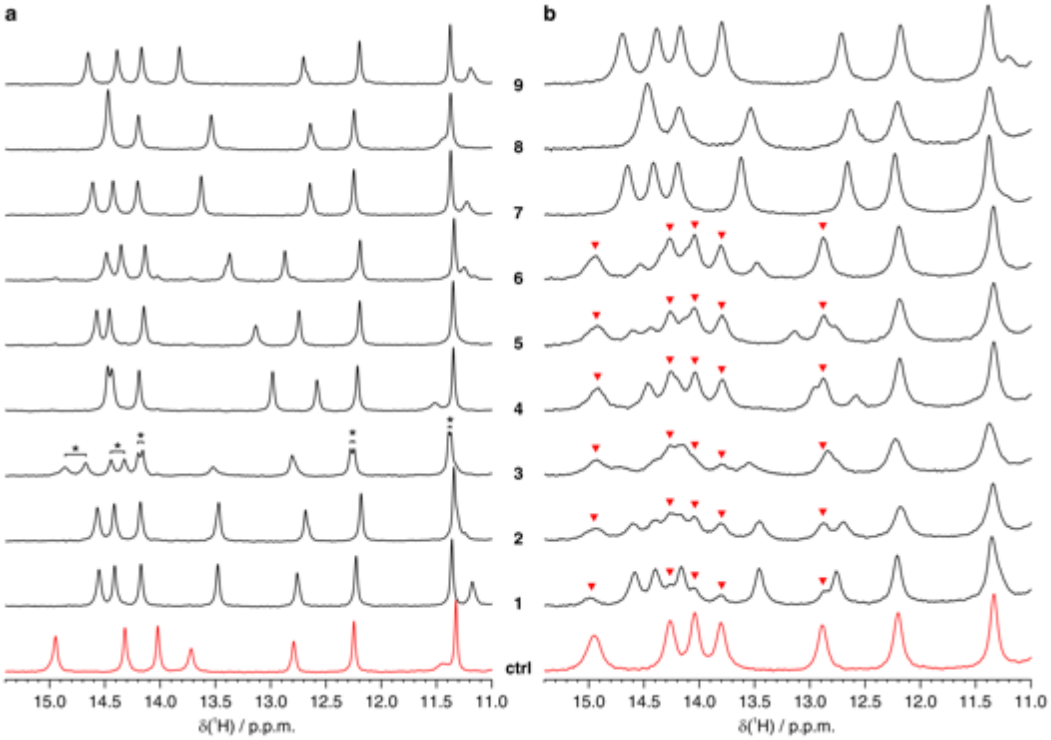


Figure 2. Imino region of the ^{1}H NMR spectra of CA II (a) in aqueous buffer and (b) in human cells in the absence (red) and in the presence (black) of the compounds investigated in this study. In vitro and in-cell NMR spectra were recorded for 15 and 30 minutes, respectively. In (a), the signals which arise from CA II bound to the two enantiomers of **3** are labeled with an asterisk. In (b), signals arising from free CA II are labeled with red arrows.

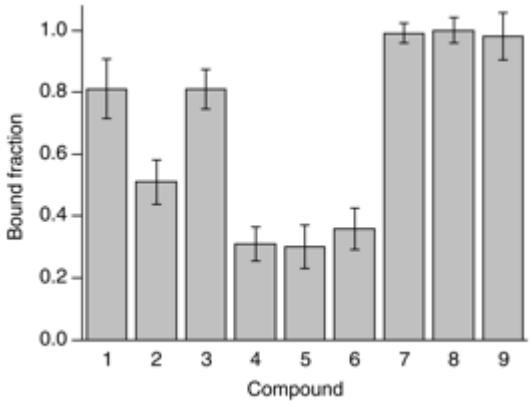


Figure 3. Bound fraction of intracellular CA II after incubation with 100 μM of each compound for 1 hour, measured as a time average over 30 minutes. Error bars were obtained from the MCR-ALS global fitting of each dose/time-dependence series (shown in Figure 4) as follows: err = 2 × [lack of fit(%)]／100.

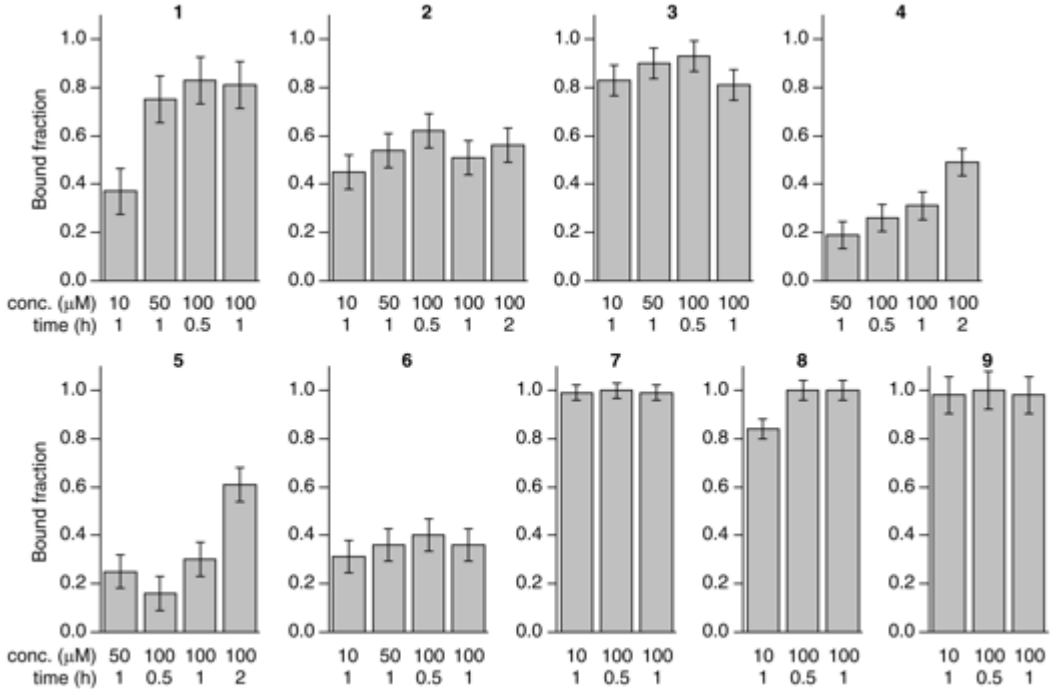


Figure 4. Dose and time dependence of the binding of each compound to intracellular CA II. Each bar plot shows the fraction of CA II bound to a given compound after incubation at different concentrations and incubation times, measured as a time average over 30 minutes. Error bars were obtained from the MCR-ALS global fitting as follows: err = 2 × [lack of fit(%)]／100.

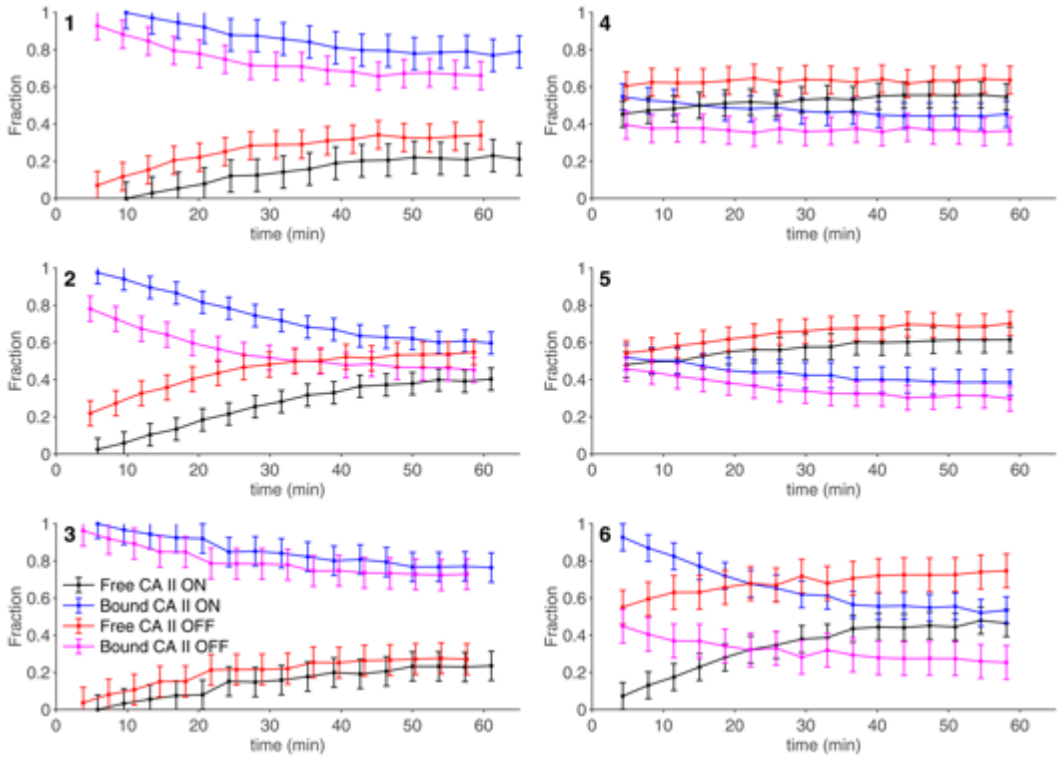


Figure 5. Time-dependence of the intracellular fractions of free (black, red) and bound (blue, magenta) CA II in the NMR spectrometer measured in the absence (red and magenta) and in the presence (black and blue) of 100 μM of external ligand. Error bars were obtained from the MCR-ALS global fitting as follows: err = 2 × [lack of fit(%)]／100.

Tables

Table 1. Data on the compounds analyzed in this study.

n	Name	MW	Main target	Used for	K_I CA II (nM)
1	Celecoxib	381.4	COX-2	Inflammation, pain	21
2	Valdecoxib	314.4	COX-2	Inflammation, pain	43
3	Indapamide	362.8	Na/Cl co-transporter	Hypertension, heart failure	2520
4	Furosemide	330.7	Na/K/Cl cotransporter	Hypertension, heart failure, edema	65
5	Hydrochlorothiazide	297.7	Na/Cl co-transporter	Hypertension, heart failure, edema, diabetes insipidus	290
6	Zonisamide	212.2	Sodium and T-type calcium channels (putative)	Epilepsy, Parkinson's disease	35
7	Ethoxzolamide	258.3	CA	Glaucoma, duodenal ulcers	8
8	SLC-0111	309.3	CA	Tumor	960
9	Methazolamide	236.3	CA	Glaucoma	14

Table 2. Permeability coefficient and fraction of bound CA II at plateau calculated for each compound from the non-linear fitting of the dose- and time-dependent binding data.

n	$K_P \cdot A$ ($dm^3\ s^{-1}$)	Bound CA II at plateau
1	$2.8 \pm 0.7 \cdot 10^{-7}$	78 ± 3 %
2	$> 6 \cdot 10^{-7}$	53 ± 2 %
3	$> 6 \cdot 10^{-7}$	86 ± 5 %
4	$1.6 \pm 0.4 \cdot 10^{-8}$	100 %
5	$1.8 \pm 0.2 \cdot 10^{-8}$	100 %
6	$> 6 \cdot 10^{-7}$	36 ± 5 %
7	$> 6 \cdot 10^{-7}$	100 %
8	$5.4 \pm 0.3 \cdot 10^{-7}$	100 %
9	$> 6 \cdot 10^{-7}$ [a]	100 %

[a]previously measured $K_P \cdot A = 1.2 \cdot 10^{-6}\ dm^3\ s^{-1}$.[13]

2.3.2 Supplementary information

Intracellular binding/unbinding kinetics of approved drugs to carbonic anhydrase II observed by in-cell NMR

Enrico Luchinat[1,2], Letizia Barbieri[1,3], Matteo Cremonini[1], Alessio Nocentini[4], Claudiu T. Supuran[4,5], Lucia Banci[1,5,*]

[1] CERM – Magnetic Resonance Center, Università degli Studi di Firenze, Via Luigi sacconi 6, 50019 Sesto Fiorentino, Italy;
[2] Consorzio per lo Sviluppo dei Sistemi a Grande Interfase – CSGI, Via della Lastruccia 3, 50019 Sesto Fiorentino, Italy;
[3] Consorzio Interuniversitario Risonanze Magnetiche di Metalloproteine, Via Luigi Sacconi 6, Sesto Fiorentino, Italy;
[4] Dipartimento Neurofarba, Sezione di Scienze Farmaceutiche, Università degli Studi di Firenze, Via Ugo Schiff 6, 50019 Sesto Fiorentino, Italy;
[5] Dipartimento di Chimica, Università degli Studi di Firenze, Via della Lastruccia 3, 50019 Sesto Fiorentino, Italy;

[*] banci@cerm.unifi.it

Supplementary Figures S1, S2

Supplementary Figures

Figure S1

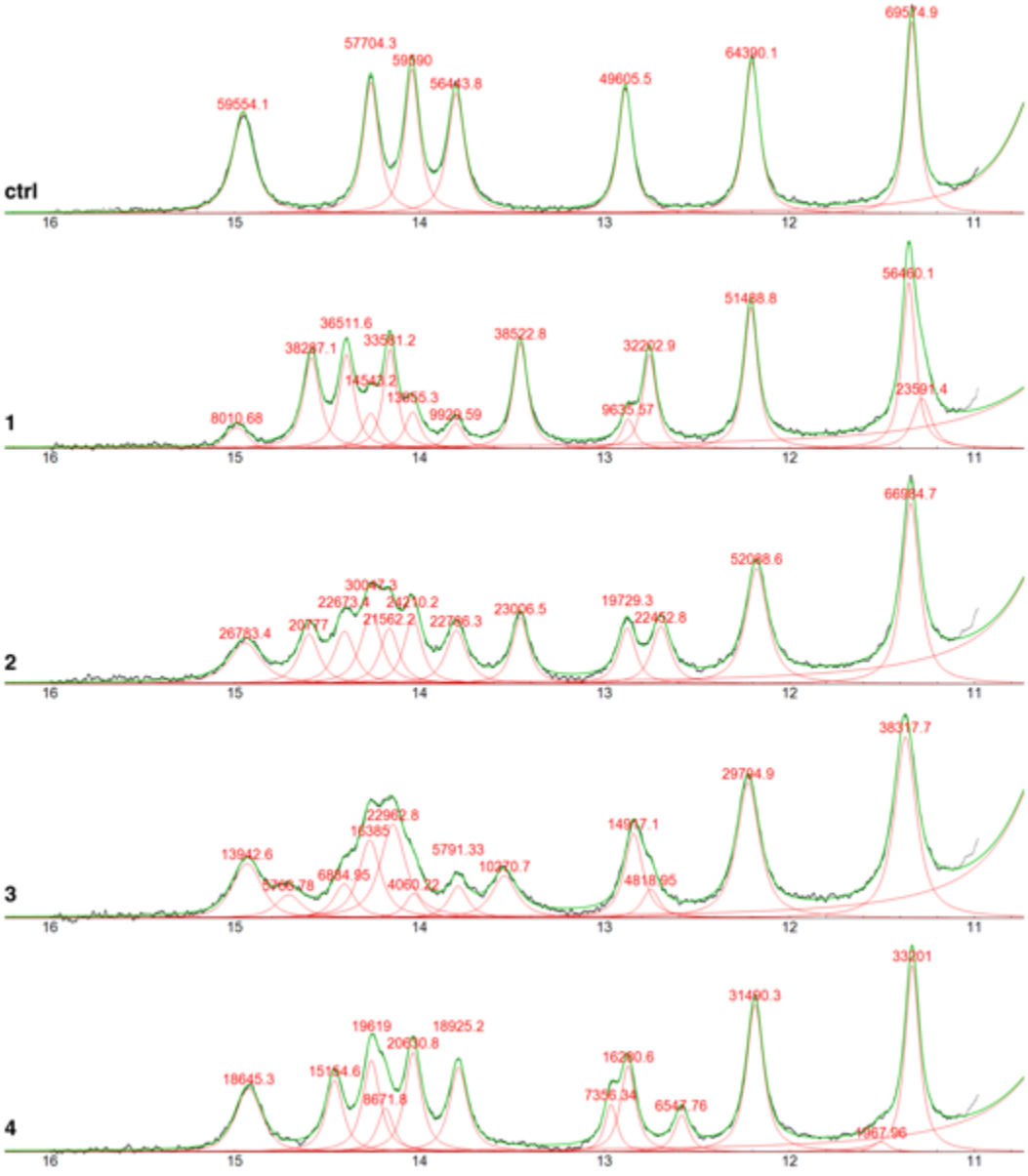


Figure S1 (cont.)

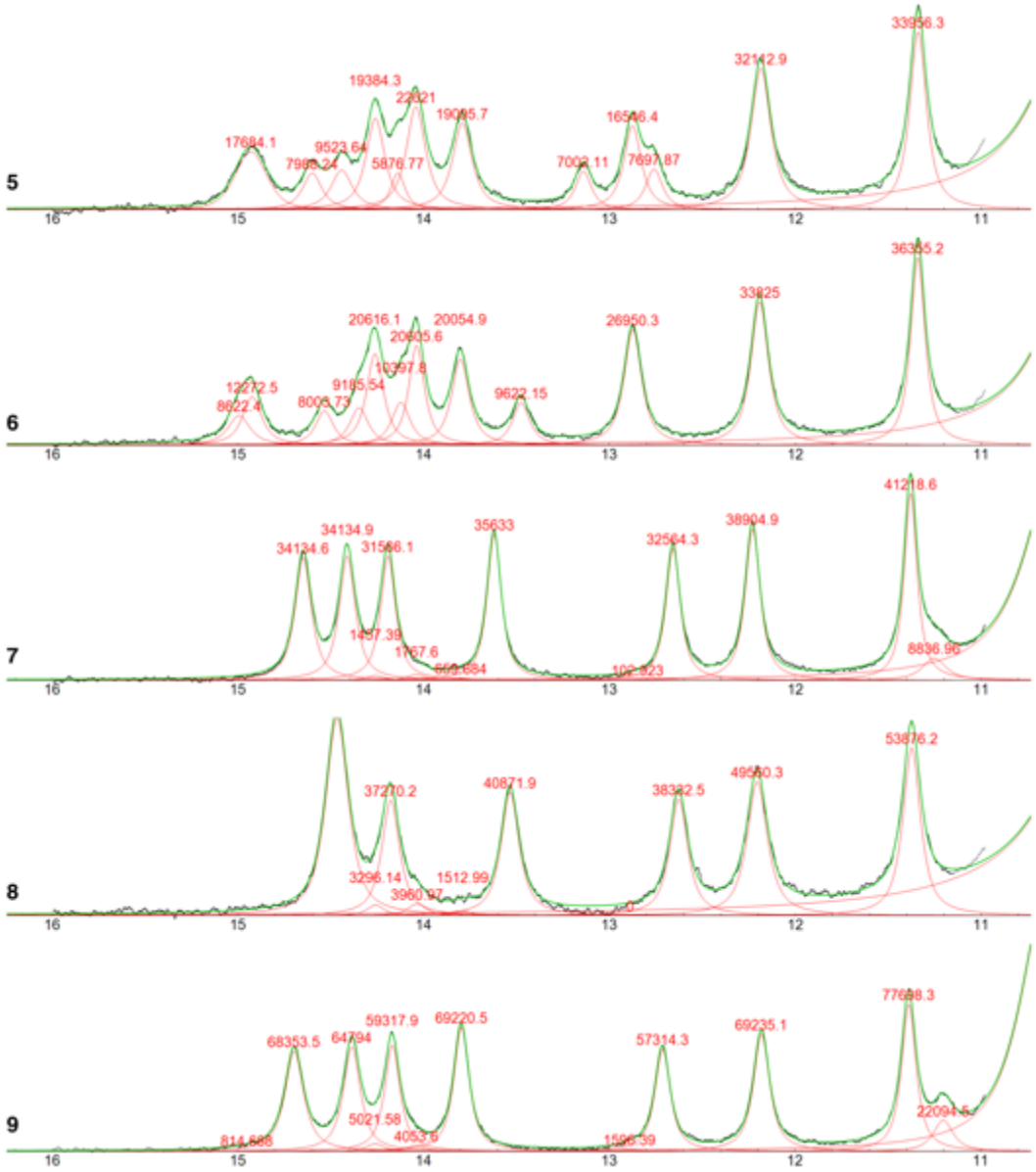


Figure S1. Spectral deconvolution in the imino region of 1D ^{1}H NMR spectra of cells expressing CA II both untreated (**ctrl**) and treated for 1 hour with 100 µM of each compound (**1-9**). The raw NMR data (black), the Pseudo-Voigt functions of each peak (red) and the overall sum (green) are shown. Peak areas are indicated in red. In some instances, peaks with non-ideal shapes were fitted as the sum of two Pseudo-Voigt functions (e.g. **6**). An additional Pseudo-Voigt function was used to correct for baseline distortion, the area of which was not used in the normalization process.

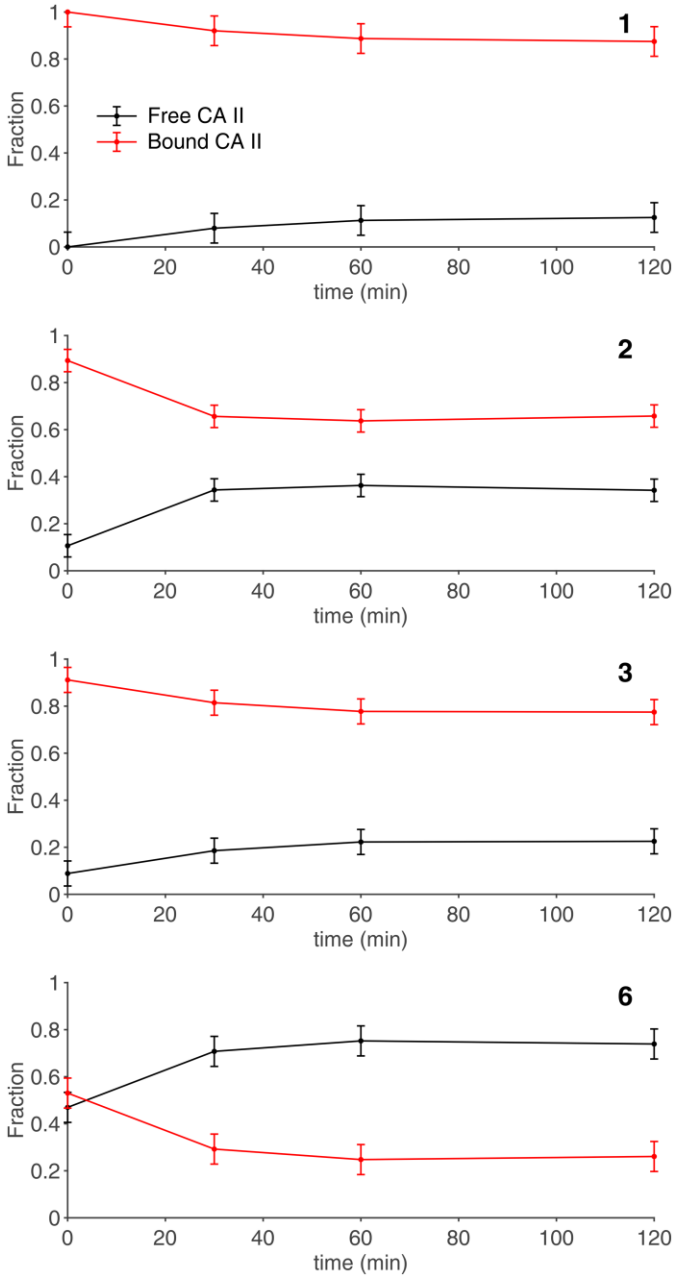


Figure S2. Time-dependence of free (black) and bound (red) CA II in cell lysates from cells treated for 1 hour with 100 μM of compounds **1-3** and **6** and subsequently washed and incubated in the absence of compound for different times. Error bars were obtained from the MCR-ALS global fitting as follows: err = 2 × [lack of fit(%)]/100.

3. CONCLUSIONS

Conclusions

During my doctoral work, I have been working on three different research projects focused on studying protein-protein and protein-ligand interactions via NMR.

The first concerned the structural characterization of the third domain of CCS, both in the homodimer and in the heterodimer with SOD1. To overcome the signal overlapping problems caused by the large size of the protein, and to make the spectrum more easily interpretable, it was decided to apply a segmental labelling technique known as protein trans splicing. To achieve this goal, different methodologies were applied, among which, a new constructs design turned out to be the most promising approach. In fact, the new design allowed us to overcome the solubility problems encountered with previous approaches and to obtain a higher yield in the trans-splicing reaction. Despite these improvements, the difficulties in purification did not allow us yet to obtain a sample of FL-CCS sufficiently pure for the NMR analysis. However, new purification strategies are possible and some parameters can be optimized to achieve better purity levels.

In parallel, a novel methodology for ligand screening directly in living human cells was developed. In-cell NMR was here applied to observe and characterize the binding and the behavior of a selected set of drugs towards an intracellular protein, i.e. CA II. From fast and cost-effective ^{1}H NMR experiments, we were able to obtain meaningful dose- and time- dependent binding curves from which it was possible to extract thermodynamic and kinetic parameters closely related to the cell permeability, and also to the affinity and selectivity of a drug for its target. Furthermore, the integration of this method with the classical drug design pipeline can provide significant information about the potency of a drug and as consequence on the required dose to exert a certain effect. This will allow to select the most promising drug candidates already at the earliest stages of the entire development process thus increasing the chances of success.

Finally, in the last research project, the previously developed method was applied to investigate the binding properties of a broader and most various selection of approved drugs towards intracellular CA II. The tested compounds included some drugs which do not have CAII as first target. Through in-cell NMR analysis, ligands were classified according to their binding behavior and three different classes emerged: I) fast-diffusing, stable binding; II) slow-diffusing, stable binding; III) fast-diffusing, unstable binding. The first two include ligands that followed a classic

diffusion-limited binding pattern, while the class three compounds showed a more complex behavior probably due to their off-target binding properties. Indeed, after an initial fast diffusion through the cell membrane and binding to CAII, they were progressively released. Although it is difficult to hypothesize the cause of this behavior, the most plausible explanation could be a competition binding effect occurring after the cells were collected, and probably induced by trypsinization.

Intracellular ligand screening by NMR revealed to be a powerful approach to accelerate the drug design process. In fact, through its application, it is not only possible to predict the specificity of a drug for its target but also to identify its possible multi-target activity. In this regard, the in-cell NMR approach may be particularly useful in the polypharmacology field in which the multi-target properties of compounds are exploited for treating complex pathological states.

4. MATERIALS AND METHODS

4.1 Cloning and ligation

CCS D1D2, CCS D3, Intein N and Intein C sequences were amplified separately trough polymerases chain reaction (PCR) using the following primers:

Primer_name	Primer Sequence
IntN_for	TGTTTAAGCTATGAAACGGAAATATTG
IntN_BamHI_rev	TATAGGATCCTTAATTCGGCAAATTATCAACCCG
Nde_6H_IntC_for	TATACATATGCATCACCATCACCATCACATCAAAATAGCCACACGTAAATATTT
IntC_rev	TGCGCTGGCCTTTTCCAGAACCCC
CCS_D3_Arg_for	TGC*TCCGCTGGCCTTTTCCAGAACC
CCS_D3_BamHI_rev	CCGGGATCCTCAAAGGTGGGCAGGGGG
D1D2_231_rev	TGCAATGATGCCACAGGCCAACCTCTC
D1D2_231_for	T7 promoter primer

BamHI Nde: restriction sites

6H: histidine tail

*mutation: - CCS_D3_Arg_for → from the original Arg (CGC) to Cys (TGC)

In red: stop codons

The PCRs were conducted with the following reagent volumes and amplification protocol.

Reagents:

H_2O: 35,5 µl
PCR buffer: 5 µl
dNTPs: 1,5 µl
Primer forward (~100 ng): 3 µl
Primer reverse (~100 ng): 3 µl
DNA (~10 ng): 1 µl
Polymerases: 1 µl
Total volume (single Eppendorf tube): 50 µl

Amplification protocol:

4 min at 95°C
30s at 95°C
45s at 60°C
30s at 72°C (30X)
5 min at 72°C
Then 4°C

The amplified fragments were ligated with T4 ligases (Thermo Fisher Scientific) at 25°C for 75' and then 65°C for 10' to deactivate the enzyme with the following reagent volumes:

Reagents

2 µl PEG
2 µl Buffer
1 µl Ligase T4
100 ng/ µl fragment 1 (different ratio was tried)
100 ng/ µl fragment 2 (different ratio was tried)
H_2O until 20 µl of total volume

The ligation product was then amplified again and cloned into a PET11a (fig. 18) plasmid between Nde and BamHI restriction site.

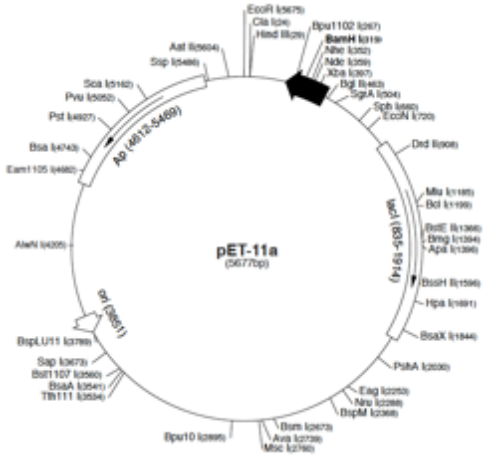


Figure 18. pET-11a expression vector map

After every phase an agarose gel electrophoresis was performed to check the correct happening of the amplification/ligation procedures.

4.2 Agarose gel electrophoresis

1.2 % agarose gel was prepared using 1 x TAE buffer (Tris base 40 mM, acetic acid 20 mM and EDTA 1 mM). Once the gel was solidified the samples previously treated with Sybr green dye were loaded. 6 µl of low/high range marker (Thermo Fisher Scientific) were used. Run was performed for 40' at 90 V.

4.3 Recombinant protein productions

4.3.1 CCS D1D2-IntN construct

E. Coli C41 DE3 strain was transformed with pET11a plasmid encoding D1D2-IntN synthetized construct and then incubated at 37°C ON on LB agar plate with ampicillin antibiotic. The day after a 1-liter LB medium flask culture was prepared starting from a selected colony and bacterial cells were incubated at 37°C 180 rpm. Protein expression was induced at OD_{600} = 0.6 with 0.5 mM of IPTG and growth temperature was lowered at 30°C. After 4h from the induction cells were harvested and centrifuged at 4500 rpm for 20'. The pellet was resuspended in Tris 20 mM, NaCl 0.5 M, Imidazole 0.5 mM, TCEP 0.5 mM pH 8 purification buffer and cell

lysis was performed through a series of sonication cycles. The soluble fraction was then extracted by centrifugation at 8000 rpm for 90' and loaded onto a previously equilibrated His-trap affinity Ni-column. The elution was performed using Tris 20 mM, NaCl 0.5 M, Imidazole 0.5 M, TCEP 0.5 mM pH 8 buffer. The final purification step consisted in a size-exclusion chromatography using a Superdex75 16/60 75 pg column in Tris 20 mM, NaCl 0.5 M TCEP 0.5 mM pH 8 buffer.
The purification under denaturing conditions was realized following the previously reported protocol at 8 M of urea concentration.

4.3.2 CCS IntC-D3 construct

E. Coli BL21 Gold strain was transformed with pET11a plasmid encoding IntC-D3 synthetized construct and then incubated at 37°C ON on LB agar plate with ampicillin antibiotic. The day after a 1-liter M9 unlabelled medium flask culture was prepared starting from a selected colony and bacterial cells were incubated at 37°C 180 rpm. Protein expression was induced at OD_{600} = 0.6 with 0.5 mM of IPTG and growth temperature was lowered at 20°C ON. The purification both in normal and under denaturing conditions was performed using the previously reported protocol.

4.3.3 Strep-MBP-TEV-CCS_D1D1-IntN

E. Coli BL21 Gold strain was transformed with plasmids encoding Strep-MBP-TEV-CCS_D1D1-IntN new synthetized construct and then incubated at 37°C ON on LB agar plate with kanamycin antibiotic. The day after a 1-liter LB medium flask culture was prepared starting from a selected colony and bacterial cells were incubated at 37°C 180 rpm. Protein expression was induced at OD_{600} = 0.6 with 1 mM of IPTG and growth temperature was lowered at 30°C ON. Bacterial cells were harvested and centrifuged at 4500 rpm for 20' and then the pellet was resuspended in Tris 20 mM, NaCl 0.5 M, TCEP 0.5 mM pH 8 buffer.

4.3.4 MBP-IntC-CCS_D3_SGV

E. Coli BL21 Gold strain was transformed with plasmids encoding MBP-IntC-CCS_D3_SGV new synthetized construct and then incubated at 37°C ON on LB agar plate with ampicillin antibiotic. The day after a 1-liter LB medium flask culture was prepared starting from a selected colony and bacterial cells were incubated at

37°C 180 rpm. Protein expression was induced at OD_{600} = 0.6 with 1 mM of IPTG and growth temperature was lowered at 30°C. After 5h from the induction cells were harvested and centrifuged at 4500 rpm for 20'. The pellet was then resuspended in Tris 20 mM, NaCl 0.5 M, TCEP 0.5 mM pH 8 buffer.

4.3.5 D1D2-IntN and IntC-D3 constructs refolding

The two constructs were refolded through a multi-step dialyses against buffers in decreasing concentration of urea. Dialyses membranes with the appropriate cut off were kept under agitation at 4°C O.N. The steps were the following:

- Urea 6 M, Tris 20 mM, NaCl 0.5 M, TCEP 0.5 mM pH 8 buffer
- Urea 4 M, Tris 20 mM, NaCl 0.5 M, TCEP 0.5 mM pH 8 buffer
- Urea 2 M, Tris 20 mM, NaCl 0.5 M, TCEP 0.5 mM pH 8 buffer
- Tris 20 mM, NaCl 0.5 M, TCEP 0.5 mM pH 8 buffer

4.4 Protein trans-splicing reaction

4.4.1 CCS D1D2-IntN and CCS IntC-D3 constructs

Protein trans-splicing reaction was performed following a reported protocol (Muona et al. 2010). CCS D1D2-IntN and CCS IntC-D3 purified proteins, suspended in Tris 20 mM, NaCl 0.5 M, TCEP 0.5 mM pH 8 buffer, were mixed equimolarly at a final concentration of 20-100 µM in a volume of 10-50 ml. The reaction was carried out for 24 h at different temperatures (25°, 30° or 37°C) without shaking. During the ligation reaction, several withdrawals (10 µl each) were made at different time to check the reaction progression through SDS-PAGE gel. Every 10 µl aliquot was mixed with 1 x SDS sample buffer and stored at -20°C until the analysis.

The protein trans-splicing reaction under denaturing conditions was performed following the previously reported protocol at 4M urea concentration.

4.4.2 Strep-MBP-TEV-CCS_D1D1-IntN and MBP-IntC-CCS_D3_SGV

Strep-MBP-TEV-CCS_D1D1-IntN and CCS IntC-D3 bacterial cell pellets were resuspended in Tris 20 mM, NaCl 0.5 M, TCEP 0.5 mM pH 8 buffer and mixed with 1:3 ratio in a total volume of ~60 ml. After some cycles of sonication, the

soluble fraction was separated from the cellular debris through centrifugation at 8000 rpm for 90'. The mixture was then incubated at room temperature for 24h without agitation. Aliquots were taken at different time for monitoring the reaction. After 24 h the reaction mixture was loaded onto a previously equilibrated His-trap affinity Ni-column and washed with Tris 20 mM, NaCl 0.5 M, Imidazole 0.5 mM, TCEP 0.5 mM pH 8 buffer. The elution was performed using Tris 20 mM, NaCl 0.5 M, Imidazole 0.5 M, TCEP 0.5 mM pH 8 buffer.

A size-exclusion chromatography using a Superdex75 16/60 column washed with Tris 20 mM, NaCl 0.5 M TCEP 0.5 mM pH 8 buffer was run. Finally, an anionic exchange chromatography was performed. The protein was eluted in Tris 20 mM, TCEP 0.5 mM pH 8 buffer and loaded onto a Hi-Prep Q HP 16/10. Tris 20 mM, NaCl 0.5 M TCEP 0.5 mM pH 8 buffer was used for the elution step.

4.5 Polyacrylamide gel electrophoresis

SDS-PAGE methodology was used to separate and thus identify expressed proteins depending on their electrophoretic mobility through a polyacrylamide gel. Protein aliquots of 10 µl were mixed with 1 x sample buffer containing sodium dodecyl phosphate (SDS) and boiled for 5 minutes. After 1' of centrifugation at 16000 x g, 6 µl of each sample was loaded on polyacrylamide gradient 4-20% mini-PROTEAN® TGX™ precast protein gel (BIORAD). Electrophoretic run was performed for at least 25' at 200 V using 25 mM Tris, 192 mM glycine, 0.1% SDS buffer in a mini-PROTEAN® vertical electrophoresis chamber (BIORAD). Proteins were then stained using Coomassie brilliant blue dye (Giotto biotech).

4.6 Western Blot analysis (WB)

Western Blot technique was performed to identify the protein of interest, i.e. FL-CCS through the use of highly specific antibody. Proteins previously separated by size thanks SDS-PAGE were transferred onto a nitrocellulose-matrix membrane (BIORAD) using a trans-Blot Turbo machine (BIORAD). The membrane was incubated for 1 h at room temperature in a 5% of non-fat dry milk Tris-buffered saline (TBS) solution with 0.005% of Tween-20 detergent to saturate all the non-specific binding sites. The blocking phase was followed by a 2h of incubation at room temperature with a monoclonal primary anti-FL_CCS antibody. After a series of 10 minutes washes with TBS 1x, a secondary antibody conjugated with

peroxidase enzyme was used to incubate the membrane at least for 1h at room temperature. A last wash with TBS 1x was then performed. The protein-antibody complexes were detected thanks to a chemiluminescent reaction obtained through the use of the EUROCLONE detection Kit. The reaction was observed at ChemiDocTMXRS.

5. REFERENCES

References

Aranko, A Sesilja, and Gerrit Volkmann. 2011. “Protein Trans- Splicing as a Protein Ligation Tool to Study” 2: 183–98. https://doi.org/10.1515/BMC.2011.014.

Arnesano, Fabio, Lucia Banci, Ivano Bertini, Manuele Martinelli, Yoshiaki Furukawa, and Thomas V. O’Halloran. 2004. “The Unusually Stable Quaternary Structure of Human Cu,Zn-Superoxide Dismutase 1 Is Controlled by Both Metal Occupancy and Disulfide Status.” *Journal of Biological Chemistry* 279 (46): 47998–3. https://doi.org/10.1074/jbc.M406021200.

Banci, Lucia, Letizia Barbieri, Ivano Bertini, Enrico Luchinat, Erica Secci, Yuguang Zhao, and A. Radu Aricescu. 2013. “Atomic-Resolution Monitoring of Protein Maturation in Live Human Cells by NMR.” *Nature Chemical Biology* 9 (5): 297–99. https://doi.org/10.1038/nchembio.1202.

Banci, Lucia, Ivano Bertini, Francesca Cantini, Nicola D’Amelio, and Elena Gaggelli. 2006. “Human SOD1 before Harboring the Catalytic Metal: Solution Structure of Copper-Depleted, Disulfide-Reduced Form.” *Journal of Biological Chemistry* 281 (4): 2333–37. https://doi.org/10.1074/jbc.M506497200.

Banci, Lucia, Ivano Bertini, Francesca Cantini, Tatiana Kozyreva, Chiara Massagni, Peep Palumaa, Jeffrey T. Rubino, and Kairit Zovo. 2012. “Human Superoxide Dismutase 1 (HSOD1) Maturation through Interaction with Human Copper Chaperone for SOD1 (HCCS).” *Proceedings of the National Academy of Sciences of the United States of America* 109 (34): 13555–60. https://doi.org/10.1073/pnas.1207493109.

Barbieri, Letizia, Enrico Luchinat, and Lucia Banci. 2015. “Protein Interaction Patterns in Different Cellular Environments Are Revealed by In-Cell NMR.” *Scientific Reports* 5 (September): 14456. https://doi.org/10.1038/srep14456.

———. 2016. “Characterization of Proteins by In-Cell NMR Spectroscopy in Cultured Mammalian Cells.” *Nature Protocols* 11 (6): 1101–11. https://doi.org/10.1038/nprot.2016.061.

Bassetto, Marcella, Alberto Massarotti, Antonio Coluccia, and Andrea Brancale. 2016. “Structural Biology in Antiviral Drug Discovery.” *Current Opinion in Pharmacology* 30: 116–30. https://doi.org/10.1016/j.coph.2016.08.014.

Becker, Holger M, and Joachim W Deitmer. 2007. “Carbonic Anhydrase II Increases the Activity of the Human Electrogenic Na+/ Cotransporter.” *Journal of Biological Chemistry* 282 (18): 13508–21. https://doi.org/10.1074/jbc.M700066200.

Berrade, Luis, and Julio A. Camarero. 2009. “Expressed Protein Ligation: A Resourceful Tool to Study Protein Structure and Function.” *Cellular and Molecular Life Sciences* 66 (24): 3909–22. https://doi.org/10.1007/s00018-009-0122-3.

Bertrand, Karl, Sergey Reverdatto, David S. Burz, Richard Zitomer, and Alexander Shekhtman. 2012. “Structure of Proteins in Eukaryotic Compartments.” *Journal of the American Chemical Society* 134 (30): 12798–806.

https://doi.org/10.1021/ja304809s.

Binolfi, Andres, Antonio Limatola, Silvia Verzini, Jonas Kosten, Francois Xavier Theillet, Honor May Rose, Beata Bekei, Marchel Stuiver, Marleen Van Rossum, and Philipp Selenko. 2016. "Intracellular Repair of Oxidation-Damaged α-Synuclein Fails to Target C-Terminal Modification Sites." *Nature Communications* 7: 1–10. https://doi.org/10.1038/ncomms10251.

Bitan, Gal, Marina D Kirkitadze, Aleksey Lomakin, Sabrina S Vollers, George B Benedek, and David B Teplow. 2003. "Amyloid β-Protein (Aβ) Assembly: Aβ40 and Aβ42 Oligomerize through Distinct Pathways." *Proceedings of the National Academy of Sciences* 100 (1): 330 LP – 335. https://doi.org/10.1073/pnas.222681699.

Blundell, Tom L., Harren Jhoti, and Chris Abell. 2002. "High-Throughput Crystallography for Lead Discovery in Drug Design." *Nature Reviews Drug Discovery* 1 (1): 45–54. https://doi.org/10.1038/nrd706.

Borchelt, D R, M K Lee, H S Slunt, M Guarnieri, Z S Xu, P C Wong, R H Brown, D L Price, S S Sisodia, and D W Cleveland. 1994. "Superoxide Dismutase 1 with Mutations Linked to Familial Amyotrophic Lateral Sclerosis Possesses Significant Activity." *Proceedings of the National Academy of Sciences* 91 (17): 8292 LP – 8296. https://doi.org/10.1073/pnas.91.17.8292.

Brown, D, T Kumpulainen, J Roth, and L Orci. 1983. "Immunohistochemical Localization of Carbonic Anhydrase in Postnatal and Adult Rat Kidney." *American Journal of Physiology-Renal Physiology* 245 (1): F110–18. https://doi.org/10.1152/ajprenal.1983.245.1.F110.

Bruijn, Lucie I, Megan K Houseweart, Shinsuke Kato, Karen L Anderson, Scott D Anderson, Eisaku Ohama, Andrew G Reaume, Rick W Scott, and Don W Cleveland. 1998. "Aggregation and Motor Neuron Toxicity of an ALS-Linked SOD1 Mutant Independent from Wild-Type SOD1." *Science* 281 (5384): 1851 LP – 1854. https://doi.org/10.1126/science.281.5384.1851.

Bruijn, Lucie I, Timothy M Miller, and Don W Cleveland. 2004. "UNRAVELING THE MECHANISMS INVOLVED IN MOTOR NEURON DEGENERATION IN ALS." *Annual Review of Neuroscience* 27 (1): 723–49. https://doi.org/10.1146/annurev.neuro.27.070203.144244.

Caruano-yzermans, Amy L, Thomas B Bartnikas, and Jonathan D Gitlin. 2006. "Mechanisms of the Copper-Dependent Turnover of the Copper Chaperone for Superoxide Dismutase *" 281 (19): 13581–87. https://doi.org/10.1074/jbc.M601580200.

Casareno, Ruby Leah B., Darrel Waggoner, and Jonathan D. Gitlin. 1998. "The Copper Chaperone CCS Directly Interacts with Copper/Zinc Superoxide Dismutase." *Journal of Biological Chemistry* 273 (37): 23625–28. https://doi.org/10.1074/jbc.273.37.23625.

Chegwidden, W Richard, and Nicholas D Carter. 2000. "Introduction to the Carbonic Anhydrases BT - The Carbonic Anhydrases: New Horizons." In , edited

by W Richard Chegwidden, Nicholas D Carter, and Yvonne H Edwards, 13–28. Basel: Birkhäuser Basel. https://doi.org/10.1007/978-3-0348-8446-4_2.

Dawson, P E, T W Muir, I Clark-Lewis, and S B Kent. 1994. "Synthesis of Proteins by Native Chemical Ligation." *Science (New York, N.Y.)* 266 (5186): 776–79. https://doi.org/10.1126/science.7973629.

Debelouchina, Galia T, and Tom W Muir. 2018. "REVIEW A Molecular Engineering Toolbox for the Structural Biologist." https://doi.org/10.1017/S0033583517000051.

Ditte, Peter, Franck Dequiedt, Eliska Svastova, Alzbeta Hulikova, Anna Ohradanova-Repic, Miriam Zatovicova, Lucia Csaderova, et al. 2011. "Phosphorylation of Carbonic Anhydrase IX Controls Its Ability to Mediate Extracellular Acidification in Hypoxic Tumors." *Cancer Research* 71 (24): 7558 LP – 7567. https://doi.org/10.1158/0008-5472.CAN-11-2520.

Fetherolf, Morgan M., Stefanie D. Boyd, Alexander B. Taylor, Hee Jong Kim, James A. Wohlschlegel, Ninian J. Blackburn, P. John Hart, Dennis R. Winge, and Duane D. Winkler. 2017. "Copper-Zinc Superoxide Dismutase Is Activated through a Sulfenic Acid Intermediate at a Copper Ion Entry Site." *Journal of Biological Chemistry* 292 (29): 12025–40. https://doi.org/10.1074/jbc.M117.775981.

Frost, Susan C. 2014. *Physiological Functions of the Alpha Class of Carbonic Anhydrases*. *Sub-Cellular Biochemistry*. Vol. 75. https://doi.org/10.1007/978-94-007-7359-2_2.

Furukawa, Yoshiaki, Andrew S Torres, and Thomas V O'Halloran. 2004. "Oxygen-Induced Maturation of SOD1: A Key Role for Disulfide Formation by the Copper Chaperone CCS." *The EMBO Journal* 23 (14): 2872–81. https://doi.org/10.1038/sj.emboj.7600276.

Hajduk, Philip J, Edward T Olejniczak, and Stephen W Fesik. 1997. "One-Dimensional Relaxation- and Diffusion-Edited NMR Methods for Screening Compounds That Bind to Macromolecules." *Journal of the American Chemical Society* 119 (50): 12257–61. https://doi.org/10.1021/ja9715962.

Hamatsu, Jumpei, Daniel O'Donovan, Takashi Tanaka, Takahiro Shirai, Yuichiro Hourai, Tsutomu Mikawa, Teppei Ikeya, et al. 2013. "High-Resolution Heteronuclear Multidimensional NMR of Proteins in Living Insect Cells Using a Baculovirus Protein Expression System." *Journal of the American Chemical Society* 135 (5): 1688–91. https://doi.org/10.1021/ja310928u.

Hayward, Lawrence J., Jorge A. Rodriguez, Ji W. Kim, Ashutosh Tiwari, Joy J. Goto, Diane E. Cabelli, Joan Selverstone Valentine, and Robert H. Brown. 2002. "Decreased Metallation and Activity in Subsets of Mutant Superoxide Dismutases Associated with Familial Amyotrophic Lateral Sclerosis." *Journal of Biological Chemistry* 277 (18): 15923–31. https://doi.org/10.1074/jbc.M112087200.

Henry, R P. 1996. "Multiple Roles of Carbonic Anhydrase in Cellular Transport and Metabolism." *Annual Review of Physiology* 58 (1): 523–38.

https://doi.org/10.1146/annurev.ph.58.030196.002515.

Inomata, Kohsuke, Ayako Ohno, Hidehito Tochio, Shin Isogai, Takeshi Tenno, Ikuhiko Nakase, Toshihide Takeuchi, et al. 2009. “High-Resolution Multi-Dimensional NMR Spectroscopy of Proteins in Human Cells.” *Nature* 458 (7234): 106–9. https://doi.org/10.1038/nature07839.

Kirby, Kim, Laran T Jensen, Janet Binnington, Arthur J Hilliker, Janella Ulloa, Valeria C Culotta, and John P Phillips. 2008. “Instability of Superoxide Dismutase 1 of Drosophila in Mutants Deficient for Its Cognate Copper Chaperone.” *Journal of Biological Chemistry* 283 (51): 35393–401. https://doi.org/10.1074/jbc.M807131200.

Laliberté, Julie, Lisa J Whitson, Jude Beaudoin, Stephen P Holloway, P John Hart, and Simon Labbé. 2004. “The Schizosaccharomyces Pombe Pccs Protein Functions in Both Copper Trafficking and Metal Detoxification Pathways.” *Journal of Biological Chemistry* 279 (27): 28744–55. https://doi.org/10.1074/jbc.M403426200.

Lamb, Audrey L., Andrew S. Torres, Thomas V. O’Halloran, and Amy C. Rosenzweig. 2001. “Heterodimeric Structure of Superoxide Dismutase in Complex with Its Metallochaperone.” *Nature Structural Biology* 8 (9): 751–55. https://doi.org/10.1038/nsb0901-751.

Lamb, Audrey L, Amy K Wernimont, Robert A Pufahl, Valeria C Culotta, Thomas V O’Halloran, and Amy C Rosenzweig. 1999. “Crystal Structure of the Copper Chaperone for Superoxide Dismutase.” *Nature Structural Biology* 6 (8): 724–29. https://doi.org/10.1038/11489.

Lamb, Audrey L, Amy K Wernimont, Robert A Pufahl, Thomas V O’Halloran, and Amy C Rosenzweig. 2000. “Crystal Structure of the Second Domain of the Human Copper Chaperone for Superoxide Dismutase,.” *Biochemistry* 39 (7): 1589–95. https://doi.org/10.1021/bi992822i.

Lashuel, Hilal A, Dean Hartley, Benjamin M Petre, Thomas Walz, and Peter T Lansbury. 2002. “Amyloid Pores from Pathogenic Mutations.” *Nature* 418 (6895): 291. https://doi.org/10.1038/418291a.

Li, Conggang, Jiajing Zhao, Kai Cheng, Yuwei Ge, Qiong Wu, Yansheng Ye, Guohua Xu, et al. 2017. “Magnetic Resonance Spectroscopy as a Tool for Assessing Macromolecular Structure and Function in Living Cells.” *Annual Review of Analytical Chemistry (Palo Alto, Calif.)* 10 (1): 157–82. https://doi.org/10.1146/annurev-anchem-061516-045237.

Li, Xiuju, Bernardo Alvarez, Joseph R Casey, Reinhart A F Reithmeier, and Larry Fliegel. 2002. “Carbonic Anhydrase II Binds to and Enhances Activity of the Na+/H+ Exchanger .” *Journal of Biological Chemistry* 277 (39): 36085–91. https://doi.org/10.1074/jbc.M111952200.

Lindberg, Mikael J., Lena Tibell, and Mikael Oliveberg. 2002. “Common Denominator of Cu/Zn Superoxide Dismutase Mutants Associated with

Amyotrophic Lateral Sclerosis: Decreased Stability of the Apo State." *Proceedings of the National Academy of Sciences of the United States of America* 99 (26): 16607–12. https://doi.org/10.1073/pnas.262527099.

Luchinat, Enrico, and Lucia Banci. 2017. "In-Cell NMR: A Topical Review," 108–18. https://doi.org/10.1107/S2052252516020625.
———. 2018. "In-Cell NMR in Human Cells: Direct Protein Expression Allows Structural Studies of Protein Folding and Maturation." Research-article. *Accounts of Chemical Research* 51: 1550–57. https://doi.org/10.1021/acs.accounts.8b00147.

Luchinat, Enrico, Letizia Barbieri, and Lucia Banci. 2017. "OPEN A Molecular Chaperone Activity of CCS Restores the Maturation of SOD1 FALS Mutants." *Scientific Reports*, no. November: 1–8. https://doi.org/10.1038/s41598-017-17815-y.

Mayer, Moriz, and Bernd Meyer. 1999. "Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy." *Angewandte Chemie International Edition* 38 (12): 1784–88. https://doi.org/10.1002/(SICI)1521-3773(19990614)38:12<1784::AID-ANIE1784>3.0.CO;2-Q.

McCord, Joe M, and Irwin Fridovich. 1969. "Superoxide Dismutase: AN ENZYMIC FUNCTION FOR ERYTHROCUPREIN (HEMOCUPREIN) ." *Journal of Biological Chemistry* 244 (22): 6049–55. http://www.jbc.org/content/244/22/6049.abstract.

Mootz, Henning D. 2009. "Split Inteins as Versatile Tools for Protein Semisynthesis." *ChemBioChem* 10 (16): 2579–89. https://doi.org/10.1002/cbic.200900370.

Muona, Mikko, A Sesilja Aranko, Vytas Raulinaitis, and Hideo Iwaï. 2010. "Segmental Isotopic Labeling of Multi-Domain and Fusion Proteins by Protein Trans-Splicing in Vivo and in Vitro." *Nature Protocols* 5 (3): 574–87. https://doi.org/10.1038/nprot.2009.240.

Ogino, Shinji, Satoshi Kubo, Ryo Umemoto, Shuxian Huang, Noritaka Nishida, and Ichio Shimada. 2009. "Observation of NMR Signals from Proteins Introduced into Living Mammalian Cells by Reversible Membrane Permeabilization Using a Pore-Forming Toxin, Streptolysin O." *Journal of the American Chemical Society* 131 (31): 10834–35. https://doi.org/10.1021/ja904407w.

Okado-Matsumoto, Ayako, and Irwin Fridovich. 2001. "Subcellular Distribution of Superoxide Dismutases (SOD) in Rat Liver: Cu,Zn-SOD IN MITOCHONDRIA ." *Journal of Biological Chemistry* 276 (42): 38388–93. https://doi.org/10.1074/jbc.M105395200.

Pellecchia, Maurizio, Ivano Bertini, David Cowburn, Claudio Dalvit, Ernest Giralt, Wolfgang Jahnke, Thomas L. James, et al. 2008. "Perspectives on NMR in Drug Discovery: A Technique Comes of Age." *Nature Reviews Drug Discovery* 7 (9): 738–45. https://doi.org/10.1038/nrd2606.

Rabizadeh, S, E B Gralla, D R Borchelt, R Gwinn, J S Valentine, S Sisodia, P Wong,

M Lee, H Hahn, and D E Bredesen. 1995. “Mutations Associated with Amyotrophic Lateral Sclerosis Convert Superoxide Dismutase from an Antiapoptotic Gene to a Proapoptotic Gene: Studies in Yeast and Neural Cells.” *Proceedings of the National Academy of Sciences* 92 (7): 3024 LP – 3028. https://doi.org/10.1073/pnas.92.7.3024.

Rae, T D, P J Schmidt, R A Pufahl, V C Culotta, and T V. O'Halloran. 1999. “Undetectable Intracellular Free Copper: The Requirement of a Copper Chaperone for Superoxide Dismutase.” *Science* 284 (5415): 805 LP – 808. https://doi.org/10.1126/science.284.5415.805.

Reaume, Andrew. G, Jeffrey L Elliott, Eric K Hoffman, Neil W Kowall, Robert J Ferrante, Donald R Siwek, Heide M Wilcox, et al. 1996. “Motor Neurons in Cu/Zn Superoxide Dismutase-Deficient Mice Develop Normally but Exhibit Enhanced Cell Death after Axonal Injury.” *Nature Genetics* 13 (1): 43–47. https://doi.org/10.1038/ng0596-43.

Rothstein, Jeffrey D, Margaret Dykes-Hoberg, Laura B Corson, Mark Becker, Don W Cleveland, Donald L Price, Valeria Cizewski Culotta, and Phillip C Wong. 1999. “The Copper Chaperone CCS Is Abundant in Neurons and Astrocytes in Human and Rodent Brain.” *Journal of Neurochemistry* 72 (1): 422–29. https://doi.org/10.1046/j.1471-4159.1999.0720422.x.

Rowland, Lewis P, and Neil A Shneider. 2001. “Amyotrophic Lateral Sclerosis.” *New England Journal of Medicine* 344 (22): 1688–1700. https://doi.org/10.1056/NEJM200105313442207.

Sakai, Tomomi, Hidehito Tochio, Takeshi Tenno, Yutaka Ito, Tetsuro Kokubo, Hidekazu Hiroaki, and Masahiro Shirakawa. 2006. “In-Cell NMR Spectroscopy of Proteins inside Xenopus Laevis Oocytes.” *Journal of Biomolecular NMR* 36 (3): 179–88. https://doi.org/10.1007/s10858-006-9079-9.

Schmidt, Paul J., Tracey D. Rae, Robert A. Pufahl, Tomoko Hamma, Jeff Strain, Thomas V. O’Halloran, and Valeria C. Culotta. 1999. “Multiple Protein Domains Contribute to the Action of the Copper Chaperone for Superoxide Dismutase.” *Journal of Biological Chemistry* 274 (34): 23719–25. https://doi.org/10.1074/jbc.274.34.23719.

Selenko, Philipp, Zach Serber, Bedrick Gadea, Joan Ruderman, and Gerhard Wagner. 2006. “Quantitative NMR Analysis of the Protein G B1 Domain in Xenopus Laevis Egg Extracts and Intact Oocytes.” *Proceedings of the National Academy of Sciences of the United States of America* 103 (32): 11904–9. https://doi.org/10.1073/pnas.0604667103.

Serber, Zach, and Volker Dötsch. 2001. “In-Cell NMR Spectroscopy.” *Biochemistry* 40 (48): 14317–23. https://doi.org/10.1021/bi011751w.

Sherman, Michael Y, and Alfred L Goldberg. 2001. “Cellular Defenses against Unfolded Proteins: A Cell Biologist Thinks about Neurodegenerative Diseases.” *Neuron* 29 (1): 15–32. https://doi.org/10.1016/S0896-6273(01)00177-5.

Sterling, Deborah, Reinhart A.F. Reithmeier, and Joseph R. Casey. 2001a. "Carbonic Anhydrase: In the Driver's Seat for Bicarbonate Transport." *Journal of the Pancreas* 2 (4): 165–70.

Sterling, Deborah, Reinhart A F Reithmeier, and Joseph R Casey. 2001b. "A Transport Metabolon: FUNCTIONAL INTERACTION OF CARBONIC ANHYDRASE II AND CHLORIDE/BICARBONATE EXCHANGERS ." *Journal of Biological Chemistry* 276 (51): 47886–94. https://doi.org/10.1074/jbc.M105959200.

Sugiki, Toshihiko, Kyoko Furuita, Toshimichi Fujiwara, and Chojiro Kojima. 2018. "Current NMR Techniques for Structure-Based Drug Discovery." *Molecules* 23 (1): 1–27. https://doi.org/10.3390/molecules23010148.

Supuran, Claudiu T. 2008. "Carbonic Anhydrases An Overview." *Current Pharmaceutical Design*. https://doi.org/http://dx.doi.org/10.2174/138161208783877884.

Swenson, Erik R. 2000. "Respiratory and Renal Roles of Carbonic Anhydrase in Gas Exchange and Acid-Base Regulation BT - The Carbonic Anhydrases: New Horizons." In *The Carbonic Anhydrases: New Horizons*, edited by W Richard Chegwidden, Nicholas D Carter, and Yvonne H Edwards, 281–341. Basel: Birkhäuser Basel. https://doi.org/10.1007/978-3-0348-8446-4_15.

Theillet, Francois Xavier, Andres Binolfi, Beata Bekei, Andrea Martorana, Honor May Rose, Marchel Stuiver, Silvia Verzini, et al. 2016. "Structural Disorder of Monomeric α-Synuclein Persists in Mammalian Cells." *Nature* 530 (7588): 45–50. https://doi.org/10.1038/nature16531.

Thiry, Anne, Jean-Michel Dogné, Bernard Masereel, and Claudiu T Supuran. 2006. "Targeting Tumor-Associated Carbonic Anhydrase IX in Cancer Therapy." *Trends in Pharmacological Sciences* 27 (11): 566–73. https://doi.org/10.1016/j.tips.2006.09.002.

Valentine, Joan Selverstone, Peter A. Doucette, and Soshanna Zittin Potter. 2005. "Copper-Zinc Superoxide Dismutase and Amyotrophic Lateral Sclerosis." *Annual Review of Biochemistry* 74 (1): 563–93. https://doi.org/10.1146/annurev.biochem.72.121801.161647.

Vila-Perelló, Miquel, and Tom W Muir. 2010. "Biological Applications of Protein Splicing." *Cell* 143 (2): 191–200. https://doi.org/10.1016/j.cell.2010.09.031.

Vince, John W, and Reinhart A F Reithmeier. 1998. "Carbonic Anhydrase II Binds to the Carboxyl Terminus of Human Band 3, the Erythrocyte Cl−/HCO3 −Exchanger ." *Journal of Biological Chemistry* 273 (43): 28430–37. https://doi.org/10.1074/jbc.273.43.28430.

———. 2000. "Identification of the Carbonic Anhydrase II Binding Site in the Cl-/HCO3- Anion Exchanger AE1." *Biochemistry* 39 (18): 5527–33. https://doi.org/10.1021/bi992564p.

Xu, R, B Ayers, D Cowburn, and T W Muir. 1999. "Chemical Ligation of Folded

Recombinant Proteins: Isotopic Labeling of Domains for NMR Studies." *Proceedings of the National Academy of Sciences of the United States of America* 96 (2): 388–93. https://doi.org/10.1073/pnas.96.2.388.

Zettler, Joachim, Vivien Schütz, and Henning D Mootz. 2009. "The Naturally Split Npu DnaE Intein Exhibits an Extraordinarily High Rate in the Protein Trans - Splicing Reaction." *FEBS Letters* 583 (5): 909–14. https://doi.org/10.1016/j.febslet.2009.02.003.

6. ACKNOWLEDGMENTS

ACKNOWLEDGMENTS

I would like to thank my tutor Prof. Lucia Banci, for giving me the opportunity to work at CERM during my doctoral years.

I would like to thank all the members of in-cell NMR group, Dr. Enrico Luchinat, Dr. Letizia Barbieri and Dr. Panagis Polykretis, for supporting me outside and inside laboratory through precious advices and suggestions.

I would like to thank all “cermians” for helping me in every moment of my doctoral studies.

I would like to thank my family for supporting me and being close to me in these years.

And finally, I would like to thank my girlfriend, Annalisa, who helped me to overcome all moments of difficulty and discouragement.