



UNIVERSITÀ
DEGLI STUDI
FIRENZE



HR EXCELLENCE IN RESEARCH

DOTTORATO DI RICERCA
INTERNATIONAL DOCTORATE IN STRUCTURAL BIOLOGY

CICLO XXXVIII

COORDINATOR Prof. Roberta Pierattelli

***^{19}F in-cell NMR to investigate protein-
ligand interactions in living human cells***

Settore Scientifico Disciplinare CHEM-03/A

PhD student
Azzurra Costantino

Supervisor
Prof. Lucia Bandi

Coordinator
Prof. Roberta Pierattelli

November 2022 – October 2025

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



*Progetto finanziato dall'Unione europea- Next Generation EU, Missione 4 Componente 1 CUP
B53C22001790006*

The consultation of the thesis is free. Unless a specific authorization is obtained from the author, the thesis can be, however, downloaded and printed only for strictly personal purposes related to study, research and teaching, with the explicit exclusion of any use that has – even indirectly – a commercial nature.

This thesis is dedicated to Francesca

Acknowledgements

I would like to express my deepest gratitude to Prof. Lucia Banci for her supervision and guidance throughout my PhD program. I am also sincerely thankful to Dr. Enrico Luchinat and Dr. Letizia Barbieri for their constant support, both in the laboratory and in the office, and for all that they have taught me over these three years.

A special thank you goes to my colleagues Martina Rosati and Lan B. T. Pham for their kindness, support, and for being there to listen when the workload became overwhelming. I am also grateful to Francesca Rossi, who, despite being at CERM only briefly to complete her thesis, provided great help, especially in completing the experiments.

I would like to thank the entire CERM and all the Cermians. In particular, Bora, Rosanna and Alessia who were by my side and with whom I shared many moments in the lab and under the CERM portico. My thanks also go to Niccolò and Linda for sharing the ScienzEstate adventure with me.

My heartfelt thanks go to my family, especially my mother and my sister Aurora, for their constant support throughout this journey, which at times felt like an uphill battle, and for helping me make important decisions along the way.

Although not a traditional acknowledgement, I would like to make a special mention of Francesca Nencetti, to whom I dedicate this thesis. You were there for me in my most difficult moments, even while facing your own challenges. I am sure you too would have reached this important milestone. Despite everything, you had found your path in the world of biology, and I know you would have faced all the challenges of the scientific world with strength and determination.

Finally, I would like to thank myself for finding the strength to face this journey, which at times felt insurmountable, and for overcoming the many challenges that these three years of doctoral study have brought.

Summary

In this PhD project, in-cell ^{19}F NMR spectroscopy approaches were developed and applied to investigate target–ligand interactions directly in living human cells. Two complementary strategies were pursued: a target-based approach and a ligand-based approach.

The target-based approach focused on expressing fluorinated proteins in HEK293T cells by controlling the incorporation of fluorinated aromatic amino acids. The study showed that fluorine incorporation occurs stochastically, resulting in a compromise between spectral sensitivity and sample homogeneity. High incorporation levels provide stronger ^{19}F signals but result in heterogeneous protein populations, causing signal splitting in the NMR spectra, whereas low incorporation enhances spectral resolution and sample homogeneity but at the cost of lower sensitivity. These findings provide useful guidance for optimizing fluorinated protein expression in future drug-discovery applications, where target–ligand interactions could be monitored through chemical-shift perturbations.

A ^{19}F NMR ligand-based approach was implemented to investigate the intracellular binding of novel fluorinated benzenesulfonamides designed to target cytosolic carbonic anhydrase (CA) isoforms. ^{19}F NMR enabled the direct observation of intracellular ligand–target complexes, revealing isoform-dependent affinities and also identifying off-target binding events for certain compounds. After selecting the most suitable compound as a “spy” ligand, real-time competitive binding experiments were performed in an NMR bioreactor, providing quantitative ratios of dissociation constants (K_d) for various compounds. The K_d values for each isoform were then estimated by comparing the spy ligand with a known reference compound. This approach could be expanded in the future to experiments capable of simultaneously monitoring multiple CA isoforms within the same cellular sample.

Overall, this work expanded the applications of in-cell ^{19}F NMR, showing that it is a powerful and physiologically relevant method for characterising drug–target interactions within living cells. The combination of silico predictions and in vitro assays with this strategy could provide a useful tool for the selection of more effective drug candidates at the preclinical stage of drug development.

Table of contents

Table of contents.....	iii
List of figures	vi
List of tables.....	viii
List of abbreviations.....	ix
 INTRODUCTION.....	 1
1.1 IN-CELL NMR SPECTROSCOPY.....	2
1.1.1 Overview of in-cell NMR spectroscopy	2
1.1.2 Isotopic labelling and NMR strategies for probing proteins in cells	4
1.1.2.a ^1H - ^{15}N In-cell NMR	6
1.1.3.b ^{19}F in-cell NMR.....	7
1.1.3 Real-Time In-cell NMR Using a Flow Bioreactor.....	8
1.1.4 In-cell NMR for Drug Discovery and Ligand Screening.....	10
1.2 CARBONIC ANHYDRASE ISOFORMS	12
1.2.1 Overview of Carbonic Anhydrase Isoforms	13
1.2.2 Carbonic Anhydrase Inhibitors	15
1.3 AIMS AND TOPICS OF THE RESEARCH	16

METHODOLOGIES	18
2.1 LABELLED PROTEIN IN MAMMALIAN CELLS.....	19
2.1.1 Approaches to Fluorinated Protein Expression.....	21
2.2 PROTEIN SAMPLE PREPARATION	23
2.2.1 Cell lysis.....	23
2.2.2 Protein purification	25
2.2.3 Protein Crystallization	27
2.3 PROTEIN QUANTIFICATION METHODS.....	29
2.3.1 Polyacrylamide gel electrophoresis (PAGE).....	29
2.3.2 Western Blot.....	30
2.3.3 UV-Visible Spectrophotometry	33
2.4 PROTEIN CHARACTERIZATION TECHNIQUES.....	34
2.4.1 X-Ray Crystallography	35
2.4.2 Mass spectrometry	36
2.4.3 NMR spectroscopy.....	37
2.4.3.a NMR Bioreactor	40
2.5 CELLULAR THERMAL SHIFT ASSAY (CETSA).....	45
RESULTS AND DISCUSSION	47
3.1 Controlling protein fluorination efficiency in human cells and its structural effects	48
3.1.1 Summary	49
3.1.2 Attached publication	51
3.2 Intracellular binding of novel fluorinated compounds to carbonic anhydrase isoforms explored by in-cell ¹⁹ F NMR	83
3.2.1 Summary	84
3.2.2 Attached publication	85

3.2.3 Additional CETSA dose-response study for IC ₅₀ determination.....	103
3.2 Quantitative competition binding of fluorinated ligands by real-time ¹⁹ F in-cell NMR to assess isoform selectivity in human cells.	107
3.3.1 Summary	108
3.3.2 Attached manuscript [in preparation].....	109
CONCLUSION AND PERSPECTIVE	137
References	140

List of figures

SECTION 1

Figure 1.1. Scheme showing the main structural studies enabled by in-cell NMR technique...	3
Figure 1.2. Scheme of competitive binding experiments in an NMR bioreactor.....	10
Figure 1.3. Scheme of the target-observed and ligand-observed approaches.	12
Figure 1.4. Human α -CA2 structure and catalytic zinc coordination.	13
Figure 1.5. Schematic representation of the catalytic mechanism in α -CA.	14
Figure 1.6. Schematic representation of the acetazolamide and methazolamide.....	15
Figure 1.7. Mechanism of CA enzyme inhibition by sulfonamide and thiocyanate.	16

SECTION 2

Figure 2.1. Methods to prepare the cells enriched with isotope-labeled proteins.....	20
Figure 2.2. Chemical structures of the fluorinated aromatic amino acids.....	22
Figure 2.3. Schematic representation of the bead mill technique.	23
Figure 2.6. Schematic representation of the ultrafiltration technique.	26
Figure 2.7. Schematic representation of the main crystallisation methods.....	28
Figure 2.8. Schematic representation of vapor diffusion crystallization method.....	28
Figure 2.9. Schematic representation of the SDS-PAGE gel electrophoresis technique.	30
Figure 2.10. Schematic representation of Western blotting and detection procedure.	31
Figure 2.11. The Trans-blot gel-membrane sandwich for wet transfer.	32
Figure 2.12. Schematic representation of a UV-Vis spectrophotometer.	34
Figure 2.13. Representation of the X-ray diffractometer instrumentation.....	35
Figure 2.14. ESI and MALDI ion sources for mass spectrometry.	37
Figure 2.17. A Gel-Encapsulated Bioreactor System for NMR Studies.	41
Figure 2.18. Experimental setup of the NMR bioreactor using a microdialysis membrane. ...	42
Figure 2.19. Schematic representation of the NMR bioreactor.....	43

SECTION 3

Figure 3.1. Comparison of CA2 dose–response curves after different pretreatment times. ..	103
Figure 3.2. Effect of temperature on CA2 dose-response curves with compounds.	104
Figure 3.3. Dose-response curve of CA2 treated with acetazolamide.	105
Figure M.1. Stability tests of fluorinated compounds in DMEM at 37°C.	112
Figure M.2. Time-resolved in-cell ^{19}F NMR competition binding between compounds 6 and 7 targeting CA1 in the NMR bioreactor.	113
Figure M.3. Fitting of competition binding data corrected for each CA–ligand complex.....	115
Figure M.4. Fitting of competition binding data between compound 6 and MZA.	117
Figure M.5. Experimental setup of the NMR bioreactor.....	121
Figure M-S1. Time-resolved in-cell ^{19}F NMR competition binding experiments.	128
Figure M-S2. Fitting of competition binding data before saturation factor correction.	129
Figure M-S3. Time-resolved in-cell ^{19}F NMR competition binding experiments with compound 6 and MZA.	130

List of tables

SECTION 3

Table 3.1. IC ₅₀ values obtained from CETSA dose–response curves.....	104
Table M.1. Binding affinities relative to compound 6 derived from saturation-corrected real-time in-cell ¹⁹ F NMR data.....	116
Table M.2. Calculated dissociation constant values of compounds for each CA isoform ...	117
Table M-S1. Durations and ligand concentrations of each step of the bioreactor runs reported in Figure M-S1.....	130
Table M-S2. Free ligand concentration ratios between the test ligand and the spy ligand. ...	133
Table M-S3. Longitudinal relaxation times and corresponding saturation factors for each CA–ligand complex.....	134
Table M-S4. Affinity Constants for CAs derived from real-time in-cell ¹⁹ F NMR Data.....	135
Table M-S5. Durations and ligand concentrations of each step of the bioreactor runs reported in Figure M-S3	135

List of abbreviations

Abbreviations

3FY – 3-fluoro-L-tyrosine

4FF – 4-fluoro-L-phenylalanine

5FW – 5-fluoro-L-tryptophan

6FW – 6-fluoro-L-tryptophan

AA – Amino acid

AAZ – Acetazolamide

BEST – Band-selective Excitation Short-Transients

BSA – Bovine serum albumin

CA – Carbonic Anhydrase

CAI – Carbonic anhydrase inhibitors

CCS-D2 – Second domain of the copper chaperone for SOD1

CETSA – Cellular Thermal Shift Assay

CF₃ – Trifluoromethyl group

CPP – Cell-penetrating peptides

CSA – Chemical shift anisotropy

DD – Dipole–Dipole interaction

DJ-1 – Parkinson-related protein deglycase

DMEM – Dulbecco's Modified Eagle Medium

DSC – Differential scanning calorimetry

E. coli – Escherichia coli

ESI – Electrospray Ionisation

FAA – Fluorinated Amino Acid

FBS – Fetal bovine serum

FID – Free induction decay

HBSS – Hanks' Balanced Salt Solution

HEK293T – Human Embryonic Kidney 293T cells

HRP – Horseradish peroxidase

HSP90 α – Human stress-inducible 90 kDa heat shock protein alpha

HSQC – Heteronuclear Single Quantum Coherence

HMQC – Heteronuclear Multiple Quantum Coherence

ITC – Isothermal titration calorimetry

K_d – Dissociation constant

K_i – Inhibitory constant

K_{off} – Dissociation rate constant

m/z – mass-to-charge ratio

MALDI – Matrix Assisted Laser Desorption/Ionisation

MS – Mass Spectrometry

MZA – Methazolamide

NMR – Nuclear Magnetic Resonance

PBS – Phosphate-buffered saline

PEG – Polyethylene glycol

PEI – Polyethylenimine

PTFE – Polytetrafluoroethylene

SDS-PAGE – Sodium Dodecyl Sulphate - PolyAcrylamide Gel Electrophoresis

SLO – Streptolysin O

SOD1 – Superoxide Dismutase 1

SOFAST – Band-Selective Optimized Flip Angle Short Transient

S – Spy ligand

ST – Medium switch time

T – Test ligand

TBS – Tris-buffered saline

TBST – TBS buffer with 0.1% Tween-20

TEMED – N,N,N',N'-tetramethylethylenediamine

TFA – Trifluoroacetic acid

TOF – Time-of-flight analyser

TROSY – Transverse Relaxation-Optimised Spectroscopy

CHAPTER 1

INTRODUCTION

1.1 IN-CELL NMR SPECTROSCOPY

The characterization of biological macromolecules, such as proteins and nucleic acids, and the study of their functional processes require nowadays to be investigated under physiological conditions in order to obtain more biologically relevant information. This approach has become increasingly important particularly in drug discovery process, since promising candidate molecules that perform well in-vitro often fail in later clinical phases due to lack of activity in cells or the occurrence of adverse effects in vivo. In this context, the nuclear magnetic resonance (NMR) spectroscopy is often employed to characterize lead compounds and evaluate their activity against targets in-vitro. Thanks to its non-destructive nature, NMR can also be extended to living cells, enabling the investigation of intermolecular interactions, such as the binding of drug candidates to their targets, in a more physiologically relevant context. In-cell NMR has thus emerged as a useful technique for studying large molecules directly in a cellular environment, evaluating their molecular interactions, post-translational modifications, and interactions with exogenous drugs. The development of NMR bioreactors has recently overcome the limitation of cellular viability, allowing prolonged monitoring and performing competitive binding measurements between different ligands. Consequently, the use of in-cell NMR and bioreactor setups in drug discovery has the potential to enhance the predictive value of preclinical research and facilitate the development of effective new drugs.

1.1.1 Overview of in-cell NMR spectroscopy

NMR spectroscopy is a powerful and versatile technique for studying biological macromolecules at atomic level. It can be used to analyse the three-dimensional structure, the dynamics, and the molecular interactions of biomolecules in solution and in cells.^{1,2} Conventional NMR studies are typically conducted and optimized in-vitro, where the biomolecules under investigation are purified and immersed in aqueous buffers, designed to increase their solubility and reduce spectral background signals. However, the cellular environment and cell crowding, which frequently affect the conformation and dynamics of biomolecules, cannot be accurately replicated in in-vitro experiments.³ As a result, in-cell NMR spectroscopy has emerged as a powerful method for investigating macromolecules directly within the cellular environment.^{1,4,5} Unlike in-vitro NMR, the aim of in-cell NMR is to monitor alterations induced by the cellular environment, such as post-translational modifications, ligand

binding, or protein folding, rather than to determine the structure of novel proteins (**Figure 1.1**).^{4,6} This technique facilitates the analysis of intricate cellular processes, like phosphorylation, acetylation, glycation and oxidation which are challenging to investigate in-vitro in the absence of enzymes and other endogenous molecules.^{1,7}

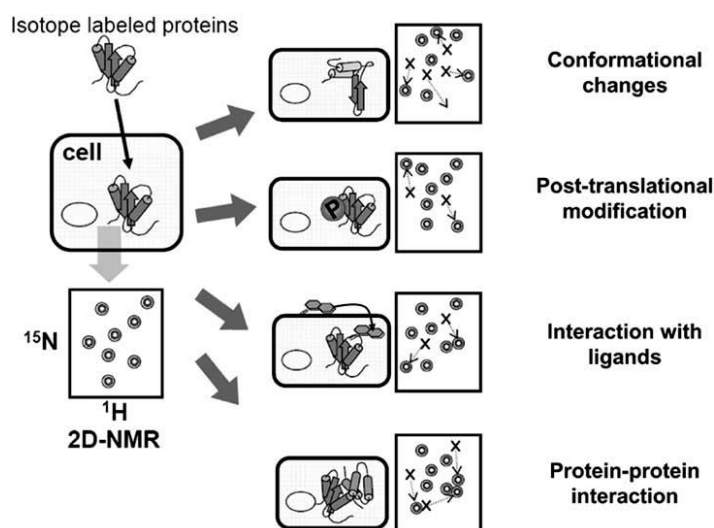


Figure 1.1. Scheme showing the main structural studies conducted on proteins and their interactions using in-cell NMR technique. The isotopically labeled protein is introduced into the cell and analyzed by in-cell NMR, providing information on conformational changes, post-translational modifications, and interactions with other molecules. Reproduced from ref.⁸

The first in-cell NMR study was conducted on *Escherichia coli* bacterial cells, which can efficiently express and label proteins isotopically.⁹ Nevertheless, bacterial systems are not always appropriate for the investigation of eukaryotic proteins, which often require specific folding mechanisms, post-translational modifications, or specific subcellular localization. This limitation has been addressed by developing alternative methods of obtaining labelled proteins inside eukaryotic models, which can either involve physical delivery or endogenous expression methods. Physical delivery methods involve producing proteins in bacterial cells and introducing them into eukaryotic cells via microinjection, electroporation or using cell-penetrating peptides (CPPs), as described in the methodologies section. These approaches ensure high labelling specificity and a low spectral background but present several limitations, such as reduced release efficiency, cell viability issues caused by potentially damaging delivery methods, and protein heterogeneity, whereby proteins can be found in free, complexed or degraded forms. These factors limit the approach's general applicability and may compromise the physiological relevance of the results obtained.¹⁰⁻¹² In contrast, endogenous intracellular

expression models involve expressing the labelled proteins directly within eukaryotic cells using systems such as *Pichia pastoris*, baculovirus-infected insect cells and mammalian expression platforms.^{13–15} This latter strategy allows proteins to be folded, matured, and undergo post-translational modifications in their natural context, facilitating studies of co-translational folding and protein maturation.

Nowadays, in-cell NMR is widely used in structural biology to characterize biomolecules in their physiological context. It allows the exploration of local perturbations induced by the cellular environment, including flexibility and orientation of secondary structure elements, which are involved in modulating protein function and are often undetectable with other structural techniques.^{16,17} In addition to structural biology, in-cell NMR is also a promising tool in drug discovery. The high-resolution screening capabilities of traditional NMR can be combined with the monitoring of cell permeability and interaction with an intracellular target thus providing the ability to select small molecules active in a cellular physiological context.^{2,4,7}

Despite its numerous applications, analysing biomolecules in complex cellular environments such as human cells is often hampered by spectral background and limited cell viability. For this reason, new labelling strategies, such as the use of ¹⁹F, and new experimental setups, such as the bioreactors, have emerged to overcome these drawbacks and obtain more highly resolved spectra, as discussed in the following sections.

1.1.2 Isotopic labelling and NMR strategies for probing proteins in cells

Isotopic labelling of macromolecules is essential to obtain high-quality NMR spectra, especially when studied in a cellular context. The principles behind in-cell NMR experiments are similar to those in-vitro, but specific adaptations are required to address some of the challenges typical of the cellular environment. These include low sensitivity, sample instability over time, significant cellular background and possible non-specific interactions of the target protein with other macromolecules.^{4,9}

One of the most commonly used experiments in in-cell NMR studies is the mono-dimensional (1D) ¹H experiment, as it provides a general overview of cellular metabolites, water and some protein signals. The ¹H nucleus is highly sensitive due to its high gyromagnetic ratio and natural abundance. Experiments based on this nucleus have proven useful for monitoring the incorporation of metal ions or drug interactions by observing the signals of the histidine ring

protons in regions of the spectrum free from cellular background.^{18–20} However, the large presence of protons in metabolites and cellular macromolecules generates very crowded 1D ^1H spectra, making it difficult to resolve the signals of a protein of interest. Consequently, NMR spectroscopy has required the development of isotope filtering and heteronuclear correlation strategies to suppress the cellular background and study specific signals of the molecule of interest.^{21,22}

In this context, labelling with stable isotopes such as ^{13}C and ^{15}N or the introduction of ^{31}P and ^{19}F nuclei play an essential role in the study of macromolecules within cells. 1D NMR spectra of these nuclei enable the rapid monitoring of intracellular metabolites and protein interactions, making them useful for screening and quantification applications.²² Specifically, ^{19}F has been extensively used for studying proteins and ligand binding, as discussed in Section 1.1.3.b, whereas ^{31}P , which serves as a sensitive probe for phosphate-containing compounds, has been employed to monitor nucleic acids and cellular metabolism in living cells.^{23,24} However, 1D spectra of ^{13}C and ^{15}N macromolecules in cells can exhibit low sensitivity and peak overlap, which makes it difficult to assign different signals. Therefore, two-dimensional (2D) NMR experiments that correlate two nuclei can be employed, providing detailed structural and spatial information crucial for understanding biomolecules in cells.⁶

The most common 2D heteronuclear experiment is the ^1H - ^{15}N or ^1H - ^{13}C HSQC (Heteronuclear Single Quantum Coherence), which provides informative, high-sensitivity data. However, for in-cell NMR studies where sample stability and the cellular environment pose significant challenges, more specific two-dimensional experiments are used. These include SOFAST-HMQC (Selective Optimized Flip Angle Short Transient - Heteronuclear Multiple Quantum Correlation), which is designed for rapid data acquisition and prioritises sensitivity, and BEST-TROSY (Band-Selective Excitation Short-Transient - Transverse Relaxation-Optimized Spectroscopy), which is optimised to maximise resolution by selecting a TROSY component of the magnetisation along both dimensions.^{25,26}

An important complementary strategy is selective isotopic labelling, which enables targeted experiments to study particular protein regions or amino acid residues. This strategy is particularly valuable for large macromolecules where uniform enrichment of the entire protein often results in more complex spectra and a stronger cellular background due to the larger number of carbon and nitrogen atoms present in cellular context.^{4,6} The most common approach involves the selective labelling of certain amino acids, such as the ^{13}C labelling of methionine

or alanine methyl groups, or the ^{15}N labelling of backbone amide groups and side-chain nitrogen in cysteine and histidine residues. This enables the generation of high-quality spectra with high sensitivity and a low background level.⁴

1.1.2.a ^1H - ^{15}N In-cell NMR

In-cell NMR requires the use of isotopically labelled molecules to study the macromolecules as previously described. Nitrogen-15 (^{15}N) is one of the most commonly used labels in NMR. This NMR-active isotope has a nuclear spin of $\frac{1}{2}$ and a low natural abundance (0.3% compared to 1% for ^{13}C). Its correlation with other nuclei, such as ^1H , using two-dimensional heteronuclear resonance experiments, such as the ^1H - ^{15}N HSQC, enables the acquisition of well-dispersed spectra in a relatively short timeframe. In these spectra, each correlation peak corresponds to an amide group, and thus to an amino acid residue, providing a fingerprint of the protein. Moreover, ^1H - ^{15}N SOFAST-HMQC and ^1H - ^{15}N BEST-TROSY experiments have the advantage of using band-selective pulses in the proton amide region, which reduces the longitudinal relaxation times and recycle delay compared to the traditional HSQC or HMQC sequences employing hard pulses. This is particularly significant in in-cell NMR spectroscopy, where cellular viability remains a limiting factor.²⁷ Specifically, the in-cell ^1H - ^{15}N NMR experiments on human cells offers significant advantages in terms of physiological relevance, but also poses significant experimental challenges. Compared to simpler systems such as *E. coli*, protein labelling techniques in human cells are more complicated in order to ensure good spectral resolution and an adequate intracellular protein concentration.²⁸ In *E. coli*, the expression of isotopically enriched proteins is easily achieved using minimal culture media containing ^{15}N -enriched ammonium salts as the exclusive source of ^{15}N . In contrast, mammalian cells require complex culture media containing most natural amino acids to support growth and efficient protein expression. Early approaches involved obtaining culture media containing labelled amino acids from an acid hydrolysate of algae or bacteria grown in $^{15}\text{NH}_4\text{Cl}$, which was then supplemented with commercially available labelled amino acids. However, several cost-effective, commercially available solutions have since been developed.^{29,30} Nevertheless, producing labelled proteins directly in mammalian cells can generate spectra with strong cellular backgrounds, and the complexity of the signals makes analysis particularly difficult, especially for large proteins. For this reason, selective ^{15}N labelling strategies have been reported for specific amino acids, such as ^{15}N -cysteine or ^{15}N -histidine.³¹ Unlike uniform ^{15}N labelling, this selective labelling improves spectral resolution and prevents overcrowding

of the spectra.^{32,33} Recent advancements in endogenous expression and exogenous delivery techniques have made the use of ^{15}N -labelling more accessible to study proteins in living human cells, opening new opportunities for the analysis of cellular processes under physiologically conditions.

1.1.3.b ^{19}F in-cell NMR

Despite the versatility of in-cell NMR, the use of traditional NMR-active nuclei such as ^1H , ^{13}C , and ^{15}N is sometimes not suitable for studying large, complex proteins. The abundance of hydrogen, carbon and nitrogen atoms in the cellular environment, along with interactions with other cellular components, can complicate a protein's in-cell NMR spectrum. This can result in signal overlap, severely broadened resonances, and reduced sensitivity, making signal assignment and interpretation difficult. To overcome these limitations, ^{19}F has emerged as a promising solution for in-cell applications.^{22,34}

The ^{19}F nucleus has a natural abundance of 100%, a high gyromagnetic ratio (83% sensitivity to ^1H), and a larger chemical shift dispersion (~400 ppm for organic compounds) than ^1H (~15 ppm), rendering it an advantageous choice for NMR studies. Despite being one of the most abundant elements on Earth, ^{19}F is very rarely found in biological compounds and biomolecules, making it well-suited to in-cell applications with virtually no spectral background. Moreover, additional ^{19}F signals from metabolic transformations are uncommon compared to other NMR-active isotopes, underscoring the lack of interference from the cellular metabolism. A further advantage lies in the small size of the fluorine atom. The presence of two or three fluorine atoms does not significantly affect protein folding or its tertiary structure, thus enabling its study in the native state.^{35–37} The chemical shift anisotropy (CSA) of ^{19}F could be a significant critical aspect, resulting in broadened resonance lines at high fields and within large biomolecules. However, spectral quality can be improved by exploiting the greater local mobility of the ^{19}F probe in specific locations, such as side chains or flexible domains.^{22,34}

Thanks to these characteristics, in recent years several methodologies have been developed that allow the incorporation of fluorine into proteins or small molecule for the study of their characteristics through in-vitro and in-cell NMR (*see Section 2.1.1*).^{38,39} One of the most widely used approaches to producing fluorinated proteins involves the use of fluorinated amino acids (FAAs). In bacterial cultures or mammalian cells, natural amino acids are replaced by FAAs or their precursors, resulting the incorporation of fluorine into proteins. Alternatively, ^{19}F -

containing groups can be directly conjugated to reactive amino acid residues, such as cysteine thiols, in proteins expressed in mammalian cells.^{39–41} These approaches and their practical implementation are discussed in deeper detail in the Methodologies section. Therefore, protein fluorine signals can be detected rapidly and efficiently in 1D ^{19}F spectra, even when in-cell ^1H – ^{15}N HSQC experiments yield few resonances. Indeed, interactions with cellular components and the crowded intracellular environment favours transverse relaxation, resulting in the broadening of spectral lines for certain ^{15}N -labelled proteins, such as HSP90 α (human stress-inducible 90 kDa heat shock protein alpha), wild-type ubiquitin, and thioredoxin, to the extent that they become unobservable.^{10,42–44} Therefore 1D ^{19}F NMR experiments have enabled the study of protein dynamics and conformation, as well as the observation of various cellular phenomena such as protein-drug and protein-protein interactions. In addition to proteins, ^{19}F can also be introduced into ligands and used as probes to monitor protein–ligand interactions. In these cases, the ^{19}F signals of the ligands provide information on binding affinity, specificity, and reversibility, as well as enabling the derivation of structural constraints in ligand–protein interactions.⁴⁰ More recently, 2D experiments such as the ^{13}C – ^{19}F TROSY have significantly improved spectral resolution, expanding the applicability of the method to complex systems. This approach selects a component of the NMR signal for which CSA and DD (Dipole-Dipole interaction) relaxation terms cancel each other out. This allows for reduced signal broadening and enables the study of proteins, RNA and large complexes that would otherwise be inaccessible using conventional NMR techniques.^{34,45}

Consequently, in-cell ^{19}F NMR has emerged as an innovative technique that not only overcomes the limitations of conventional in-cell NMR, but also expands the potential for investigating of structural biology in complex cellular contexts.^{36,42,46}

1.1.3 Real-Time In-cell NMR Using a Flow Bioreactor

Despite its many applications and ongoing development, in-cell NMR spectroscopy suffers from some intrinsic limitations, such as the short cell viability and reduced NMR sensitivity. These characteristics make it difficult to monitor prolonged biological events, as the cells deteriorate progressively due to cellular hypoxia and nutrient deprivation.^{46,47} To overcome these limitations, in-cell NMR bioreactors have been developed (*see Section 2.4.3 a*).^{23,48,49} In these setups, cells within the NMR spectrometer are maintained in a metabolically active state by a continuous flow of culture medium, which provides oxygen and metabolites and removes

waste products. Various flow parameters can be adjusted in some bioreactor configurations, such as pH, medium composition and dissolved gas concentration. Drugs and metal ions can also be introduced into the system during spectral acquisition.⁴⁶

The main application of the NMR bioreactor is the real-time monitoring of cellular events, such as changes in metabolite concentration under stress conditions, or conformational changes in proteins following interaction with cofactors, drugs, or other proteins. This setup has been used to monitor different metabolites concentrations, such as ATP, ADP and NAD(P)H using ^{31}P NMR spectroscopy, to analyse different metabolic pathways by studying the accumulation and degradation of ^{13}C -labelled substrates and to evaluate the cytotoxic effects of pharmacological compounds on cell viability.^{50–52} Recently, Luchinat et al. used bioreactor NMR to monitor real-time drug binding to well-known pharmacological targets such as carbonic anhydrase 2 or Hsp90 α .^{23,43}

Bioreactor NMR has proven very useful for measuring the relative dissociation constants (K_d) between two ligands in a cellular context. When a ligand has a very high affinity ($K_d < 1\ \mu\text{M}$) and completely saturates the target at the beginning of the titration, it becomes impossible to measure its dissociation constant directly. Therefore, an indirect measurement of the affinity by competition binding is necessary, in which a reference ligand, called a spy ligand, and the test ligand compete for the same target (**Figure 1.2**). In this context, bioreactor NMR offers significant advantages in terms of cost and time efficiency. By maintaining the cells metabolically active for extended periods, bioreactor NMR facilitates the acquisition of multiple titration points from a single sample. This reduces the number of experiments required compared to traditional in-cell NMR in a closed tube, where low cell viability caused by long acquisition times results in the need for separate and multiple samples for each titration point.⁵³

Overall, the ability to monitor conformational changes and ligand-target interactions in real time under physiological conditions makes the bioreactor an important advance in in-cell NMR approaches.

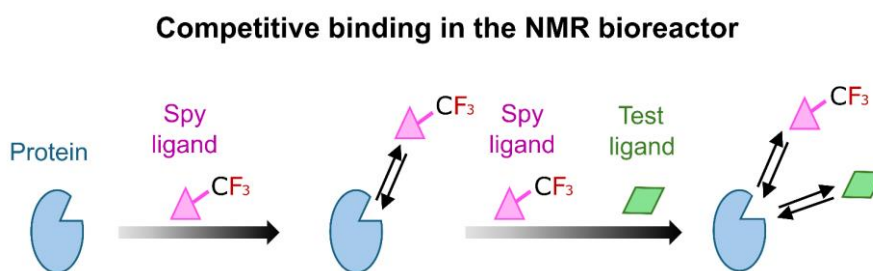


Figure 1.2. Schematic representation of competitive binding experiments in an NMR bioreactor. A reference ligand, referred to as the spy ligand, containing a fluorine tag, binds to the protein target. Subsequently, the introduction of another ligand, the test ligand, competes for the same binding site on the target, establishing an equilibrium of competition between the two ligands. Created based on concepts from ref.⁵³

1.1.4 In-cell NMR for Drug Discovery and Ligand Screening

Drug discovery and development is a time-consuming and resource-intensive process, requiring significant financial investment before a new compound can be approved for clinical trials. This process consists of various phases, from identifying a target and optimising a compound to conducting preclinical and clinical studies. Despite this organised pipeline, only a small fraction of drug candidates actually reach clinical approval.⁵⁴ Many compounds are discontinued in the clinical trials owing to limited efficacy, poor pharmacological properties, or hazardous toxicity, since current preclinical models have limited predictive capacity for human responses to drug candidates.^{55,56} Therefore, the integration of more complex and physiologically relevant models together with in silico prediction tools is crucial in order to bridge this translational gap, enhance predictive value, and minimise reliance on animal models.⁵⁷

In this context, real-time in-cell NMR spectroscopy can be an effective tool for drug discovery and screening, with in-vitro data being integrated with data acquired in a more physiological cellular environment. As discussed previously, the NMR bioreactor has the capability for real-time monitoring of processes occurring within the cell, such as the interactions between intracellular targets and potential drugs. Two different approaches can be used to study these interactions through in-cell NMR, as also used for in-vitro NMR: the target-observed approach, in which the labelled intracellular target is observed, and the ligand-observed approach, in which the labelled ligand is used to track its binding to the unlabelled intracellular target (**Figure 1.3**). Both strategies are well-established in-vitro NMR experiments and have also been

successfully applied to in-cell NMR, which presents additional challenges such as spectral crowding and cellular background noise.^{20,58–60} However, conducting experiments directly within living cells provides the benefit of investigating protein-ligand interactions within a more physiological environment. The cellular environment allows the preservation of the protein native folding state, post-translational modifications, and intracellular localisation, while also providing information on ligand permeability and competition with endogenous molecules. The target-observed approach provides not only information on ligand binding affinity, but also detailed insights into the location of binding sites within the protein. However, it requires selective isotopic labelling of the target, which must yield detectable and assignable NMR signals, which can be particularly challenging in in-cell experiments.^{20,61,62} On the other hand, the ligand-based approach can also provide information on off-target binding and enable the study of interactions with target proteins that remain undetectable with conventional in-cell NMR experiments, such as ^1H – ^{15}N HSQC or SOFAST-HMQC. In ligand-based experiments, linewidths and relaxation properties depend significantly on binding kinetics and the dissociation rate constant (k_{off}). In the fast exchange regime, signals remain sharp and can be used to determine binding affinity, whereas in the intermediate or slow exchange regimes line broadening or separated signals may appear, which provide additional information about the interaction.^{53,63} In in-vitro experiments, ligand labelling is not necessarily required, but it often becomes indispensable in cells due to endogenous ^1H signals in the cellular environment. In these cases, specific labels such as ^{19}F or other sensitive nuclei are used. Interestingly, fluorine atoms are present in about 20% of approved pharmaceutical drugs, which makes in-cell ^{19}F NMR spectroscopy particularly useful for drug screening where many candidate drugs can be directly monitored by in-cell ^{19}F NMR without the need for isotopic enrichment or chemical modification. The selective fluorination of a small therapeutic molecule can enhance its pharmacokinetic and physicochemical properties by increasing its lipophilicity and membrane permeability. At the same time, since fluorine is only slightly larger than hydrogen, several studies have shown that fluorine can reasonably mimic hydrogen and cause only minimal steric perturbations with respect to the compound's binding mode to a receptor.^{64,65}

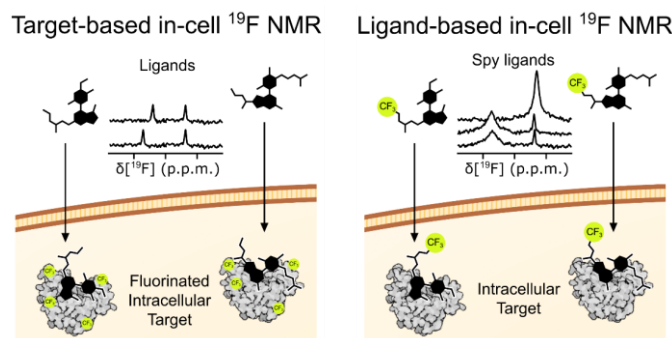
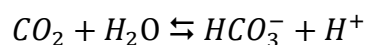


Figure 1.3. Schematic representation of the target-observed and ligand-observed approaches. The target-observed approach (on the left) employs a labelled intracellular target to study protein-ligand interactions, whereas the ligand-observed approach (on the right) uses a labelled ligand to monitor its binding to the unlabelled intracellular target. Adapted from ref.⁴³

As a result, in-cell ^{19}F NMR analysis of ligand-target interactions could be very useful in the later stages of in-vitro screening, especially for characterising the most promising lead compounds. In the early stages of drug development, studying the compound permeability and stability inside the cells, as well as its interaction with the target under normal physiological conditions could help to identify more effective drugs with a higher clinical success rate.

1.2 CARBONIC ANHYDRASE ISOFORMS

Carbonic anhydrase (CA) is a class of metalloenzymes containing a zinc ion in their active site and catalyses the reversible hydration reaction of carbon dioxide (CO_2) to produce a bicarbonate ion and a proton:



These enzymes are found in a variety of organisms, including animals, plants, bacteria and archaea. They are present in many tissues and play a role in various processes, including ion transport, pH regulation, and the production of gastric secretions and cerebrospinal fluid. They have also been linked to certain tumours and other diseases, including glaucoma, epilepsy, and obesity. Therefore, studying and understanding these proteins is crucial in order to develop new drugs and therapies. Sulfonamides are a class of CA inhibitor that is frequently used to treat diseases involving CA isoforms. Nevertheless, developing selective drugs for specific CA isoforms remains challenging due to the large number of these isoforms.

1.2.1 Overview of Carbonic Anhydrase Isoforms

CA enzymes are divided into several classes. Mammals have CA isoforms that belong to the α class, while those in plants, diatoms, eubacteria, and archaea are divided into the β and γ classes.⁶⁶ In humans, there are 16 different CAs, differing in their amino acid sequences outside the catalytic site, enzymatic properties, and organ and tissue distributions. These are separated into cytosolic (CA1, CA2, CA3, CA7, and CA13), membrane-bounded (CA4, CA9, and CA12), and mitochondrial (CA5A and CA5B) isoforms, with the exception of CA6, which is secreted in saliva.⁶⁷ Human CA isoforms exhibit distinct tissue distributions: CA2 is widely expressed in all cells; CA1 and CA3 are mainly found in skeletal muscle and erythrocytes; CA4 is found in the kidney, pancreas, and colon; CA5A and CA5B are found in the liver, kidney, and gastrointestinal mucosa; CA7 is found in the central nervous system; and CA13 is found in the small intestine, and prostate. In addition, CA9 and CA12 are present in gastrointestinal tract, colon, kidney, and some tumours.⁶⁸ The majority of these isoenzymes exhibit a high degree of structural similarity, characterised by highly conserved residues in the catalytic site and three-dimensional structures. Their tertiary structure consists of six α -helices located on the surface and ten β -sheet strands that enclose a 15 Å deep catalytic cavity. The imidazoles of three conserved histidine residues (His94, His96 and His119) coordinate the zinc ion tetrahedrally, along with a water molecule or hydroxide ion (**Figure 1.4**).^{68,69}

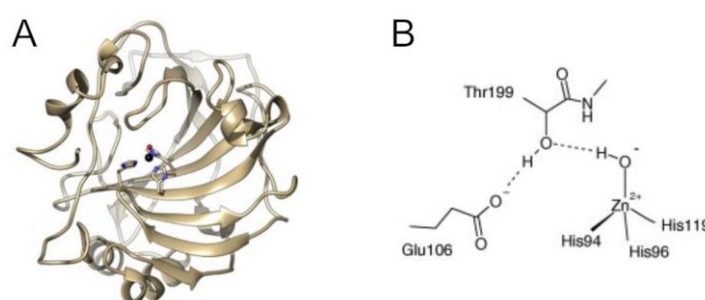


Figure 1.4. Human α -CA2 structure and catalytic zinc coordination. A) Monomeric human α -CA2 (PDB 1F2W); B) Schematic representation of the coordination of the zinc ion in α -CA with the three histidine residues (His94, His96 and His119) and the hydroxide ion forming a hydrogen bond network with Thr199 and Glu106. Adapted from ref.^{66,69}

In the active site, a water molecule bound to zinc is involved in hydrogen bonding interactions with the hydroxyl group of Thr199, which is linked to the carboxyl group of Glu106. To generate the basic and active form of the enzyme, a proton transfer reaction occurs from the

water molecule in the active site to its surroundings. This reaction can be assisted by some residues, such as the proton shuttle His64, or by histidine residues involved in proton transfer pathways as occurs in CA2. The resulting hydroxide ion attacks the CO_2 molecule present in a hydrophobic pocket, forming bicarbonate. This ion is then replaced by a water molecule and released into the solution, resulting in the enzyme's acidic and inactive form, which is able to undergo another catalytic cycle (**Figure 1.5**).⁶⁹

Different carbonic anhydrase isoforms play a role in various processes, including taste and smell perception, adaptation to cellular stress, and protection of the gastrointestinal tract against extreme pH conditions. Some isoenzymes, like CA5, are implicated in molecular signalling processes, including insulin secretion by pancreatic β -cells, while other isoenzymes, like CA9 and CA12, have been connected to tumour development and oncogenesis. Thanks to their diverse physiological functions and involvement in various pathological processes, CA isoforms are significant pharmacological targets, with a substantial number of inhibitors already developed and others under investigation for potential clinical use.

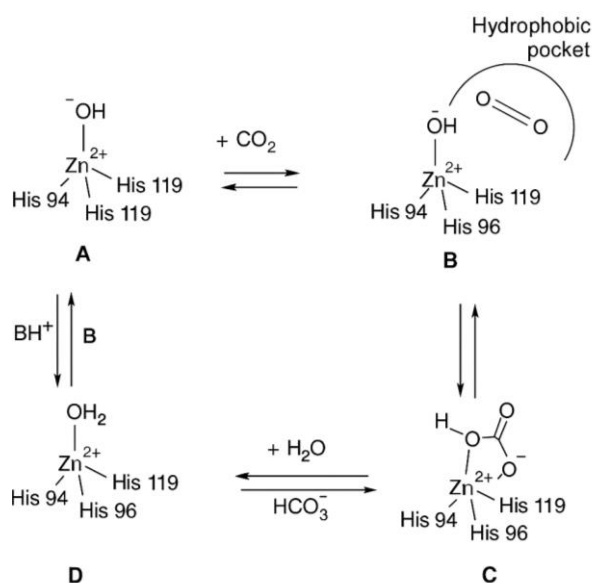


Figure 1.5. Schematic representation of the catalytic mechanism for CO_2 hydration in α -CA. (A) The Zn^{2+} ion in the active site is coordinated with three histidine residues (His94, His96 and His119) and the hydroxide ion. (B) CO_2 is bound in the hydrophobic pocket near the catalytic site, ready to be attacked by the hydroxide ion. (C) The ion bicarbonate is formed. (D) Bicarbonate is released and replaced by a water molecule to regenerate the active state of the enzyme. Reproduced from ref.⁶⁹

1.2.2 Carbonic Anhydrase Inhibitors

Thanks to their many functions and presence in various organs and tissues, CAs are used as indicators of tumour development in many organs or pathologies such as glaucoma, epilepsy, and obesity, and their inhibition allows for the treatment of these diseases.⁷⁰

Carbonic anhydrase inhibitors (CAIs) are divided into two main classes: metal complex anions and unsubstituted sulfonamides. The latter bind to the Zn^{2+} ion present in the enzyme's active site, replacing the water/hydroxide ion in the metal's coordination sphere. The binding of metal-coordinating inorganic anions or more complex species such as carboxylates is less efficient than that of sulfonamides, such as acetazolamide or methazolamide (**Figure 1.6**).⁷¹ The common structure of these compounds is characterised by a sulfonamide group ($-\text{SO}_2\text{NH}_2$) in the binding region and an aromatic or heterocyclic chain (R).

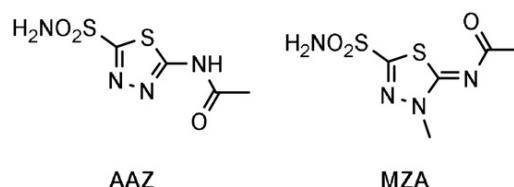


Figure 1.6. Schematic representation of the acetazolamide (AAZ) and methazolamide (MZA) molecules.

The coordination of the nitrogen atom of the sulfonamide moiety to Zn^{2+} occurs in a tetrahedral geometry, with the formation of an extensive network of hydrogen bonds with residues Thr199 and Glu106. At the same time, the R group interacts with the hydrophilic and hydrophobic residues of the cavity (**Figure 1.7**). Since CA isoforms have subtle variations in their active sites, sulfonamide compounds often fail to selectively target a particular isoform, resulting in low efficacy and a higher likelihood of side effects, such as fatigue and nausea.⁷² Consequently, developing more isoform-specific agents remains challenging requiring the development of new inhibitors and the application of cutting-edge techniques for their characterization given the large number of CAs present in the human body.

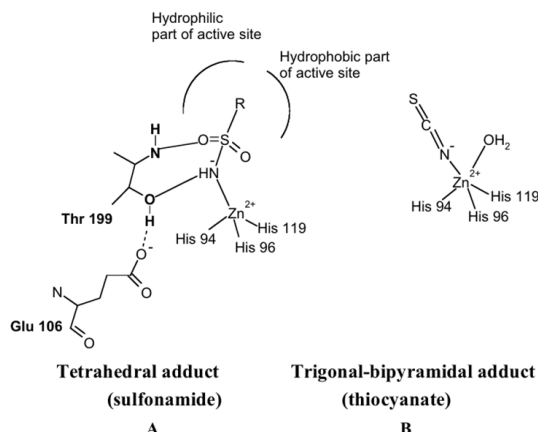


Figure 1.7. Mechanism of CA enzyme inhibition by sulfonamide (A) and thiocyanate (B). The sulfonamide group ($-SO_2NH_2$), which is coordinated with Zn^{2+} , creates a large hydrogen bond network involving residues Thr199 and Glu106, while the organic region of the inhibitor interacts with hydrophilic and hydrophobic residues within the cavity. Reproduced from ref.⁶⁹

1.3 AIMS AND TOPICS OF THE RESEARCH

The aim of this PhD project was to develop and evaluate in-cell ^{19}F NMR methodology for drug screening applications by studying ligand-protein interactions within living human cells. The model protein system was chosen to be CA, a well-known drug target that can be inhibited by approved drugs such as sulfonamides. The research focused on applying in-cell ^{19}F NMR using two complementary approaches: a target-based approach involving fluorinated proteins, and a ligand-based approach involving fluorinated ligands.

The incorporation of fluorine atoms into proteins expressed in human cells was a requirement for target-based methodology. A preliminary study was therefore conducted to evaluate the efficiency with which fluorinated aromatic amino acids (FAAs) were incorporated into proteins expressed in HEK293T cells using the expression method introduced by Pham et al.³⁹ The aim was to assess the feasibility of conducting ligand-interaction studies of these fluorinated proteins, with a particular focus on CA2, the ubiquitous isoform of CAs possessing multiple aromatic amino acid residues in its sequence. The results demonstrated that FAAs were incorporated stochastically and independently of the site. High incorporation efficiency conditions produced spectra with overlapping and split signals due to the formation of CA2 molecules with variable number and distribution of incorporated fluorine atoms. On the other hand, low incorporation efficiency resulted in improved signal resolution and reduced overlap.

However, the latter condition has proven unsuitable for in-cell NMR applications due to poor signal-to-noise ratios in cellular environments. Overall, the in-cell ^{19}F NMR approach, based on a fluorinated target expressed under these conditions, is not well-suited to studying ligand-protein interactions when the target has numerous fluorine incorporation sites, as is the case with CA2. Conversely, proteins with a limited number of fluorine sites may be better suited to this target-based approach, as it generates well-separated and easily interpretable ^{19}F signals.

The second part of this thesis considered the ligand-based methodology. This approach has the advantage of not requiring target labelling and exploiting the relaxation properties of trifluoromethyl ($-\text{CF}_3$) groups introduced into small fluorinated benzenesulfonamide compounds. These groups provide an intense and well-resolved ^{19}F signal due to the magnetic equivalence of the three fluorine atoms. Therefore, in-cell NMR spectroscopy was used to observe the binding interactions between different cytosolic CA isoforms and novel fluorinated compounds. Interestingly, in-cell ^{19}F NMR revealed off-target interactions for some compounds, indicating a lack of binding specificity and consequently their unsuitability for pharmacological testing. Initially these interactions could only provide a qualitative ranking of affinities between the different ligands and CA isoforms. However, combining these results with the uptake kinetics enabled the identification of an optimal spy ligand for competitive binding studies between fluorinated compounds, using bioreactor real-time in-cell ^{19}F NMR spectroscopy. These investigations provided the previously unexplored quantitative information on the dissociation constant ratios between various ligands by tracking the displacement of the spy ligand from CAs and the formation of complexes with the test ligands, ultimately enabling the calculation of their K_d values.

This methodology has enabled the characterisation of novel fluorinated compounds and their binding to different cytosolic CA isoforms within living human cells, as well as the quantitative assessment of relative dissociation constants of the compounds. Overall, in-cell ^{19}F NMR spectroscopy could provide a more physiologically relevant screening platform for lead compounds, bridging the gap between clinical efficacy and in-vitro predictions and offering potential benefits to the pharmaceutical development industry.

CHAPTER 2

METHODOLOGIES

2.1 LABELLED PROTEIN IN MAMMALIAN CELLS

In-cell NMR in eukaryotic cells requires the expression or delivery of isotopically labelled proteins into the cells (**Figure 2.1**). The process of protein delivery begins with the synthesis of proteins within bacterial cells, followed by their introduction into eukaryotic cells. The simplicity of producing labelled proteins in bacterial cells is offered as the principal advantage of this method, which involves the growth of bacteria in an unlabelled medium and their subsequent transfer to an isotopically labelled minimal medium to induce protein expression.⁶² The resultant labelled proteins are then purified and delivered to eukaryotic cells via cell-penetrating peptides (CPPs) or bacterial toxins such as streptolysin O (SLO). CPPs are short amino acid sequences that characteristically consist of lysine and arginine residues, which have the capacity to form a complex with the protein and translocate it into the cytoplasm via endosomal pathways, without causing any impairment to the cell membrane.^{73,74} In contrast, SLO forms pores to temporarily permeabilize the membrane, allowing the labelled proteins to diffuse into the cytoplasm. The process involves calcium-induced membrane resealing, which restores cell viability.⁷⁵ Moreover, electroporation can be an additional method for introducing labelled proteins into mammalian cells, whereby electrical pulses are used to permeabilise the cell membrane and introduce a wide range of proteins into various mammalian cell lines.^{12,76} Finally, volumetric mechanical cell transfection (VECT) has recently emerged as a versatile method for cytosolic delivery of macromolecules through transient mechanical deformations of the cell membrane.⁷⁷

In parallel, the direct intracellular expression of labelled proteins within eukaryotic cells has emerged as a viable alternative to physical delivery methods. Various methods of producing recombinant labelled proteins endogenously have been developed. These include viral infection, in which viruses are used to infect host cells; transient transfection, which involves the internalisation of a DNA molecule and subsequent protein expression; and the integration of expression constructs into the host genome. Although transient transfection produces high protein levels within the first 48 hours and allows co-expression of multiple proteins, it requires the transfection of new cells for each experiment. In contrast, stable transfection with integration of the constructs into the genome provides a constant source of recombinant proteins. However, this method necessitates a lengthy preliminary process to select individual cells that achieve the highest expression levels.^{15,32,78}

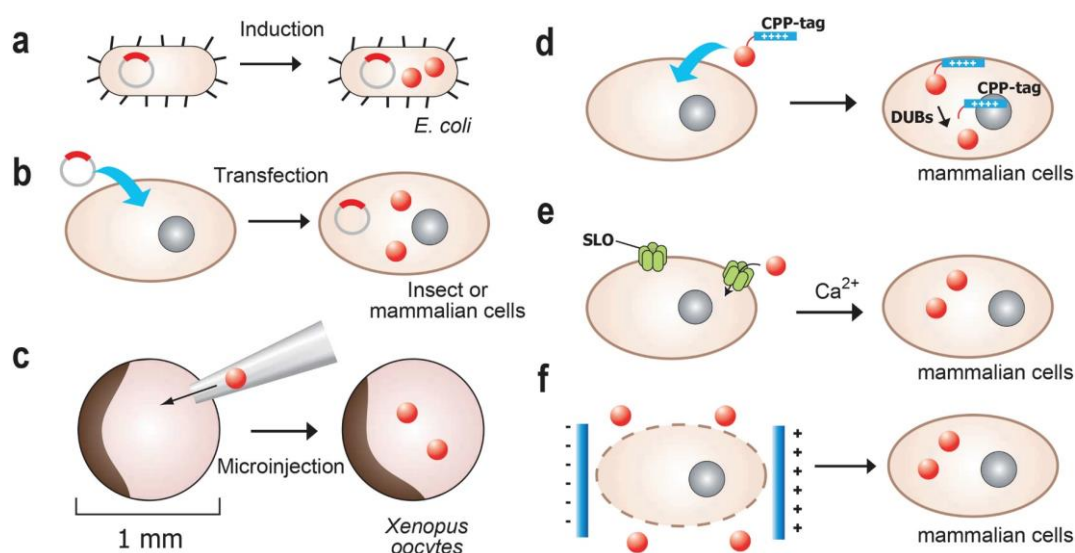


Figure 2.1. Methods to prepare the cells enriched with isotope-labelled proteins. (a–b) Overexpression in bacterial or eukaryotic cells grown in isotope-enriched medium. (c) Microinjection into *X. laevis* oocytes. (d) Penetrant peptide (CPP)-mediated delivery. (e) Insertion through SLO-induced pores. (f) Cell permeabilization by electroporation. Reproduced from ref.⁷⁹

This doctoral research used transient transfection to express proteins in human HEK293T cells, following the protocol developed by Aricescu et al.^{15,32} This cell line exhibits a high growth rate, excellent transfection efficiency and high recombinant protein expression capacity. The cells were cultured in high-glucose Dulbecco's Modified Eagle Medium (DMEM, Gibco), supplemented with 1% penicillin–streptomycin (Gibco) and 10% fetal bovine serum (FBS, Gibco), and maintained in a humidified incubator at 37 °C with 5% CO₂. Plasmids encoding the proteins of interest were obtained from the pHLsec mammalian expression vector after the secretion sequence was removed, as previously described.^{15,20,80} Mammalian cells were transfected with branched polyethylenimine (PEI), a reagent effective at forming complexes with DNA and capable of penetrating the cellular membrane. Transfection was performed using a 1:2 DNA:PEI mixture and for each 75 cm² flask, 25 µg of DNA containing the gene encoding the protein of interest and 50 µg of PEI were used. To produce the unlabelled protein, expression was carried out for 48 hours in commercial DMEM supplemented with 1% penicillin–streptomycin and 2% FBS at 37 °C with 5% CO₂. An analogous protocol was followed to produce the ¹⁵N-labelled protein, using ¹⁵N-labelled DMEM instead of commercial unlabelled DMEM, whereas to produce fluorinated or selectively labelled proteins, a laboratory-reconstituted expression medium (homemade DMEM) was used. In this medium all

components present in the DMEM composition were added, except for certain amino acids, which must be replaced with isotopically labelled ones or fluorinated homologues. After 48 hours of protein expression, the cells were detached from the flask using trypsin, suspended in DMEM containing 10% FBS, and washed with phosphate-buffered saline (PBS). Different protocols were then followed depending on the analysis to be performed. For NMR analysis, the cells were resuspended in 185 μ L of NMR buffer (DMEM, 90 mM glucose, 70 mM HEPES, and 20% (v/v) D₂O), after which the cell suspension is transferred to a 3 mm Shigemi tube following a well-established protocol.³²

2.1.1 Approaches to Fluorinated Protein Expression

In recent years, several techniques have been developed to incorporate fluorine into proteins, including the introduction of fluorinated probes or fluorinated amino acids (FAAs) into the primary structure at specific positions. The first method involves the use of reactive trifluoromethyl derivatives such as 2,2,2-trifluoroethanethiol or fluorinated haloacetamides, which can undergo a chemical reaction post-translation with the nucleophilic side chains of the protein of interest, such as cysteine and lysine.^{81,82} Conversely, the introduction of FAA into proteins remains the most widely used technique, performed through both chemical synthesis and biosynthetic incorporation by bacteria and mammalian cells. Chemical synthesis allows peptides containing up to 50 amino acids to be synthesised and linked together to form proteins using the Dawson-Kent thioester ligation method, which occurs at a cysteine residue. In this process, the C-terminal thioester of one peptide reacts with the thiol group of the N-terminal cysteine of the other peptide to form a stable peptide bond. While this procedure enables the introduction of FAAs at specific positions, it is only suitable for small proteins and is time-consuming and expensive.^{38,82} On the other hand, the biosynthetic incorporation of fluorine into prokaryotic expression systems is a simple and versatile method that allows FAAs to be introduced into proteins of any size. The use of auxotrophic strains allows the expression of fluorinated proteins, since the addition of fluorinated analogue such as fluorophenylalanine, trifluoromethionine, fluoroproline or trifluoroleucine to the culture medium enables their incorporation during translation in place of the natural residues. Although this method theoretically allows all sites of a specific amino acid to be replaced by its fluorinated analogue, full incorporation is never achieved, as the bacteria must first be propagated in the presence of all natural amino acids.^{38,82} An alternative to the use of auxotrophic strains consists in adding glyphosate to the bacterial culture, which inhibits the shikimate pathway and consequently the

synthesis of aromatic amino acids within the bacterial culture. The culture medium is enriched with the natural aromatic amino acids necessary for cell growth, except for the one that must be replaced with its fluorinated analogue or its precursor, such as 5-fluoroindole.^{41,83} Furthermore, our group recently proposed an approach to the biosynthesis of fluorinated proteins in human HEK293T cells.³⁹ In this approach, FAAs substitute natural amino acids in the expression medium 24 hours after transfection and the beginning of protein expression. This method ensures high cell viability and good incorporation of FAAs into the expressed proteins, overcoming the limitations imposed by prokaryotic expression such as the absence of chaperones, transcriptional machinery and co-translational processes.

In this doctoral project, fluorinated proteins were produced following the protocol of Pham et al.³⁹ Various proteins were expressed with four different FAA, such as 6-fluoro-L-tryptophan (6FW, Sigma-Aldrich), 5-fluoro-L-tryptophan (5FW, Fluorochem), 4-fluoro-L-phenylalanine (4FF, Thermo Fisher Scientific), and 3-fluoro-L-tyrosine (3FY, Fluorochem) (**Figure 2.2**).

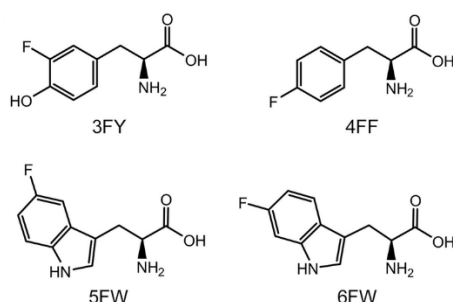


Figure 2.2. Chemical structures of the fluorinated aromatic amino acids (FAAs).

Two procedures were employed to express the fluorinated proteins. In the first procedure, the transfected cells were placed in commercial DMEM. After 8, 16 or 24 hours, the culture medium was replaced with homemade DMEM containing the FAA instead of the corresponding natural amino acid (AA). Protein expression continued for a further period, reaching a total of 48 hours. In the second set of experiments, the cells were placed in homemade DMEM containing a mixture of the corresponding natural and FAA in a ratio of 1/2 FAA and 1/2 AA or 3/4 FAA and 1/4 AA, immediately after transfection, and protein expression was carried out for 48 hours.

2.2 PROTEIN SAMPLE PREPARATION

In order to verify the correct expression and analyse proteins for functional or structural studies, cells or tissues frequently require disintegration and lysis. The protein of interest can then be analysed directly in the extract, or it can be purified to remove other components, such as nucleic acids, carbohydrates, lipids and small molecules. The purified protein can then be studied in solution or crystallised for further structural analysis.

2.2.1 Cell lysis

After cellular expression, soluble proteins are recovered through cell lysis. Gentle lysis methods are generally used for cultured or blood cells, while more vigorous methods are used for more resistant cells, such as bacterial or plant cells. Extraction methods can be classified as either mechanical or non-mechanical.

Mechanical lysis involves the physical destruction of the cell membrane using crushing and grinding techniques. These techniques include the use of high-pressure homogenisers or bead mills, which compress or mix the cells at high speed with small glass or ceramic beads, thereby breaking the cell membrane (**Figure 2.3**).⁸⁴

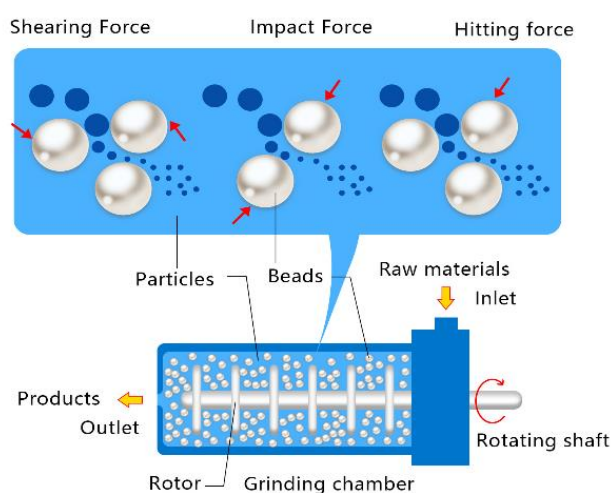


Figure 2.3. Schematic representation of the bead mill technique. Cells are mixed with ceramic beads in a grinding chamber where a rotor spins at high speed. Reproduced from ref.⁸⁵

On the other hand, non-mechanical lysis uses physical methods involving an external force, such as heat, ultrasound or pressure, to disrupt the cell membrane. In thermal lysis, the cell

suspension is repeatedly frozen in liquid nitrogen and thawed at room temperature. Ultrasonic cavitation, meanwhile, uses ultrasonic vibrations to generate sound pressure waves. Finally, osmotic shock lysis exploits intracellular pressure: cells are first placed in a hypertonic solution, causing them to lose their contents to the water. They then swell and burst when transferred to a hypotonic solution (**Figure 2.4**).⁸⁴

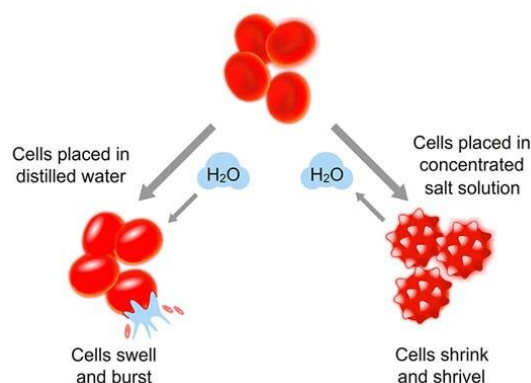


Figure 2.4. Schematic representation of the osmotic shock lysis method. In this method the different osmolarity of the solutions is exploited to break the cell membrane. Reproduced from ref.⁸⁶

In addition to physical method, non-mechanical lysis uses two other methods: the chemical method, which solubilises and disrupts the cell membrane using detergents and alters its pH; and the biological method, which solubilises the peptidoglycan layer present in the bacteria using enzymes such as lysozyme, zymolase, or proteases.⁸⁴ The samples are finally subjected to centrifugation to separate the soluble protein from the other components in the sample, after which the supernatant is preserved.⁸⁴

Disruption of the HEK293T cells and subsequent recovery of the soluble protein were achieved by thermal lysis. In brief, the pellet from a 75 cm² flask was resuspended in 150 µL of PBS. This cell suspension was then flash-frozen in liquid nitrogen and thawed at room temperature. This freeze-thaw process was repeated nine times to disrupt the plasma membrane and extract the intracellular contents. The samples were then centrifuged at 14000 g and 4 °C for approximately one hour. Finally, the supernatant containing the soluble proteins was separated from the pellet and stored in a new eppendorf tube.

2.2.2 Protein purification

Protein purification is a complex process that usually involves several steps. The first step in this process involves using fractionation methods, such as salting out or precipitation with an organic solvent, to remove most of the proteins present in the cell lysate. These methods exploit the differing solubilities of proteins when the salt concentration increases, or organic solvents are present. Without their hydration sphere, the proteins interact with each other via hydrophobic residues and begin to precipitate. Selective denaturation fractionation can also be used to isolate heat-stable proteins by exploiting high temperatures. The protein precipitates are then removed by centrifugation.^{87,88}

The protein of interest can then be isolated using various chromatographic techniques. These techniques are based on the protein's interaction with the stationary and mobile phases. There are various types of chromatography, such as ion-exchange, gel permeation, and affinity. In ion-exchange chromatography, electrostatic interactions occur between charged protein groups and the solid support material, which has an opposite ionic charge to the protein being separated. In gel permeation chromatography, the stationary phase consists of inert particles with small pores that retain smaller molecules, such as salts, while allowing larger molecules to elute (**Figure 2.5**). On the other hand, affinity chromatography is used to purify enzymes, hormones, nucleic acids and specific proteins. The column packing material is bound with a ligand, such as dextran, polyacrylamide or cellulose, which forms a complex with the desired proteins, allowing the others to elute. Then, by changing the ionic strength through pH adjustment or adding a salt solution, the bound protein can leave the column.⁸⁹

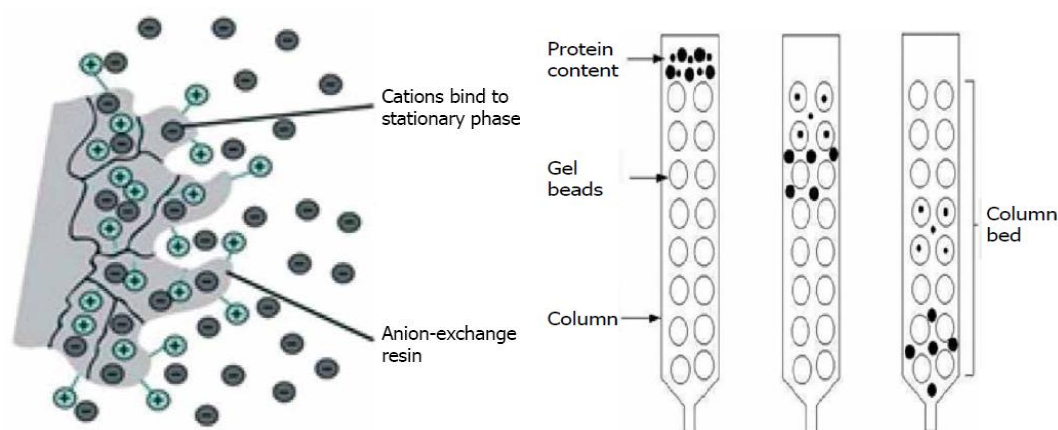


Figure 2.5. Schematic representation of ion exchange chromatography and gel permeation chromatography. Reproduced from ref.⁸⁹

Following purification in a chromatographic column, the protein is typically solubilised in a saline buffer, which is often incompatible with subsequent analytical techniques, such as NMR spectroscopy. Therefore, the final purification step involves changing the buffer and subsequently concentrating the protein through dialysis and ultrafiltration.^{87,88} In dialysis, the protein solution is placed inside a semipermeable membrane and immersed in a low-salt solution. Salts and low-molecular-weight substances diffuse through the membrane, while high-molecular-weight molecules remain inside the dialysis bag. Ultrafiltration, on the other hand, uses semipermeable membranes and a pressure difference. The protein solution is forced through the membrane, separating the salts and low-molecular-weight compounds from the high-molecular-weight molecules (**Figure 2.6**).⁸⁷

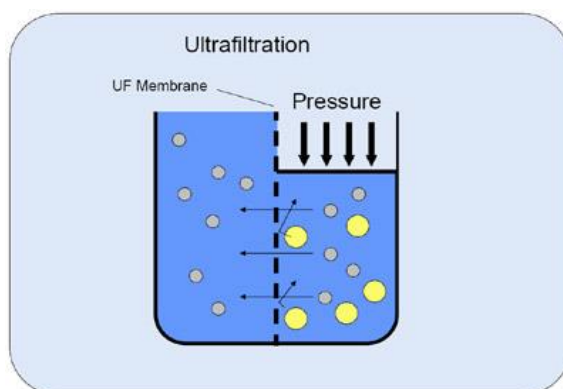


Figure 2.6. Schematic representation of the ultrafiltration technique. Reproduced from ref. ⁹⁰

In this doctoral research project, we purified fluorinated CA2 produced by HEK293T cells. Due to the small number of cells obtained from each culture, lysates from three 75 cm² flasks, prepared using the same procedure, were purified. The fluorinated CA2 was purified by relying on the work of Banerjee et al. (2004),⁹¹ which demonstrated that human CA2 binds to metal-loaded resins in the absence of a histidine tag and can be purified using a ZnSO₄-loaded iminodiacetate-Sepharose affinity chromatography column. Approximately 2.5 mL of binding buffer (20 mM Tris-H₂SO₄ at pH 8) was added to the cell lysate (approximately 500–600 µL), reaching a total volume of 3 mL. This solution was then loaded onto a 5 mL HisTrap nickel-chelating column (GE Healthcare) that had been previously washed with 10 mL of binding buffer. The protein was eluted using an increasing concentration gradient of an imidazole solution, with the different eluted fractions collected from the column. Protein elution was monitored by SDS-PAGE, which demonstrated that the protein eluted at a concentration of 20 mM imidazole in 20 mM Tris- H₂SO₄ at pH 8. The protein was placed in a new buffer solution

(10 mM HEPES at pH 6.8) and concentrated by ultrafiltration using a Centricon with a 10 kDa cut-off membrane (Amicon Ultra-4 with Ultracel-PL PLGC membrane, 10 kDa).

2.2.3 Protein Crystallization

A crystal is a solid structure composed of atoms, molecules, or ions that possesses an ordered crystal lattice, or a geometrically regular arrangement that is repeated in three spatial dimensions. Unlike crystals of low-molecular-weight compounds, macromolecular crystals such as proteins, nucleic acids, viruses and ribosomes are usually smaller and softer, and more easily shattered, since they typically consist of around 50% solvent. The high solvent content, weak lattice forces and structural complexity cause macromolecules in the crystal to vary slightly within the unit cells, resulting in limited diffraction patterns.⁹²

The crystallisation of biological molecules depends on various factors, including chemical and physical conditions such as temperature and pH, as well as the time required for crystallisation and the presence of precipitating agents. These substances can be divided into three broad categories: salts, such as ammonium sulfate; organic solvents, such as ethyl alcohol; and long-chain polymers, such as polyethylene glycol (PEG), and low-molecular-weight polymers, such as methylpentanediol. Furthermore, impurities resulting from the protein purification process could prevent the formation of an ordered crystal lattice.⁹²

The crystallisation of a macromolecule requires the creation of a supersaturated state. This is a non-equilibrium state in which the macromolecule remains dissolved in the solution despite exceeding its solubility limit. Under these conditions, crystals gradually form in the solution and begin to grow until the system regains complete equilibrium. This supersaturated state can be created using various techniques. In microbatch crystallisation experiments, for example, droplets of protein solution are combined with droplets of precipitant, buffer and additives in miniaturised wells under oil. In vapour diffusion crystallisation methods, however, droplets of protein solution and precipitant solution are placed on a pedestal in a well containing the precipitant solution (**Figure 2.7**).⁹³

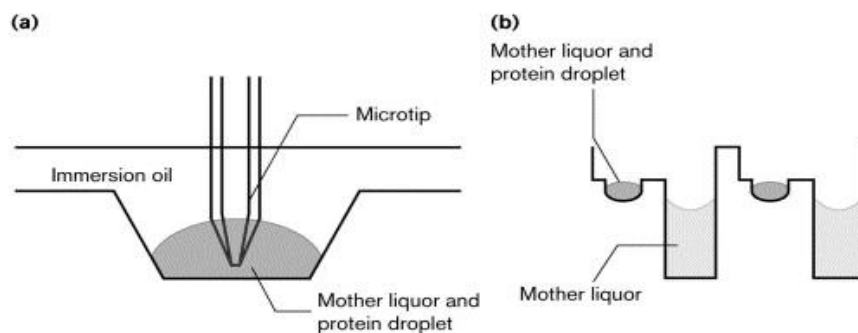


Figure 2.7. Schematic representation of the main crystallisation methods. a) Microbatch method; b) Vapour diffusion method. Reproduced from ref.⁹³

To crystallize native CA2 and fluorinated CA2, the vapor diffusion technique using the sitting drop method was employed (**Figure 2.8**). A screening kit (Hampton Research, solution for crystal growth) was used to determine the ideal crystallisation conditions. Specifically, a crystallisation plate containing 24 pedestals and 24 wells was used, to which 1 ml of crystallisation buffer (reservoir) was added to each well and 2 μ l of protein solution (15 mg/ml in 10 mM HEPES, pH 6.8), mixed with 2 μ l of reservoir, was pipetted onto the corresponding pedestal. The plate was then left at room temperature for a week. The optimal precipitating solution consists of 100 mM HEPES, 1.4 M sodium citrate, and 1 mM 4-(hydroxymercuri)benzoic acid, pH 7.5. This condition was then applied to crystallise all fluorinated CA2s.

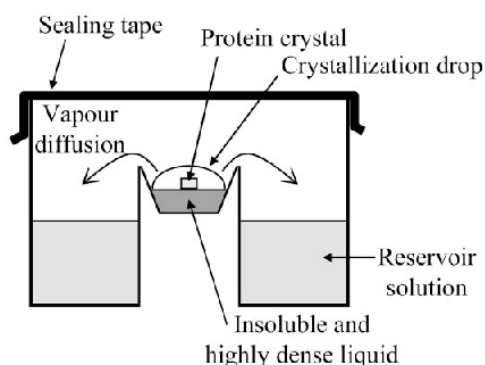


Figure 2.8. Schematic representation of the sitting drop vapor diffusion crystallization method. Reproduced from ref.⁹⁴

2.3 PROTEIN QUANTIFICATION METHODS

Several methods are available to visualise and quantify purified proteins or those in cell extracts, including polyacrylamide gel electrophoresis, Western blot analysis and UV-visible spectrophotometry. The protein of interest can be visualised by polyacrylamide gel electrophoresis even in the presence of numerous proteins in a cell extract, and low-abundance proteins can be detected by Western blot analysis. UV-visible spectrophotometry can be used to quantify proteins in solution by determining the absorbance at 280 nm of the aromatic amino acids, such as tryptophan, tyrosine and phenylalanine, which are present in their structures.

2.3.1 Polyacrylamide gel electrophoresis (PAGE)

Polyacrylamide gel electrophoresis (PAGE) is one of the most widely used techniques for analysing complex protein mixtures. This fast and cost-effective method requires only a few micrograms of the protein to be analysed. Polyacrylamide gels are formed through the polymerisation of acrylamide in the presence of an initiator (ammonium persulfate), a catalyst (N,N,N',N'-tetramethylethylenediamine (TEMED)) and a cross-linking agent (N,N'-methylenebisacrylamide). When an electric field is applied, charged proteins pass through the matrix and their migration speed is proportional to the ratio of their charge to their mass.⁹⁵ However, the complex, three-dimensional structure of proteins can interfere with their migration speed. Therefore, the proteins to be analysed are first denatured and coated with a negative charge using sodium dodecyl sulfate (SDS). The SDS-protein complexes move towards the anode after a voltage is applied, and proteins are separated by the porous gel based exclusively on their molecular weight. Smaller molecules migrate faster than larger molecules (**Figure 2.9**).⁹⁵ Proteins separated on a polyacrylamide gel can be detected using dyes such as Coomassie blue, which can detect up to 0.2–0.6 µg of protein, or silver dyes, which are more commonly used in two-dimensional gels.

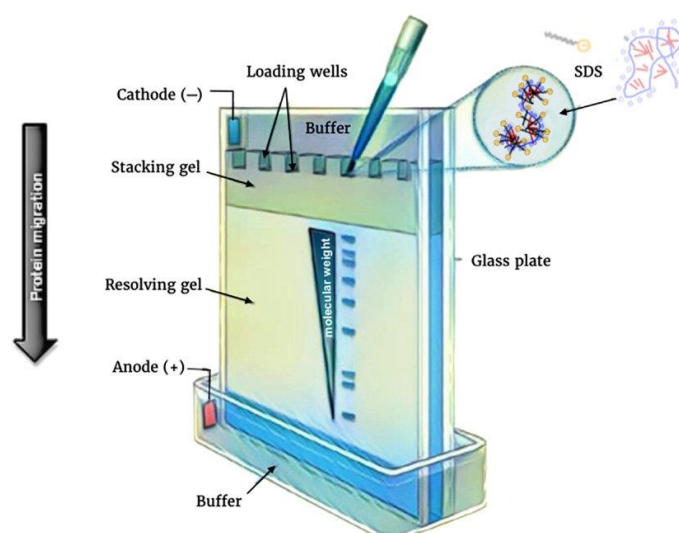


Figure 2.9. Schematic representation of the SDS-PAGE polyacrylamide gel electrophoresis technique. A two-part polyacrylamide gel composed of a stacking (top) and a resolving (bottom) gel is enclosed between two glass plates and submerged in running buffer. After loading the samples into the gel wells, a voltage is applied so that the samples migrate toward the anode. The porous polyacrylamide gel performs a sieving action that separates the proteins based on their size. Reproduced from ref.⁹⁶

In this project, the SDS-PAGE technique was used several times to check for correct protein expression and to verify the presence of proteins of interest in the different samples. In brief, 10 μ L of the lysate sample and 5 μ L of the 3x loading buffer solution (containing 300 mM Tris-HCl at pH 6.8, 6% SDS, 30% glycerol, 0.3% bromophenol blue and 15% dithiothreitol) were added to an eppendorf tube. The samples were then heated to 90 °C for 5 minutes and centrifuged at 14,000 g for 1 minute. Then, 7 μ L of each sample was added to each well of a precast cross-linked polyacrylamide gel (BIO-Rad). The gel was run at 220 V for about 30 minutes using a running buffer (25 mM Tris, 192 mM glycine, 0.1% SDS, pH 8.3). The resulting gel was then either stained with ProBlue Safe Stain containing Coomassie Blue dye or transferred to a Western blot membrane.

2.3.2 Western Blot

The Western blot technique enables the identification and quantification of specific proteins, even at low concentrations. Proteins from an SDS-PAGE gel are transferred to an absorbent membrane to form an exact replica, which is then detected using antibody probes (**Figure 2.10**).^{97,98} WB membranes can be constituted from nitrocellulose or polyvinylidene difluoride

(PVDF). Nitrocellulose is the most commonly used membrane support; however, PVDF membranes have a higher protein binding capacity and greater chemical stability.

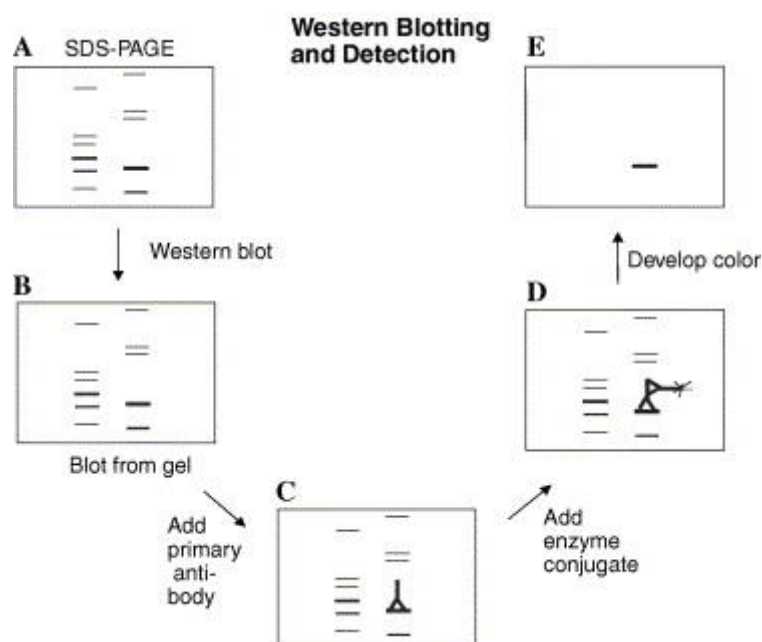


Figure 2.10. Schematic representation of Western blotting and detection procedure. (A) Unstained SDS-PAGE gel before Western blot. (B) Exact replica of the SDS-PAGE gel obtained as a blot after Western transfer. (C) Primary antibody binding to a specific band on the blot. (D) Secondary antibody binding to the primary antibody. (E) Color development. Reproduced from ref.⁹⁷

Protein transfer from an SDS-PAGE gel to a membrane can be achieved using either simple diffusion or electroblotting. In the simple diffusion procedure, the gel is placed under a membrane, a layer of dry filter paper and a glass plate and incubated at 37 °C in a humid container. This system allows several transfers to be obtained from a single gel onto multiple membranes, but the protein transfer is not quantitative.⁹⁹ In contrast, electroblotting uses an electric field perpendicular to the gel's surface to transfer proteins. The gel and membrane are placed inside a stack of blotting paper prior to electroblotting, which can be performed in two ways. The first method is a wet transfer, where the gel-membrane sandwich is immersed in a container with the buffer and platinum wire electrodes (**Figure 2.11**). The second method is a semi-dry transfer, where the gel-membrane sandwich is placed between blotting papers soaked in transfer buffer. In this case, two plate electrodes are used to create a uniform electric field.

Due to the speed and completeness of the transfer process, electroblotting remains the most commonly used procedure for transferring proteins from a gel to a membrane.^{97,100}

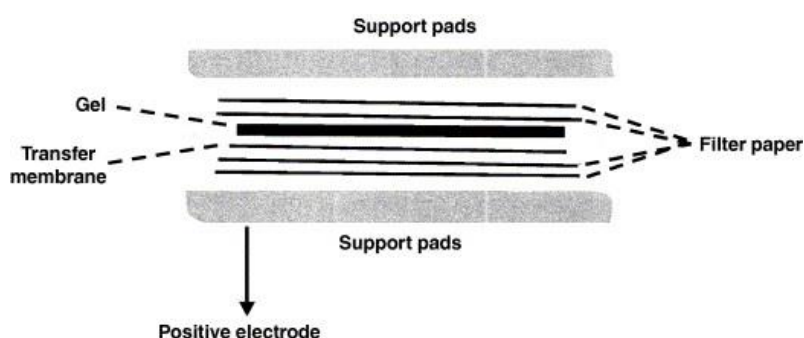


Figure 2.11. The Trans-blot gel-membrane sandwich for wet transfer. The transfer membrane is inserted between the gel, filter paper, and support pads. The cassette containing the assembled gel-membrane sandwich is inserted into a Trans-blot apparatus containing cold transfer buffer, and electroblotting is performed. Reproduced from ref.⁹⁷

The membranes are blocked using either skimmed milk or 5% bovine serum albumin (BSA) to prevent non-specific binding of antibodies to the membrane and reduce background noise. The membranes are then incubated with a primary antibody and washed with PBS or Tris-buffered saline (TBS) to remove unbound antibody.¹⁰¹ The most common method for detecting proteins in Western blots is chemiluminescence. In this approach, the membrane is incubated with a secondary antibody conjugated to horseradish peroxidase (HRP), which catalyses the oxidation of luminol in the presence of hydrogen peroxide, a component of chemiluminescent reagents. The excited luminol then emits light during decay and this signal is recorded using a darkroom acquisition system.^{97,101} However, protein detection in Western blots is generally semi-quantitative, so only a relative comparison of protein levels is possible due to variations in sample loading and transfer, and the non-linear detection of the chemiluminescent signal across the concentration range.

In this project, Western blot analysis was employed to visualise and quantify the endogenous CA2 protein produced by HEK293T cells, which was invisible using standard SDS-PAGE stained with Coomassie blue. Specifically, semi-dry electroblotting was used to transfer the gel onto a membrane. In brief, the polyacrylamide gel is placed on a pad soaked in transfer buffer, with a 0.2 μm nitrocellulose membrane (Bio-Rad) and another pad on an upper layer. The Trans-Blot Turbo Transfer System (Bio-Rad) was used for the transfer at 2.5 A and 25 V for 3 minutes. Correct transfer is then verified by staining the membrane with Ponceau Red. After rinsing the

membrane with TBST buffer (TBS buffer with 0.1% Tween-20), it was blocked for 1 hour with 5% skimmed milk in TBST buffer. The membrane was then washed with TBST three times (each for 15 min) and incubated overnight at 4°C with the primary rabbit anti-CA2 antibody (Proteintech 16961-1-AP; Lot #00082518) diluted 1:2000 in a buffer containing 2.0% (w/v) skimmed milk in TBST. The membrane is then washed with TBST to remove any unbound primary antibody before being incubated for one hour at 25 °C with secondary antibody goat anti-rabbit IgG (whole molecule)-peroxidase (Sigma: A0545) diluted 1:70,000. This step is also followed by washing in TBST. LiteAblot EXTEND chemiluminescent substrates (EuroClone) were used for detection, and images were acquired with the ChemiDoc XRS imaging system (Bio-Rad).

2.3.3 UV-Visible Spectrophotometry

Ultraviolet/visible (UV/Vis) absorption spectrophotometry is a quantitative analytical chemistry technique commonly used for kinetic studies and titrations. It is based on absorption spectra in the ultraviolet/visible region, derived from electronic transitions between a molecule's energy states. Quantitative analyses are performed using monochromatic rays, which consist of radiation of a single frequency. When monochromatic radiation of intensity I_0 crosses an unknown concentration solution, part of the radiation will be absorbed by the sample, while the remaining radiation will be transmitted and detected by the detector. The amount of radiation transmitted depends on the solution's concentration and the thickness of the absorbent layer. Since measurements are made by comparing with a standard, the contribution of radiation reflected by the cell walls and scattered is neglected.^{102,103}

Transmittance (T) is defined as the ratio of the light transmitted to the incident light (I) and can take values between 0 and 1.

$$T = \frac{I}{I_0}$$

The logarithm of the reverse of the transmittance is defined as absorbance (A)

$$A = \log \frac{1}{T}$$

Lambert-Beer's law can be used to calculate the unknown sample concentration:

$$A = \varepsilon \cdot C \cdot d$$

Where A is the absorbance, ϵ is the molar extinction coefficient, C is the sample concentration, and d is the optical pathlength. The absorbance is proportional to both the concentration of the absorbing substance and the thickness of the layer traversed. Therefore, the greater the number of molecules that transit from the ground state to the excited state, the greater the absorbance.

A UV-VIS spectrophotometer consists of a radiation source, a monochromator that selects monochromatic light, a cell in which the sample is housed and a detector (**Figure 2.12**). Instruments that use ultraviolet and visible light have two radiation sources: an incandescent lamp for visible light and a deuterium or hydrogen discharge lamp for UV light. The monochromator is an optical system that disperses polychromatic light into monochromatic bands. It can be a filter that blocks a portion of the light or a prism that separates the various radiation components.

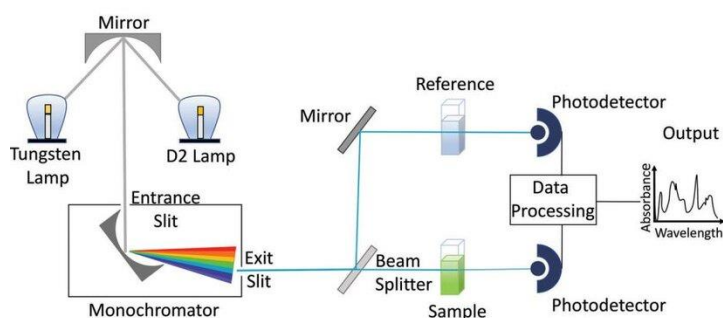


Figure 2.12. Schematic representation of a UV-Vis spectrophotometer. The light from one of the two lamps strikes a mirror, entering the monochromator, which selects a specific wavelength. The light then strikes a splitter, which divides it into two parts, allowing it to reach the two cuvettes containing the sample and the reference. Finally, the transmitted light is detected by the photodetector. Reproduced from ref.¹⁰⁴

2.4 PROTEIN CHARACTERIZATION TECHNIQUES

Many techniques are available for the characterisation of proteins, including X-ray crystallography, mass spectrometry, and NMR spectroscopy. X-ray crystallography requires the protein under study to be crystallised and allows its structure to be determined, but it does not provide information on its dynamics since the molecules are immobilised in specific positions in the crystal. Mass spectrometry and NMR spectroscopy, on the other hand, are widely used to study protein conformation and dynamics. This allows conformational changes and interactions with ligands to be analysed.

2.4.1 X-Ray Crystallography

X-ray crystallography is one of the most widely used techniques for determining the structure of proteins and other biological macromolecules. This methodology involves exposing a crystal to X-rays and deriving a three-dimensional molecular structure from it. X-rays are a form of ionising radiation belonging to the portion of the electromagnetic spectrum between 10 nm and 1 pm and they can be generated in two ways: by accelerating an electron beam in a synchrotron, or by exciting a copper anode. They are then focused and collimated into a parallel beam through a series of adjustable slits with a diameter of 0.1–0.3 mm.¹⁰⁵ One instrument commonly used in X-ray crystallography is the X-ray diffractometer (**Figure 2.13**). This consists of an X-ray source on one side, a platform in the centre where the crystal under analysis is placed and a motorised detector on the opposite side that detects the scattered intensity after the crystal has been exposed to the X-rays.

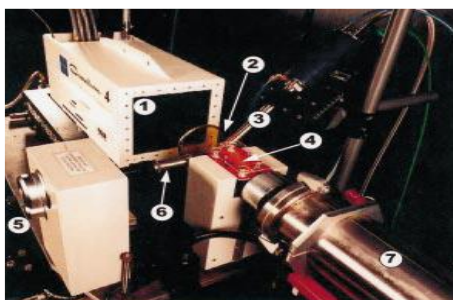


Figure 2.13. Representation of the X-ray diffractometer instrumentation. 1) Detector; 2) Lead sphere mounted on plastic film and supported by a ring to prevent the intense X-ray source from reaching the detector and damaging it; 3) Nozzle for the flow of nitrogen at 100 K; 4) Slits on the panel to adjust the diameter of the incident beam; 5) Device for rotating the crystal; 6) Crystal housing; 7) X-ray beam source. Reproduced from ref.¹⁰⁵

X-ray diffraction produces a diffraction pattern that reflects the arrangement of atoms and molecules in unit cells. The spacing of the spots and the symmetry of the diffraction pattern are determined by the size and shape of the unit cell, respectively. The first data processing step involves determining the crystal structure and the dimensions of the unit cell. Afterwards, each reflection is assigned a Miller index and the intensities of a diffracted spot are measured, which depend on the amplitude of the reflection and the phase angle between the diffracted waves. This data allows to determine the structure factor, from which the arrangement of atoms in the unit cell can be calculated.

In the case of protein crystals, the phase angle cannot normally be calculated directly, so it is determined using indirect methods, such as isomorphous substitution and molecular substitution. The isomorphous substitution method compares the differences between data sets derived from the protein crystal and those derived from the same protein crystal in which heavy atoms have been included. By determining the position of the heavy atoms within the modified crystal, the phase angles and amplitudes of the crystal are calculated and then applied to the protein crystal under examination. The molecular substitution method, on the other hand, is used in cases where a protein structure closely related to the structure under examination is present. The phases and amplitudes are derived from the model protein structure and then applied to the structure under examination.¹⁰⁵

2.4.2 Mass spectrometry

Mass spectrometry is a technique used to identify and quantify molecules based on their mass-to-charge ratio (m/z). It is widely used in proteomics to study the structure, function and dynamics of proteins, and in the analysis of food and soil contaminants. It has also recently been used in drug discovery for metabolite screening.¹⁰⁶

Mass spectrometers consist of an ion source, a mass analyser and an ion detector. Samples loaded into the spectrometer in liquid or gaseous form are vaporised and ionised by an ion source. These sources can be of various types, depending on the sample and the desired degree of ionisation and fragmentation. The electrospray ionisation (ESI) source produces a spray of charged droplets which are desolvated through several pressure steps. This process creates ions with multiple charges and increases the number of masses that can be analysed. In contrast the matrix-assisted laser desorption/ionization (MALDI) source uses an organic matrix, such as sinapinic acid or α -cyano-4-hydroxycinnamic acid, which is photoexcited and ionises the analyte via proton transfer. This process individually charges the analyte ions in the gas phase (**Figure 2.14**).^{106,107}

Once the sample has been ionised, the ion beam is focused and directed towards the mass analysers. These use electric and/or magnetic fields to separate the ions based on their mass-to-charge ratio (m/z). The performance levels of mass analyzers vary depending on the maximum measurable m/z ratio, the fraction of ions reaching the detector, and their resolution, which is the ability to distinguish between two closely spaced m/z values. There are various types of mass analyser, like time-of-flight (TOF) analysers and ion traps.¹⁰⁸ In time-of-flight analysers,

ions are initially accelerated in an electric field and then enter a field-free flight tube, where they are separated according to their mass. The detector measures the time it takes for the individual ions to reach it. In ion traps, the ions are trapped in an electric field, whose frequencies are increased to expel ions with higher m/z values that will reach the detector. The ion trap is highly sensitive, but its resolution is lower than that of the TOF.¹⁰⁶ Upon reaching the detector, the ions strike a metal plate, generating a cascade of electron emissions and a measurable current. Data analysis programs detect and organise the ions based on their individual m/z values and relative abundance. These ions can then be identified using consolidated databases containing libraries of various molecules organised by m/z value.

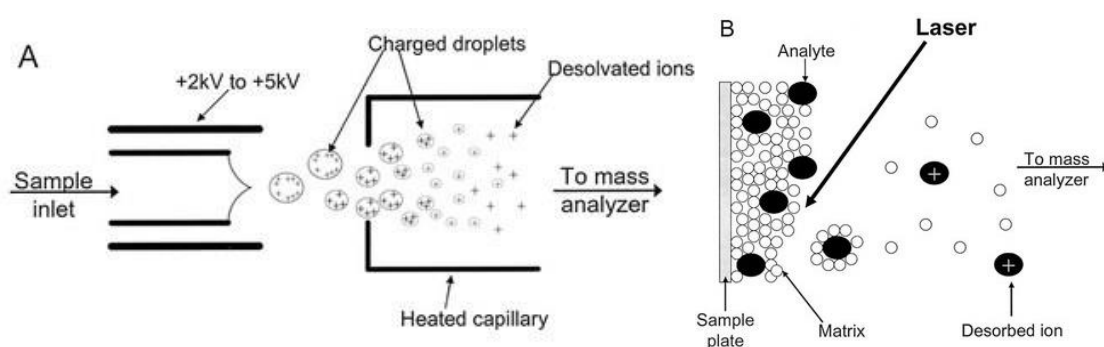


Figure 2.14. ESI and MALDI ion sources for mass spectrometry. A) In ESI source, the liquid sample exits the capillary and a voltage is applied, ionizing the sample and forming small droplets. B) In the MALDI source, the sample molecules co-crystallized with the matrix are pulsed by a laser. Reproduced from ref.¹⁰⁶

2.4.3 NMR spectroscopy

As previously described, NMR spectroscopy is a non-destructive analytical technique that enables us to characterise molecules by observing the behaviour of atomic nuclei in a magnetic field.¹⁰⁹

During this PhD project, NMR technique was used to characterise fluorinated proteins under different experimental conditions, both in-vitro and in cell lysates. It was also used to investigate the molecular interactions between fluorinated ligands and unlabelled targets directly in cell. All experiments were performed at 310 K, which is a physiologically relevant temperature and were carried out using 900 or 950 MHz spectrometers Bruker Avance NEO equipped with 5 mm TCI cryoprobes for ^1H - ^{15}N HSQC experiments or SEL-HP or QCI-F probes for ^{19}F experiments. Following protein expression, the cells were harvested, washed and

resuspended in 185 μL of NMR buffer (DMEM, 70 mM HEPES and 20% D_2O) and subsequently transferred to 3 mm Shigemi tubes. Before NMR acquisition, the cell suspension was gently centrifuged to form a soft pellet. For lysate analysis, the cells were recovered separated from the buffer by centrifugation and lysed by freeze-thawing in 150 μL of PBS (Life Technologies) at the end of in-cell experiments. The soluble lysates obtained by centrifugation were analyzed using ^1H - ^{15}N or ^{19}F NMR spectroscopy in the presence of 10% (v/v) D_2O . Purified proteins and lysates were analyzed in 5 mm and 3 mm NMR tubes, respectively. Cell lysates provide an intermediate system between in-vitro and in-cell NMR. They preserve some intracellular complexity while reducing variables related to cell viability, thus facilitating more accurate optimisation of experimental conditions. Furthermore, lysates have proven to be an effective control for verifying correct protein expression in samples that cannot be detected in cells, as well as allowing the generation of narrower, more resolved signals than those observed in live cells.

One set of experiments included classical two-dimensional ^1H - ^{15}N HSQC spectra, often optimised using sequences such as SOFAST-HMQC and BEST-TROSY to allow data acquisition in shorter times with higher resolution. These approaches have been employed to monitor protein expression and stability in cells and their corresponding lysates. Specifically, 2D ^1H - ^{15}N SOFAST-HMQC spectra were employed to analyse the expression and stability of various ^{15}N -labelled CA isoforms in live cells and cell lysates. This sequence (sfhmqcf3gpqh) is optimised for the rapid acquisition of ^1H - ^{15}N spectra, making it ideal for cellular samples where temporal stability is limited. Shaped pulses (Pc9_4_90.1000 and Reburp.1000) were used for proton-selective excitation and refocusing with a short interscan delay of 0.3 s, and the sequence's -DCALC_SP option was used to automatically calculate the duration and power of these pulses. For each sample, 64 initial scans and 128 increments were acquired, giving a total duration of around 58 minutes per spectrum. To minimise background signals resulting from uniform ^{15}N incorporation into other cellular components, each 2D spectrum obtained from cells or lysates was processed in TopSpin (Bruker) by subtracting the corresponding spectrum obtained from cells transfected with an empty vector. In addition, to analyse the most stable lysates in depth for extended periods of time, the ^1H - ^{15}N BEST-TROSY sequence was adopted. This sequence exploits the TROSY component selection in both dimensions, enabling the acquisition of high-resolution spectra that are particularly well-suited to identifying specific cross-peaks. The 2D BEST-TROSY spectra were used to analyse cellular lysates of fluorinated proteins expressed in the presence of ^{15}N -Histidine, focusing on the histidine spectral region.

The sequence (b_trosyf3gp-ph.2) used selects the TROSY component to maximise resolution and uses Eburp2.1000/Eburp2tr.1000 and Reburp.1000 pulses. For each sample, three independent spectra were acquired with 256 scans and an interscan delay of 0.25 s for a total duration of approximately 9 hours. The data were processed and summed in TopSpin and the cross-peaks were integrated by keeping the integration region constant for each histidine residue. The sets of cross-peaks with the highest intensity and clearest separation were selected. To calculate the error, the background noise was estimated by sampling areas equivalent to the integration regions and calculating the standard deviation.

Alongside the ^1H - ^{15}N -based experiments, ^{19}F NMR analyses have been conducted. As previously discussed, this nucleus offers distinctive advantages: it is completely absent in biological systems and provides a wide chemical shift dispersion. As a result, the spectra are essentially free from cellular background and sufficiently resolved to reveal signals that are otherwise invisible in cells when using traditional isotopic labelling and ^1H - ^{15}N HSQC experiments. This aspect is particularly relevant for detecting proteins such as CA7 and CA13, as reported later in this work. In this context, ^{19}F NMR spectroscopy was employed using two complementary approaches: the protein-observed approach, in which the protein of interest was expressed in fluorinated form; and the ligand-observed approach, in which fluorinated ligands were employed as molecular probes to monitor interactions with intracellular targets. Protein-observed experiments were performed on purified fluorinated proteins, transferred into 5 mm NMR tubes at a total volume of 550 μL containing 10% D_2O . Each sample was analysed using a series of 28-minute acquisitions, comprising 1280 scans with an interscan delay of 1 second, to allow sensitive and selective monitoring of the fluorinated protein under controlled conditions. On the other hand, ligand-observed experiments were conducted directly in cells. After 48 hours of protein expression in a non-isotopically labelled medium, the cells were treated with either 100 μM of one fluorinated ligand or 50 μM of two different ligands for 1 hour at 37 $^\circ\text{C}$ and 5% CO_2 . Following incubation, the cells were harvested and resuspended in 185 μL of NMR buffer (DMEM, 70 mM HEPES and 20% D_2O). They were then transferred to 3 mm Shigemi tubes and gently centrifuged to form a soft pellet. To ensure the physiological relevance of the data, in-cell acquisitions were limited to a maximum of two hours, as previous studies have shown that cell viability remains high within this timeframe. Following NMR acquisition, the pellet was recovered and the cell suspension was centrifuged. Therefore, the supernatant was analysed by ^{19}F NMR spectroscopy and the pellet was lysed in 150 μL of PBS to generate a soluble lysate, which was also subjected to ^{19}F analysis with the addition of 10%

(v/v) D₂O. ¹⁹F nuclear magnetic resonance (NMR) experiments were performed using a simple one-dimensional sequence (zg, Bruker pulse programme), based on a single 90° pulse followed by free induction decay (FID) acquisition. Acquisitions were performed at 310 K on spectrometers tuned for ¹⁹F (564.6 MHz, corresponding to a 600 MHz frequency for ¹H), typically accumulating 1280 scans per spectrum, with a delay of 1 s between scans. The ¹⁹F chemical shift scale was referenced to trichlorofluoromethane (CFCl₃) by setting the trifluoroacetic acid (TFA) signal in an external sample to −76.55 ppm. The data were processed in TopSpin by applying 10 Hz exponential line broadening and phase correction, as well as summing consecutive spectra. Where necessary, a polynomial baseline correction was used to eliminate distortions caused by the polytetrafluoroethylene (PTFE) components within the probe. In protein-observed experiments, the spectral window width was kept constant at 50.3 ppm while the offset was adjusted according to the expected positions of signals from different fluorinated protein residues, at approximately −108 ppm for 5FW, 6FW, and 4FW and at approximately −126 ppm for 3FY. In ligand-observed experiments, however, the offset was set to approximately −54 ppm to centre the signals of the trifluoromethyl CF₃ groups present in the fluorinated ligands.

2.4.3.a NMR Bioreactor

NMR bioreactors are devices designed to resemble fermenters or perfusion systems that allow cells to be kept alive for longer periods of time inside NMR spectrometers. The first bioreactors were designed for growing large quantities of biomass and studying their metabolic fluxes, but they required wide-bore magnets able to fit 10 or 20 mm probes. In these early configurations, yeast and bacterial cell cultures were grown in suspension in an external reactor connected to a pump that flowed the cells into the NMR tube inside the magnet for measurement. Subsequently, in-magnet bioreactors were developed, in which microorganisms could be grown and maintained metabolically active directly inside the NMR magnet for extended periods of time.⁴⁶

Nowadays, new configurations allow NMR bioreactors to be used with the most common 5 mm probes. The first example was reported by Kubo et al.,⁴⁹ where cells were encapsulated in a gel in the form of a coil-shaped string, and DMEM medium was continuously supplied to the bottom of a 5 mm NMR tube through a glass capillary (**Figure 2.17**). This system was used to examine time-dependent changes in the intracellular ATP concentration of HeLa cells by ³¹P NMR spectroscopy.

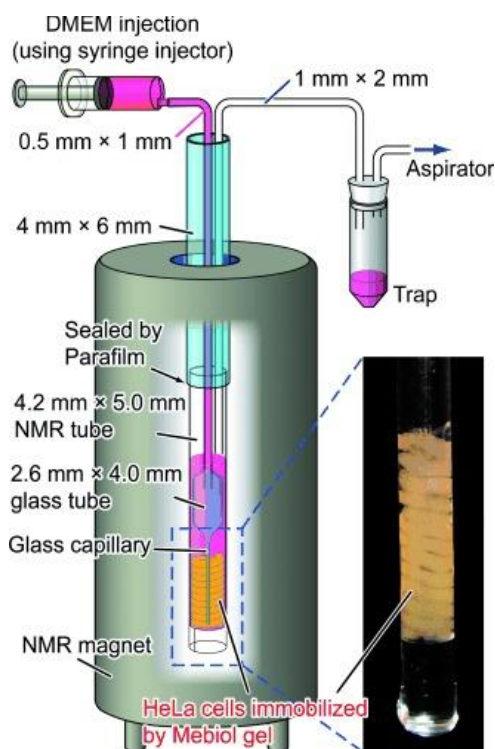


Figure 2.17. A Gel-Encapsulated Bioreactor System for NMR Studies of Protein–Protein Interactions in Living Mammalian Cells. Reproduced from ref.⁴⁹

Variants of this model also rely on continuous-flow systems to immobilise mammalian cells in different gel matrices. The encapsulated cells are confined in an NMR tube and a constant flow of fresh medium is applied to allow the exchange of nutrients, metabolites and gases. The most commonly used gels are Mebiol gel, which is a copolymer that gels when heated, and low-melting agarose gel, which gels when cooled. The cell suspension in the gel is introduced into the NMR tube as a thin, spiral-shaped thread.¹¹⁰ Alternatively, cells can be maintained in suspension and perfused using microdialysis membranes that membranes allow the free exchange of small molecules and nutrients, while also removing waste bioproducts (**Figure 2.18**).¹¹¹

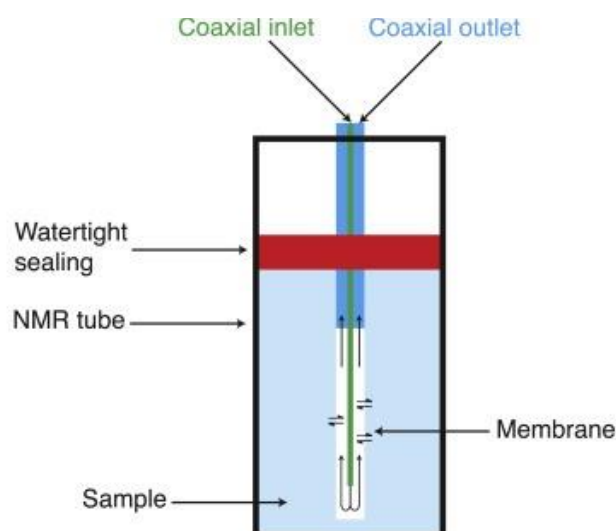


Figure 2.18. Experimental setup of the NMR bioreactor using a microdialysis membrane. Inside the microdialysis membrane, inserted into the NMR tube, are the inlet and outlet tubes. The cell sample is confined within the NMR tube thanks to a watertight seal (red in the diagram). The culture medium solution flows into the membrane through the inlet, exchanges with the solution outside the membrane, and flows out through the outlet (light blue) thanks to the pressure created by the pump. Reproduced from ref.¹¹¹

This PhD project used the same bioreactor model as that described by Luchinat et al.,²³ which enables cell viability to be maintained at above 90% for up to 72 hours. In brief, cells that had been transiently transfected in a 75 cm² flask were heated to 37 °C and resuspended in 400 µL of low-gelling agarose solution (agarose dissolved at 1.5% (w/v) in PBS). The cell-agarose suspension was then aspirated with a syringe into a PEEK chromatography tube with an internal diameter of 0.75 mm and left to cool to room temperature for 2 minutes. The agarose strands were then inserted into a 5 mm NMR tube containing PBS and connected to the holder tube of the bioreactor flow unit. This thermostated transfer line ensures the flow of nutrients and metabolites through a PEEK inlet capillary and PTFE outlet tubing secured to the top of the NMR tube. The inlet capillary of the transfer line is connected to bottles containing fresh DMEM (Sigma-Aldrich, powder, reconstituted in sterile-filtered Milli-Q H₂O and supplemented with 2% FBS, 10 mM NaHCO₃, antibiotics and 2% D₂O, pH 7.4) via a peristaltic pump. The bottles were placed in a 37 °C water bath and sealed with a cap fitted with two hose nozzles: one was connected to the pump via PTFE tubing with an internal diameter of 1.0 mm, and the other was connected to a 0.22 µm PTFE syringe filter for air extraction (**Figure 2.19**).

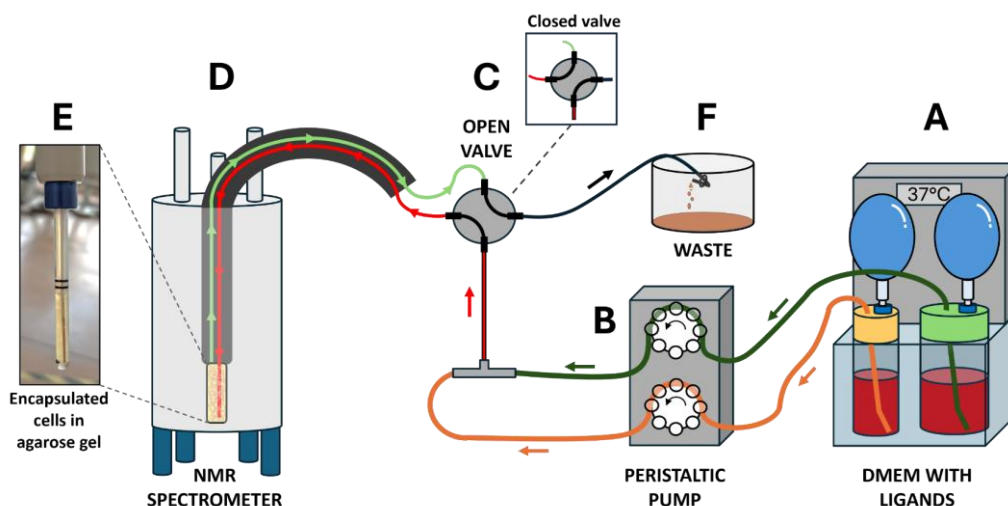
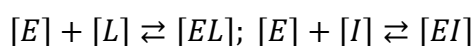


Figure 2.19. Schematic representation of the NMR bioreactor. Bottles containing DMEM and ligands are kept at 37 °C in a temperature-controlled bath (A). The solutions are pumped through a two-channel peristaltic pump (B) to the inlet capillary (red line), passing through a valve (C). This valve can be set to “open” mode to connect the peristaltic pump to the flow unit (in black) or to “closed” mode to convey the liquids directly to the waste tank (F). The inlet capillary leads into the NMR tube containing the agarose-encapsulated spaghetti cells (D, E). The outlet (green line) allows waste products to be eliminated and collected in the waste tank (F).

The bioreactor was used to perform competitive binding measurements between two different fluorinated ligands that target the intracellular CA protein. The spy ligand was added to both bottles of DMEM at a concentration of 10 μM , whereas the test ligand was only added to the second bottle, bringing its final concentration to 20 μM . During the in-cell NMR experiments, the dual-channel peristaltic pump, operating at a flow rate of 0.1 mL/min, enabled the concentration of the test ligand to be varied while keeping the spy ligand concentration constant.

In these competition binding experiments, a pure competition mechanism was assumed, in which:



$$K_{dL} = \frac{[E][L]}{[EL]}; K_{dI} = \frac{[E][I]}{[EI]} \quad (1)$$

where $[E]$ is the concentration of free protein, $[L]$ and $[I]$ are the concentrations of the free spy ligand and the test ligand, respectively, $[EL]$ and $[EI]$ are the concentrations of the protein-spy and protein-test complexes, and K_{dL} and K_{dI} are their corresponding dissociation constants.

For competition binding experiments by in vitro NMR, the fraction of protein bound to the spy ligand (FS) and the displacement of the spy ligand in the presence of a competing molecule depend on the concentration of the protein-spy complex and the total protein concentration $[E_t]$, according to the following equation:

$$F_S = \frac{[EL]}{[E_t]} \quad (2)$$

From equation (1), the ratio between the dissociation constants and, consequently, between the two complexes can be expressed as a function of the dissociation constant ratios:

$$\frac{K_{dL}}{K_{dI}} = \frac{[E][L]}{[EL]} \div \frac{[E][I]}{[EI]} = \frac{[E][L]}{[EL]} \frac{[EI]}{[E][I]} = \frac{[L]}{[I]} \frac{[EI]}{[EL]} \rightarrow \frac{[EI]}{[EL]} = \frac{K_{dL}[I]}{K_{dI}[L]} \quad (3)$$

Taking into account the depletion of free ligand upon binding, this relationship holds as long as the concentration of free protein $[E]$ is negligible — which is valid for strong ligands when their total concentrations are in molar excess relative to the protein. Under these conditions:

$$[E_T] = [E] + [EL] + [EI] \cong [EL] + [EI] \text{ when } L_T + I_T > E_T$$

Therefore, the following equation is obtained:

$$F_S = \frac{[EL]}{[E_t]} \cong \frac{[EL]}{[EL] + [EI]} = \frac{1}{1 + \frac{[EI]}{[EL]}} = \frac{1}{1 + \frac{K_{dL}[I]}{K_{dI}[L]}} \quad (4)$$

Equation (4) is derived from equations (2) and (3) by assuming that $[E]$ is negligible. It can be applied when the concentration of the free ligand is constant and known (for example, in the NMR bioreactor), or when both ligands are in large molar excess relative to the protein (for example, in the culture flask before cell harvesting for in-cell NMR experiments). Under these conditions, the following approximations hold:

$$[L] \cong [L_t] \text{ and } [I] \cong [I_t] \text{ when } L_t \gg E_t$$

Because the intracellular free ligand concentrations cannot be directly quantified in the NMR bioreactor, owing to the low cell filling factor and possible exchange broadening, it is assumed at equilibrium that intracellular and extracellular free ligand concentrations are equal. Both

ligands are continuously supplied in excess by the perfused medium, ensuring saturation of all intracellular binding sites. Therefore, [L] and [I] are taken as the concentrations in the perfusion medium to calculate the ratio of dissociation constants in equation (4).

2.5 CELLULAR THERMAL SHIFT ASSAY (CETSA)

The cellular thermal shift assay (CETSA) is a technique that monitors the interaction between a ligand and its protein target directly within cells. The assay is based on the principle that the ligand induces thermal stabilisation of the target protein by raising its melting temperature (T_m).^{112,113} During a CETSA experiment, cells are pretreated with ligands at 37 °C for 30–60 minutes and then divided into PCR tubes. When samples are exposed to a temperature gradient, unbound proteins denature and precipitate, while those that interact with the ligand remain in solution. Following cell lysis and centrifugation, the soluble fraction is quantified using Western blotting or ELISA. Binding can be confirmed by comparing the melting curves of the treated and control samples. Indeed, when the interaction between the ligand and the native protein persists at high temperatures, target stabilisation is observed with a $\Delta T_m > 0$.^{114,115} Once the temperature at which the unbound protein precipitates has been established, an isothermal dose-response curve (ITDRF_{CETSA}) can be generated, in which protein stabilisation can be observed as a function of ligand concentration.^{113,116} If the protein doesn't go through binding-related conformational changes, the difference between two ITDRF_{CETSA} curves obtained with different compounds shows the different amounts of compound needed to reach the same effective concentration on the target.

Several parameters must be optimised for each CETSA experiment, such as the heating time of the sample, the concentration of the ligand, and the temperature at which the dose-response curves are produced. Most studies adhere to the original protocol, whereby cells are harvested following incubation and washed in a saline buffer solution to eliminate any excess compound prior to the heat pulse. While this procedure reduces potential artefacts from poorly permeable ligands, it can also lead to disruption of cellular equilibrium and loss of some of the bound compound.¹¹⁷ CETSA has proven useful in numerous contexts, including validating drug-target binding, monitoring drug transport and activation, studying off-target effects and drug resistance in tumour cell lines.^{112,118}

However, the CETSA response depends not only on the ligand's affinity for the protein, but also on the thermodynamics and kinetics of binding. For instance when ligand-protein interactions only occur at physiological temperatures, stabilisation may be masked, resulting in a false negative.¹¹⁷ For this reason, CETSA must be combined with other techniques for characterising protein stability, such as differential scanning calorimetry (DSC), ThermoFluor and isothermal titration calorimetry (ITC).^{113,119,120} A fundamental limitation of Western blot-based CETSA is the small number of testable experimental conditions. To overcome this limitation, mass spectrometry (MS)-based CETSA has been recently emerged to allow the analysis of ligand-induced changes in protein-protein complexes and the metabolome.¹¹⁷

CHAPTER 3

RESULTS AND DISCUSSION

3.1 Controlling protein fluorination efficiency in human cells and its structural effects

*Azzurra Costantino*¹ | *Lan B. T. Pham*¹ | *Letizia Barbieri*^{1,2} | *Vito Calderone*^{1,3} | *Gili Ben-Nissan*⁴ | *Michal Sharon*⁴ | *Lucia Banci*^{1,2,3} |
Enrico Luchinat^{1,2,3}

- ¹ Magnetic Resonance Center - CERM, Università degli Studi di Firenze, Sesto Fiorentino, Italy
- ² Consorzio Interuniversitario Risonanze Magnetiche di Metallo Proteine – CIRMMP, Sesto Fiorentino, Italy
- ³ Chemistry Department, Università degli Studi di Firenze, Sesto Fiorentino, Italy
- ⁴ Department of Biomolecular Sciences, Weizmann Institute of Science, Rehovot, Israel

Protein Science. 2024 Mar;33(3):e4910.

3.1.1 Summary

In this study, we investigated the incorporation efficiency of FAAs into proteins expressed directly in human HEK293T cells. Using the expression protocol previously developed by our group,³⁹ we analysed this incorporation using NMR spectrometry, mass spectrometry (MS) and X-ray crystallography. We chose carbonic anhydrase type 2 (CA2), which contains 12 phenylalanine, 8 tyrosine and 7 tryptophan residues, as the model protein for the analysis. We used four different aromatic FAAs: 3-fluoro-tyrosine (3FY), 4-fluoro-phenylalanine (4FF), 5-fluoro-tryptophan (5FW) and 6-fluoro-tryptophan (6FW).

The first experimental condition used for FFA incorporation aimed to maximise their incorporation into the protein, thereby increasing the NMR signal. This condition used a medium switch time (ST), since the FAA addition immediately after transfection resulted in poor cell viability and protein expression. The natural amino acid (AA) was replaced with its corresponding FAA in the culture medium 8, 16 or 24 hours after transfection, producing ST8h, ST16h and ST24h samples. This effect was observed for all FAAs tested except 3FY, which could be added immediately after transfection without compromising cell survival, thereby generating the ST0h condition. The second condition analysed attempted to achieve the incorporation of the FAA on a single position in the protein, thereby reducing potential fluorine-induced structural perturbations due to interactions of fluorine atoms. The protein was then expressed in the presence of a medium containing a mixture of FAA and the corresponding natural amino acid at ratios of 1/2 FAA + 1/2 AA or 3/4 FAA + 1/4 AA. These samples were called Mix50 and Mix75 respectively. To determine FAA incorporation into different targets, two other proteins were also considered: Parkinson-related protein deglycase (DJ-1) and the second domain of the copper chaperone for SOD1 (CCS-D2).

NMR spectroscopy and MS analysis indicated that the ST samples contained multiple proteoforms corresponding to proteins with different levels of fluorination. This resulted in signal splitting and multiple overlaps in the ¹⁹F NMR spectra. By contrast, the Mix50 and Mix75 samples, which had lower incorporation, exhibited signals of higher resolution, albeit at reduced intensity. Notably, the incorporation probability of the ST samples was comparable to that of 3FY-CA2 (ST0h), suggesting that achieving 100% fluorine incorporation is challenging even when FAA is added immediately post-transfection. Furthermore, the results suggest that the distribution of fluorinated proteoforms is binomially distributed, indicating stochastic incorporation independent of residue site. The incorporation probability can be determined

using also NMR spectroscopy by integrating the H107 and H96 signals from the ^1H - ^{15}N BEST-TROSY NMR spectrum of samples expressed in the presence of ^{15}N -His, alongside FAA. These results enabled us to calculate the FAA incorporation probability for each condition and develop a predictive model for the level of FAA incorporation in proteins expressed in HEK293T.

Finally, X-ray structures revealed that the overall folding of fluorinated CA2 is conserved in the presence of fluorinated amino acids. However, there were some discrepancies in the probabilities predicted by the binomial distribution model at sites Y114 for 3FY-CA2 and W6 for 5FW-CA2. This suggests that the incorporation of FAA at specific positions can induce subtle conformational changes, resulting in the selective exclusion of certain protein forms during crystallisation.

3.1.2 Attached publication

Received: 22 December 2023 | Revised: 8 January 2024 | Accepted: 9 January 2024

DOI: 10.1002/pro.4910

METHODS AND APPLICATIONS



Controlling the incorporation of fluorinated amino acids in human cells and its structural impact

Azzurra Costantino¹ | Lan B. T. Pham¹ | Letizia Barbieri^{1,2} |
Vito Calderone^{1,3} | Gili Ben-Nissan⁴ | Michal Sharon⁴ | Lucia Banci^{1,2,3} |
Enrico Luchinat^{1,2,3}

¹CERM – Magnetic Resonance Center,
Università degli Studi di Firenze, Sesto
Fiorentino, Italy

²Consorzio Interuniversitario Risonanze
Magnetiche di Metallo Proteine –
CIRMMP, Sesto Fiorentino, Italy

³Dipartimento di Chimica, Università
degli Studi di Firenze, Sesto Fiorentino,
Italy

⁴Department of Biomolecular Sciences,
Weizmann Institute of Science, Rehovot,
Israel

Correspondence

Lucia Banci and Enrico Luchinat,
CERM – Magnetic Resonance Center,
Università degli Studi di Firenze, Via
Luigi Sacconi 6, Sesto Fiorentino 50019,
Italy.
Email: banci@cerm.unifi.it and
eluchinat@cerm.unifi.it

Funding information

H2020 European Research Council;
Horizon 2020 Framework Programme,
Grant/Award Number: 871037; Ministero
dell'Università e della Ricerca,
Grant/Award Number: 20177XJCHX

Review Editor: Hideo Akutsu

Abstract

Fluorinated aromatic amino acids (FAAs) are promising tools when studying protein structure and dynamics by NMR spectroscopy. The incorporation of FAAs in mammalian expression systems has been introduced only recently. Here, we investigate the effects of FAAs incorporation in proteins expressed in human cells, focusing on the probability of incorporation and its consequences on the ¹⁹F NMR spectra. By combining ¹⁹F NMR, direct MS and x-ray crystallography, we demonstrate that the probability of FAA incorporation is only a function of the FAA concentration in the expression medium and is a pure stochastic phenomenon. In contrast with the MS data, the x-ray structures of carbonic anhydrase II reveal that while the 3D structure is not affected, certain positions lack fluorine, suggesting that crystallization selectively excludes protein molecules featuring subtle conformational modifications. This study offers a predictive model of the FAA incorporation efficiency and provides a framework for controlling protein fluorination in mammalian expression systems.

KEYWORDS

¹⁹F incorporation, direct mass spectrometry, HEK293T, in-cell NMR, mammalian expression, nuclear magnetic resonance spectroscopy, x-ray crystallography

1 | INTRODUCTION

Nuclear magnetic resonance (NMR) spectroscopy is a highly versatile and powerful tool for investigating proteins and providing detailed information about their structures, interactions, and dynamics both in vitro and in cells (Hu et al., 2021; Luchinat et al., 2022; Sekhar & Kay, 2019; Theillet, 2022). Despite its versatility, the use of traditional NMR-active nuclei like ¹H, ¹³C, and ¹⁵N can be particularly challenging for the evaluation of large

and complex proteins due to signal overlap and reduced sensitivity, which limit the quality of data gained from NMR experiments. Furthermore, the presence of numerous hydrogen, carbon, and nitrogen atoms makes spectra more complicated and signals difficult to assign. To address these issues, the incorporation of fluorine-19 (¹⁹F) into proteins and nucleic acids has emerged as a promising solution for the study of these molecules through ¹⁹F NMR spectroscopy (Arntson & Pomerantz, 2016; Chen et al., 2013; Divakaran

et al., 2019; Gronenborn, 2022; Puffer et al., 2009; Sochor et al., 2016; Sudakov et al., 2023). ^{19}F is an attractive alternative to the traditional NMR-active nuclei in protein studies thanks to its high sensitivity (83% of ^1H). Additionally, the ^{19}F NMR spectra exhibits a broad shift range, thus enabling resolved signals for individual fluorinated amino acids within the protein's primary sequence, despite the potential broadness of the fluorine resonances (Gerig, 1994; Kitevski-LeBlanc & Prosser, 2012). Furthermore, in-cell ^{19}F NMR studies of fluorinated proteins and nucleic acids have recently emerged as an appealing area of research, owing to the scarcity of natural fluorine abundance within biological systems, which results in a virtual absence of the fluorine background signal (Jackson et al., 2007; Krafčík et al., 2021; Li et al., 2010; Luchinat & Banci, 2022; Zhu et al., 2022).

In recent years, the development of techniques to incorporate fluorinated amino acids into proteins, either through chemical synthesis or the most widely used biosynthetic incorporation by organisms, has enabled the study of fluorinated proteins in vitro, in bacteria and in mammalian cells. The chemical synthesis method involves the concatenation of small peptides via the thioester ligation method, offering precise control over the introduction of fluorinated amino acids. While this method allows introducing a wide variety of fluorinated synthetic amino acids, is better suited to smaller proteins due to increasing time and costs with the molecular size (Gimenez et al., 2021; Yoder & Kumar, 2002). On the contrary, biosynthetic incorporation in prokaryotic expression systems has been widely employed to introduce fluorinated amino acids in proteins of any size due to its versatility and simplicity (Crowley et al., 2012; Gee et al., 2016; Li et al., 2010; Yoder & Kumar, 2002). We have recently proposed an approach for the biosynthesis of fluorinated proteins in human HEK293T cells, which further expanded the potential applications of NMR to fluorinated proteins (Pham et al., 2023). This approach exploits the distinctive cellular environment of eukaryotic systems, thus overcoming limitations imposed by prokaryotic expression, such as the absence of chaperones, transcriptional machinery, and co-translational processing.

In this work, we show how different expression conditions allow controlling the incorporation efficiency of fluorinated aromatic amino acids into proteins expressed in human cells. We identify two optimal expression conditions: one for maximizing the incorporation of multiple fluorinated amino acids per protein molecule, which results in higher intensity signals in ^{19}F NMR spectra and therefore is beneficial for applications where sensitivity is critical (e.g., in-cell NMR), and the other for limiting the incorporation to a single fluorinated amino acid per molecule, which minimizes structural perturbations that may

arise from the interaction of fluorine atoms with their neighbors within the protein fold. For each condition, the overall incorporation of fluorinated amino acids is studied using NMR spectroscopy and the frequency distribution of each fluorinated protein form is investigated by direct mass spectrometry (MS). Furthermore, we explore the potential structural changes induced by fluorinated amino acids on a model protein, carbonic anhydrase 2 (CA2), through x-ray crystallography. Finally, we provide a predictive model to anticipate the level of incorporation of a specific fluorinated amino acid in proteins expressed in HEK293T cells, based on the amount of fluorinated amino acid added to the expression medium. This predictive model holds the potential to provide an invaluable mammalian fluorinated protein expression protocol for researchers, further advancing the frontiers of protein studies through ^{19}F NMR spectroscopy analysis.

2 | RESULTS AND DISCUSSION

The incorporation of fluorinated aromatic amino acids (FAAs) 3-fluoro-L-tyrosine (3FY), 4-fluoro-L-phenylalanine (4FF), 5-fluoro-L-tryptophan (5FW) and 6-fluoro-L-tryptophan (6FW) (Figure 1a) in a protein expressed in HEK293T cells transiently transfected with the gene of interest is made possible by replacing a specific amino acid with its corresponding FAA in the medium during protein expression (Pham et al., 2023). As protein expression decreases when cells are supplemented with FAAs immediately after transfection, a medium switch time (ST) between 8 and 24 h post-transfection was introduced. Prior to the ST, the cells were maintained in normal medium to allow the internalization of DNA and transcription in non-fluorinated medium. Subsequently, the medium was replaced with FAA-containing medium to allow the incorporation of FAAs into the expressed protein. The total expression time was kept constant at 48 h (Figure 1b) (Pham et al., 2023). Here, the optimal conditions to maximize the incorporation of 3FY, 4FF, 5FW, and 6FW were investigated with STs of 8, 16, and 24 h (see Materials and Methods). Conversely, to optimize the diluted incorporation of FAAs, that is, such that each molecule contains only one fluorine atom, cells were incubated immediately after transfection with an expression medium containing a mixture of FAA and natural amino acid at two different ratios (Mix50 and Mix75 in Figure 1b). Unlike 4FF, 5FW, and 6FW, protein expression levels in cells incubated with 3FY-containing medium immediately after transfection was comparable to that achieved with the medium switch condition (Figure S1), albeit resulting in a reduced total

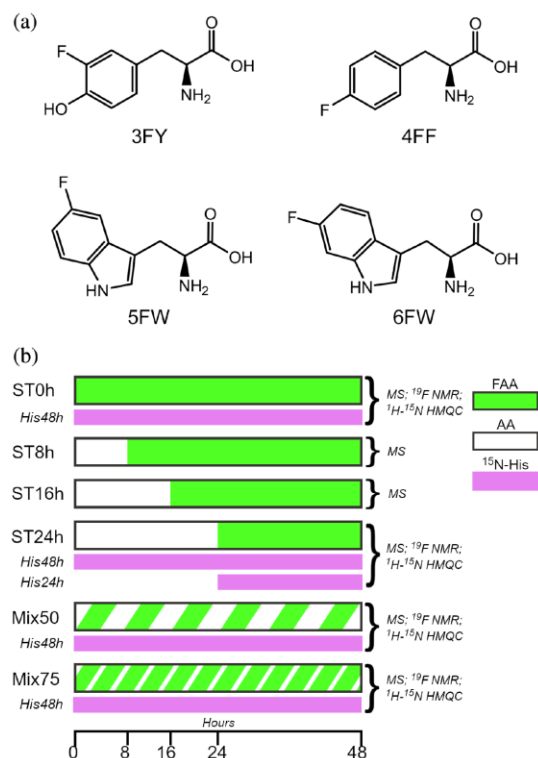


FIGURE 1 Fluorinated amino acids and expression conditions. (a) Chemical structures of the fluorinated aromatic amino acids (FAAs) used in this study; (b) Schematic representation of the different expression conditions for FAA incorporation. ST0h, ST8h, ST16h, ST24h: switch time of 0, 8, 16, 24 h post-transfection, respectively; Mix50, Mix75: mixed media containing 50% FAA + 50% natural amino acid or 75% FAA + 25% natural amino acid, respectively. His48h, His24h: ^{15}N -histidine provided immediately after transfection or 24 h post-transfection, respectively. Each FAA was tested in all conditions except for ST0h, which was only employed for 3FY. Samples for MS analysis were not incubated with ^{15}N -histidine.

number of cells. Therefore, this condition for the incorporation of 3FY was also investigated.

The effect of different expression conditions on the protein ^{19}F NMR spectrum was investigated on samples of pure FAA-CA2. These samples were detected using a room-temperature ^1H probe tuned for ^{19}F detection (see Materials and Methods section). The ^{19}F spectra of FAA-CA2 (Mix50) exhibit well-resolved and distinct peaks, whereas those of FAA-CA2 (ST24h) and, to less extent, FAA-CA2 (Mix75) display several additional signals causing multiple overlaps (Figure 2). In the Mix50 and Mix75 conditions, the overall signal intensity was markedly decreased compared to the ST24h, indicating that the

fluorine incorporation is much lower when the FAA is mixed with the natural amino acid. However, the ST24h spectra suggest that this condition results in higher but still incomplete fluorine incorporation, causing the signal from each fluorine nucleus to split in many components, each arising from a unique combination of fluorine atoms present in other positions within the protein structure. In the Mix50 and Mix75 conditions fewer fluorine atoms are present in each protein molecule, causing less interference and resulting in a more resolved and cleaner spectrum.

The fluorine incorporation in each expression condition was assessed by direct MS directly on cell lysates. The high resolving power of direct MS enabled us to identify in a single spectrum the populations of all coexisting fluorinated states, which differ in the number of incorporated fluorine atoms (n_F) obtained in each expression condition (Figures S2, S3 and Tables S2–S5). MS analysis of FAA-CA2 samples obtained with medium switch conditions revealed two distinct and well-separated populations, namely the non-fluorinated protein ($n_F = 0$) expressed prior the medium switch and the fluorinated proteoforms with different number of incorporated fluorine atoms ($n_F \geq 1$) expressed after the medium switch (Figure S4). Conversely, the distributions obtained from the FAA-CA2 (Mix50) and (Mix75) samples showed a single population comprising proteoforms with lower n_F values (Figure S5). Both the post-switch and the Mix distributions have the characteristic shape of binomial distributions, indicating that the incorporation of FAAs is a purely stochastic phenomenon in all the expression conditions tested, that is, that the probability of incorporation (p_F) is independent of the position along the protein sequence. We therefore derived p_F for all the expression conditions tested (Figure 3 and Table S6).

The p_F values of the Mix samples were much lower than those obtained with the medium switch, consistent with the lower signal intensity observed in the ^{19}F NMR spectra. Furthermore, analysis of the ST samples shows that p_F remains relatively constant regardless of the ST value, indicating that the medium switch time only changes the amount of total fluorinated protein expressed, but not the probability of fluorine incorporation. The p_F of the 3FY-CA2 (ST0h) sample is also comparable to the p_F of the ST8h, ST16h, and ST24h samples, consistent with the ^{19}F NMR spectrum of 3FY-CA2 (ST0h), which is superimposable to that of 3FY-CA2 (-ST24h) (Figure S6). Therefore, introducing the FAA at the beginning of protein expression or after an arbitrarily long incubation time has no effect on the incorporation efficiency in the protein expressed post-switch. Overall, these data indicate that 100% fluorine incorporation cannot be accomplished even when the FAA is administered

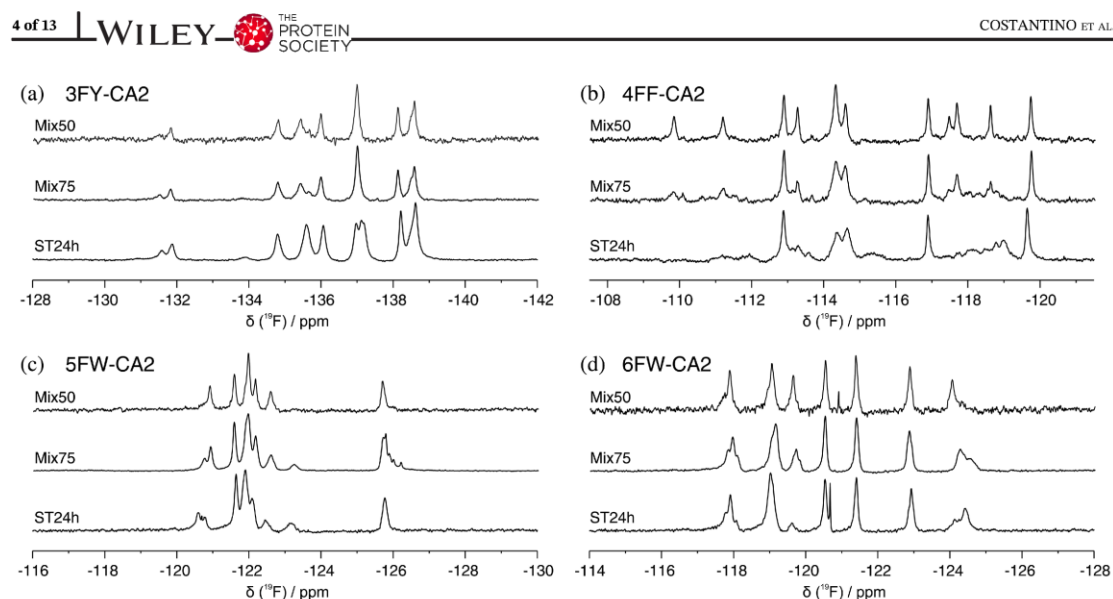


FIGURE 2 ^{19}F NMR spectra of fluorinated CA2. 1D ^{19}F NMR spectra of purified FAA-CA2 incorporating: 3FY (a), 4FF (b), 5FW (c), and 6FW (d) in different expression conditions: Mix50 (top), Mix75 (middle) and ST24h (bottom). The spectra were arbitrarily scaled along the Y-axis for clarity.

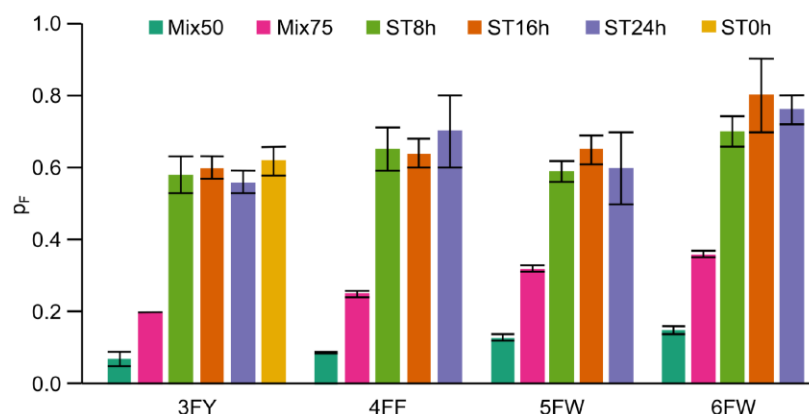


FIGURE 3 Probability of incorporation from MS. Probability of fluorine incorporation (p_F) obtained by MS data for FAA-CA2 in each expression condition.

immediately after transfection, and suggest that a pool of natural amino acid remains inside the cells after the medium switch, retaining the potential to produce proteins containing both non-fluorinated and fluorinated amino acids.

To determine whether the fluorine incorporation could also be assessed by NMR spectroscopy, CA2 was expressed in the presence of both FAA and ^{15}N -labeled histidine. The histidine side-chain ^1H and ^{15}N chemical shifts were used to monitor the presence of nearby fluorine atoms, which induce a change in the chemical surroundings. In the medium switch condition, labeled histidine was introduced to the expression medium either

immediately after transfection or 24 h post-transfection together with the FAA. In the mix condition, labeled histidine was introduced immediately after transfection along with the FAA + non-FAA mixture (Figure 1b). Each expression condition was analyzed by 2D ^1H - ^{15}N NMR directly on the cell lysate. The spectra of FAA-CA2 (ST24h + His48h) showed two distinct signals with different intensities for each histidine (Figure S7) as a consequence of being either in the non-fluorinated protein form or in the fluorinated one. Conversely, in the spectra of FAA-CA2 (ST24h + His24h) the signals arising from the non-fluorinated protein form were much weaker but still clearly detected (Figure 4). In the spectra of FAA-

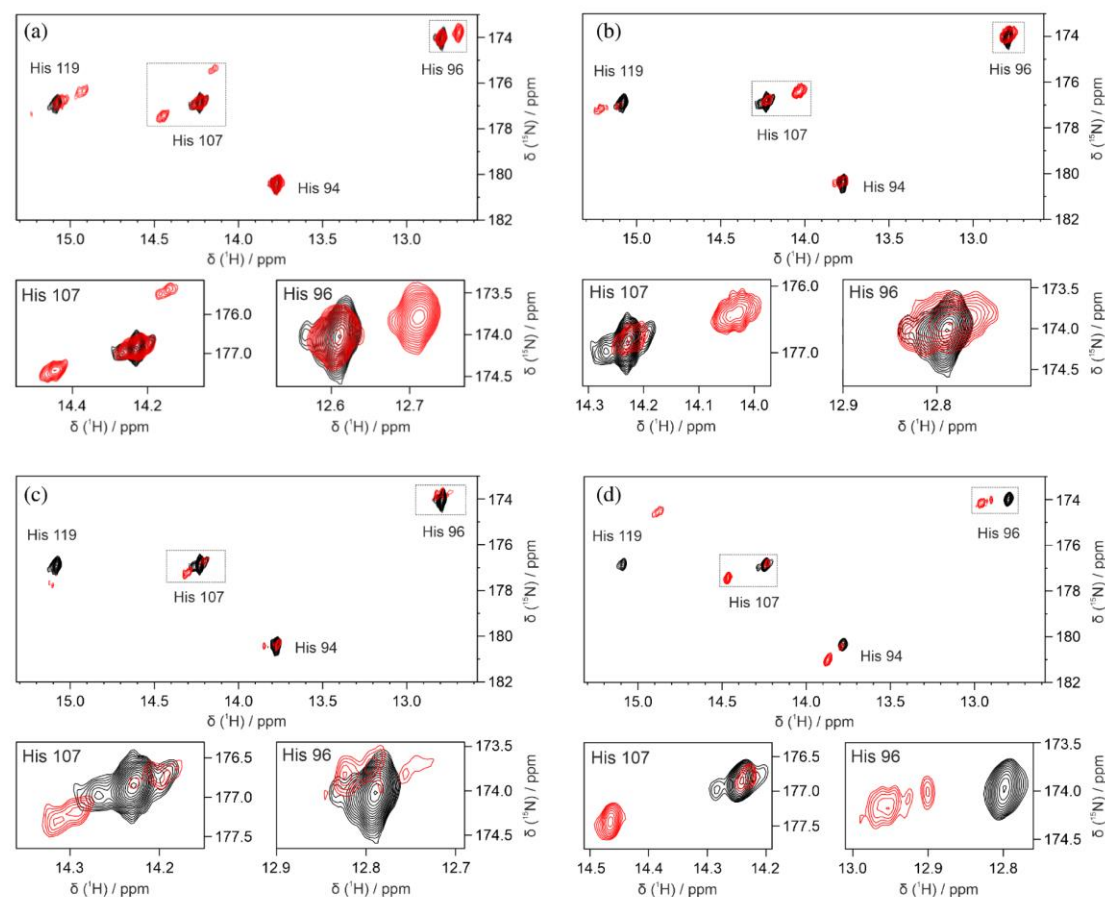


FIGURE 4 ^1H - ^{15}N NMR spectra of fluorinated CA2. 2D ^1H - ^{15}N BEST-TROSY NMR cell lysate spectra of 3FY- (a), 4FF- (b), 5FW- (c), and 6FW-CA2 (d) (ST24h + His24h, red contour lines) overlaid to a reference cell lysate spectrum of non-fluorinated CA2 (black contour lines). The bottom panels show enlarged spectral regions of His 107 and His 96.

CA2 (Mix50) and (Mix75), the signals from the fluorinated protein were barely detected, as expected given very low fluorine incorporation reported by MS (Figures S8 and S9).

Assuming that the position of the cross peak corresponding to the fluorinated protein is primarily influenced by a single fluorine atom, with minimal influence from other fluorine atoms, the extent of fluorine incorporation within each sample was estimated by signal integration. ^{15}N -histidine added immediately after transfection provides the overall fluorine content in the total protein population, whereas labeled histidine added 24 h post-transfection allows focusing on the protein expressed after the medium switch, and thus can be used to estimate the probability of fluorine incorporation in the FAA-medium (Table S7 and Figure 5a). The good

agreement between p_F values derived from NMR on lysates and direct MS suggests that, albeit with higher uncertainty, the incorporation of FAAs can also be estimated by NMR spectroscopy by resorting to ^{15}N -histidine or similar amino acid-type selective labeling schemes (Figure 5b).

In order to examine the probability of FAA incorporation across different proteins, we analyzed the distribution of fluorinated proteoforms of two additional proteins: the Parkinson-related protein deglycase (DJ-1), and the second domain of the copper chaperone for SOD1 (CCS-D2) (Figures S10 and S11). Overall, the p_F values obtained for DJ-1 and CCS-D2 were comparable to those of CA2 expressed under the same conditions, thus demonstrating that the probability of FAA incorporation in our HEK293T expression system does not depend on

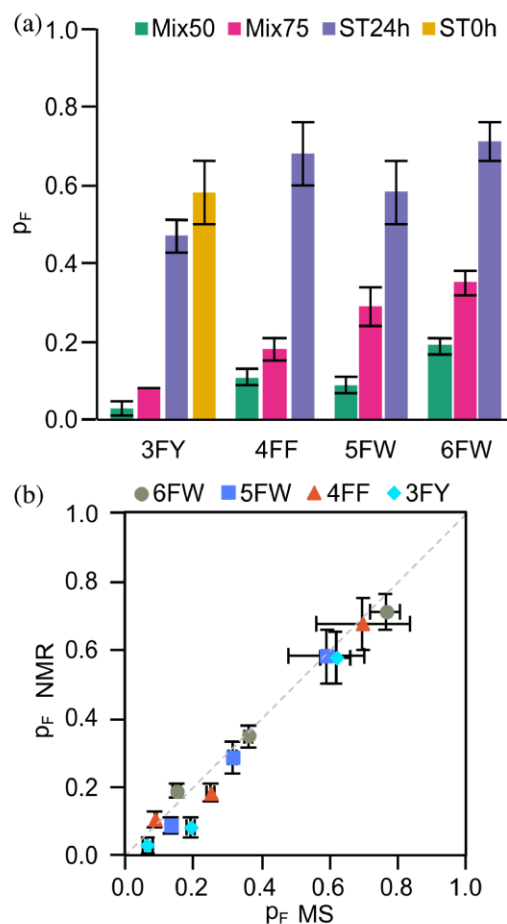


FIGURE 5 Probability of incorporation from NMR. (a) Incorporation of fluorine in each sample estimated by integrating the cross peaks of the 2D ^1H - ^{15}N BEST-TROSY NMR of samples FAA-CA2 (ST24h + His24h). (b) Comparison of p_F values obtained by direct MS and by NMR for FAA-CA2 expressed in each expression condition.

the specific protein being expressed, rather is solely a function of the amount of FAA in the expression medium (Figure 6). Based on this finding, the p_F values obtained by MS were globally fitted with a unified model of FAA incorporation efficiency (Equation 3 in Materials and Methods), which for each FAA depends solely on (i) the total amino acid provided (FAA + AA) in the medium, (ii) a “selectivity coefficient” c which accounts for the fact that the expression machinery recognizes FAAs less efficiently than their corresponding natural AAs, and (iii) the residual endogenous natural AA (AA_{int}), which remains available for protein synthesis after replacing the

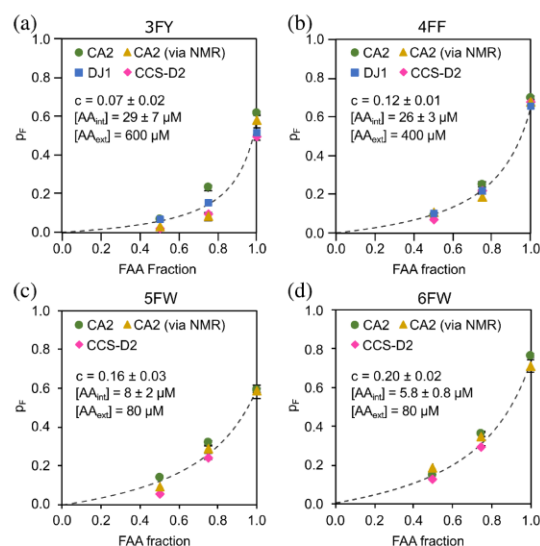


FIGURE 6 Probability of incorporation as a function of fluorinated aromatic amino acid (FAA) fraction. Probability of fluorine incorporation (p_F) in different proteins as a function of the fraction of FAA in the expression medium. The values at FAA fraction = 1 corresponds to the p_F calculated from 4FF(ST24h) (b), 5FW(ST24h) (c), and 6FW(ST24h) (d), while for samples with 3FY (a) the condition 3FY(ST0h) was considered. p_F values for CA2 (MS data, blue) CA2 (NMR data, yellow), DJ1 (blue) and CCS-D2 (magenta) are reported. Since CCS-D2 contains only one tryptophan the protein populations synthesized before and after the medium switch cannot be separated, therefore the probability for 6FW and 5FW fractions = 1 was not calculated. Curve fitting with Equation (3) is shown (gray dashed line). The values of c and $[\text{AA}_{\text{int}}]$ obtained from the fitting are reported together with $[\text{AA}_{\text{ext}}]$ for each FAA.

medium. Using the values of c and AA_{int} obtained for each FAA (Figure 6), equation 3 can be used as a predictive model of FAA incorporation in HEK293T cells.

To evaluate the impact of FAA incorporation on protein structure, we determined the structure of CA2-FAA (ST24h) by x-ray crystallography. The structures of 5FW-CA2 (ST24h), 4FF-CA2 (ST24h) and 3FY-CA2 (ST0h) were determined by x-ray crystallography through molecular replacement, obtaining structures with a resolution of 1.3 Å each (Figure 7 and S12). The structures of FAA-CA2 were perfectly superimposable with the structure of CA2 previously deposited in the PDB and share the same space group and cell parameters (backbone RMSD 0.21 Å) as well as with that of 6FW-CA2 already reported by us (Pham et al., 2023). Fluorine occupancy was then evaluated by letting it free to refine along with the B-factors during structure refinement. In 6FW-CA2, fluorine atoms were present in all seven tryptophan residues

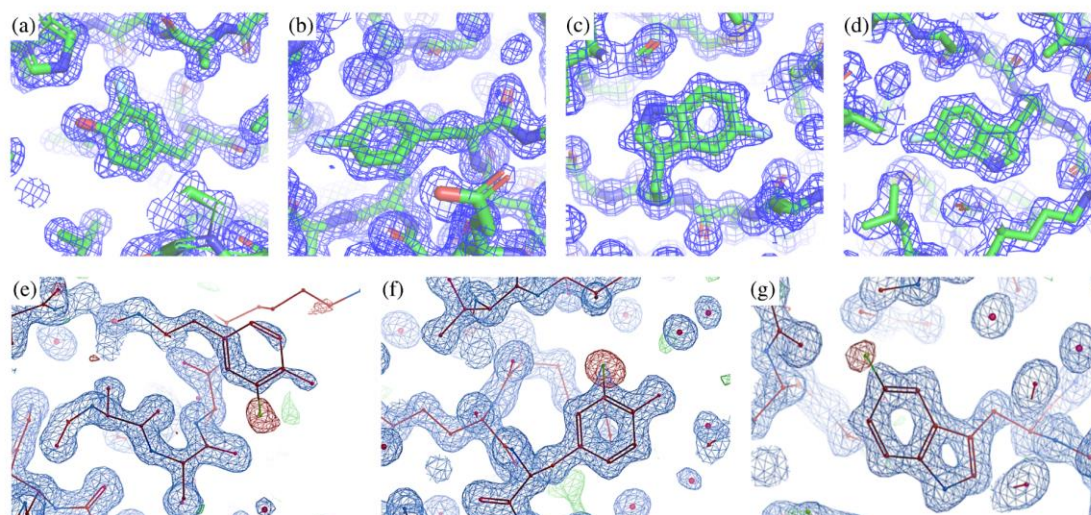


FIGURE 7 Fluorine occupancy from x-ray crystallography. Details of the electronic maps of various fluorinated residues. (a–c) 2mFo-DFc density maps (1σ , blue) of FAA-CA2 structures refined with variable occupancy showing representative fluorinated residues; (a) Tyr 51 in 3FY-CA2; (b) Phe 231 in 4FF-CA2; (c) Trp 97 in 5FW-CA2. (d) Trp 97 in 6FW-CA2. (e–g) 2mFo-DFc (1σ , blue) and mFo-DFc ($\pm 3\sigma$, green/red) density maps of structures refined with fixed occupancy from MS data, showing residues with negative mFo-DFc density on the fluorine atom; (e) Tyr 40 in 3FY-CA2; (f) Tyr 114 in 3FY-CA2; (g) Trp 16 of 5FW-CA2. Note that Tyr 40 is solvent-exposed and therefore has not a well-defined electron density.

with occupancy values ranging from 0.4 to 0.6 (Pham et al., 2023), while in 5FW-CA2 (ST24h) the fluorine occupancy values ranged from 0.12 to 0.47. In 4FF-CA2 (ST24h) fluorine atoms were present in all phenylalanine residues with an occupancy factor ranging from 0.32 to 0.60. In contrast, 3FY-CA2 (ST0h) exhibited greater variability, with position Y114 lacking fluorine atom and residues 3FY128, 3FY191, and 3FY194 presenting two different rotamers due to the ring flip of tyrosine. Notably, the different occupancy values obtained by x-ray crystallography seem at odds with the stochastic fluorine incorporation inferred by MS data.

However, unlike the B-factors, which are rigorously treated during crystallographic structure refinement, occupancies are notoriously unreliable parameters (Bhat, 1989; Deller & Rupp, 2015; Pearce, Bradley, et al., 2017; Pearce, Krojer, & von Delft, 2017; Tronrud, 2004). Therefore, to assess the reliability of the above discrepancies, the FAA-CA2 structures were refined again by fixing the occupancy of the fluorine atoms to the average fluorine content obtained from MS. The weighted difference density (mFo-DFc) maps of the structures refined with fixed occupancy were then used to highlight any discrepancy between the expected (from MS) and the observed occupancy of each fluorine atom. This analysis highlighted three residues with

significant negative fluorine density ($< 3.5\sigma$) in the mFo-DFc map: Tyr 40 and Tyr 114 in 3FY-CA2, and Trp 16 in 5FW-CA2 (Figure 7e–g), whereas the remaining residues of 3FY-CA2 and 5FW-CA2, as well as those of 4FF-CA2 and 6FW-CA2, did not display significant difference in density in the mFo-DFc maps, indicating that for these residues the occupancy values were compatible with the average fluorine content obtained from MS data.

Based on these observations, we hypothesize that while fluorination preserves the overall folding of CA2, the introduction of fluorine atoms at certain positions induces some conformational changes. A selection process then occurs during crystallization, where the crystal lattice imposed by the non-perturbed CA2 leads to the selective exclusion of perturbed CA2 molecules, resulting in crystallographic structures in which fluorine is absent in the positions where it causes conformational changes. For 5FW-CA2, this hypothesis is corroborated by the observation that the fluorine atom of Trp 16 (Figure 7g), if present, would be sterically hindered due to the proximity of the backbone oxygen of Thr 199 ($2.3\text{ \AA}$). For 3FY-CA2, the tyrosine ring flip further complicates the analysis. However, the fluorine atom of Tyr 114 (Figure 7f) would be relatively close to a methyl carbon of Val 109 ($3.0\text{ \AA}$), also suggesting a steric effect.

3 | CONCLUSIONS

In this work, we combined NMR spectroscopy, direct MS, and x-ray crystallography to thoroughly investigate the incorporation of fluorinated aromatic amino acids into proteins expressed in HEK293T cells under different expression conditions that aimed at either maximizing fluorine incorporation or minimizing structural perturbation. We showed that, for all tested proteins, FAA incorporation follows a predictable pattern which is dependent on the FAA concentration in the expression medium, demonstrating the robustness of the method. Fluorine incorporation occurs stochastically, that is, independent on the amino acid position along the protein sequence, and its efficiency is inherently limited by the cellular system, reaching 60%–70% at most, regardless of the medium switch time, as observed in the 3FY-CA2(ST0h) sample. Moreover, the good agreement between p_F values derived from NMR and those from MS show that it is possible to estimate the incorporation of FAAs by NMR spectroscopy by resorting to ^{15}N -histidine and, in principle, other amino acid-type selective labeling schemes. Finally, the lack of fluorine in distinct positions in the x-ray structure of CA2 suggests that fluorine incorporation at those positions induces subtle conformational changes, which result in the selective exclusion of those protein forms during crystallization. Thus, x-ray crystallography combined with MS and/or NMR may unveil hidden consequences of fluorine incorporation on apparently unperturbed protein structures. In summary, this work provides a predictive model of FAAs incorporation in proteins expressed in human cells, and contributes to understand and control protein fluorination and its structural consequences, further advancing the frontiers of protein study through ^{19}F NMR spectroscopy.

4 | MATERIALS AND METHODS

4.1 | Gene constructs

Plasmids encoding carbonic anhydrase 2 (CA2, NP_000058.1), DJ-1 (NP_009193.2) and the second domain of the copper chaperone for SOD (CCS-D2; CCS 84-234, NP_005116.1) were obtained from the pHLsec mammalian expression vector (Aricescu et al., 2006) after removing the secretion sequence, as described in previous works (Barbieri et al., 2018; Luchinat et al., 2017; Luchinat et al., 2020).

4.2 | Protein expression in human cells

The HEK293T (ATCC CRL-3216) cell line was used for transient transfection of the vector, following a protocol described previously (Barbieri et al., 2016). Cells were cultured in T75 flask in high-glucose Dulbecco's Modified Eagle Medium (DMEM, Gibco), supplemented with 1% penicillin–streptomycin (Gibco) and 10% fetal bovine serum (Gibco) and were maintained in a humidified incubator at 37 °C with 5% CO_2 . The expression of fluorinated protein was achieved by using a custom-formulated DMEM, prepared in-house following the reported composition of high-glucose DMEM, in which a desired amino acid was replaced with its fluorinated homolog(s). The following fluorinated amino acids were employed (the concentration in DMEM is reported for each): 6-fluoro-L-tryptophan (6FW, 80 μM , Sigma-Aldrich), 5-fluoro-L-tryptophan (5FW, 80 μM , Fluorochem), 4-fluoro-L-phenylalanine (4FF, 400 μM , Thermo Fisher Scientific), 3-fluoro-L-tyrosine (3FY, 600 μM , Fluorochem). The in-house DMEM was supplemented with 100 $\mu\text{g}/\text{mL}$ of penicillin–streptomycin (Life Technologies) and 2% (v/v) of fetal bovine serum (FBS, Life Technologies), as previously described (Pham et al., 2023). The medium was additionally supplemented with 10 μM ZnSO_4 for the expression of CA2 and CCS-D2. The cells were transiently transfected with vector using polyethylenimine (PEI, Sigma-Aldrich) at a ratio of 1:2 (25 μg DNA:50 μg PEI) (Barbieri et al., 2016). Subsequently, two sets of experimental conditions were employed to investigate the incorporation of fluorinated amino acids on protein expression. In the first set of experiments, cells were initially incubated in in-house DMEM containing non-fluorinated amino acids. After an interval of time ranging from 8 to 24 h, the medium was replaced with in-house DMEM, in which one nonfluorinated amino acid was replaced with the corresponding fluorinated amino acid. The cells were then incubated in the fluorinated medium for an additional period of time, reaching a total expression time of 48 h. For the 2D ^1H - ^{15}N NMR experiments, ^{13}C , ^{15}N -Histidine (^{13}C , ^{15}N -His) was supplemented to the in-house DMEM and two different conditions were employed, in which the labeled histidine was added to the medium either during the medium switch or immediately after transfection for the entire duration of protein expression. In the second set of experiments, cells were immediately incubated after transfection in in-house DMEM containing a mixture of corresponding non-fluorinated and fluorinated amino acid in different ratios (see Results). Cells were incubated in this medium for a total of 48 h. For the 2D ^1H - ^{15}N experiments, ^{13}C , ^{15}N -His

was added to the medium at the same time as the mixture of amino acids, and it was present for the entire duration of protein expression. An additional expression condition was conducted in which cells were incubated in-house DMEM containing 3FY instead of natural tyrosine immediately after transfection for a period of 48 h.

4.3 | Preparation for cell lysate NMR spectroscopy

Proteins were expressed in HEK293T cells for a period of 48 h. Following expression, cells were harvested, and the pellet was lysed in 150 μ L PBS using the freeze–thaw lysis method. Subsequently the soluble lysate was transferred into a 3 mm NMR tube and then analyzed by NMR spectroscopy in presence of 10% of D₂O.

4.4 | CA2 purification

Fluorinated CA2 was expressed from HEK293T and purified by affinity chromatography, by adapting a previously reported protocol (Banerjee et al., 2004). Briefly, transfected cells from three identical T75 flasks were lysed in 150 μ L each by the freeze–thaw method, and the resulting lysates were pooled together and diluted with binding buffer (20 mM Tris, pH 8) to a final volume of 3 mL. The protein was purified using a Ni-NTA column; the elution was performed in steps, each consisting of 3 mL of binding buffer solution containing an increasing concentration of imidazole. CA2 was eluted at 20 mM of imidazole. The protein was buffer-exchanged into 10 mM HEPES at pH 6.8 and concentrated to a final volume of 500 μ L.

4.5 | ¹⁹F NMR spectroscopy on purified protein

Each purified protein was transferred into a 5 mm NMR tube with 10% of D₂O for a total volume of 550 μ L for ¹⁹F NMR analysis. ¹⁹F NMR experiments were recorded at 310 K on a 14.1 T (600 MHz ¹H) Bruker Avance III spectrometer equipped with a room-temperature SEL-HP probe tuned at 564.6 MHz for ¹⁹F detection. The zg Bruker pulse program was used consisting of a single 90 pulse followed by FID acquisition. Each sample was analyzed with a series of 28-min spectra with 1280 scans each, and an interscan delay of 1 s. TopSpin 4.2.0 (Bruker) was employed to process the spectra. The 1D spectra were processed using 10 Hz exponential line broadening and phase correction, and were summed together, followed by polynomial baseline correction to

remove strong baseline distortion arising from polytetrafluoroethylene (PTFE) components inside the probe. The ¹⁹F chemical shift scale was referenced to trichlorofluoromethane (CFCl₃) by setting the signal of trifluoroacetic acid (TFA) in an external reference sample to 76.55 ppm.

4.6 | ¹H-¹⁵N cell lysate NMR spectroscopy

NMR spectra of cell lysates containing CA2 labeled with ¹³C,¹⁵N-His were collected at 310 K on a 900 MHz Bruker Avance HD spectrometer or a 950 MHz Bruker Avance III spectrometer, both equipped with a 5-mm TCI Cryo-Probe. Each sample was analyzed with three ¹H-¹⁵N BEST-TROSY spectra with 256 scan and an interscan delay of 0.25 s each, resulting in a total acquisition time of 9 h. The 2D BEST-TROSY spectra were processed and summed together in Topspin 4.2.0.

4.7 | Integration of histidine cross peaks in 2D BEST-TROSY spectra

The integration of cross peaks in the 2D spectra was performed using TopSpin 4.2.0 software (Bruker). The integration region for both cross peaks of each histidine was kept consistent. The sets of cross peaks with higher intensity and better separation were chosen for integration. As a result, for samples 4FF-CA2, 5FW-CA2, and 6FW-CA2, the cross peaks arising from histidine 107 were integrated, whereas for sample CA2-3FF, the cross peaks arising from histidine 96 were integrated. For error calculation, the background noise of the spectra was sampled with the integration region the same size as above, and the standard deviation of the noise was computed using Excel (Microsoft).

4.8 | Sample preparation for MS analysis

The levels of non-fluorinated and different fluorinated proteoforms of CA2, DJ1, and CCS-D2 were assessed by direct MS on soluble cell lysates under denaturing conditions (Vimer et al., 2020). CA2, DJ1, and CCS-D2 were expressed in HEK293T using a T25 flask (Greiner Bio-One). For all incorporation conditions, the used materials and reagents were scaled down three times with respect to the expression in T75 flask. Cells were harvested 48 h after transfection. Cell pellets were collected, washed once in PBS and then in 500 mM ammonium acetate. Prior to MS analysis, each cell pellet was lysed in 290 μ L

of 500 mM ammonium acetate by freeze-thaw lysis. The soluble lysates were subsequently separated by centrifugation at 14,000 g at 4 °C for 15 min.

4.9 | Direct MS and LC-MS

For CA2 and DJ1 analysis, the collected lysates were diluted 10- to 50-fold in 20% acetic acid and directly sprayed into a modified Q-Exactive Plus Orbitrap EMR mass spectrometer (ThermoFisher Scientific) (Ben-Nissan et al., 2017), using gold-coated nano-ESI capillaries prepared in-house, as previously described (Kirshenbaum et al., 2010). The spectra were recorded using the following instrumental parameters: capillary voltage 1.7 kV, the source was operated at a constant energy of 2 V in the flatpole bias and interflatpole lens, resolution of 10,000 m/Δm, interflatpole lens 2.0 V, bent flatpole DC bias 1.8 V with gradient of 11.8 V, high collision dissociation (HCD) energy up to 15 V, and trapping gas pressure of 1.0, corresponding to fore vacuum (FV) of 1.56 mbar, high vacuum (HV) of 3.12e-05 mbar, and ultra-high vacuum (UHV) of 1.46e-10 mbar.

For CCS-D2 measurements, the cell lysates were analyzed by liquid chromatography, using a nanoAcquity UPLC (Waters) coupled to the mass spectrometer, using the HESI source. Proteins from 5 μL of each lysate were separated over a linear gradient of 20%–50% acetonitrile (0.05% formic acid), using a reversed-phase monolithic column (Rozen et al., 2013), at 60 °C and at a flow rate of 15 μL/min, over 15 min. MS spectra were recorded using the following settings: capillary voltage 2.9 kV, the source was operated at a constant energy of 2 V in the flatpole bias and interflatpole lens, resolution of 10,000 m/Δm, interflatpole lens 1.9 V, bent flatpole DC bias 1.7 V with gradient of 10 V, HCD energy 15 V, and trapping gas pressure of 0.8 corresponding to FV of 1.64 mbar, HV of 2.67e-05 mbar, and UHV of 1.24e-10 mbar.

4.10 | MS data analysis

Mass deconvolution of recorded spectra were carried out using Unidec (v.5.2.1) (Marty et al., 2015). Data were processed with normalization and bin 2.0, and with following Unidec parameters: sample mass every 1 Da, smoothing nearby points—some, suppressing artifacts—some, Gaussian peak shape function, charge smooth width of 1.0, point smooth width of 1.0, maximum 50 iterations, m/z to mass transformation—smart, adduct mass of 1.0 Da, average mass output and native charge offset of ±1000. Deconvoluted masses of CA2, DJ1, and CCS-D2

proteoforms were obtained from the m/z ranges of 850–1200, 750–1400 and 700–3000, and searched in the mass ranges of 29,130–29,400 Da, 19,770–19,990 Da, and 15,900–16,400 Da, respectively. For each expression condition, the intensities of deconvoluted masses of interest were averaged from three technical repetitions. Deconvoluted masses with D-score <60 were inspected manually. The sum of intensities of non-incorporated and incorporated proteoforms in each data set were normalized to the total intensity of all detected proteoforms and expressed as percentage.

4.11 | Calculation of fluorine incorporation probability

The probability of incorporating the fluorinated amino acid (FAA) within the protein was derived from MS data, which provided the distribution of different fluorinated proteoforms within the same sample. The probability of fluorine incorporation was calculated using the equation:

$$p_F = \frac{\sum f \cdot n_F}{\sum f} \cdot \frac{1}{n_F} \quad (1)$$

where p_F indicates the probability of fluorine incorporation, f indicates the frequency distribution obtained from the MS data, and n_F indicates the number of residues replaced with the FAA. In the medium switch conditions, the frequency data in which the protein had zero fluorine was eliminated to calculate the probability of incorporation, while in the mixture conditions all frequencies of the distribution, including the one with zero fluorine, were considered to calculate the probability. Using the calculated probability values, the theoretical binomial distribution for each condition was back-calculated using the binomial distribution formula:

$$P(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad (2)$$

Here, n represents the number of residues to be replaced with FAA, x denotes the number of successful incorporations of fluorine within each molecule, p represents the calculated probability of success, and $1-p$ represents the probability of failure. The p_F values obtained from the MS data for different proteins and different expression conditions were globally fitted for each FAA in OriginPro 8 using the following function of incorporation probability:

$$p_F(X) = \frac{cX}{1 + X(c-1) + \frac{[AA]_{int}}{[AA]_{ext}}} \quad (3)$$

where $p_F(X)$ is the probability of fluorine incorporation, c is a selectivity coefficient ($0 \leq c \leq 1$) that accounts for the fact that the expression machinery recognizes each FAA less efficiently than the corresponding natural AA, $[AA_{int}]$ is the apparent concentration of endogenous natural AA calculated over the total cell culture volume, $[AA_{ext}]$ is the total concentration of FAA + AA, and X is the fraction of FAA over the total supplemented AA:

$$X = \frac{[FAA]}{[AA_{ext}]} = \frac{[FAA]}{[FAA] + [AA]} \quad (4)$$

$[AA_{ext}]$ was kept constant at 600 μ M for tyrosine, 400 μ M for phenylalanine, and 80 μ M for tryptophan. The derivation of Equation (3) is described in the Supporting Information S1.

4.12 | FAA-CA2 crystallization, data collection, and structure calculation

3FY-CA2 (ST0h), 4FF-CA2 (ST24h) and 5FW-CA2 (ST24h) were crystallized using the sitting drop vapor diffusion technique. A crystallization plate containing 24 pedestals and 24 wells was used. An aliquot of 1 mL of reservoir buffer (100 mM of HEPES, 1.4 M sodium citrate and 1 mM 4-(hydroxymercuri)benzoic acid, pH 7.5) was introduced into each well, providing an optimized environment for crystal growth. Subsequently, 2 μ L of a purified protein solution (15 mg/mL in 10 mM HEPES, pH 6.8) was added to the corresponding pedestal, along with 2 μ L of the reservoir buffer. The crystallization plate was stored at 20 °C and after 3 days crystals had successfully formed.

The dataset was collected in-house, using a BRUKER D8 Venture diffractometer equipped with a PHOTON III detector, at 100 K; the crystal used for data collection was cryo-cooled using 30% ethylene glycol in the mother liquor. The crystal diffracted up to 1.1 Å resolution, but all the structures have been refined at 1.3 Å: they all belong to space group P21 with one molecule in the asymmetric unit (consistent with the vast majority of CA2 entries deposited on the PDB), a solvent content of about 50%, and a mosaicity of 0.3°.

The data were processed using the program XDS (Kabsch, 2010), reduced and scaled using XSCALE and amplitudes were calculated using XDSCONV (Kabsch, 2010). The structures were solved by molecular replacement using a published crystal structure of 6FW-CA2 (ST24h) (PDB 8B29). The successful orientation hand translation of the molecule within the crystallographic unit cell was determined with MOLREP (Vagin & Teplyakov, 2010). The refinement and water

molecule fitting were carried out using PHENIX taking full anisotropic displacement parameters into account (Adams et al., 2010). Fluorine occupancies were set to 0.50 and left free to refine along with the B-factors. In between the refinement cycles, the model was subjected to manual rebuilding using COOT (Emsley et al., 2010). The quality of the refined structure was assessed using the program MOLPROBITY (Chen et al., 2010). Data processing and refinement statistics are shown in Table S1. Coordinates and structure factors have been deposited in the PDB under the accession codes 8P6U (5FW-CA2), 8PHL (4FF-CA2), 8Q0C (3FY-CA2). To assess the reliability of the fluorine occupancy, all the structures, including 6FW-CA2, were refined again by fixing the occupancy values to the average fluorine content (3FY-CA2: 0.62; 4FF-CA2: 0.33; 5FW-CA2: 0.28; 6FW-CA2: 0.48), which was determined for each expression condition with Equation (1) using all frequencies of the distribution from MS, including the one with zero fluorine atoms.

AUTHOR CONTRIBUTIONS

Enrico Luchinat: Writing – original draft; writing – review and editing; supervision; conceptualization; methodology; project administration. **Azzurra Costantino:** Formal analysis; investigation; writing – original draft; writing – review and editing; methodology; visualization. **Lan B. T. Pham:** Writing – review and editing; formal analysis; investigation; methodology. **Letizia Barbieri:** Writing – review and editing; investigation; methodology. **Vito Calderone:** Writing – review and editing; formal analysis; investigation. **Gili Ben-Nissan:** Writing – review and editing; investigation; supervision; methodology. **Michal Sharon:** Writing – review and editing; funding acquisition; supervision; methodology; resources. **Lucia Banci:** Writing – original draft; writing – review and editing; funding acquisition; supervision; conceptualization; resources.

ACKNOWLEDGMENTS

This work was supported by Instruct-ERIC, a Landmark ESFRI project, and specifically by the CERM/CIRMMP Italian Instruct Centre, by iNEXT-Discovery, grant agreement No. 871037, funded by the Horizon 2020 research and innovation program of the European Commission, by Fragment-Screen, grant agreement No. 101094131, funded by the Horizon Europe program of the European Commission, and by the Ministero dell'Università e della Ricerca (MUR) PRIN grant No. 20177XJCHX. We acknowledge the project “Potentiating the Italian Capacity for Structural Biology Services in Instruct-ERIC,” Acronym “ITACA.SB” (Project no. IR0000009) within the

call MUR 3264/2021 PNRR M4/C2/L3.1.1, funded by the European Union – NextGenerationEU. Michal Sharon is grateful for the support of an Advanced Grant European Research Council (ERC) (Horizon 2020)/ERC Grant Agreement no. 101092725. Michal Sharon is the incumbent of the Aharon and Ephraim Katzir Memorial Professorial Chair.

ORCID

Lucia Banci  <https://orcid.org/0000-0003-0562-5774>

Enrico Luchinat  <https://orcid.org/0000-0003-4183-4311>

REFERENCES

- Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, et al. PHENIX: a comprehensive python-based system for macromolecular structure solution. *Acta Cryst D*. 2010; 66(2):213–21. <https://doi.org/10.1107/S0907444909052925>
- Aricescu AR, Lu W, Jones EY. A time- and cost-efficient system for high-level protein production in mammalian cells. *Acta Cryst D*. 2006;62(10):1243–50. <https://doi.org/10.1107/S0907444906029799>
- Arntson KE, Pomerantz WCK. Protein-observed fluorine NMR: a bioorthogonal approach for small molecule discovery. *J Med Chem*. 2016;59(11):5158–71. <https://doi.org/10.1021/acs.jmedchem.5b01447>
- Banerjee AL, Swanson M, Mallik S, Srivastava DK. Purification of recombinant human carbonic anhydrase-II by metal affinity chromatography without incorporating histidine tags. *Protein Expr Purif*. 2004;37(2):450–4. <https://doi.org/10.1016/j.pep.2004.06.031>
- Barbieri L, Luchinat E, Banci L. Characterization of proteins by in-cell NMR spectroscopy in cultured mammalian cells. *Nat Protoc*. 2016;11(6):1101–11. <https://doi.org/10.1038/nprot.2016.061>
- Barbieri L, Luchinat E, Banci L. Intracellular metal binding and redox behavior of human DJ-1. *J Biol Inorg Chem*. 2018;23(1): 61–9. <https://doi.org/10.1007/s00775-017-1509-5>
- Ben-Nissan G, Belov ME, Morgenstern D, Levin Y, Dym O, Arkind G, et al. Triple-stage mass spectrometry unravels the heterogeneity of an endogenous protein complex. *Anal Chem*. 2017; 89(8):4708–15. <https://doi.org/10.1021/acs.analchem.7b00518>
- Bhat TN. Correlation between occupancy and temperature factors of solvent molecules in crystal structures of proteins. *Acta Cryst A*. 1989;45(1):145–6. <https://doi.org/10.1107/S0108767388008967>
- Chen H, Viel S, Ziarelli F, Peng L. 19F NMR: a valuable tool for studying biological events. *Chem Soc Rev*. 2013;42(20):7971–82. <https://doi.org/10.1039/c3cs60129c>
- Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Cryst D*. 2010;66(1):12–21. <https://doi.org/10.1107/S0907444909042073>
- Crowley PB, Kyne C, Monteith WB. Simple and inexpensive incorporation of 19F-tryptophan for protein NMR spectroscopy. *Chem Commun (Camb)*. 2012;48(86):10681–3. <https://doi.org/10.1039/c2cc35347d>
- Deller MC, Rupp B. Models of protein–ligand crystal structures: trust, but verify. *J Comput Aided Mol Des*. 2015;29(9):817–36. <https://doi.org/10.1007/s10822-015-9833-8>
- Divakaran A, Kirberger SE, Pomerantz WCK. SAR by (protein-observed) 19F NMR. *Acc Chem Res*. 2019;52(12):3407–18. <https://doi.org/10.1021/acs.accounts.9b00377>
- Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Cryst D*. 2010;66(4):486–501. <https://doi.org/10.1107/S0907444910007493>
- Gee CT, Arntson KE, Urlick AK, Mishra NK, Hawk LML, Wisniewski AJ, et al. Protein-observed 19F-NMR for fragment screening, affinity quantification and druggability assessment. *Nat Protoc*. 2016;11(8):1414–27. <https://doi.org/10.1038/nprot.2016.079>
- Gerig JT. Fluorine NMR of proteins. *Prog Nucl Magn Reson Spectrosc*. 1994;26:293–370. [https://doi.org/10.1016/0079-6565\(94\)80009-X](https://doi.org/10.1016/0079-6565(94)80009-X)
- Gimenez D, Phelan A, Murphy CD, Cobb SL. 19F NMR as a tool in chemical biology. *Beilstein J Org Chem*. 2021;17(1):293–318. <https://doi.org/10.3762/bjoc.17.28>
- Gronenborn AM. Small, but powerful and attractive: 19F in biomolecular NMR. *Structure*. 2022;30(1):6–14. <https://doi.org/10.1016/j.str.2021.09.009>
- Hu Y, Cheng K, He L, Zhang X, Jiang B, Jiang L, et al. NMR-based methods for protein analysis. *Anal Chem*. 2021;93(4):1866–79. <https://doi.org/10.1021/acs.analchem.0c03830>
- Jackson JC, Hammill JT, Mehl RA. Site-specific incorporation of a 19F-amino acid into proteins as an NMR probe for characterizing protein structure and reactivity. *J Am Chem Soc*. 2007; 129(5):1160–6. <https://doi.org/10.1021/ja064661t>
- Kabsch W. XDS. *Acta Crystallogr D Biol Crystallogr*. 2010;66(2): 125–32. <https://doi.org/10.1107/S0907444909047337>
- Kirshenbaum N, Michalevski I, Sharon M. Analyzing large protein complexes by structural mass spectrometry. *J Vis Exp*. 2010;40: 1954.
- Kitevski-LeBlanc JL, Prosser RS. Current applications of 19F NMR to studies of protein structure and dynamics. *Prog Nucl Magn Reson Spectrosc*. 2012;62:1–33. <https://doi.org/10.1016/j.pnmrs.2011.06.003>
- Krařík D, Ištvanek E, Džatko Š, Viřkova P, Foldynova-Trantřrkova S, Trantřrek L. Towards profiling of the G-quadruplex targeting drugs in the living human cells using NMR spectroscopy. *Int J Mol Sci*. 2021;22(11):6042. <https://doi.org/10.3390/ijms22116042>
- Li C, Wang G-F, Wang Y, Creager-Allen R, Lutz EA, Scronce H, et al. Protein 19F NMR in *Escherichia coli*. *J Am Chem Soc*. 2010;132(1):321–7. <https://doi.org/10.1021/ja907966n>
- Luchinat E, Banci L. In-cell NMR: from target structure and dynamics to drug screening. *Curr Opin Struct Biol*. 2022;74: 102374. <https://doi.org/10.1016/j.sbi.2022.102374>
- Luchinat E, Barbieri L, Banci L. A molecular chaperone activity of CCS restores the maturation of SOD1 fALS mutants. *Sci Rep*. 2017;7(1):17433. <https://doi.org/10.1038/s41598-017-17815-y>
- Luchinat E, Barbieri L, Cremonini M, Nocentini A, Supuran CT, Banci L. Drug screening in human cells by NMR spectroscopy allows the early assessment of drug potency. *Angew Chem Int Ed Engl*. 2020;59(16):6535–9. <https://doi.org/10.1002/anie.201913436>
- Luchinat E, Cremonini M, Banci L. Radio signals from live cells: the coming of age of in-cell solution NMR. *Chem Rev*. 2022; 122(10):9267–306. <https://doi.org/10.1021/acs.chemrev.1c00790>

- Marty MT, Baldwin AJ, Marklund EG, Hochberg GKA, Benesch JLP, Robinson CV. Bayesian deconvolution of mass and ion mobility spectra: from binary interactions to polydisperse ensembles. *Anal Chem*. 2015;87(8):4370–6. <https://doi.org/10.1021/acs.analchem.5b00140>
- Pearce NM, Bradley AR, Krojer T, Marsden BD, Deane CM, von Delft F. Partial-occupancy binders identified by the pan-dataset density analysis method offer new chemical opportunities and reveal cryptic binding sites. *Struct Dyn*. 2017;4(3):032104. <https://doi.org/10.1063/1.4974176>
- Pearce NM, Krojer T, von Delft F. Proper modelling of ligand binding requires an ensemble of bound and unbound states. *Acta Cryst D*. 2017;73(3):256–66. <https://doi.org/10.1107/S2059798317003412>
- Pham LBT, Costantino A, Barbieri L, Calderone V, Luchinat E, Banci L. Direct expression of fluorinated proteins in human cells for ^{19}F in-cell NMR spectroscopy. *J Am Chem Soc*. 2023; 145(2):1389–99. <https://doi.org/10.1021/jacs.2c12086>
- Puffer B, Kreutz C, Rieder U, Ebert M-O, Konrat R, Micura R. 5-fluoro pyrimidines: labels to probe DNA and RNA secondary structures by 1D ^{19}F NMR spectroscopy. *Nucleic Acids Res*. 2009;37(22):7728–40. <https://doi.org/10.1093/nar/gkp862>
- Rozen S, Tieri A, Ridner G, Stark A-K, Schmalzer T, Ben-Nissan G, et al. Exposing the subunit diversity within protein complexes: a mass spectrometry approach. *Methods*. 2013;59(3):270–7. <https://doi.org/10.1016/j.ymeth.2012.12.013>
- Sekhar A, Kay LE. An NMR view of protein dynamics in health and disease. *Annu Rev Biophys*. 2019;48(1):297–319. <https://doi.org/10.1146/annurev-biophys-052118-115647>
- Sochor F, Silvers R, Müller D, Richter C, Fürtig B, Schwalbe H. ^{19}F -labeling of the adenine H2-site to study large RNAs by NMR spectroscopy. *J Biomol NMR*. 2016;64(1):63–74. <https://doi.org/10.1007/s10858-015-0006-9>
- Sudakov A, Knezic B, Hengesbach M, Fürtig B, Stirnal E, Schwalbe H. Site-specific labeling of RNAs with modified and ^{19}F -labeled nucleotides by chemo-enzymatic synthesis. *Chem A Eur J*. 2023;29(25):e202203368. <https://doi.org/10.1002/chem.202203368>
- Theillet F-X. In-cell structural biology by NMR: the benefits of the atomic scale. *Chem Rev*. 2022;122(10):9497–570. <https://doi.org/10.1021/acs.chemrev.1c00937>
- Tronrud DE. Introduction to macromolecular refinement. *Acta Cryst D*. 2004;60(12):2156–68. <https://doi.org/10.1107/S090744490402356X>
- Vagin A, Teplyakov A. Molecular replacement with MOLREP. *Acta Cryst D*. 2010;66(1):22–5. <https://doi.org/10.1107/S0907444909042589>
- Vimer S, Ben-Nissan G, Sharon M. Direct characterization of over-produced proteins by native mass spectrometry. *Nat Protoc*. 2020;15(2):236–65. <https://doi.org/10.1038/s41596-019-0233-8>
- Yoder NC, Kumar K. Fluorinated amino acids in protein design and engineering. *Chem Soc Rev*. 2002;31(6):335–41. <https://doi.org/10.1039/b201097f>
- Zhu W, Guseman AJ, Bhinderwala F, Lu M, Su X-C, Gronenborn AM. Visualizing proteins in mammalian cells by ^{19}F NMR spectroscopy. *Angew Chem Int Ed*. 2022;61(23):e202201097. <https://doi.org/10.1002/anie.202201097>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Costantino A, Pham LBT, Barbieri L, Calderone V, Ben-Nissan G, Sharon M, et al. Controlling the incorporation of fluorinated amino acids in human cells and its structural impact. *Protein Science*. 2024;33(3): e4910. <https://doi.org/10.1002/pro.4910>

Supplementary Materials

Controlling the incorporation of fluorinated amino acids in human cells and its structural impact

Azzurra Costantino¹, Lan B. T. Pham¹, Letizia Barbieri^{1,2}, Vito Calderone^{1,3}, Gili Ben-Nissan⁴, Michal Sharon⁴, Lucia Banci^{1,2,3*}, Enrico Luchinat^{1,2,3*}

¹CERM – Magnetic Resonance Center, Università degli Studi di Firenze, 50019 Sesto Fiorentino, Italy;

²Consorzio Interuniversitario Risonanze Magnetiche di Metallo Proteine – CIRMMMP, Sesto Fiorentino, Italy;

³Dipartimento di Chimica, Università degli Studi di Firenze, Sesto Fiorentino, Italy;

⁴Department of Biomolecular Sciences, Weizmann Institute of Science, Rehovot, 7610001, Israel;

Correspondence

Lucia Banci and Enrico Luchinat, CERM – Magnetic Resonance Center, Università degli Studi di Firenze, Via Luigi Sacconi 6, Sesto Fiorentino 50019, Italy.

Email: banci@cerm.unifi.it and eluchinat@cerm.unifi.it

Supplementary Methods

Supplementary Figures S1-S12

Supplementary Tables S1, S2

Supplementary Methods

Derivation of the incorporation probability function

The fluorine incorporation probability, p_F , as a function of the fraction of fluorinated amino acid (FAA) over the total amount of amino acid supplemented in the expression medium (AA_{ext}) was derived as follows.

It is assumed that p_F is purely stochastic, therefore has the same value for the incorporation of a given FAA in any position along the sequence of any protein. It can be defined as:

$$p_F = \frac{V_F}{V_F + V_N} \quad (s1)$$

Where $V_F = eff_F \cdot [FAA]$ and $V_N = eff_N \cdot [AA]$ are the rates of incorporation of the FAA and the normal AA, respectively, which depend on the amino acid concentrations ($[FAA]$ and $[AA]$) and on how efficiently the protein expression machinery recognizes them (eff_F and eff_N). eff_F is lower than eff_N due to less efficient recognition of the FAA (e.g., due to slower uptake and/or poor recognition by the endogenous aminoacyl-tRNA synthetases and/or by the ribosome). Therefore:

$$eff_F = c \cdot eff_N \quad (s2)$$

Where c is a 'selectivity coefficient' comprised between 0 (= the FAA is not recognized by the cell) and 1 (= no difference between FAA and AA). Eq. s1 becomes:

$$p_F = \frac{c[FAA]}{c[FAA] + [AA]} \quad (s3)$$

If we account for the presence of a residual pool of natural AA in the cell culture, which is available for protein synthesis, $[AA_{int}]$, then eq. s3 becomes:

$$p_F = \frac{c[FAA]}{c[FAA] + [AA] + [AA_{int}]} \quad (s4)$$

The fraction of FAA over AA_{ext} is:

$$X = \frac{[FAA]}{[AA_{ext}]} = \frac{[FAA]}{[FAA] + [AA]} \quad (s5)$$

Therefore, $[FAA] = X[AA_{ext}]$ and $[AA] = (1 - X)[AA_{ext}]$

Substituting in eq. s4 and rearranging gives:

$$p_F(X) = \frac{cX}{1 + X(c - 1) + \frac{[AA_{int}]}{[AA_{ext}]}} \quad (s6)$$

Which is eq. 3 in the main text.

Supplementary Figures

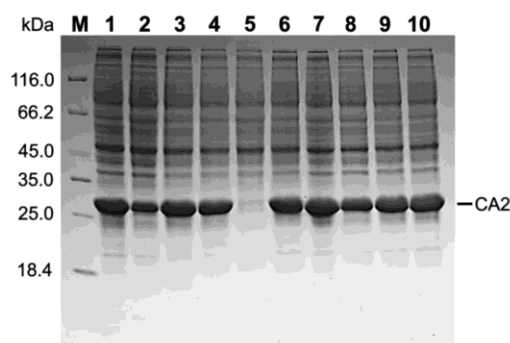


Figure S1. Expression of CA2 in HEK293T cells supplemented with FAA-containing media at different doses and times. Coomassie-stained SDS-PAGE of lysates from HEK293T cells expressing CA2 in different media formulations with different FAA medium switch times (ST): M = Marker; 1) commercial DMEM and lysed in 150 μ L of PBS; 2) in-house DMEM supplemented with 5FW at ST = 0 and lysed in 80 μ L of PBS; 3) in-house DMEM supplemented with a mixture of 50% of 5FW and 50% of W at ST = 0 and lysed in 150 μ L of PBS; 4) in-house DMEM supplemented with 5FW at ST = 24h and lysed in 150 μ L of PBS; 5) in-house DMEM supplemented with 4FF at ST = 0 and lysed in 100 μ L of PBS; 6) in-house DMEM supplemented with a mixture of 50% of 4FF and 50% of F at ST = 0 and lysed in 150 μ L of PBS; 7) in-house DMEM supplemented with 4FF at ST = 24h and lysed in 80 μ L of PBS; 8) in-house DMEM supplemented with 3FY at ST = 0 and lysed in 80 μ L of PBS; 9) in-house DMEM supplemented with a mixture of 50% of 3FY and 50% of Y at ST = 0 and lysed in 150 μ L of PBS; 10) in-house DMEM supplemented with 3FY at ST = 24h and lysed in 150 μ L of PBS.

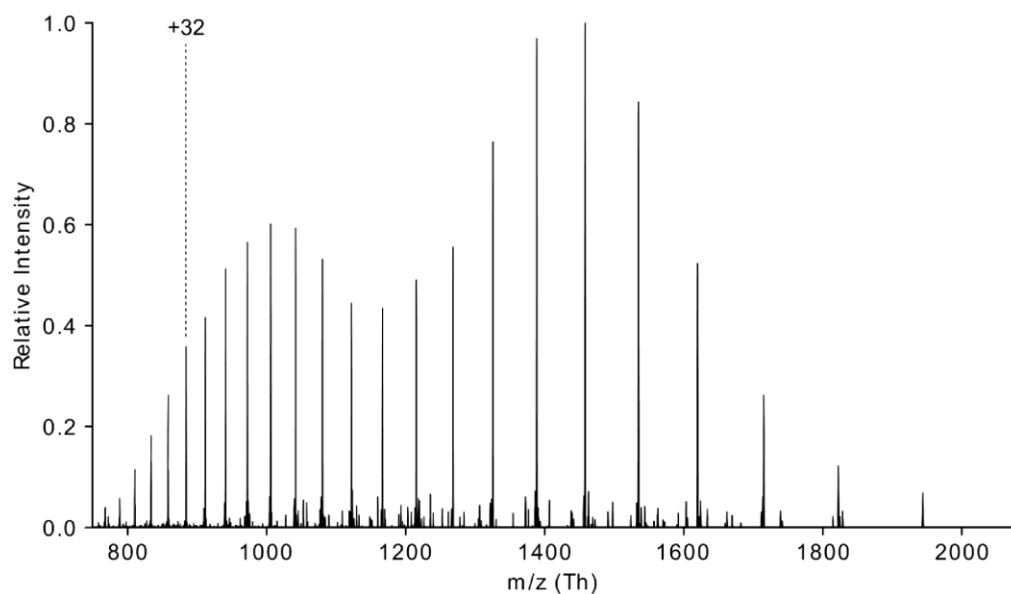
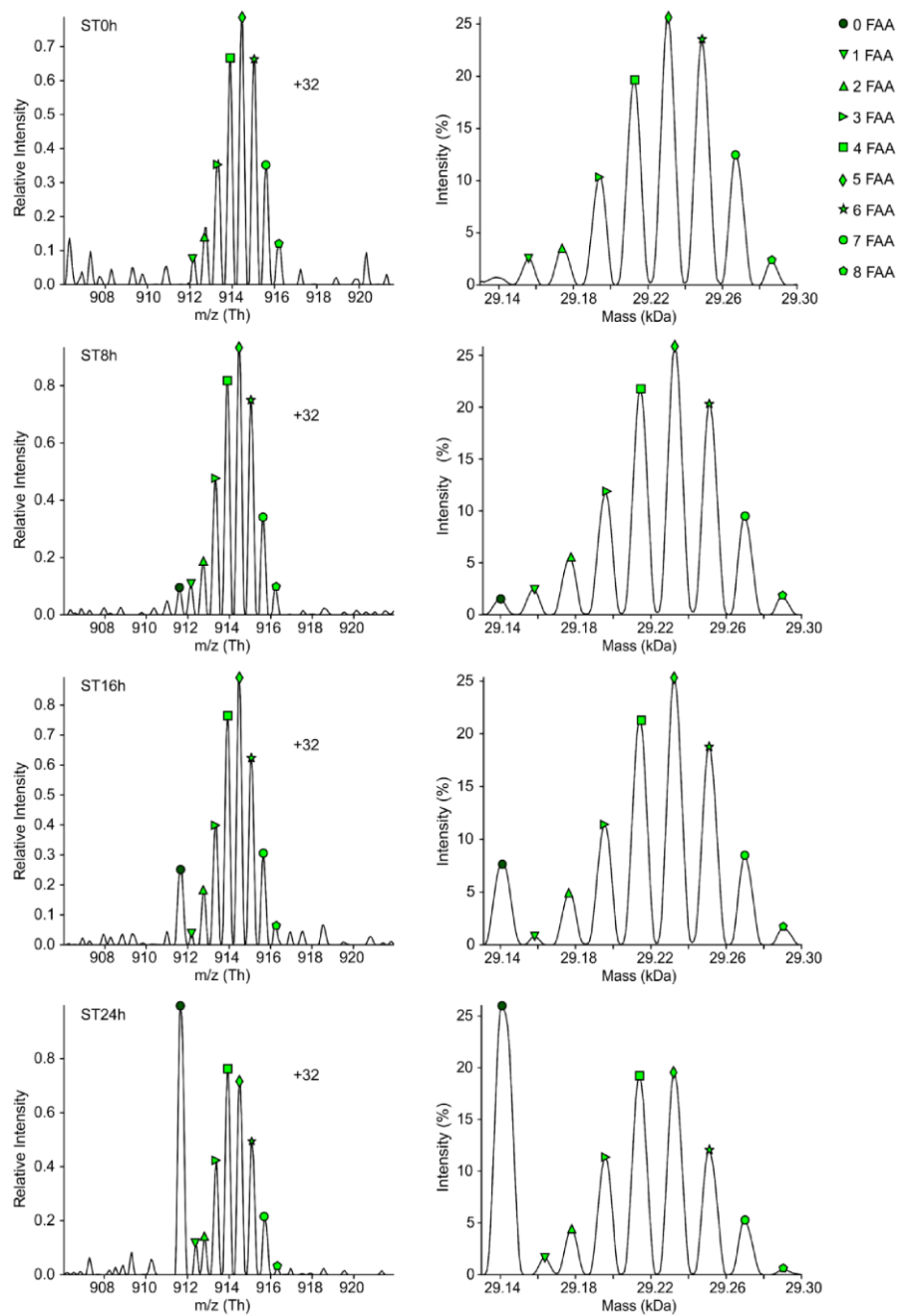


Figure S2. Direct MS spectrum of non-fluorinated and denatured CA2 in HEK293T cell lysate. CA2 was expressed in medium without FAA. The spectrum displays the distribution of CA2 charge states ranging from +38 to +14 (left-to-right), in a m/z range of 700-2200. The charge state +32 is labelled and selected for illustrating the fluorination of CA2 resulting from different expression conditions in Figures S3 and S4. The figure was generated by Unidec (5.2.1).



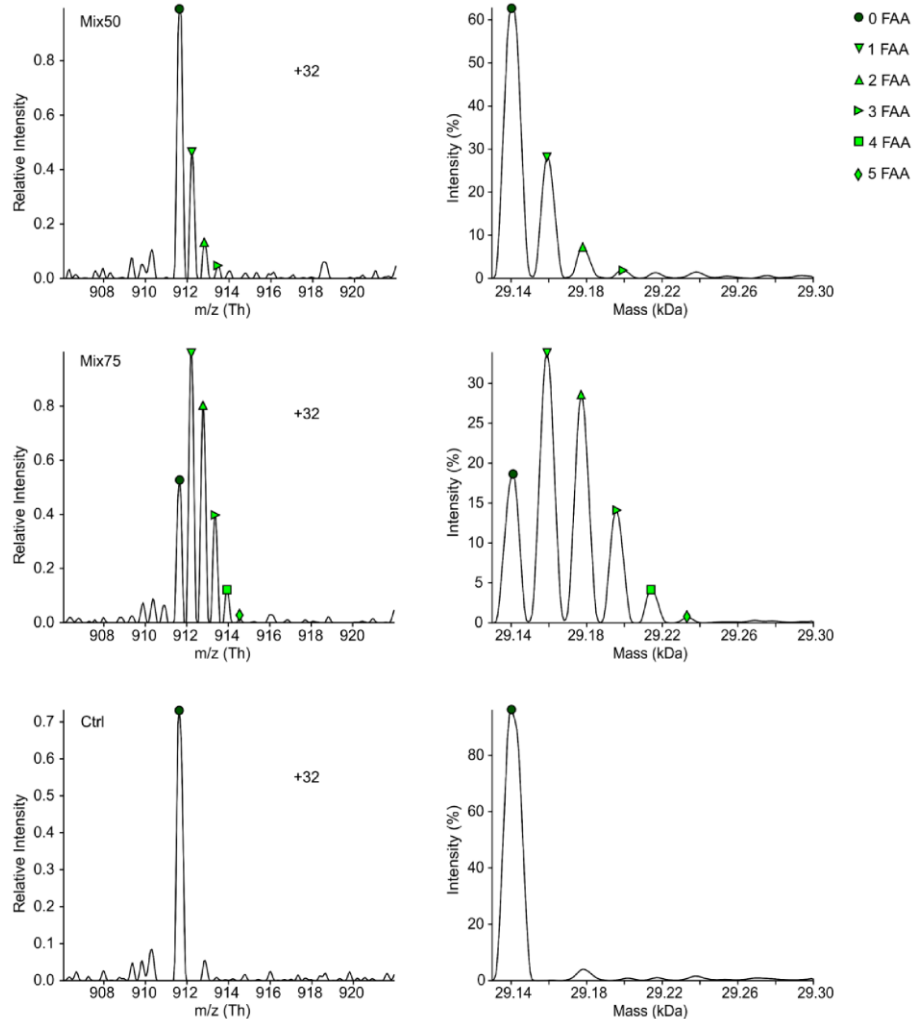


Figure S3. Direct MS spectra of denatured CA2 (charge state +32) in cell lysates (left) and the corresponding deconvoluted mass spectra (right) showing 3FY-CA2 grown under different expression conditions. ST: switch time; Ctrl: control sample containing CA2 expressed in medium without FAA. The deconvoluted masses of CA2 proteoforms were obtained and selected from a m/z range of 850-1200 as described in Materials and Methods, y axis indicates the percentage intensities of selected masses, of which the sum is hundred percent. The figures were generated by Unidec (5.2.1).

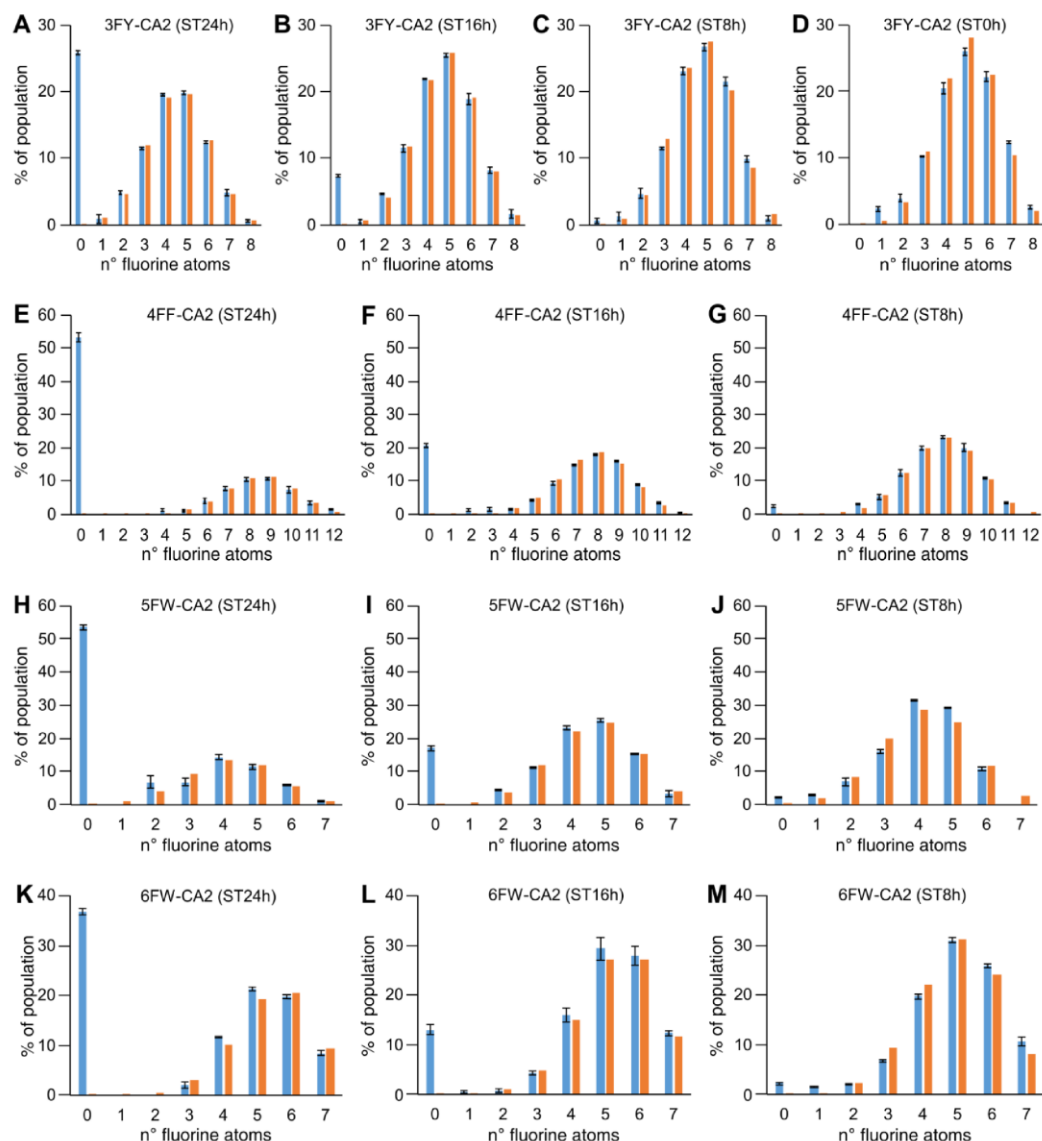


Figure S4. Experimental distribution of fluorinated proteoforms obtained by MS (blue) of: 3FY-CA2 A. (ST24h), B. (ST16h), C. (ST8h) and D. (ST0h); 4FF-CA2 E. (ST24h), F. (ST16h) and G. (ST8h); 5FW-CA2 H. (ST24h), I. (ST16h) and J. (ST8h); 6FW-CA2 K. (ST24h), L. (ST16h) and M. (ST8h). The binomial distribution back-calculated from the p_F values of the fluorinated population (orange, see Materials and Methods) shows good agreement with the experimental data, indicating that the incorporation of FAAs is purely stochastic.

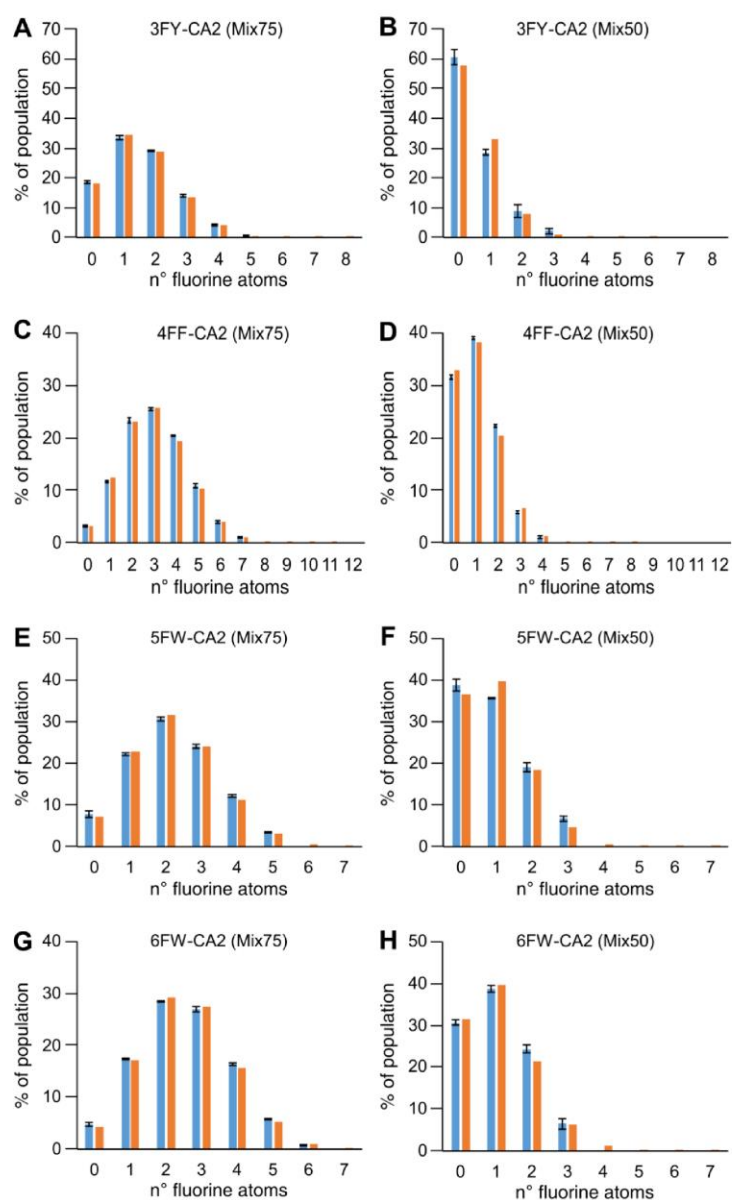


Figure S5. Experimental distribution of fluorinated proteoforms obtained by MS (blue) of: 3FY-CA2 A. (Mix75) and B. (Mix50); 4FF-CA2 C. (Mix75) and D. (Mix50); 5FW-CA2 E. (Mix75) and F. (Mix50); 6FW-CA2 G. (Mix75) and H. (Mix50). The binomial distribution back-calculated from the p_F values (orange, see Materials and Methods) shows good agreement with the experimental data, indicating that the incorporation of FAAs is purely stochastic.

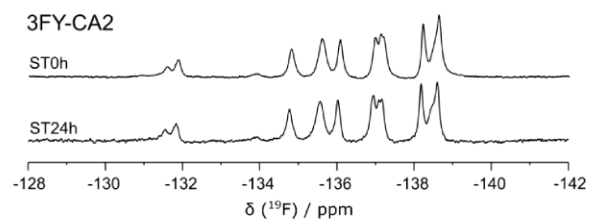


Figure S6. 1D ${}^{19}\text{F}$ NMR spectra of purified 3FY-CA2 (ST0h) (top) and 3FY-CA2 (ST24h) (bottom).

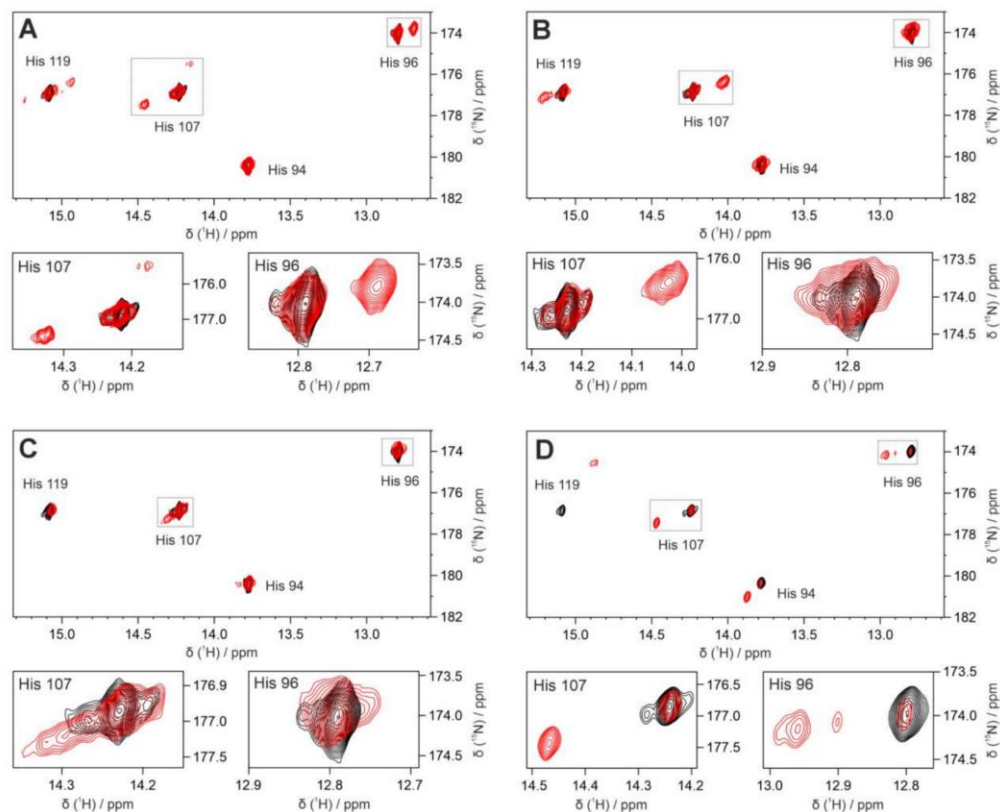


Figure S7. 2D ^1H - ^{15}N BEST-TROSY NMR cell lysate spectra of 3FY- (A), 4FF- (B), 5FW- (C) and 6FW-CA2 (D) (ST24h + His48h, red contour lines) overlaid to a reference cell lysate spectrum of non-fluorinated CA2 (black contour lines). The bottom panels show enlarged spectral regions of His 107 and His 96.

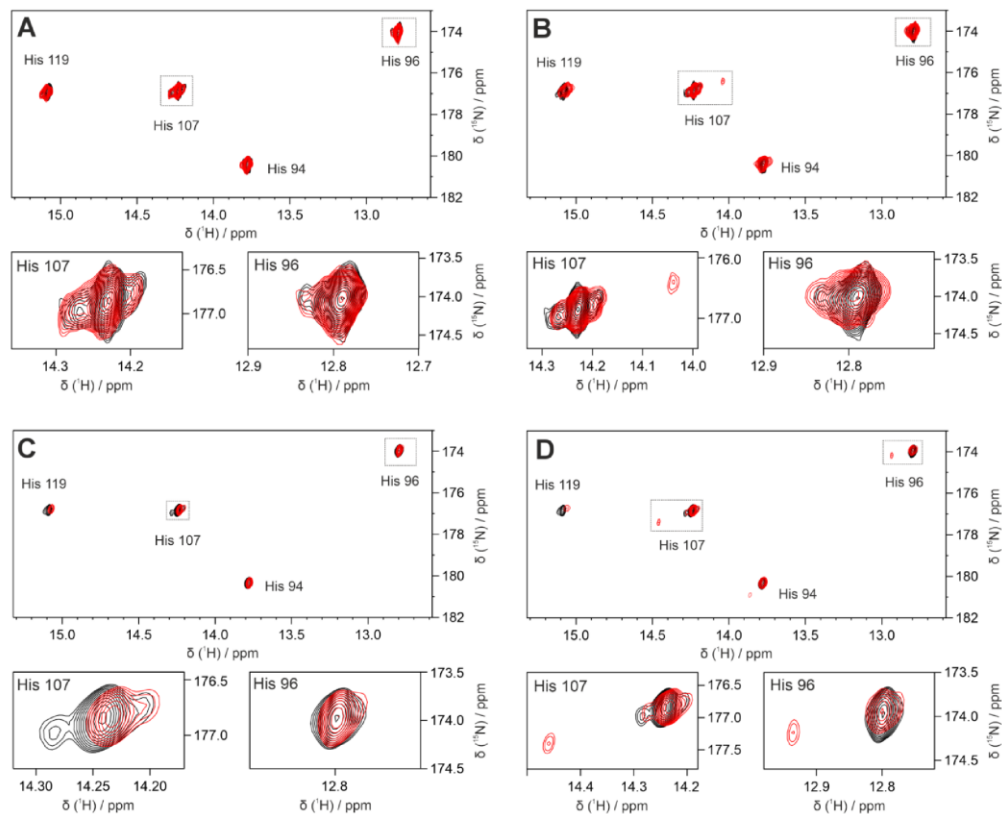


Figure S8. 2D ^1H - ^{15}N BEST-TROSY NMR cell lysate spectra of 3FY- (A), 4FF- (B), 5FW- (C) and 6FW-CA2 (D) (Mix50 + His48h, red contour lines) overlaid to a reference cell lysate spectrum of non-fluorinated CA2 (black contour lines). The bottom panels show enlarged spectral regions of His 107 and His 96.

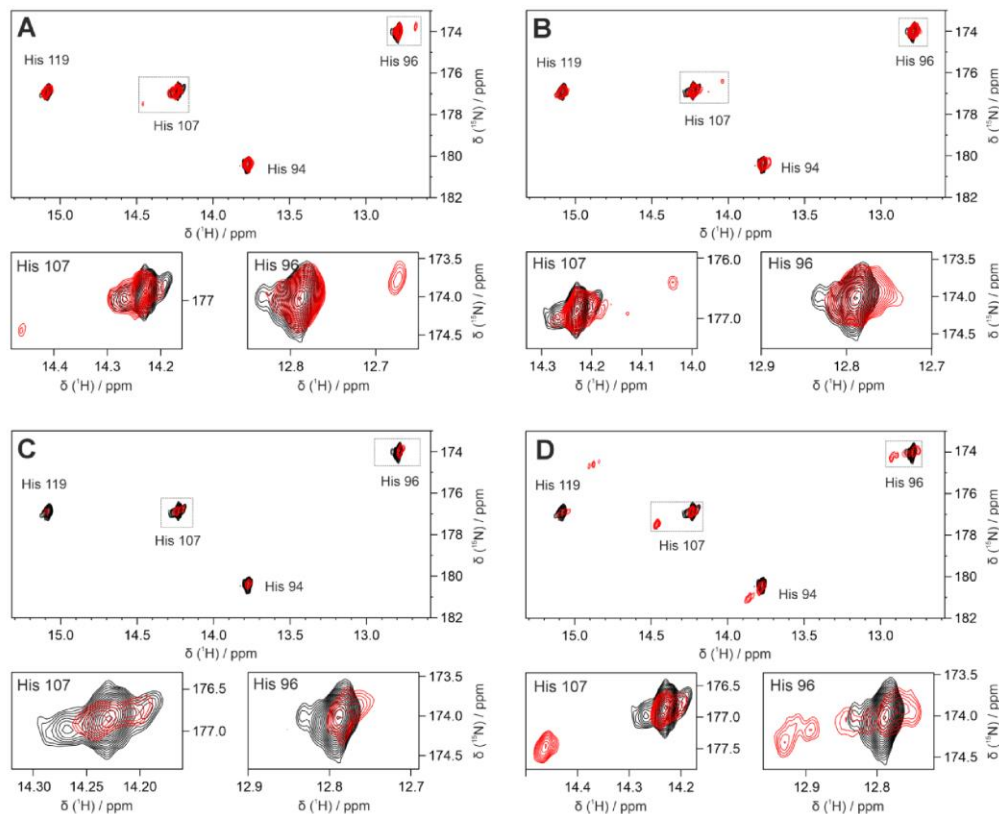


Figure S9. 2D ^1H - ^{15}N BEST-TROSY NMR cell lysate spectra of 3FY- (A), 4FF- (B), 5FW- (C) and 6FW-CA2 (D) (Mix75 + His48h, red contour lines) overlaid to a reference cell lysate spectrum of non-fluorinated CA2 (black contour lines). The bottom panels show enlarged spectral regions of His 107 and His 96.

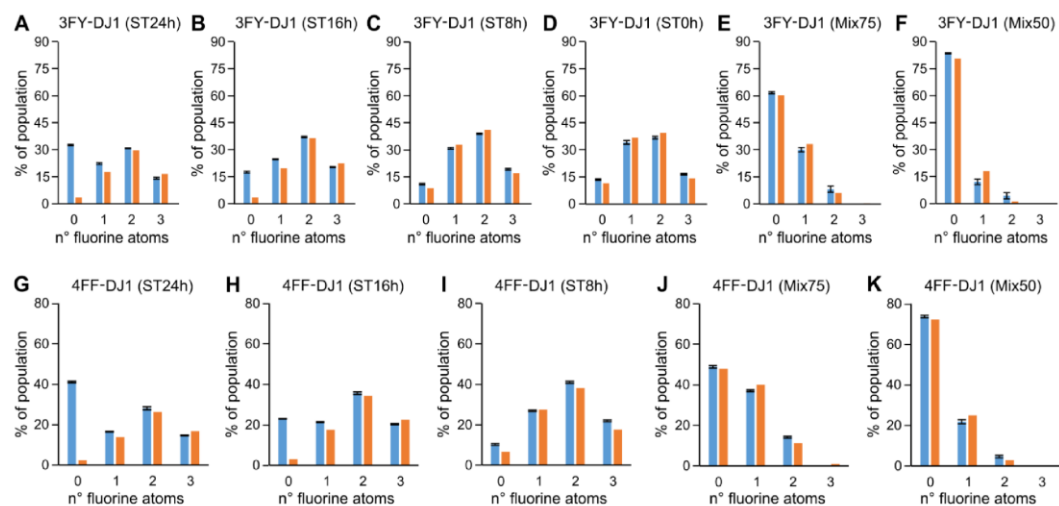


Figure S10: Experimental distribution of fluorinated proteoforms obtained by MS (blue) of DJ-1 under different expression conditions. DJ-1 has 3 Tyr, 3 Phe and no Trp residues. The binomial distribution back-calculated from the p_F values (orange, see Materials and Methods) is shown.

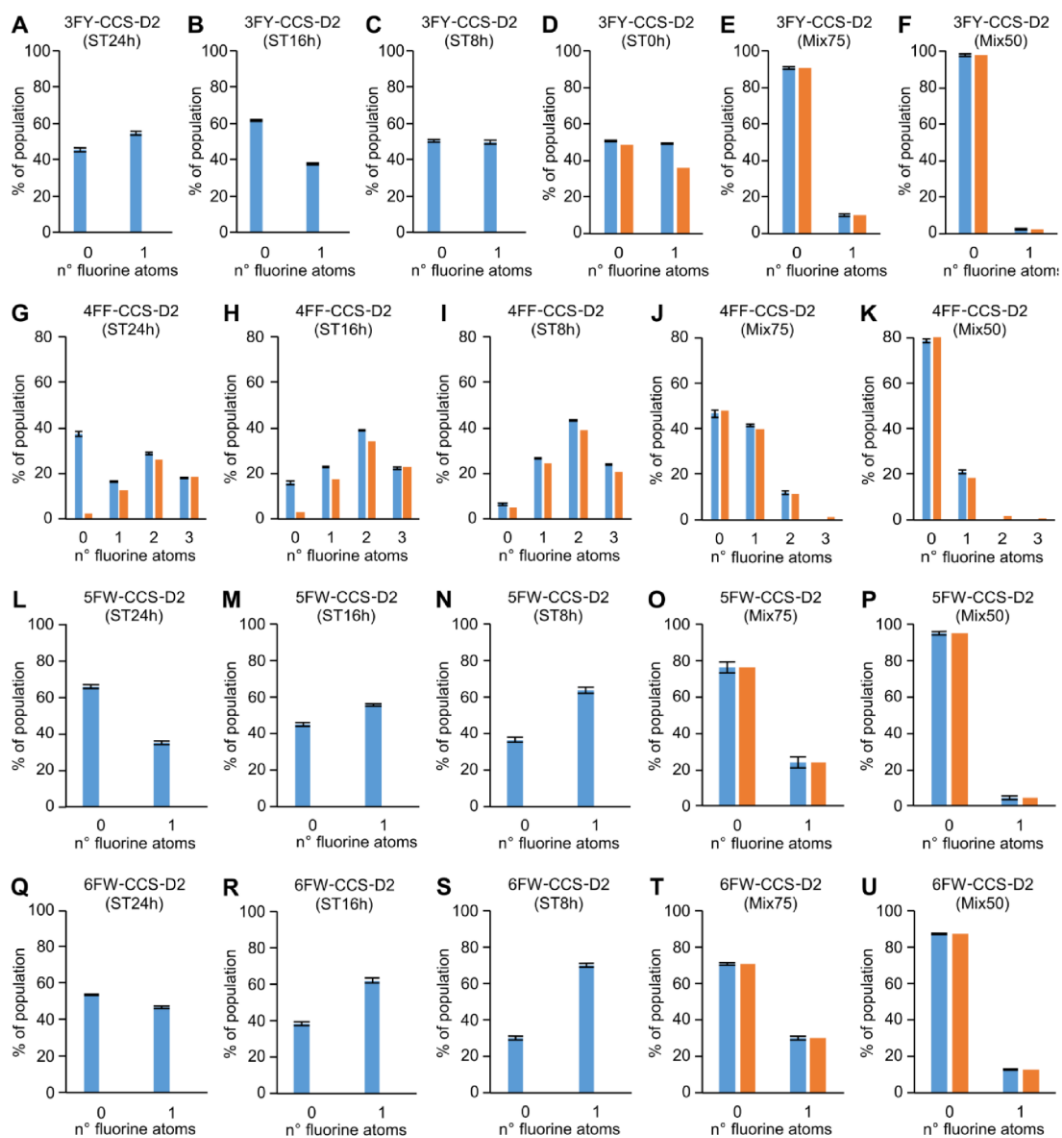


Figure S11: Experimental distribution of fluorinated proteoforms obtained by MS (blue) and back-calculated binomial distribution (orange) of CCS-D2 under different expression conditions. CCS-D2 has 1 Tyr, 3 Phe and 1 Trp. Plots of 3FY, 5FW and 6FW were omitted as they were not employed to calculate p_F . The binomial distribution back-calculated from the p_F values (orange, see Materials and Methods), where available, is shown.

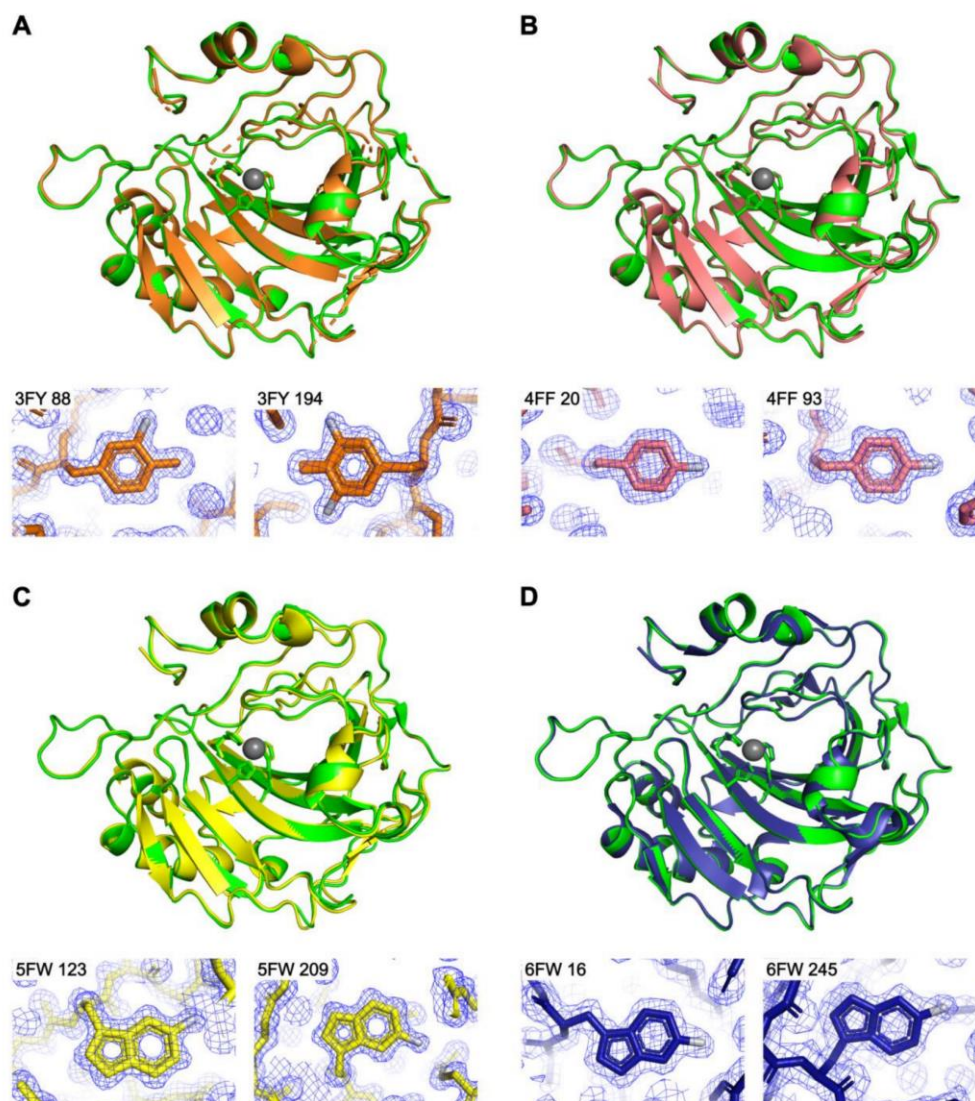


Figure S12. X-ray structure of A. 3FY-CA2 (PDB 8Q0C, orange), B. 4FF-CA2 (PDB 8PHL, salmon), C. 5FW-CA2 (PDB 8P6U, yellow) and D. 6FW-CA2 (PDB 8B29, blue) superimposed to that of native, nonfluorinated CA2 (PDB 1CA2, green), showing negligible differences between the structures. The insets show the 2mFo-DFc maps of the different fluorinated proteins contoured at 1σ (blue) showing the presence of fluorine atoms on different residues.

Supplementary tables

Table S1: Data collection and refinement statistics. Values in parenthesis refer to the highest resolution shell.

	3FY-CA2	4FF-CA2	5FW-CA2
Wavelength	1.541	1.541	1.541
Resolution range	23.01 - 1.3 (1.32 - 1.30)	31.5 - 1.3 (1.32 - 1.30)	20.9 - 1.3 (1.32 - 1.30)
Space group	P2 ₁	P2 ₁	P2 ₁
Unit cell	41.99 41.14 71.21 90.0 104.2 90.0	41.82 40.9 70.87 90.0 104.7 90.0	41.85 41.28 71.72 90.0 104.1 90.0
Total reflections	360810 (12607)	403210 (13045)	105971 (4124)
Unique reflections	112511 (4638)	109267 (5596)	69301 (3123)
Multiplicity	3.2 (2.7)	3.7 (2.3)	3.5 (2.3)
Completeness (%)	99.46 (90.56)	93.73 (87.22)	96.16 (92.92)
Mean I/sigma(I)	6.17 (1.15)	6.43 (1.13)	7.42 (2.76)
Wilson B-factor	11.88	10.81	7.07
R-merge	0.08701 (0.9582)	0.09176 (0.6669)	0.04966 (0.1618)
R-pim	0.05484 (0.6688)	0.04847 (0.4775)	0.04511 (0.1577)
CC1/2	0.997 (0.383)	0.997 (0.547)	0.995 (0.928)
Reflections used in refinement	57867 (2494)	53616 (2607)	56383 (2705)
Reflections used for R-free	2893 (125)	2680 (131)	2818 (135)
R-work	0.1651 (0.2999)	0.1878 (0.2700)	0.1442 (0.1883)
R-free	0.1962 (0.3170)	0.2014 (0.2496)	0.1754 (0.2845)
Number of non-hydrogen atoms	2324	2328	2393
RMSD (bonds)	0.010	0.010	0.017
RMSD (angles)	1.36	1.52	1.34
Ramachandran favored (%)	96.55	96.40	95.74
Ramachandran allowed (%)	3.45	3.60	4.26
Ramachandran outliers (%)	0.00	0.00	0.00
Rotamer outliers (%)	0.47	0.95	0.93
Average B-factor	17.27	16.28	11.47
macromolecules	16.22	15.33	10.19
ligands	13.50	14.75	19.16
solvent	26.68	24.02	19.56

Table S2. Primary sequences of CA2, DJ1 and CCS-D2.

Protein	Primary sequence
Carbonic anhydrase II (CA2)	MAHHWGYGKHNGPEHWHKDFPIAKGERQSPVDIDHTAKYDPS LKPLSVSYDQATSLRILNNGHAFNVEFDDSQDKAVLKGGPLDGT RLIQFHFHWGSLDGGQSEHTVDKKKYAAELHLVHWNTKYGDFG KAVQQPDGLAVLGIFLKVGSAPGLQKVVDVLDSIKTKGKSADFT NFDPRGILLPESLDYWTYPGSLTTPPLECVTWIVLKEPISVSSEQV LKFRKLNFNGEPEELMVDNWRPAQPLKNRQIKASFK
Protein deglycase DJ-1	MASKRALVILAKGAEEMETVIPVDVMRRAGIKVTVAGLAGKDPVQ CSRVDVICPDASLEDAKKEGPDYVVLPGGNLGAQNLSESAVK EILKEQENRKGLIAAICAGPTALLAHEIGFGSKVTTHPLAKDKMMN GGHYTSENREKDGILTSRPGTSFEFALAIVEALNGKEVAAQ VKAPLVVKD
Copper chaperone for SOD, domain 2 (CCS-D2)	MQNLGAAVAAILGGPGTVQGVVRFLLQTPERCLIEGTIDGLEPGLH GLHVHQYGDLTNNCNSCGNHFNPDGASHGGPQDSRHRGDLG NVRADADGRAIFRMEDEQLKVWDVIGRSLIIDEGEDDLGRGGHPL SKITGNSGERLACGIIARSA

Table S3. Calculated average masses and means of experimental deconvoluted masses of CA2 and its fluorinated proteoforms. PTMs: post-translational modifications; $\bar{x}$: mean calculated from all experimental masses of 3FY, 4FF, 6FW and 5FW incorporations.

Protein and PTMs	No. of F	Calculated average masses	$\bar{x}$ of experimental masses
CA2 without Met1 and with N-acetyl	0	29140.93	29140.40 ± 0.91
	1	29158.92	29158.70 ± 2.03
	2	29176.91	29176.60 ± 1.76
	3	29194.90	29195.36 ± 2.40
	4	29212.89	29213.12 ± 0.97
	5	29230.88	29231.28 ± 0.81
	6	29248.87	29249.64 ± 0.84
	7	29266.86	29268.74 ± 1.76
	8	29284.85	29287.05 ± 2.31
	9	29302.84	29303.34 ± 0.36
	10	29320.83	29321.43 ± 0.50
	11	29338.82	29340.16 ± 0.72
	12	29356.81	29361.20 ± 1.34

Table S4. Calculated average masses and means of experimental deconvoluted masses of DJ1 and its fluorinated proteoforms. PTMs: post-translational modifications; $\bar{x}$: mean calculated from all experimental masses of 3FY, 4FF, 6FW and 5FW incorporations.

Protein and PTMs	No. of F	Calculated average masses	$\bar{x}$ of experimental masses
DJ1 without Met1 and with N-acetyl	0	19801.95	19801.80 $\pm$ 0.16
	1	19819.94	19820.23 $\pm$ 0.35
	2	19837.93	19838.41 $\pm$ 0.82
	3	19855.92	19856.40 $\pm$ 0.27

Table S5. Calculated average masses and means of experimental deconvoluted masses of CCS-D2 and its fluorinated proteoforms. PTMs: post-translational modifications; $\bar{x}$: mean calculated from all experimental masses of 3FY, 4FF, 6FW and 5FW incorporations.

Protein and PTMs	No. of F	Calculated average masses	$\bar{x}$ of experimental masses
CCS-D2 with N-acetyl	0	16127.95	16127.27 $\pm$ 0.36
	1	16145.94	16145.43 $\pm$ 0.43
	2	16163.93	16163.25 $\pm$ 0.49
	3	16181.92	16181.07 $\pm$ 0.48

Table S6. Values of p_F calculated for each sample of FAA-CA2 by mass spectrometry data.

		6FW	5FW	4FF	3FY
Expr. condition	ST0h	-	-	-	0.62 ± 0.04
	ST24h	0.76 ± 0.04	0.6 ± 0.1	0.7 ± 0.1	0.56 ± 0.03
	ST16h	0.8 ± 0.1	0.65 ± 0.04	0.64 ± 0.04	0.60 ± 0.03
	ST8h	0.7 ± 0.04	0.59 ± 0.03	0.65 ± 0.06	0.58 ± 0.05
	Mix75	0.36 ± 0.01	0.32 ± 0.01	0.25 ± 0.01	0.23 ± 0.01
	Mix50	0.15 ± 0.01	0.13 ± 0.01	0.088 ± 0.003	0.07 ± 0.02

Table S7. Values of p_F calculated for each sample of FAA-CA2 by integrating the cross peaks of the 2D ^1H - ^{15}N BEST-TROSY NMR of samples.

		6FW	5FW	4FF	3FY
Expr. condition	ST0h	-	-	-	0.58 ± 0.08
	ST24h	0.71 ± 0.05	0.58 ± 0.08	0.68 ± 0.08	0.47 ± 0.04
	Mix75	0.35 ± 0.03	0.29 ± 0.05	0.18 ± 0.03	0.08 ± 0.03
	Mix50	0.19 ± 0.02	0.09 ± 0.02	0.11 ± 0.02	0.03 ± 0.02

3.2 Intracellular binding of novel fluorinated compounds to carbonic anhydrase isoforms explored by in-cell ^{19}F NMR

*Azzurra Costantino*¹ | *Letizia Barbieri*^{1,2} | *Simone Giovannuzzi*³ |
*Alessio Nocentini*³ | *Claudiu T. Supuran*³ | *Mindaugas Raitelaitis*⁴ |
*Pär Nordlund*⁴ | *Enrico Luchinat*^{1,2,5} | *Lucia Banci*^{1,2,5}

- ¹ Magnetic Resonance Center - CERM, University of Florence, via Luigi Sacconi, 6, 50019, Sesto Fiorentino, Italy.
- ² Consorzio Interuniversitario Risonanze Magnetiche di Metallo Proteine – CIRMMP, via Luigi Sacconi, 6, 50019, Sesto Fiorentino, Italy.
- ³ NEUROFARBA Department, Section of Pharmaceutical and Nutraceutical Sciences, University of Florence, via Ugo Schiff, 6, 50019, Sesto Fiorentino, Italy.
- ⁴ Department of Oncology-Pathology, Karolinska Institutet, Solnavägen 30, 171 64 Solna, Stockholm, Sweden.
- ⁵ Chemistry Department, University of Florence, Via della Lastruccia, 3, 50019, Sesto Fiorentino, Italy.

Journal of Medicinal Chemistry. 2025. Oct 27;68(21):23363–23374.

3.2.1 Summary

In this study, novel fluorinated benzenesulfonamides were synthesised and characterised using in-cell ^{19}F NMR spectroscopy. Benzenesulfonamides are a well-known class of human carbonic anhydrase (CA) inhibitors. These compounds were designed by introducing one or more CF_3 groups onto the scaffold to enable detection by ^{19}F NMR. Different cytosolic CA isoforms (CA1, CA2, CA3, CA7 and CA13) were then considered.

HEK293T cells that had been transiently transfected to express CAs were treated with the fluorinated compounds for one hour at 37 °C and 5% CO_2 , after which they were analysed using in-cell ^{19}F NMR. The ^{19}F NMR spectra clearly show that the ligands are bound to CA, thus confirming the interaction between the benzenesulfonamides and the various tested isoforms, with the exception of CA3. This protein is known not to bind to this class of inhibitor. Notably, the CA7 and CA13 isoforms, which are invisible in ^1H - ^{15}N HMQC spectra when uniformly ^{15}N -labeled, display well-resolved signals in the ^{19}F spectra of the CA-bound ligands. This result highlights the advantage of a ligand-observed approach over the traditional protein-observed one, as it enables the study of isoforms that would otherwise remain undetectable using a target-detection approach. A qualitative affinity ranking was determined for CA1, CA2 and CA13. For this purpose, the cells were treated with two ligands (at a concentration of 50 μM each) simultaneously for one hour at 37 °C and 5% CO_2 . The in-cell ^{19}F NMR spectra obtained for each ligand pair combination showed one or more distinct sets of signals attributable to the ligands bound to the protein. A qualitative comparison of the intensity of these signals in each spectrum enabled us to determine the affinities of each isoform. The cellular uptake and membrane permeabilities of these ligands were then analysed using an NMR bioreactor. Transiently transfected cells were perfused with DMEM containing the fluorinated compounds at a concentration of 5 μM . The data showed that compound 6 exhibited the highest cellular permeability, while the other compounds exhibited similar permeability. Finally, CETSA experiments were performed to confirm the interaction between the fluorinated ligands and the endogenous CA2 enzyme.

3.2.2 Attached publication

Intracellular Binding of Novel Fluorinated Compounds to Carbonic Anhydrase Isoforms Explored by In-Cell ^{19}F NMR

Azzurra Costantino, Letizia Barbieri, Simone Giovannuzzi, Alessio Nocentini, Claudiu T. Supuran, Mindaugas Raitelaitis, Pär Nordlund, Lucia Banci, and Enrico Luchinat*

Cite This: <https://doi.org/10.1021/acs.jmedchem.5c02227>

Read Online

ACCESS |

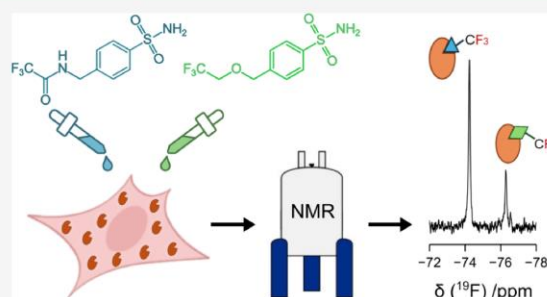
Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: One of the most important challenges of drug development is identifying whether a candidate molecule can efficiently reach and engage intracellular targets. Cell-based assays are usually employed to address this issue; nevertheless, many compounds that appear promising in vitro fail in cellular or in vivo contexts due to low intracellular activity or off-target effects. In this work, we employed ligand-based in-cell ^{19}F NMR to characterize ligand–target interactions in the cellular environment. We investigated the binding of newly synthesized fluorinated sulfonamide derivatives to various cytosolic isoforms of carbonic anhydrase in living human cells. We assessed target engagement and membrane permeability of each ligand and provided a qualitative affinity ranking of the compounds toward each isoform.

The fluorinated compounds characterized in this study are promising scaffolds for the development of isoform-selective inhibitors and may serve as chemical probes for future competitive screenings of nonfluorinated drug candidates in cells.



■ INTRODUCTION

Drug design and implementation is a complex, multistep process in which candidate molecules undergo multiple steps of screening, selection, and optimization before they can reach the clinical trial phases. Starting from target identification, numerous ligand-based screening approaches can be used to identify compounds that interact with the target of interest, starting from various types of libraries. These screenings, integrated with protein-based structural studies to characterize the binding to the target protein, reduce the set of molecules, which could allow the rational design of lead compounds with increased binding affinity and possibly selectivity to the target. An essential further step is the assessment of the compound's ability to reach intracellular targets, which is usually evaluated with ad hoc cell-based assays. Frequently, drug candidates that are effective in vitro fail on this aspect or at later stages in the cellular evaluation.^{1,2}

Nuclear magnetic resonance (NMR) spectroscopy in solution is often employed to characterize lead compounds and screen for activity against their targets in vitro. Thanks to its nondestructive nature, NMR can also investigate intermolecular interactions, such as the binding of drug candidates to their target, directly in living cells.^{3–5} The in-cell NMR approach, applied to living cells overexpressing the target of interest and exposed to the candidate drugs, enables the assessment of target engagement in living human cells.^{6,7} The combination of in-cell NMR studies with a flow system

further allows real-time monitoring of ligand binding, providing insights into the compound permeability by measuring membrane diffusion kinetics.^{8–10} In-cell NMR studies on ligand–target interactions can exploit both a target-based approach and a ligand-based approach. The target-based approach employs isotope-labeled intracellular targets and provides information on cell permeability and on the binding site location on the target protein.^{6,8,11} The ligand-based approach can provide information on cell permeability, target binding, and can also reveal off-target binding.^{8,12} Furthermore, by observing the ligand, it is possible to detect binding to target proteins difficult to observe directly by NMR due to interactions with the cellular environment. While ligand-based NMR in vitro relies on ^1H spectra, in-cell detection requires additional means for selective observation of the molecule of interest. Recently, fluorine-19 (^{19}F) NMR has shown its potential usefulness in structural biology and in drug development.^{13,14} ^{19}F is 100% abundant, has high sensitivity (83% of ^1H), and is naturally absent from biological systems. These characteristics make it a powerful tool for probing

Received: August 6, 2025

Revised: September 30, 2025

Accepted: October 20, 2025

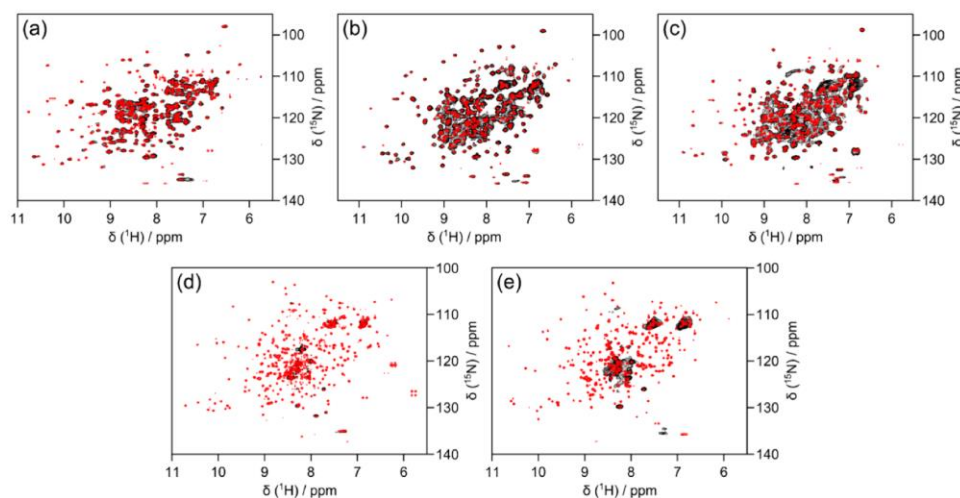


Figure 1. ^1H – ^{15}N NMR spectra of CA isoforms. 2D ^1H – ^{15}N BEST-TROSY in-cell NMR spectra (black) and cell lysate NMR spectra (red) of CA1 (a), CA2 (b), CA3 (c), CA7 (d), and CA13 (e).

molecular interactions in living cells, while greatly reducing background interference.^{14,15} Indeed, in-cell ^{19}F NMR has been applied to detect both intracellular proteins,^{16–20} nucleic acids,²¹ and small molecules carrying one or more fluorine atoms.^{10,22,23}

In this study, we employed in-cell ^{19}F NMR spectroscopy to characterize novel sulfonamide-based inhibitors of carbonic anhydrase (CA). CAs are zinc-containing enzymes found across all domains of life that catalyze the reversible hydration of carbon dioxide to bicarbonate. This reaction plays a crucial role in several physiological processes, including pH regulation, CO_2 transport, and respiration. In humans, 15 α -class CA isoforms are known, each with distinct tissue distribution and subcellular localization. Cytosolic isoforms such as CA1, CA2, and CA3 are highly expressed in erythrocytes, muscle, and various organs, while mitochondrial isoform CA5 is localized in the kidney, liver, and pancreas, where it contributes to biosynthetic pathways. Membrane-bound isoforms such as CA9 and CA12, frequently overexpressed in tumors, are found in the gastrointestinal tract and reproductive tissues, whereas the secreted isoform CA6 is present in saliva. Two additional cytosolic isoforms are expressed in the central nervous system (CA7) and in the small intestine, prostate, and colon (CA13).^{24,25} Their functional diversity and isoform-specific tissue distribution make CAs valuable therapeutic targets. Several CA inhibitors are already used clinically, while others are currently in development for several therapeutic applications, where achieving isoform-selective inhibition is critical.^{24–26} The compounds reported here were specifically designed for in-cell detection; thereby, they incorporate one or more trifluoromethyl groups ($-\text{CF}_3$) to enable detection with high sensitivity via ^{19}F NMR. The binding of these compounds to a panel of cytosolic CA isoforms (CA1, CA2, CA3, CA7, and CA13) was assayed in human cells by in-cell ^{19}F NMR. These cytosolic CA isoforms share a highly conserved catalytic domain, with the Zn^{2+} coordinated by His94, His96, and His119, and a key proton shuttle involving His64. CA2, the most ubiquitous and extensively studied isoform, along with CA1, CA7, and CA13, displays high catalytic activity. In contrast, CA3, primarily expressed in skeletal muscle and its

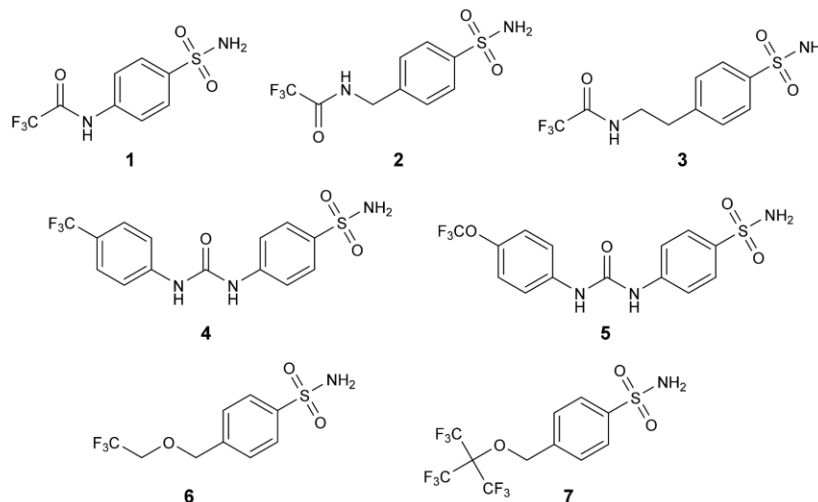
precursors, exhibits a significantly reduced enzymatic efficiency due to a lysine substitution at the His64 position.^{27,28} To explore selective binding across these highly similar cytosolic isoforms, we established a qualitative affinity ranking of the compounds for each CA isoform and assessed their cellular uptake kinetics. Additionally, we applied the cellular thermal shift assay (CETSA) to independently confirm the target engagement of the compounds to endogenously expressed CA2. These compounds could be used as scaffolds for the development of novel isoform-selective CA inhibitors. Furthermore, one of the compounds, which shows markedly faster membrane diffusion kinetics, will be a promising chemical probe for screening nonfluorinated CA inhibitors via competition binding in-cell ^{19}F NMR.¹⁰

RESULTS AND DISCUSSION

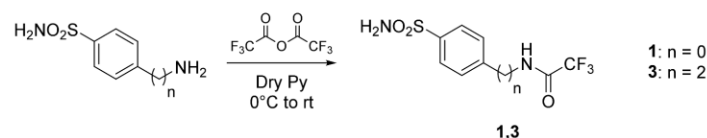
The expression level of five different human cytosolic CA isoforms—CA1, CA2, CA3, CA7, and CA13—in transiently transfected cultured human cells was evaluated. CA1, CA2, and CA3 were significantly overexpressed, reaching average intracellular concentrations between 130 and 170 μM , whereas CA7 and CA13 were expressed at lower levels, with concentrations of ~ 50 and ~ 80 μM , respectively, but still within the sensitivity range of in-cell NMR experiments (Figure S1). In principle, such target concentrations are compatible with target-based in-cell NMR ligand binding studies.⁶ However, slow tumbling caused by interactions with cellular components may hamper NMR detection of the target proteins. The intracellular behavior of each isoform was therefore evaluated by in-cell ^1H – ^{15}N NMR. CA1, CA2, and CA3 resulted in well-resolved signals, whereas no signals were detected for CA7 and CA13, except for those arising from the highly dynamic FLAG tag (Figure 1). All isoforms were clearly detected in the NMR spectra of the corresponding cell lysates, confirming that in intact cells, CA7 and CA13 experience line broadening beyond detection due to intracellular interactions.

In contrast to target-based approaches based on the detection of the protein backbone (e.g., by ^1H – ^{15}N 2D spectra), ligand-observed in-cell ^{19}F NMR experiments are less

Chart 1. Structures of the Fluorinated Benzenesulfonamide Derivatives Tested in This Study



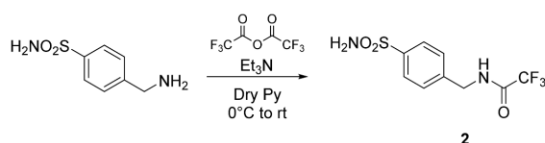
Scheme 1. Synthesis of Compounds 1 and 3



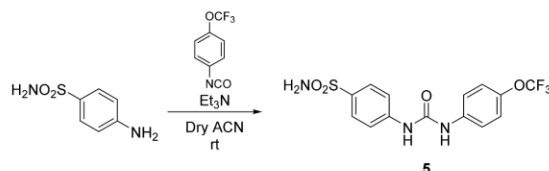
crowded and less affected by the tumbling rate of the complex, allowing ligand screening studies also for slow-tumbling targets.¹⁰ Furthermore, the binding of multiple ligands can be investigated simultaneously thanks to the high sensitivity of the ¹⁹F chemical shift to the local environment.

For this study, a novel series of fluorinated benzenesulfonamides (compounds 1–7, Chart 1) was designed as potential ligands for different CA isoforms. In these compounds, a trifluoromethyl group (–CF₃) was incorporated onto the substituents of the benzenesulfonamide moiety. Sulfanilamide and SLC-0111, two established and potent carbonic anhydrase inhibitors (CAIs), were the lead structures for the design of the fluorinated ligands. Compound 1 is a derivative of sulfanilamide previously reported as a CAI,²⁹ while compounds 2 and 3 are congeners of compound 1, distinguished by the presence of extended linkers between the CAI core and the –CF₃ group (Schemes 1 and 2). Compounds 4 (reported previously³⁰) and 5 (Scheme 3) are analogues of SLC-0111, incorporating a –CF₃ or –OCF₃ group in place of a single fluorine atom. Compound 6 represents an analogue of compound 2 that features an ether linker, as opposed to the amide found in its predecessor. Compound 7 is structurally related to compound 6 and is characterized by the inclusion of

Scheme 2. Synthesis of Compound 2



Scheme 3. Synthesis of Compound 5



additional –CF₃ groups on its terminal segment (Scheme 4). This focused structural diversification aimed to explore the impact of linker type and degree of fluorination on carbonic anhydrase inhibition. Compared to other groups containing nonequivalent fluorine atoms, the –CF₃ provides a 3-fold increase in NMR signal intensity and favorable relaxation properties. This property was maximized in compound 7, which contains three equivalent CF₃ groups.

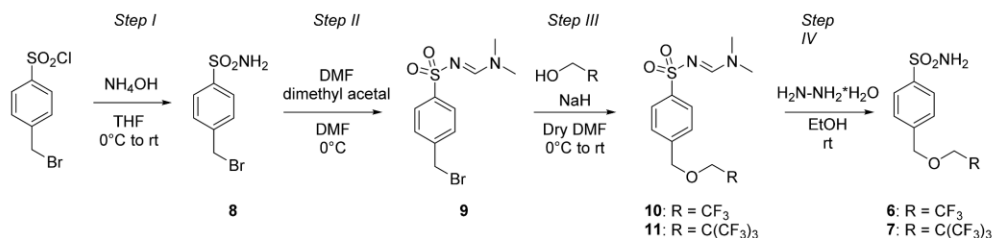
The inhibition constants (*K_s*) of the compounds against each cytosolic CA isoform were evaluated *in vitro* through a stopped-flow assay (Table 1) and compared to that of the reference compound acetazolamide (AAZ), a clinically established CA inhibitor. All the compounds exhibited inhibitory activity against CA1, CA2, CA7, and CA13, with comparable *K_i* values, while none of the compounds inhibited the CA3 isoform (*K_i* > 100 μM), including AAZ. This is consistent with the known resistance of CA3 to benzenesulfonamide-based ligands, which arises from the substitution of a leucine with phenylalanine in the active site that makes the CA3 ligand binding pocket inaccessible to aromatic sulfonamides.³¹

The stability of the fluorinated ligands in the cell culture medium was assessed through ¹⁹F NMR (Figure S2).

C

<https://doi.org/10.1021/acs.jmedchem.5c02227>
J. Med. Chem. XXXX, XXX, XXX–XXX

Scheme 4. Synthesis of Compounds 6 and 7

Table 1. In Vitro Inhibition Data (K_i Values in nM) of Human Cytosolic CA Isoforms by Compounds 1–7 and the Reference Drug Acetazolamide (AAZ)

compound	K_i (nM) ^a				
	CA1	CA2	CA3	CA7	CA13
1	6550	145	>100 μ M	178	90.2
2	1260	90.9	>100 μ M	107	146
3	988	121	>100 μ M	76.9	55.8
4	9.7 ^b	1150 ^b	>100 μ M	552	153
5	142	203	>100 μ M	289	82.0
6	49.3	78.0	>100 μ M	101	230
7	81.6	57.6	>100 μ M	69.3	124
AAZ	250.0	12.0	>100 μ M	5.7	17.0

^aMean from three different assays, by a stopped-flow technique (errors were in the range of ± 5 –10% of the reported values). ^bTaken from Pacchiano et al.³⁰

Compounds 2–7 were stable for one hour at 37 °C, while compound 1 showed the appearance of an additional signal over time, with the simultaneous decrease of the initial signal, suggesting a spontaneous hydrolysis of the trifluoroacetamide moiety. The hydrolysis of compound 1 in water and the consequent formation of sulfanilamide were confirmed by LC-MS (Figures S3–S5).

The binding of the compounds to each cytosolic CA isoform was then analyzed by in-cell ¹⁹F NMR spectroscopy. Incubating the cells for 1 h with an excess of each compound resulted in the formation of the intracellular complex with CA1, CA2, CA7, and CA13, giving rise to a signal characteristic for each isoform (Figure 2). These signals were absent in control samples of nonoverexpressing cells treated with the same compounds, confirming that they arise from the binding to the targets. Notably, compound 1 also formed a stable complex with intracellular CA isoforms, despite being gradually hydrolyzed in the external medium during the incubation phase. As expected, no complex formation occurred in the cells expressing CA3, consistent with the lack of inhibition observed through the stopped-flow assay. An additional signal, weaker but sharper, was also observed in some of the spectra. The same signal is also present in the ¹⁹F NMR spectra of the supernatant fractions from each cell sample, and its chemical shift corresponds to that of the free compound, thus suggesting that a small fraction of the compound is released from the cells during the in-cell NMR analysis (Figure S6). The ¹⁹F NMR spectra of the cell lysates showed clear signals at the same chemical shifts observed in the intact cells, indicating that the complexes formed in the cells were preserved after cell lysis (Figure S7).

While the spectra of control cells treated with compounds 1, 2, 3, 4, and 6 did not contain any signal (Figure 2a–d,f),

compounds 5 and 7 gave rise to broad signals, which were also present in the CA-expressing cells, indicating that these compounds bind nonspecifically to other cellular components (Figure 2e, g and Figure S8). This suggests that compounds 5 and 7 may not be suitable for therapeutic targeting of CA isoforms due to their lack of specificity.

Based on the nano- to micromolar K_i values determined in vitro, all the compounds are expected to saturate the intracellular target proteins. In these experimental conditions, the signal intensities in the NMR spectra are determined by the level of target protein in each sample and do not report on the intracellular binding affinity of the compounds. Therefore, to determine the relative binding affinities of the fluorinated compounds for each cytosolic CA isoform, we performed intracellular competition binding experiments. To this aim, cells expressing CA1, CA2, and CA13 were treated for 1 h with two compounds simultaneously, followed by in-cell ¹⁹F NMR analysis as above. CA3 and CA7 were excluded from this analysis due to a lack of binding and low expression levels, respectively. Compounds 2, 4, 5, 6, and 7 were selected, while compound 1 was excluded due to its low stability in the culture medium, and compound 3 was omitted due to its structural similarity with compound 2. The resulting spectra showed two distinct signals with different intensities, corresponding to the two compounds in complex with intracellular CA (Figure 3). Assuming that the incubation phase is sufficiently long to allow both compounds to reach the intracellular target, we reached a dynamic equilibrium such that the ratio of the signal intensities (normalized by the number of equivalent fluorine atoms) is directly proportional to the ratio of the binding affinities of the two compounds. This allowed us to establish a qualitative affinity ranking for each CA isoform. Notably, the affinity ranking profiles varied across the CA isoforms. Specifically, compounds 2 and 6 had lower affinity for CA2 and CA13 with respect to 5 and 4, whereas the opposite was observed for CA1, while compound 7 showed the highest affinity for all three CA isoforms. Strikingly, the in-cell rankings did not correlate with the K_i s measured in vitro. For example, compound 4 ranked the highest in vitro and was the second weakest in cells against CA1, while against CA2, it had the opposite behavior. These results suggest that the cellular context affects each ligand-target interaction differently in complex ways.

When recorded in real time using an NMR flow bioreactor,³² in-cell ¹⁹F NMR allows evaluating the cell permeability of fluorinated compounds. To this aim, cells overexpressing CA2 were encapsulated in agarose gel and perfused in the NMR bioreactor with a constant flow of medium containing 5 μ M each ligand, while ¹⁹F NMR spectra were continuously recorded to measure the rate of intracellular complex formation. As the cellular uptake is the slow step in

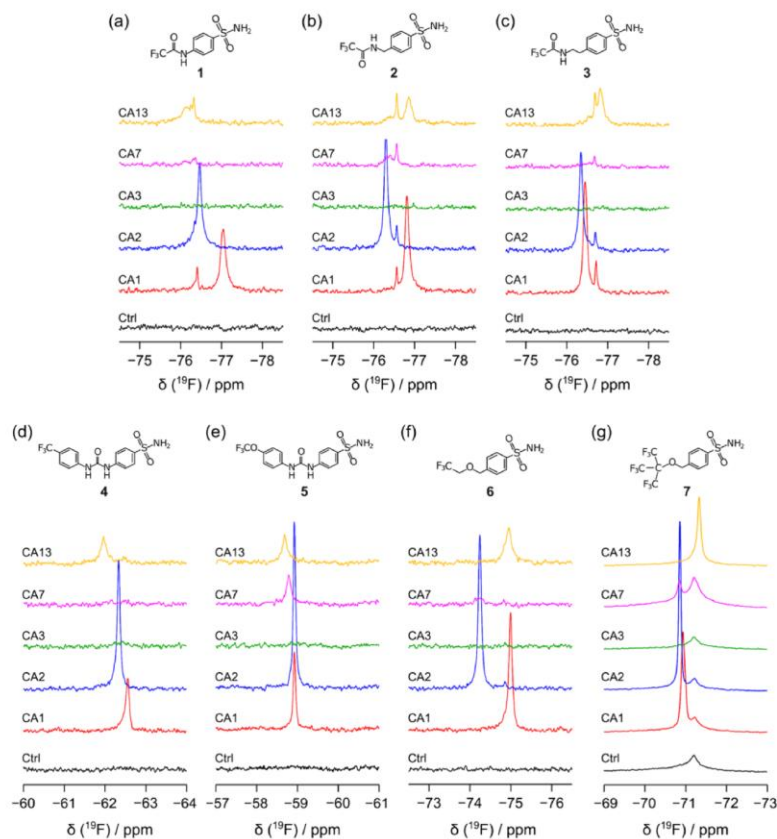


Figure 2. Binding of the fluorinated compounds to intracellular CA isoforms. 1D in-cell ^{19}F NMR spectra of cells expressing different CA isoforms treated for 1 h with 100 μM of compound 1 (a), 2 (b), 3 (c), 4 (d), 5 (e), 6 (f), and 7 (g). The spectra are color-coded based on the CA isoform: CA1 (red), CA2 (blue), CA3 (green), CA7 (purple), and CA13 (yellow). The control spectra (black) of cells transfected with an empty vector are also shown.

the ligand binding kinetics, the rate of complex formation is dominated by the membrane permeability, rather than the association rate constant to the intracellular target.⁶ Linear fitting of the fluorine signal intensity over time revealed that compound 6 had a markedly faster cellular uptake, whereas the other compounds had similar uptake rates (Figure 4).

Finally, CETSA experiments were performed to independently confirm the interaction between the fluorinated compounds and the endogenous CA2. CETSA is based on the principle that ligand binding stabilizes proteins against thermal denaturation in living cells.^{33,34} The assay was performed on compounds 2 and 4–7 ranked above (Figure 5). In cells treated with compounds 4–7, the melting curves of endogenous CA2 showed melting temperatures (T_m) ranging from ~ 53 to ~ 63 $^{\circ}\text{C}$, which are noticeably higher than the T_m of unbound CA2 (50–52 $^{\circ}\text{C}$). These findings are consistent with the in-cell ^{19}F NMR spectra, further confirming that binding also occurred with endogenous protein levels. Interestingly, even though a protein-bound signal appeared in the in-cell ^{19}F NMR spectrum, the melting curve for cells treated with compound 2 differed only slightly from the control (Figure 5a). This suggests that although compound 2 binds to CA2, as clearly seen by NMR, the interaction does not increase the thermal stability of the protein to a large extent. As

our ligand-based NMR approach gives a strong signal, this highlights the complementarity of CETSA and in-cell NMR. While in the case of CETSA the protein reports on the interaction by being stabilized by the ligand, NMR provides a direct indication of the interaction regardless of the thermal stability of the resulting complex.

CONCLUSIONS

In this work, we characterized the binding of newly synthesized fluorinated benzenesulfonamide-based inhibitors targeting cytosolic CA isoforms using in-cell ^{19}F NMR spectroscopy. By employing this ligand-based approach, protein–ligand interactions were observed directly in the cellular environment, providing a physiologically relevant binding assay compared to classical in vitro binding studies. Among the five CA isoforms analyzed, CA1, CA2, CA7, and CA13 clearly exhibited binding interactions with compounds, whereas CA3 displayed no interaction, as expected due to the increased steric hindrance in its active site. In the intracellular affinity ranking, compound 7 consistently showed the highest affinity among the isoforms; however, along with compound 5, it also displayed additional NMR signals likely arising from nonspecific interactions, which may be detrimental for future therapeutic applications and should be investigated further. The in-cell affinity rankings

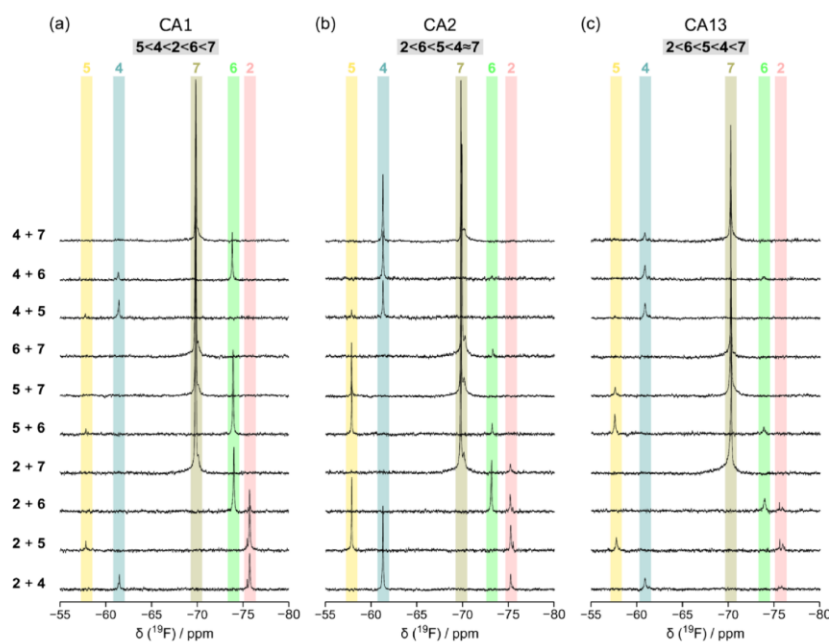


Figure 3. Competition binding and affinity ranking of the compounds to intracellular CA isoforms. 1D in-cell ^{19}F NMR spectra of cells expressing different CA isoforms: CA1 (a), CA2 (b), and CA13 (c), treated with different pairs of fluorinated compounds (indicated on the left of each spectrum), 50 μM each for 1 h. The qualitative affinity rankings are reported at the top of each panel.

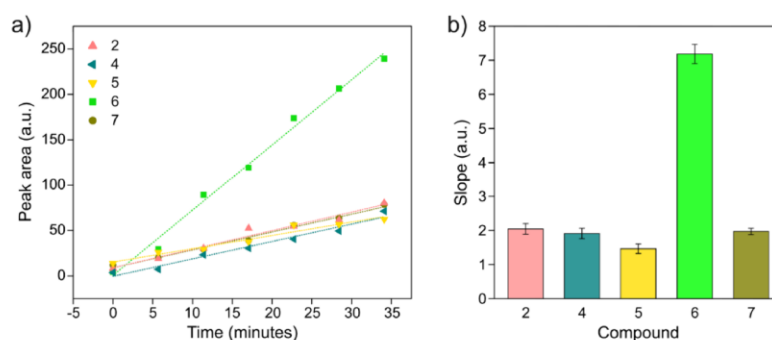


Figure 4. Uptake kinetics of fluorinated compounds in HEK293T cells expressing CA2. (a) Time-dependent ^{19}F NMR signal integrals for each compound, acquired in real time by using the NMR bioreactor. (b) Histogram of the relative slopes obtained from fitting the fluorine signal integrals as a function of time, indicating the comparative uptake rates of the compounds.

were significantly different from those based on enzyme inhibition *in vitro*, further highlighting the importance of performing early-stage screenings in a cellular model, as doing so will affect the choice of promising lead compounds, with important consequences for any further drug development. Uptake kinetics analysis by real-time NMR revealed that compound 6 has higher membrane permeability compared to the others, making it a promising hit compound for future drug development, and the best candidate as a “spy” ligand for affinity ranking studies of other hit/lead compounds based on competition binding and ^{19}F NMR detection. CETSA assays revealed ligand-induced thermal stabilization of native CA2, which further validates the binding data acquired through in-cell NMR. Although compound 2, which showed binding in NMR spectra, only showed a weak thermal stabilization of the

protein, compounds 4–7 increased the melting temperature of CA2, confirming binding to endogenous protein levels. This underscores the complementarity between CETSA and NMR: the former is more sensitive and can also be applied to the endogenous protein, while the latter provides direct evidence of the interaction regardless of thermal stability and provides further relevant data from affinity rankings and uptake kinetics. The fluorinated ligands developed in this work represent promising scaffolds for further optimization and development of specific CA inhibitors. Moreover, thanks to their fluorine content, some of these compounds may also be used as probes for competitive screening of nonfluorinated candidates in cellular environments.

Finally, from a methodological point of view, this approach overcomes some of the limitations of target-based in-cell NMR

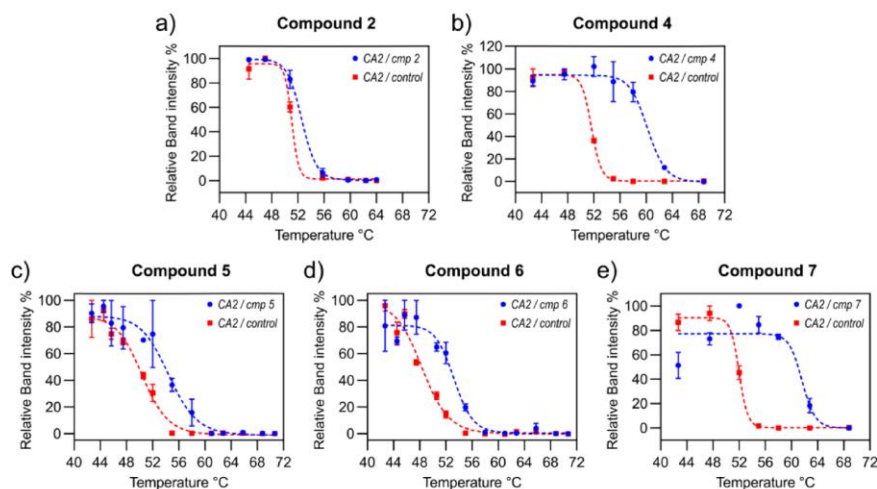


Figure 5. CETSA thermal stability curves of endogenous CA2 in cells treated with fluorinated compounds. CETSA profiles show the thermal stability of CA2 in cells treated with fluorinated compounds (blue curves) compared with DMSO-treated controls (red curves). Band intensities from Western blots using an anti-CA2 antibody were quantified and plotted against the temperature at which the cells were heated. Each panel corresponds to a different fluorinated compound used for treatment: (a) compound 2, (b) compound 4, (c) compound 5, (d) compound 6, and (e) compound 7.

methods, in which some targets may not be detected due to interactions with cellular components, causing extensive broadening of the protein NMR signals. The strategy employed here allows a relatively fast ranking of the compounds based on their binding affinity in cells. However, this ranking is to be taken qualitatively, as several factors can affect the intracellular binding equilibrium, such as the release of ligand from the cells during the analysis or additional nonspecific interactions with cellular components. Performing the same experiments in an NMR bioreactor would allow higher control of the ligand concentrations and of the sample conditions, enabling a quantitative affinity ranking purely based on the intracellular K_d s. In general, these results highlight the potential of in-cell ^{19}F NMR as a powerful and versatile technique for directly investigating ligand–target interactions in living cells.

EXPERIMENTAL SECTION

Gene Constructs. Codon-optimized sequences encoding different cytosolic carbonic anhydrase isoforms (CA1, NP_001729; CA2, NP_000058.1; CA3, NP_005172.1; CA7, NP_001014435.1; CA13, NP_940986.1), with the secretion sequence removed as previously described,⁶ were obtained by gene synthesis and cloned into the pHLsec mammalian expression vector³⁵ (Twist Bioscience). A FLAG tag (GSSDYKDDDDK) was added to the C-terminus of each CA isoform.

Protein Expression in Human Cells. The HEK293T (ATCC CRL-3216) cell line was used for transient transfection, following a protocol described previously.³⁶ Cells were cultured in a T75 flask using high-glucose Dulbecco's modified Eagle medium (DMEM, Gibco), supplemented with 1% penicillin–streptomycin (Gibco) and 10% fetal bovine serum (Gibco), and maintained at 37 °C in a humidified incubator with 5% CO_2 . The cells were transiently transfected using polyethylenimine (PEI, Sigma-Aldrich) at a 1:2 ratio (25 μg DNA:50 μg PEI).³⁶ After transfection, protein expression was carried out for 48 h in DMEM supplemented with 100 $\mu\text{g}/\text{mL}$ of penicillin–streptomycin (Life Technologies), 2% (v/v) of fetal bovine serum (FBS, Life Technologies), and 10 μM of ZnSO_4 .

Protein Quantification via SDS-PAGE. The expression levels of cytosolic CA isoforms were assessed by using SDS-polyacrylamide gel electrophoresis (SDS-PAGE). Transfected cells from T25 flasks were lysed in 150 μL of phosphate-buffered saline (PBS), and the soluble lysates were collected. As a negative control, a cell lysate was prepared from cells transfected with the same procedure but with a plasmid lacking the gene of interest. Each cell lysate sample was loaded at three different dilutions for quantification. To generate a linear regression standard curve for quantification, purified CA2 protein was loaded at different concentrations (20, 15, 10, 5, and 2.5 μM). Both samples and standards were mixed 2:1 with SDS loading buffer [300 mM Tris-HCl (pH 6.8), 6% (w/v) SDS, 30% (v/v) glycerol, 0.3% (w/v) bromophenol blue, and 150 mM dithiothreitol]. Proteins were resolved on precast gels (Bio-Rad) at 200 V for 30 min and stained with Coomassie dye (ProBlue Safe Stain, Giotto Biotech). Gels were imaged by using a ChemiDoc XRS system (Bio-Rad). Image analysis was performed with Image Lab software, version 5.0 (Bio-Rad). For quantification, the background intensity, calculated from the average of the areas without bands, was subtracted from each standard band intensity. Similarly, the intensity of each CA band in the transfected samples was corrected by subtracting the corresponding band intensity from the negative control at the same dilution. A calibration curve was generated in Excel (Microsoft) by plotting the intensities of the standard bands against their known concentrations. The protein concentrations of the lysate samples were then interpolated from this calibration curve based on their corrected band intensities. The final concentration of each sample was calculated as the average of the concentrations obtained from the different sample dilutions that were run on the same gel.

In-Cell and Cell Lysate ^1H – ^{15}N NMR Spectroscopy. After 48 h of expression in a ^{15}N -labeled medium, cells were harvested, resuspended in 185 μL of NMR buffer (DMEM, 70 mM HEPES, and 20% (v/v) D_2O), and transferred to a 3 mm Shigemitsu, following a previously established protocol.³⁶ Before NMR acquisition, the suspension was gently centrifuged to form a soft pellet.

For the analysis of cell lysates, cells were recollected and lysed after the in-cell NMR experiments. Pellets were separated from the NMR buffer by centrifugation and lysed in 150 μL of phosphate-buffered saline (PBS, Life Technologies) by the freeze–thaw method. Soluble lysates were obtained by centrifugation and analyzed by ^1H – ^{15}N NMR spectroscopy with 10% (v/v) of D_2O .

^1H – ^{15}N NMR spectra on cells and cell lysates were recorded at 310 K at a 950 MHz Bruker Avance NEO spectrometer equipped with a 5 mm TCI Cryoprobe. 2D ^1H – ^{15}N SOFAST-HMQC spectra (Bruker pulse sequence sfhmqc3gpph) were recorded on cell samples and on the corresponding lysates with frequency offsets of 4.7 ppm (^1H) and 118 ppm (^{15}N), spectral windows of 16 ppm (^1H) and 50 ppm (^{15}N), acquisition times of 67.4 ms (^1H) and 13.3 ms (^{15}N), and an interscan delay of 0.3 s using the shaped pulses Pc9_4. 90.1000 and Reburp.1000 for selective ^1H inversion and refocusing, respectively.

The excitation width and offset were set to 6 and 8.7 ppm, respectively. Shaped pulse lengths and power levels were automatically calculated (-DCALC_SP option in the pulse sequence). 64 initial scans and 128 increments were employed, resulting in a total experimental time of 58 min. To remove the background signals arising from the incorporation of ^{15}N in other cellular components, each 2D NMR spectrum recorded on cells and lysates was further processed in Topspin (Bruker) by subtracting a spectrum recorded using identical parameters on cells transfected with an empty vector and on the corresponding lysate, as previously described.³⁶

Synthesis and Characterization of Fluorinated Compounds.

Anhydrous solvents and all reagents were purchased from Merck, Fluorochem, and TCI. All reactions involving air- or moisture-sensitive compounds were performed under a nitrogen atmosphere using dried glassware and syringe techniques to transfer solutions. ^1H NMR, ^{13}C NMR, and ^{19}F NMR spectra were recorded using a Bruker Avance III 400 MHz spectrometer in DMSO- d_6 . Chemical shifts are reported in parts per million (ppm), and the coupling constants (J) are expressed in Hertz (Hz). Splitting patterns are designated as follows: s, singlet; d, doublet; t, triplet; q, quadruplet; m, multiplet; and dd, doublet of doublets. The assignment of exchangeable protons was confirmed by the addition of D_2O . Analytical thin-layer chromatography (TLC) was carried out on Sigma-Aldrich silica gel F-254 plates. Flash chromatography purifications were performed on Sigma-Aldrich Silica gel 60 (230–400 mesh ASTM) as the stationary phase, and ethyl acetate/*n*-hexane or MeOH/DCM were used as eluents. All spectra were in accord with the assigned structures. All final compounds were >95% pure by HPLC. High-resolution mass spectrometry (HRMS) analysis was performed by using an LTQ Orbitrap XL hybrid mass spectrometer (HRMS system) equipped with an Ion Max electrospray ionization (ESI) source operating in the positive ion mode at 60,000 fwhm mass resolution (referred at 400 m/z). ESI source and ion optics parameters were optimized on the protonated molecule of each analyte as follows: ESI voltage was 4 kV, capillary voltage was 13 V, tube lens voltage was 60 V, and the capillary temperature was 290 °C. Sheath and auxiliary gas were nitrogen at 15 and 8 arbitrary units, respectively. The stock solution of each analyte was prepared at 1 mg/mL in methanol containing 10% DMSO, and then, by a proper dilution, the respective working solution at 20 mg/mL in acetonitrile (ACN)/water 1:1 containing 0.1% formic acid (FA) was prepared. The accurate m/z ratio measurement was carried out by introducing into the HRMS system via a syringe pump at 10 $\mu\text{L}/\text{min}$ the working solution of each analyte, monitoring the m/z range between 160 and 1000. The purity of target compounds was assessed by HPLC using an Agilent 1200 Series gradient HPLC system with a Luna PFP column (3 μm , 2 mm $\times$ 30 mm) at a flow rate of 0.25 mL/min and a linear gradient of the mobile phase, i.e., 10 mM FA and 5 mM ammonium formate in ultrapure water solution (solvent A) and 10 mM FA and 5 mM ammonium formate in ACN (solvent B). Representative HPLC traces are shown in Figure S9.

The synthesis of 4-(3-(4-(trifluoromethyl)phenyl)ureido)benzenesulfonamide (4) was carried out as described by Pacchiano et al. (referred to as compound 11).³⁰

Synthesis of Compounds 2,2,2-Trifluoro-*N*-(4-sulfamoylphenyl)acetamide (1) and 2,2,2-Trifluoro-*N*-(4-sulfamoylphenethyl)acetamide (3) (Scheme 1). Compounds 1 and 3 were synthesized using a similar synthetic approach to that employed previously for compound 1 by Scozzafava et al.²⁹ Trifluoroacetic anhydride (1.2 equiv) was carefully added at 0 °C

to a solution of sulfanilamide or 4-sulfamoylphenethylamine (0.2 g, 1 equiv) in anhydrous pyridine (5 mL) under a nitrogen atmosphere. The reaction mixture was warmed at rt and stirred for 0.5 h. After monitoring by TLC (EtOAc/hexane), slush and HCl 12 M were added until pH = 4, and the resulting solution was extracted in EtOAc (25 mL $\times$ 3). The collected organic phases were dried with Na_2SO_4 , filtered, and evaporated to yield a yellow oil. It was purified by a silica gel chromatography column (EtOAc/hexane), achieving 1 and 3 as white powders.

2,2,2-Trifluoro-*N*-(4-sulfamoylphenyl)acetamide (1). Yield 45%; silica gel TLC R_f 0.54 (MeOH/DCM 6% v/v); δ_{H} (400 MHz, DMSO- d_6): 11.59 (s, 1H, exchange with D_2O , CONH), 7.89 (s, 4H, 4 $\times$ Ar-H), 7.40 (s, 2H, exchange with D_2O , SO_2NH_2); δ_{C} (100 MHz, DMSO- d_6): 156.4 ($q^2J_{\text{C-F}} = 41.9$ Hz), 141.9, 140.2, 127.8, 122.0, 120.9 ($q^1J_{\text{C-F}} = 287.2$ Hz); δ_{F} (376 MHz, DMSO- d_6): -73.93; HRMS (m/z): calcd for $\text{C}_8\text{H}_7\text{F}_3\text{N}_2\text{O}_3\text{S}$ ($[\text{M} - \text{H}]^+$), 268.2102, found: 268.2109.

2,2,2-Trifluoro-*N*-(4-sulfamoylphenethyl)acetamide (3). Yield 45%; silica gel TLC R_f 0.54 (MeOH/DCM 6% v/v); δ_{H} (400 MHz, DMSO- d_6): 9.55 (bs, 1H, exchange with D_2O , CONH), 7.79 (d, $J = 7.9$ Hz, 2H, 2 $\times$ Ar-H), 7.44 (d, $J = 7.9$ Hz, 2H, 2 $\times$ Ar-H), 7.30 (bs, 2H, exchange with D_2O , SO_2NH_2), 3.94 (t, $J = 6.1$ Hz, 2H, CH_2), 2.92 (t, $J = 6.1$ Hz, 2H, CH_2); δ_{C} (100 MHz, DMSO- d_6): 157.8 ($q^2J_{\text{C-F}} = 39.9$ Hz), 143.8, 143.0, 130.1, 126.7, 121.3 ($q^1J_{\text{C-F}} = 289.3$ Hz), 34.8; δ_{F} (376 MHz, DMSO- d_6): -74.38; HRMS (m/z): calcd for $\text{C}_{10}\text{H}_{11}\text{F}_3\text{N}_2\text{O}_3\text{S}$ ($[\text{M} - \text{H}]^+$), 296.2642, found: 296.2630.

Synthesis of 2,2,2-Trifluoro-*N*-(4-sulfamoylbenzyl)-acetamide (2) (Scheme 2). Et₃N (1.2 equiv) and trifluoroacetic anhydride (1.2 equiv) were carefully added at 0 °C to a solution of 4-sulfamoylbenzylamine (0.2 g, 1 equiv) in anhydrous pyridine (5 mL) under a nitrogen atmosphere. The reaction mixture was warmed at rt and stirred for 0.5 h. After monitoring by TLC (EtOAc/hexane), slush and HCl 12 M were added until pH = 4, and the resulting solution was extracted in EtOAc (25 mL $\times$ 3). The collected organic phases were dried with Na_2SO_4 , filtered, and evaporated to yield a yellow oil. It was purified by a silica gel chromatography column (EtOAc/hexane), achieving 2 as a white powder.

Yield 45%; silica gel TLC R_f 0.54 (MeOH/DCM 6% v/v); δ_{H} (400 MHz, DMSO- d_6): 10.12 (s, 1H, exchange with D_2O , CONH), 7.85 (d, $J = 7.1$ Hz, 2H, 2 $\times$ Ar-H), 7.48 (d, $J = 7.1$ Hz, 2H, 2 $\times$ Ar-H), 7.38 (s, 2H, exchange with D_2O , SO_2NH_2), 4.50 (s, 2H, CH_2); δ_{C} (100 MHz, DMSO- d_6): 158.1 ($q^2J_{\text{C-F}} = 42.9$ Hz), 144.2, 142.4, 128.8, 126.9, 121.3 ($q^1J_{\text{C-F}} = 287.1$ Hz), 43.3; δ_{F} (376 MHz, DMSO- d_6): -74.27; HRMS (m/z): calcd for $\text{C}_9\text{H}_8\text{F}_3\text{N}_2\text{O}_3\text{S}$ ($[\text{M} - \text{H}]^+$), 282.2372, found: 282.2380.

Synthesis of 4-(3-(4-(Trifluoromethoxy)phenyl)ureido)benzenesulfonamide (5) (Scheme 3). Sulfanilamide (0.2 g, 1 equiv) and DIPEA (0.05 equiv) were added at rt to a solution of 4-trifluoromethoxyphenyl isocyanate (1.1 equiv) in anhydrous ACN under a nitrogen atmosphere. The reaction mixture was then stirred on. After monitoring by TLC (MeOH/DCM), the resulting precipitate was filtered and washed with cold ACN, achieving 5 as a white powder.

Yield 45%; silica gel TLC R_f 0.54 (MeOH/DCM 6% v/v); δ_{H} (400 MHz, DMSO- d_6): 9.14 (s, 1H, exchange with D_2O , NHCONH), 9.03 (s, 1H, exchange with D_2O , NHCONH), 7.77 (d, $J = 9.1$ Hz, 2H, 2 $\times$ Ar-H), 7.64 (d, $J = 8.7$ Hz, 2H, 2 $\times$ Ar-H), 7.60 (d, $J = 9.1$ Hz, 2H, 2 $\times$ Ar-H), 7.34 (d, $J = 8.7$ Hz, 2H, 2 $\times$ Ar-H), 7.24 (s, 2H, exchange with D_2O , SO_2NH_2); δ_{C} (100 MHz, DMSO- d_6): 153.3, 143.9, 143.7, 139.6, 138.1, 127.8, 125.1 ($q^1J_{\text{C-F}} = 255.8$ Hz), 122.7, 120.7, 118.6; δ_{F} (376 MHz, DMSO- d_6): -57.07; HRMS (m/z): calcd for $\text{C}_{14}\text{H}_{12}\text{F}_3\text{N}_3\text{O}_4\text{S}$ ($[\text{M} - \text{H}]^+$), 375.3222, found: 375.3229.

Synthesis of 4-((2,2,2-Trifluoroethoxy)methyl)-benzenesulfonamide (6) and 4-((3,3,3-Trifluoro-2,2-bis(trifluoromethyl)propoxy)methyl)-benzenesulfonamide (7) (Scheme 4). *Step 1.* NH_4OH 28–30% (2 equiv) was slowly added at 0 °C to a solution of 4-bromomethylbenzenesulfonyl chloride (0.3 g, 1 equiv) in THF, and then the reaction mixture was warmed at rt and stirred for 2 h. After monitoring by TLC (EtOAc/hexane), the solvent was evaporated, and the resulting crude product was

suspended in H₂O. HCl (12M) was added until pH = 4, and it was filtered and washed with H₂O and Et₂O, obtaining **8** as a white powder. It was used in the next step without further purification.

Yield 95%; silica gel TLC *R_f* 0.46 (EtOAc/hexane 50% v/v); δ_{H} (400 MHz, DMSO-*d*₆): 7.84 (d, *J* = 8.3 Hz, 2H, 2 × Ar-H), 7.67 (d, *J* = 8.3 Hz, 2H, 2 × Ar-H), 7.42 (s, 2H, exchange with D₂O, SO₂NH₂), 4.80 (s, 2H, CH₂); δ_{C} (100 MHz, DMSO-*d*₆): 144.7, 142.8, 130.7, 127.0, 33.9.³⁷

Step II. DMF dimethyl acetal (1.2 equiv) was carefully added at 0 °C to a solution of **8** (0.3 g, 1 equiv) in DMF (1 mL), and then the reaction mixture was warmed to rt and stirred for 0.25 h. After monitoring by TLC (EtOAc/hexane), slush was added, and the resulting suspension was filtered, achieving **9** as a white powder. It was used in the next step without further purification.

Yield 78%; silica gel TLC *R_f* 0.18 (EtOAc/hexane 50% v/v); δ_{H} (400 MHz, DMSO-*d*₆): 8.26 (s, 1H, CH), 7.80 (d, *J* = 8.3 Hz, 2H, 2 × Ar-H), 7.63 (d, *J* = 8.3 Hz, 2H, 2 × Ar-H), 4.79 (s, 2H, CH₂), 3.18 (s, 3H, CH₃), 2.95 (s, 3H, CH₃).

Step III. NaH (1.5 equiv) was carefully added at 0 °C to a solution of 2,2,2-trifluoroethanol or 3,3,3-trifluoro-2,2-bis(trifluoromethyl)propan-1-ol (1.5 equiv) in anhydrous DMF (2 mL) under a nitrogen atmosphere, then the resulting suspension was stirred for 0.5 h at 0 °C. After that, **9** (0.3 g, 1 equiv) was added, and the reaction mixture was stirred at rt. After monitoring by TLC (EtOAc/hexane), slush was added, and it was extracted in EtOAc (25 mL × 3). The collected organic phases were washed with brine (15 mL × 3) and dried with Na₂SO₄, and then it was filtered and evaporated, achieving **10** or **11** as yellow powders. It was used in the next step without further purification.

(E)-N,N-Dimethyl-N'-((4-((2,2,2-trifluoroethoxy)methyl)phenyl)sulfonyl)formimidamide (10). Yield 81%; silica gel TLC *R_f* 0.21 (EtOAc/hexane 50% v/v); δ_{H} (400 MHz, DMSO-*d*₆): 8.25 (s, 1H, CH), 7.81 (d, *J* = 8.1 Hz, 2H, 2 × Ar-H), 7.52 (d, *J* = 8.1 Hz, 2H, 2 × Ar-H), 4.76 (s, 2H, CH₂), 4.18 (m, 2H, CH₂), 3.18 (s, 3H, CH₃), 2.94 (s, 3H, CH₃).

(E)-N,N-Dimethyl-N'-((4-((3,3,3-trifluoro-2,2-bis(trifluoromethyl)propoxy)methyl)phenyl)sulfonyl)formimidamide (11). Yield 63%; silica gel TLC *R_f* 0.27 (EtOAc/hexane 50% v/v); δ_{H} (400 MHz, DMSO-*d*₆): 8.26 (s, 1H, CH), 7.80 (d, *J* = 7.9 Hz, 2H, 2 × Ar-H), 7.62 (d, *J* = 7.9 Hz, 2H, 2 × Ar-H), 4.78 (s, 2H, CH₂), 3.18 (s, 3H, CH₃), 2.94 (s, 3H, CH₃).

Step IV. Hydrazine hydrate (10 equiv) was carefully added at rt to a solution of **10** or **11** (0.27 g, 1 equiv) in EtOH (5 mL). The reaction mixture was stirred for 1 h. After monitoring by TLC (EtOAc/hexane), the solvent was evaporated and the resulting crude was suspended in cold H₂O, and then it was filtered and purified by silica gel chromatography column (EtOAc/hexane), achieving **6** or **7** as white powders.

4-((2,2,2-Trifluoroethoxy)methyl)benzenesulfonamide (6). Yield 53%; silica gel TLC *R_f* 0.42 (EtOAc/hexane 80% v/v); δ_{H} (400 MHz, DMSO-*d*₆): 7.86 (d, *J* = 7.4 Hz, 2H, 2 × Ar-H), 7.56 (d, *J* = 7.4 Hz, 2H, 2 × Ar-H), 7.41 (s, 2H, exchange with D₂O, SO₂NH₂), 4.78 (s, 2H, CH₂), 4.19 (m, 2H, CH₂); δ_{C} (100 MHz, DMSO-*d*₆): 144.6, 142.1, 129.6 (*q*₁*J*_{C-F} = 276.8 Hz), 128.8, 126.8, 73.4, 68.4 (*q*₂*J*_{C-F} = 32.7 Hz); δ_{F} (376 MHz, DMSO-*d*₆): -72.76; HRMS (*m/z*): calcd for C₉H₁₀F₃NO₃ ([M - H]⁺), 269.2382, found: 269.2371.

4-((3,3,3-Trifluoro-2,2-bis(trifluoromethyl)propoxy)methyl)benzenesulfonamide (7). Yield 45%; silica gel TLC *R_f* 0.54 (MeOH/DCM 6% v/v); δ_{H} (400 MHz, DMSO-*d*₆): 7.91 (d, *J* = 8.1 Hz, 2H, 2 × Ar-H), 7.63 (d, *J* = 8.1 Hz, 2H, 2 × Ar-H), 7.45 (s, 2H, exchange with D₂O, SO₂NH₂), 5.28 (s, 2H, CH₂); δ_{C} (100 MHz, DMSO-*d*₆): 145.5, 139.4, 129.3, 127.0, 125.4 (*q*₁*J*_{C-F} = 294.1 Hz), 81.0 (*q*₂*J*_{C-F} = 31.1 Hz), 71.8; δ_{F} (376 MHz, DMSO-*d*₆): -69.77; HRMS (*m/z*): calcd for C₁₂H₁₀F₉NO₃ ([M - H]⁺), 419.2616, found: 419.2608.

CA Inhibition Assay. An Applied Photophysics stopped-flow instrument has been used for assaying the CA-catalyzed CO₂ hydration activity.³⁸ Phenol red (at a concentration of 0.2 mM) has been used as an indicator, working at the absorbance maximum of 557 nm, with 20 mM HEPES (pH 7.5) as buffer, and 20 mM Na₂SO₄ (to

maintain a constant ionic strength), following the initial rates of the CA-catalyzed CO₂ hydration reaction for a period of 10–100 s. The CO₂ 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 1 h at room temperature prior to assay, in order to allow for the formation of the complex. The inhibition constants were obtained by nonlinear least-squares methods using PRISM 3 and the Cheng-Prusoff equation, as reported earlier,^{39,40} and represent the mean from at least three different determinations. All CA isoforms were recombinant ones obtained in-house as reported earlier.⁴¹

Ligand Stability Assessed by ¹⁹F NMR Spectroscopy. Fluorinated compounds 1–7 were dissolved in DMSO at a final concentration of 80 mM. To assess their stability in cell culture medium, an aliquot of each compound was diluted in DMEM supplemented with 2% fetal bovine serum (FBS), 1% penicillin/streptomycin, and 3% D₂O, yielding a final concentration of 100 μM. Each fluorinated compound in DMEM was analyzed by ¹⁹F NMR spectroscopy at 37 °C, immediately after preparation and again after 1 h of incubation at 37 °C.

¹⁹F NMR spectra were acquired at 310 K at a 600 MHz Bruker Avance NEO spectrometer equipped with a QCI-F 5 mm Cryoprobe operating at 564.6 MHz. A single 90° pulse was used (zg Bruker pulse program), followed by 1.2 s of acquisition, with a frequency offset of -66.0 ppm and a spectral width of 49.2 ppm. Each spectrum was recorded with 32 scans and a 10 s interscan delay, resulting in a total acquisition time of approximately 7 min per sample.

In-Cell and Cell Lysate ¹⁹F NMR Spectroscopy. To perform the in-cell ¹⁹F NMR experiments, after 48 h of protein expression in an unlabeled medium, cells were treated with 100 μM of one of the seven fluorinated ligands. For competition binding experiments, cells were incubated with two ligands simultaneously, each at a concentration of 50 μM. Incubation was carried out for 1 h at 37 °C with 5% CO₂. Following incubation, cells were then harvested, resuspended in 185 μL of NMR buffer (DMEM, 70 mM HEPES, and 20% (v/v) D₂O), transferred to a 3 mm Shigemitsu tube, and gently centrifuged to form a soft pellet. Following NMR analysis, the cell pellet was resuspended, collected in a tube, and centrifuged. The supernatant was collected and analyzed by ¹⁹F NMR spectroscopy, while the cell pellet was lysed in 150 μL of PBS. The resulting soluble lysate was subsequently analyzed by ¹⁹F NMR spectroscopy with 10% (v/v) of D₂O.

¹⁹F NMR spectra were recorded at 310 K at a 600 MHz Bruker Avance III equipped with a room-temperature SEL-HP probe tuned at 564.6 MHz for ¹⁹F detection. The ¹⁹F chemical shift scale was referenced to trichlorofluoromethane by setting the signal of trifluoroacetic acid in an external reference sample to -76.55 ppm. A single 90° pulse was employed, followed by 0.3 s acquisition with a frequency offset of -54.3 ppm and a spectral window of 50.3 ppm. Cell samples and the corresponding lysates and supernatants, a set of four spectra with 1280 scans each and an interscan delay of 1 s were recorded on each sample for a total acquisition time of 112 min. The spectra were processed in Topspin with 10 and 5 Hz exponential line broadening for cells and lysates, respectively, phase-corrected, and summed together. A polynomial baseline correction was applied to the sum spectrum to remove a strong baseline distortion arising from poly(tetrafluoroethylene) (PTFE) components inside the probe.

Uptake Kinetics Assessed by ¹⁹F NMR Bioreactor Experiments. Transfected cells were encapsulated in agarose threads, composed of 1.5% (w/v) low-gelling agarose (Sigma-Aldrich) dissolved in PBS, following a previously described protocol.³² Briefly, cells from one 75 cm² flask were prewarmed to 37 °C and mixed with 400 μL of agarose solution at the same temperature. The suspension was drawn into PEEK tubing (i.d. 0.75 mm) using a syringe, allowed

to cool at room temperature for 2 min, and then loaded into the NMR bioreactor prefilled with PBS.

The bioreactor setup was assembled as previously reported.³² Briefly, gel-encapsulated cells were maintained at 37 °C in an NMR flow unit (InsightMR, Bruker) using a circulating water bath (Julabo) for temperature control. A PEEK capillary (i.d. 0.5 mm) delivered medium to the bottom of the NMR tube, while a PTFE outlet at the top ensured continuous flow. Unlabeled DMEM supplemented with 2% FBS, 10 mM NaHCO₃, 1% antibiotics, 3% D₂O, and the fluorinated compound was pumped at a constant flow rate of 0.1 mL/min using a peristaltic pump (Reglo ICC, Ismatec) through a temperature-controlled circuit.

For time-resolved in-cell NMR experiments performed in the bioreactor, a series of ¹⁹F NMR spectra was acquired at 310 K on a 600 MHz Bruker Avance NEO spectrometer equipped with a QCI-F 5 mm Cryoprobe operating at 564.6 MHz. Each spectrum was recorded with 256 scans and a 1 s interscan delay, yielding a time resolution of 5 min per spectrum over a total acquisition time of 1 h. The resulting time-series data were analyzed using the Dynamics Center (Bruker) to quantify peak areas as a function of time. The rate of uptake for each compound was obtained by linear regression of the first 7 time points of the binding curve using Origin.

Cellular Thermal Shift Assay. For the CETSA assay, cultured cells from a T75 flask (approximately 30 million cells are required to establish one melting curve) were treated with 25 μ L of DMSO stock of fluorinated compound (final compound concentration of 100 μ M, DMSO concentration of 0.00125%) or with 25 μ L of DMSO as the control for 1 h at 37 °C and 5% CO₂. Cells were harvested with Accutase (Sigma-Aldrich), washed once with Hanks' balanced salt solution (HBSS, Thermo Fisher Scientific), and resuspended in 800 μ L of HBSS. The cell suspension was divided into 50 μ L aliquots in PCR 0.2 mL tubes and heated at different temperatures (42.7–44.5–45.7–47.5–50.6–52.0–55.0–58.0–61.0–62.6–65.8–67.7–68.8–70.7 °C) for 3 min in a PCR machine (MyCycler Thermal Cycler, Bio-Rad) followed by a cooling at 4 °C. 50 μ L of lysis buffer (HEPES 40 mM, pH 7.5, NaCl 150 mM) was added to each heat-treated aliquot, which were then frozen–thawed three times using liquid nitrogen and a heating block set at 25 °C. The resulting cell lysates were centrifuged at 14,000 $\times$ g for 1 h at 4 °C. Each supernatant was transferred to a new microtube and analyzed by Western blot. Samples were run on SDS-PAGE as described above and subsequently transferred to a 0.2 μ m nitrocellulose membrane (Trans-Blot Turbo Transfer Pack, Bio-Rad) using the Bio-Rad transfer System. The membrane was blocked for 1 h at room temperature in blocking buffer [5.0% low-fat dry milk in tris-buffered saline (TBS) with 0.1% Tween-20], washed three times with TBS containing 0.1% Tween-20 (TBST). The primary rabbit anti-CA2 antibody (Proteintech 16961–1-AP; Lot #00082518) was diluted 1:2000 in a buffer prepared with 2.0% (w/v) low-fat dry milk in TBST and incubated overnight at 4 °C. After washing in TBST, the membrane was incubated with goat antirabbit IgG (whole molecule)-peroxidase secondary antibody (Sigma: A0545) diluted 1:70000 for 1 h at room temperature, followed by three washes with TBST. LiteAblot EXTEND chemiluminescence substrate (EuroClone) was used for detection. Images were acquired using the ChemiDoc XRS imaging system (Bio-Rad) and quantified using Image Lab software version 5.0 (Bio-Rad). Band intensities were normalized such that the maximum intensity was set to 100% and the minimum to 0%, generating relative intensity values for each temperature point.

To construct the thermal stability curve, the normalized intensity data were plotted against the temperature and analyzed using GraphPad Prism. Nonlinear regression was applied using the four-parameter logistic (4PL) sigmoidal model, selecting the "[Agonist] vs response – Variable slope" function. In this model, the temperature serves as the X-variable, and the relative band intensity serves as the Y-response. Curve fitting was performed to estimate the inflection point, which corresponds to the apparent melting temperature (T_m), defined as the temperature at which 50% of the protein is denatured or insoluble.

HPLC-MS Analysis. A solution of compound 1 in Milli-Q water (100 μ M) was incubated at 37 °C for 7 days and subsequently analyzed by HPLC-MS³. The resulting MS³ spectra were compared with those of reference samples of compound 1 and sulfanilamide. The analyses were performed using an HPLC-MS instrument composed of an HPLC Accela, with autosampler and column oven, coupled to an LTQ XL linear quadrupole ion trap mass spectrometer equipped with an IonMax ESI interface (Thermo Scientific). Chromatographic separation was done using a Gemini C18 column (100 $\times$ 2 mm, 5 μ m, Phenomenex) using the gradient elution program reported in Table S1. Mobile phase A was LC-MS grade water, and B was LC-MS grade ACN, both containing 0.1% FA. The column was maintained at 25 °C. Injection volume was 20 μ L. MS and MSⁿ parameters were optimized by infusion of H₂O/ACN 50/50 0.1% FA solution of the standards of the two molecules. Data acquisition was done in positive polarity for the first 5.5 min of the chromatographic run, then the polarity switched to negative ion. Sheath gas, auxiliary gas, and sweep gas were 30, 15, and 1 (arbitrary units), respectively, while the spray voltage value was 4.7 kV in the two polarities: the ion transfer capillary was maintained at 275 °C. Capillary and tube lens voltages were 2 and 25 V in positive ion mode, and –20 and –60 V in negative ion mode. Data were acquired in full scan in MS and in MS/MS/MS mode; for sulfanilamide, the proton adduct ion at 173 m/z was isolated (isolation window 2.2), fragmented at 72 normalized collision energy and then the fragment ion at m/z 156 was isolated (isolation window 3.0) and fragmented at 96 normalized collision energy, while for compound 1 the deprotonated ion at m/z 267.1 was isolated (isolation window 2.8), fragmented at 37 normalized collision energy and then the fragment ion at m/z 197 was isolated (isolation window 3.0) and fragmented at 80 normalized collision energy. Activation Q and activation time were 0.25 and 30 ms, respectively. The mass spectrometer was calibrated following the manufacturer's instructions before the analysis. Data acquisition and analysis were done using the Xcalibur software (Thermo, ver. 2.2).

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jmedchem.5c02227>.

Figures S1–S9: quantification of intracellular CA isoforms, stability of compounds, HPLC-MS of compound 1, ¹⁹F NMR spectra of cell lysates and supernatants, ¹⁹F NMR spectra of control samples, representative HPLC traces; Table S1: HPLC elution program (PDF)

Molecular formula strings and inhibitory constants (CSV)

■ AUTHOR INFORMATION

Corresponding Author

Enrico Luchinat – CERM – Magnetic Resonance Center, University of Florence, Sesto Fiorentino 50019, Italy; Consorzio Interuniversitario Risonanze Magnetiche di Metallo Proteine – CIRMMMP, Sesto Fiorentino 50019, Italy; Chemistry Department, University of Florence, Sesto Fiorentino 50019, Italy; orcid.org/0000-0003-4183-4311; Email: eluchinat@cerm.unifi.it

Authors

Azzurra Costantino – CERM – Magnetic Resonance Center, University of Florence, Sesto Fiorentino 50019, Italy
Letizia Barbieri – CERM – Magnetic Resonance Center, University of Florence, Sesto Fiorentino 50019, Italy; Consorzio Interuniversitario Risonanze Magnetiche di Metallo Proteine – CIRMMMP, Sesto Fiorentino 50019, Italy

Simone Giovannuzzi – NEUROFARBA Department, Section of Pharmaceutical and Nutraceutical Sciences, University of Florence, Sesto Fiorentino 50019, Italy; orcid.org/0000-0002-0669-7481

Alessio Nocentini – NEUROFARBA Department, Section of Pharmaceutical and Nutraceutical Sciences, University of Florence, Sesto Fiorentino 50019, Italy; orcid.org/0000-0003-3342-702X

Claudio T. Supuran – NEUROFARBA Department, Section of Pharmaceutical and Nutraceutical Sciences, University of Florence, Sesto Fiorentino 50019, Italy; orcid.org/0000-0003-4262-0323

Mindaugas Raitelaitis – Department of Oncology-Pathology, Karolinska Institutet, Stockholm 171 11, Sweden

Pär Nordlund – Department of Oncology-Pathology, Karolinska Institutet, Stockholm 171 11, Sweden

Lucia Banci – CERM – Magnetic Resonance Center, University of Florence, Sesto Fiorentino 50019, Italy; *Consorzio Interuniversitario Risonanze Magnetiche di Metallo Proteine – CIRMMP*, Sesto Fiorentino 50019, Italy; Chemistry Department, University of Florence, Sesto Fiorentino 50019, Italy; orcid.org/0000-0003-0562-5774

Complete contact information is available at:
<https://pubs.acs.org/10.1021/acs.jmedchem.5c02227>

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work has been supported by Fragment-Screen, grant agreement no. 101094131, funded by the Horizon Europe program of the European Commission, by Prin 2022 DD 104-02-02-2022 project "TiReD – time-resolved magnetic resonance to investigate dynamic events in biological systems and biotransformations" (grant 2022WANFHS, CUP: B53D23013990006) of the Italian Ministry for University and Research, and by ITACA.SB "Potentiating the Italian Capacity for Structural Biology Services in Instruct-ERIC (Project no. IR0000009) within the call MUR 3264/2021 PNRR M4/C2/L3.1.1, funded by the European Union – Next Generation EU. The study has also been supported by funding to P.N. from the Swedish Cancer Society (22 2413 Pj 01 H) and Swedish Research Council (621-2014-5583), as well as Radiumhemmets Research Funds (224072), and by an EMBO Scientific Exchange Grant to A.C. We acknowledge the CISM - Mass Spectrometry Center of the University of Florence and Dr. Giuseppe Pieraccini for the LC-MS analysis.

ABBREVIATIONS USED

AAZ, acetazolamide; ACN, acetonitrile; CA, carbonic anhydrase; CETSA, cellular thermal shift assay; DMEM, Dulbecco's modified Eagle medium; FA, formic acid; FBS, fetal bovine serum; fwhm, full width at half-maximum; HEK293T, human embryonic kidney 293T; HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; MeOH, methanol; PEI, polyethylenimine; SDS-PAGE, SDS-polyacrylamide gel electrophoresis; T_m , melting temperature

REFERENCES

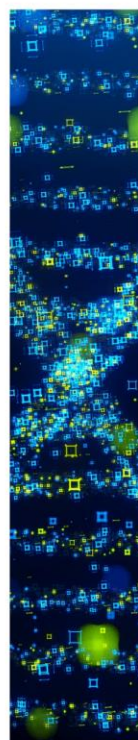
- (1) Sun, D.; Gao, W.; Hu, H.; Zhou, S. Why 90% of Clinical Drug Development Fails and How to Improve It? *Acta Pharmaceutica Sinica B* **2022**, 12 (7), 3049–3062.
- (2) Dowden, H.; Munro, J. Trends in Clinical Success Rates and Therapeutic Focus. *Nat. Rev. Drug Discov* **2019**, 18 (7), 495–496.
- (3) Theillet, F.-X. In-Cell Structural Biology by NMR: The Benefits of the Atomic Scale. *Chem. Rev.* **2022**, 122 (10), 9497–9570.
- (4) Luchinat, E.; Cremonini, M.; Banci, L. Radio Signals from Live Cells: The Coming of Age of In-Cell Solution NMR. *Chem. Rev.* **2022**, 122 (10), 9267–9306.
- (5) Luchinat, E.; Banci, L. In-Cell NMR: From Target Structure and Dynamics to Drug Screening. *Curr. Opin. Struct. Biol.* **2022**, 74, No. 102374.
- (6) Luchinat, E.; Barbieri, L.; Cremonini, M.; Nocentini, A.; Supuran, C. T.; Banci, L. Drug Screening in Human Cells by NMR Spectroscopy Allows the Early Assessment of Drug Potency. *Angew. Chem., Int. Ed.* **2020**, 59 (16), 6535–6539.
- (7) Luchinat, E.; Barbieri, L.; Cremonini, M.; Pennestri, M.; Nocentini, A.; Supuran, C. T.; Banci, L. Determination of Intracellular Protein-Ligand Binding Affinity by Competition Binding in-Cell NMR. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2021**, 77 (Pt 10), 1270–1281.
- (8) Siegal, G.; Selenko, P. Cells, Drugs and NMR. *J. Magn. Reson.* **2019**, 306, 202–212.
- (9) Luchinat, E.; Banci, L. In-Cell NMR: From Target Structure and Dynamics to Drug Screening. *Curr. Opin. Struct. Biol.* **2022**, 74, No. 102374.
- (10) Luchinat, E.; Barbieri, L.; Davis, B.; Brough, P. A.; Pennestri, M.; Banci, L. Ligand-Based Competition Binding by Real-Time ^{19}F NMR in Human Cells. *J. Med. Chem.* **2024**, 67 (2), 1115–1126.
- (11) Inomata, K.; Ohno, A.; Tochio, H.; Isogai, S.; Tenno, T.; Nakase, I.; Takeuchi, T.; Futaki, S.; Ito, Y.; Hiroaki, H.; Shirakawa, M. High-Resolution Multi-Dimensional NMR Spectroscopy of Proteins in Human Cells. *Nature* **2009**, 458 (7234), 106–109.
- (12) Burz, D. S.; Breindel, L.; Shekhtman, A. Chapter Seventeen - Improved Sensitivity and Resolution of in-Cell NMR Spectra. In *Methods in Enzymology*; Shukla, A. K., Ed.; Chemical and Synthetic Biology Approaches To Understand Cellular Functions - Part A; Academic Press: 2019; Vol. 621, pp 305–328.
- (13) Buchholz, C. R.; Pomerantz, W. C. K. ^{19}F NMR Viewed through Two Different Lenses: Ligand-Observed and Protein-Observed ^{19}F NMR Applications for Fragment-Based Drug Discovery. *RSC Chem. Biol.* **2021**, 2 (5), 1312–1330.
- (14) Gronenborn, A. M. Small, but Powerful and Attractive: ^{19}F in Biomolecular NMR. *Structure* **2022**, 30 (1), 6–14.
- (15) Dalvit, C.; Vulpetti, A. Ligand-Based Fluorine NMR Screening: Principles and Applications in Drug Discovery Projects. *J. Med. Chem.* **2019**, 62 (5), 2218–2244.
- (16) Li, C.; Wang, G.-F.; Wang, Y.; Creager-Allen, R.; Lutz, E. A.; Scronce, H.; Slade, K. M.; Ruf, R. A. S.; Mehl, R. A.; Pielak, G. J. Protein (^{19}F) NMR in Escherichia Coli. *J. Am. Chem. Soc.* **2010**, 132 (1), 321–327.
- (17) Ye, Y.; Liu, X.; Xu, G.; Liu, M.; Li, C. Direct Observation of Ca^{2+} -Induced Calmodulin Conformational Transitions in Intact *Xenopus* Laevis Oocytes by (^{19}F) F NMR Spectroscopy. *Angew. Chem., Int. Ed. Engl.* **2015**, 54 (18), 5328–5330.
- (18) Zhu, W.; Guseman, A. J.; Bhinderwala, F.; Lu, M.; Su, X.-C.; Gronenborn, A. M. Visualizing Proteins in Mammalian Cells by ^{19}F NMR Spectroscopy. *Angew. Chem., Int. Ed.* **2022**, 61 (23), No. e202201097.
- (19) Pham, L. B. T.; Costantino, A.; Barbieri, L.; Calderone, V.; Luchinat, E.; Banci, L. Direct Expression of Fluorinated Proteins in Human Cells for ^{19}F In-Cell NMR Spectroscopy. *J. Am. Chem. Soc.* **2023**, 145 (2), 1389–1399.
- (20) Costantino, A.; Pham, L. B. T.; Barbieri, L.; Calderone, V.; Ben-Nissan, G.; Sharon, M.; Banci, L.; Luchinat, E. Controlling the Incorporation of Fluorinated Amino Acids in Human Cells and Its Structural Impact. *Protein Sci.* **2024**, 33 (3), No. e4910.

- (21) Bao, H.-L.; Ishizuka, T.; Sakamoto, T.; Fujimoto, K.; Uechi, T.; Kenmochi, N.; Xu, Y. Characterization of Human Telomere RNA G-Quadruplex Structures in Vitro and in Living Cells Using 19F NMR Spectroscopy. *Nucleic Acids Res.* **2017**, *45* (9), 5501–5511.
- (22) Dalvit, C.; Veronesi, M.; Vulpetti, A. Fluorine NMR Functional Screening: From Purified Enzymes to Human Intact Living Cells. *J. Biomol NMR* **2020**, *74* (10), 613–631.
- (23) Krafčík, D.; Istvánková, E.; Džatko, Š.; Višková, P.; Foldynová-Trantírková, S.; Trantírek, L. Towards Profiling of the G-Quadruplex Targeting Drugs in the Living Human Cells Using NMR Spectroscopy. *International Journal of Molecular Sciences* **2021**, *22* (11), 6042.
- (24) Supuran, C. Carbonic Anhydrases as Drug Targets - An Overview. *CTMC* **2007**, *7* (9), 825–833.
- (25) Supuran, C. T. Emerging Role of Carbonic Anhydrase Inhibitors. *Clinical Science* **2021**, *135* (10), 1233–1249.
- (26) Mboge, M.; Mahon, B.; McKenna, R.; Frost, S. Carbonic Anhydrases: Role in pH Control and Cancer. *Metabolites* **2018**, *8* (1), 19.
- (27) Imtaiyaz Hassan, Md.; Shajee, B.; Waheed, A.; Ahmad, F.; Sly, W. S. Structure, Function and Applications of Carbonic Anhydrase Isozymes. *Bioorg. Med. Chem.* **2013**, *21* (6), 1570–1582.
- (28) Supuran, C. T.; Scozzafava, A. Carbonic Anhydrases as Targets for Medicinal Chemistry. *Bioorg. Med. Chem.* **2007**, *15* (13), 4336–4350.
- (29) Vullo, D.; Franchi, M.; Gallori, E.; Antel, J.; Scozzafava, A.; Supuran, C. T. Carbonic Anhydrase Inhibitors. Inhibition of Mitochondrial Isozyme V with Aromatic and Heterocyclic Sulfonamides. *J. Med. Chem.* **2004**, *47* (5), 1272–1279.
- (30) Pacchiano, F.; Carta, F.; McDonald, P. C.; Lou, Y.; Vullo, D.; Scozzafava, A.; Dedhar, S.; Supuran, C. T. Ureido-Substituted Benzenesulfonamides Potently Inhibit Carbonic Anhydrase IX and Show Antimetastatic Activity in a Model of Breast Cancer Metastasis. *J. Med. Chem.* **2011**, *54* (6), 1896–1902.
- (31) Duda, D. M.; Tu, C.; Fisher, S. Z.; An, H.; Yoshioka, C.; Govindasamy, L.; Laipis, P. J.; Agbandje-McKenna, M.; Silverman, D. N.; McKenna, R. Human Carbonic Anhydrase III: Structural and Kinetic Study of Catalysis and Proton Transfer. *Biochemistry* **2005**, *44* (30), 10046–10053.
- (32) Barbieri, L.; Luchinat, E. Monitoring Protein-Ligand Interactions in Human Cells by Real-Time Quantitative In-Cell NMR Using a High Cell Density Bioreactor. *JoVE* **2021**, *169*, No. 62323.
- (33) Molina, D. M.; Jafari, R.; Ignatushchenko, M.; Seki, T.; Larsson, E. A.; Dan, C.; Sreekumar, L.; Cao, Y.; Nordlund, P. Monitoring Drug Target Engagement in Cells and Tissues Using the Cellular Thermal Shift Assay. *Science* **2013**, *341* (6141), 84–87.
- (34) Martinez Molina, D.; Nordlund, P. The Cellular Thermal Shift Assay: A Novel Biophysical Assay for In Situ Drug Target Engagement and Mechanistic Biomarker Studies. *Annu. Rev. Pharmacol. Toxicol.* **2016**, *56* (1), 141–161.
- (35) Aricescu, A. R.; Lu, W.; Jones, E. Y. A Time- and Cost-Efficient System for High-Level Protein Production in Mammalian Cells. *Acta Cryst. D* **2006**, *62* (10), 1243–1250.
- (36) Barbieri, L.; Luchinat, E.; Banci, L. Characterization of Proteins by In-Cell NMR Spectroscopy in Cultured Mammalian Cells. *Nat. Protoc.* **2016**, *11* (6), 1101–1111.
- (37) Angeli, A.; di Cesare Mannelli, L.; Trallori, E.; Peat, T. S.; Ghelardini, C.; Carta, F.; Supuran, C. T. Design, Synthesis, and X-Ray of Selenides as New Class of Agents for Prevention of Diabetic Cerebrovascular Pathology. *ACS Med. Chem. Lett.* **2018**, *9* (5), 462–467.
- (38) Khalifah, R. G. The Carbon Dioxide Hydration Activity of Carbonic Anhydrase. I. Stop-Flow Kinetic Studies on the Native Human Isoenzymes B and C. *J. Biol. Chem.* **1971**, *246* (8), 2561–2573.
- (39) Nocentini, A.; Alterio, V.; Bua, S.; Micheli, L.; Esposito, D.; Buonanno, M.; Bartolucci, G.; Osman, S. M.; AlOthman, Z. A.; Cirilli, R.; Pierini, M.; Monti, S. M.; Di Cesare Mannelli, L.; Gratteri, P.; Ghelardini, C.; De Simone, G.; Supuran, C. T. Phenyl(Thio)phosphon(amid)ate Benzenesulfonamides as Potent and Selective

Inhibitors of Human Carbonic Anhydrases II and VII Counteract Allodynia in a Mouse Model of Oxaliplatin-Induced Neuropathy. *J. Med. Chem.* **2020**, *63* (10), 5185–5200.

(40) Bua, S.; Nocentini, A.; Bonardi, A.; Palma, G.; Ciampi, G.; Giliberti, A.; Iannelli, F.; Bruzzese, F.; Supuran, C. T.; de Nigris, F. Harnessing Nitric Oxide-Donating Benzofuroxans for Targeted Inhibition of Carbonic Anhydrase IX in Cancer. *J. Med. Chem.* **2024**, *67* (17), 15892–15907.

(41) Bonardi, A.; Micheli, L.; Di Cesare Mannelli, L.; Ghelardini, C.; Gratteri, P.; Nocentini, A.; Supuran, C. T. Development of Hydrogen Sulfide-Releasing Carbonic Anhydrases IX- and XII-Selective Inhibitors with Enhanced Antihyperalgesic Action in a Rat Model of Arthritis. *J. Med. Chem.* **2022**, *65* (19), 13143–13157.



CAS BIOFINDER DISCOVERY PLATFORM™

**STOP DIGGING
THROUGH DATA
—START MAKING
DISCOVERIES**

CAS BioFinder helps you find the
right biological insights in seconds

Start your search



<https://doi.org/10.1021/acs.jmedchem.5c02227>
J. Med. Chem. XXXX, XXX, XXX–XXX

L

SUPPLEMENTARY INFORMATION

Intracellular binding of novel fluorinated compounds to carbonic anhydrase isoforms explored by in-cell ^{19}F NMR

Azzurra Costantino^a, Letizia Barbieri^{a,b}, Simone Giovannuzzi^c, Alessio Nocentini^c, Claudio T. Supuran^c, Mindaugas Raitelaitis^d, Pär Nordlund^d, Lucia Banci^{a,b,e}, Enrico Luchinat^{a,b,e,*}

^a CERM – Magnetic Resonance Center, University of Florence, via Luigi Sacconi, 6, 50019, Sesto Fiorentino, Italy;

^b Consorzio Interuniversitario Risonanze Magnetiche di Metallo Proteine – CIRMMMP, via Luigi Sacconi, 6, 50019, Sesto Fiorentino, Italy;

^c NEUROFARBA Department, Section of Pharmaceutical and Nutraceutical Sciences, University of Florence, Sesto Fiorentino, Italy;

^d Department of Oncology-Pathology, Karolinska Institutet, Stockholm, Sweden;

^e Chemistry Department, University of Florence, Via della Lastruccia, 3, 50019, Sesto Fiorentino, Italy.

* Email: eluchinat@cerm.unifi.it

Table of Contents

Supplementary Figures S1-S5.

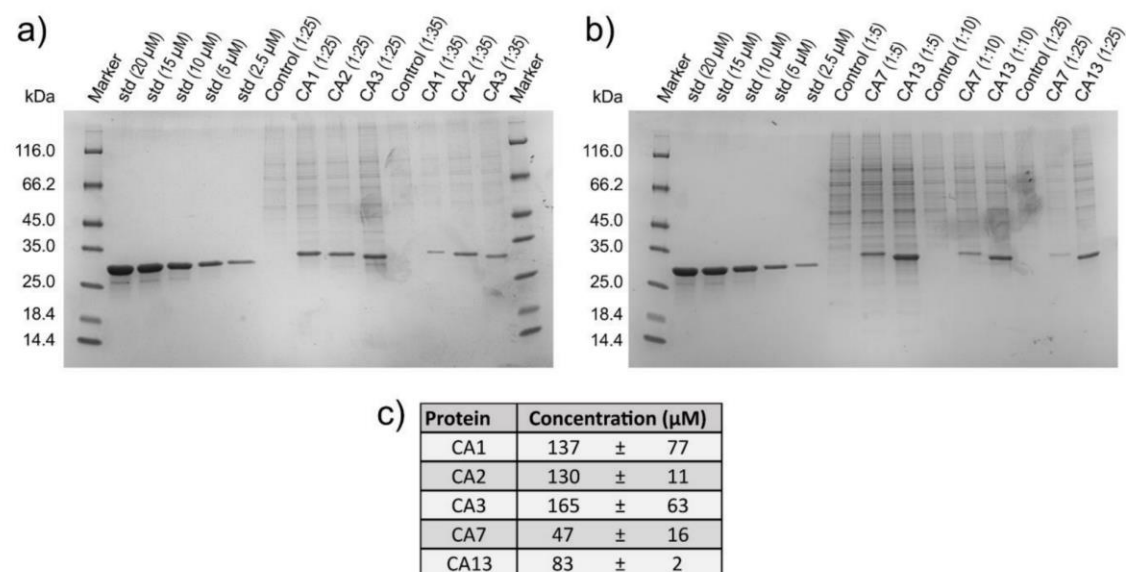
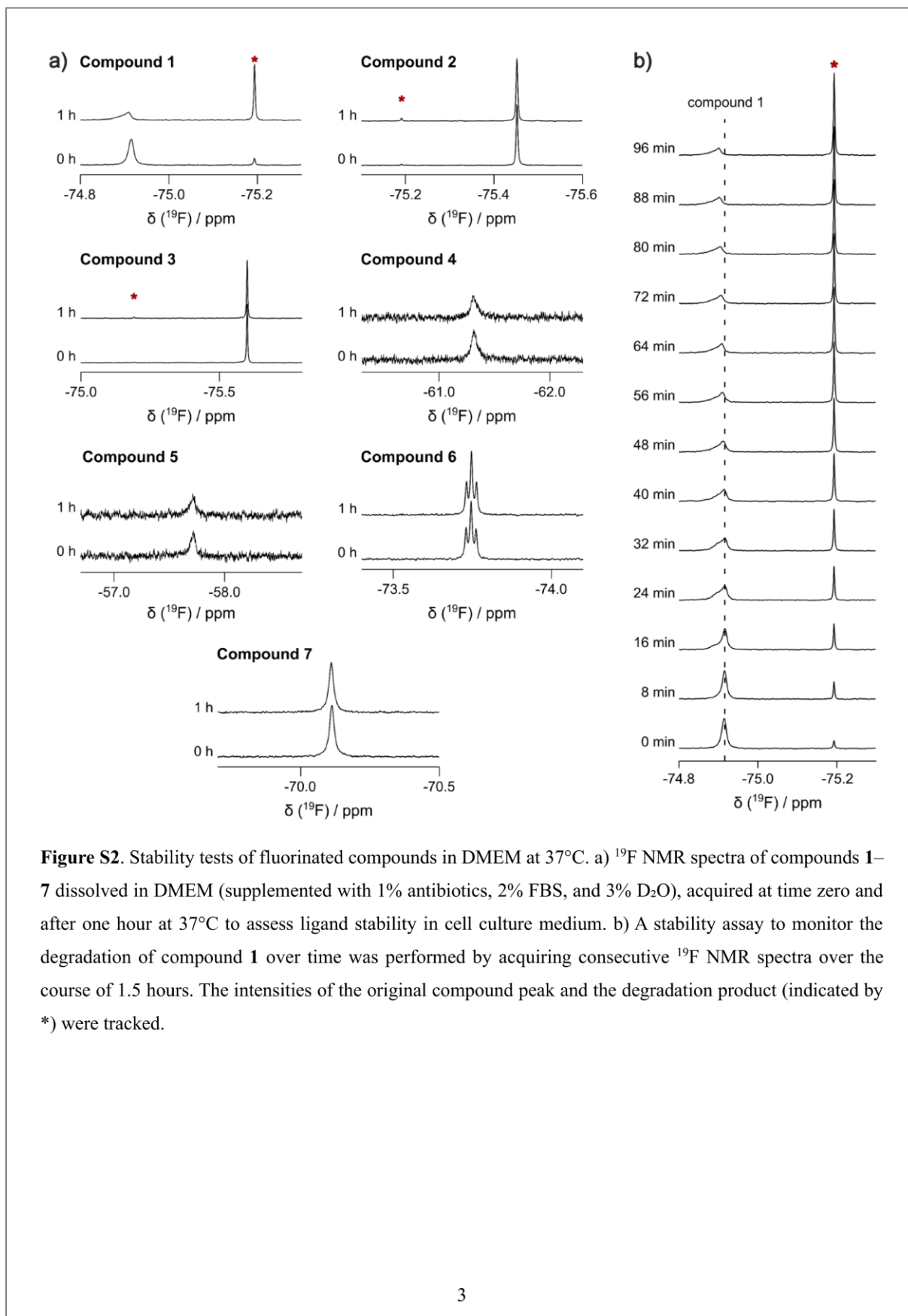


Figure S1. Quantification of cytosolic carbonic anhydrase isoforms expressed in HEK293T cells. Coomassie-stained SDS-PAGE gels showing the expression levels of various cytosolic carbonic anhydrase isoforms (CA1, CA2, CA3, CA7, and CA13) in HEK293T cells. Cell lysates were obtained from T25 flasks, lysed in 150 μL of PBS, and serially diluted before gel loading. Known concentrations of purified CA2 protein were used as standards for quantification. a) SDS-PAGE showing standards and lysates expressing CA1, CA2, and CA3. b) SDS-PAGE showing standards and lysates expressing CA7 and CA13. c) Table summarizing the calculated concentrations of each cytosolic CA isoform based on band intensity analysis.



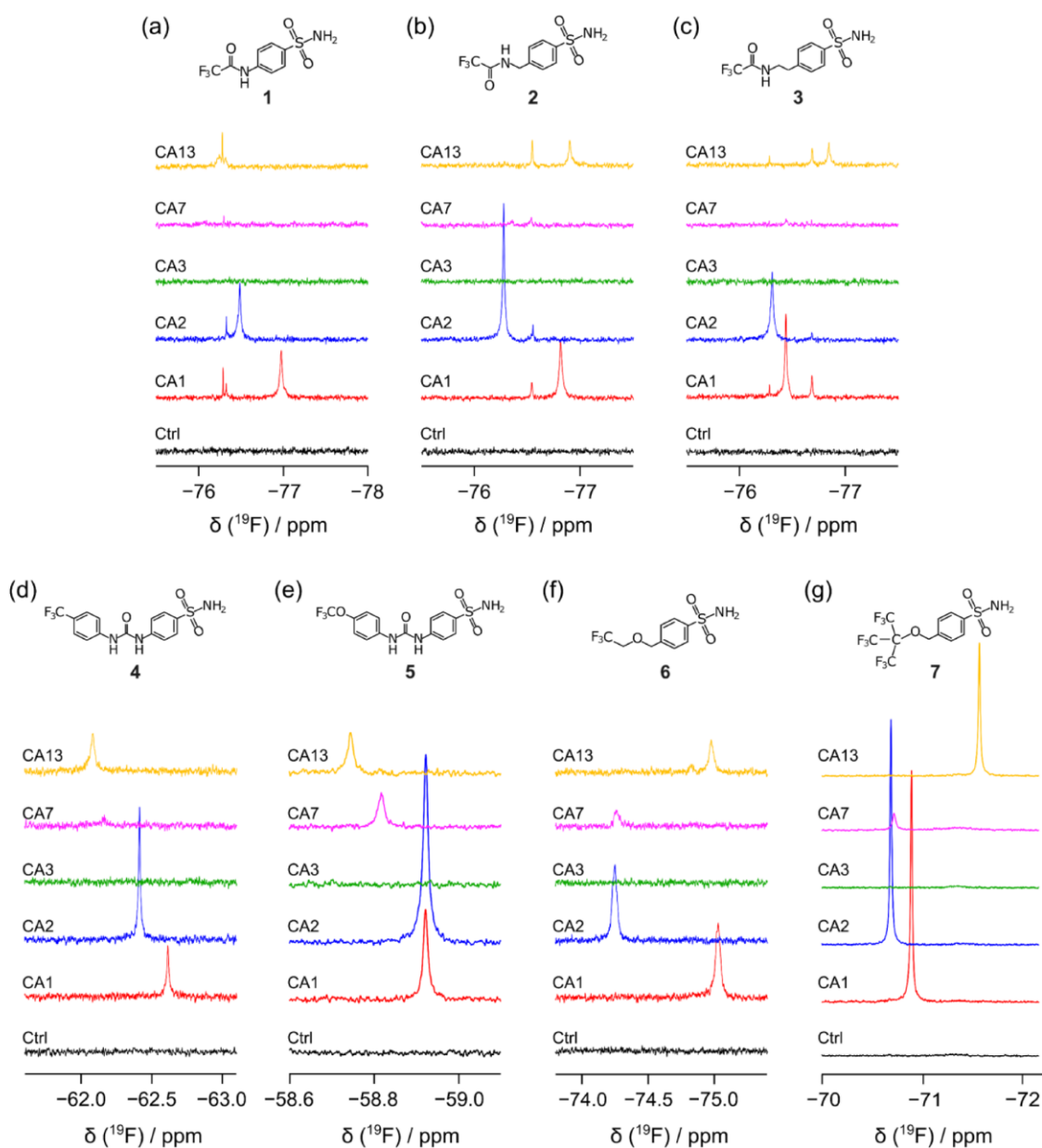
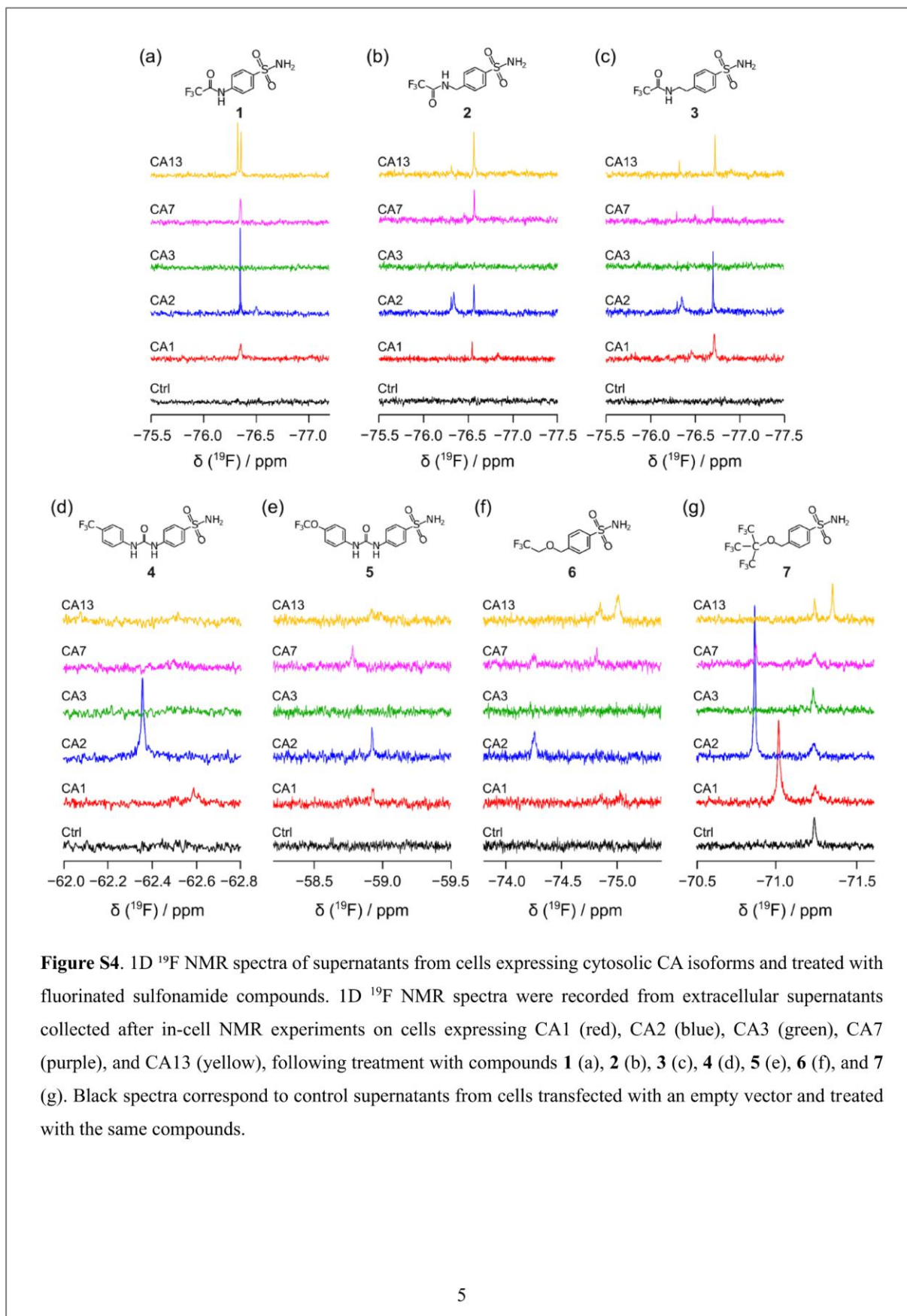


Figure S3. 1D ^{19}F NMR spectra of cell lysates from cells expressing cytosolic CA isoforms and treated with fluorinated sulfonamide compounds. Spectra were recorded from cell lysates after treatment with compound 1 (a), 2 (b), 3 (c), 4 (d), 5 (e), 6 (f), and 7 (g). Each spectrum shows the complex with the compound and a specific CA isoform: CA1 (red), CA2 (blue), CA3 (green), CA7 (purple), and CA13 (yellow). Control spectra (black) correspond to lysates from cells transfected with an empty vector and treated with the same compounds.



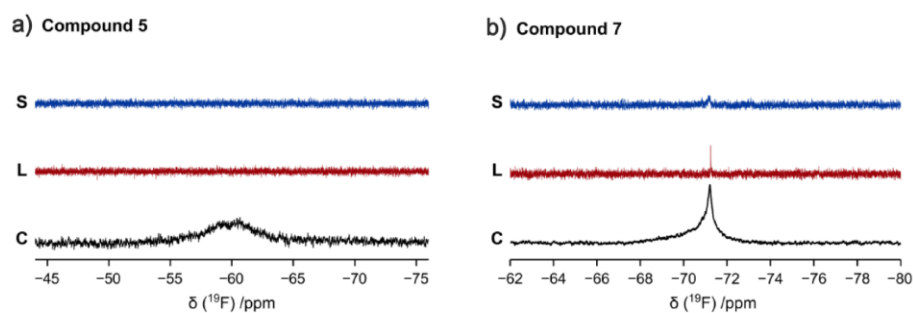


Figure S5. Detailed ^{19}F NMR spectra of in-cell, cell lysate, and supernatant samples from control cells treated with compounds **5** and **7**. 1D ^{19}F NMR spectra were acquired from HEK293T cells transfected with a vector without the gene encoding the CA isoform and treated with compound **5** (a) or compound **7** (b). Spectra are shown for intact cells (C, black), corresponding cell lysates (L, red), and extracellular supernatants collected after NMR experiments (S, blue).

3.2.3 Additional CETSA dose-response study for IC₅₀ determination

In addition to the CETSA experiments reported in the main manuscript, which determined CA2 melting curves in the presence of various compounds, dose-response curves were established to evaluate the binding of fluorinated compounds to the CA2 isoform and to estimate IC₅₀ values. These values represent the apparent concentration of ligand required to reduce the soluble protein fraction detected under the chosen experimental conditions by 50%.

The initial investigation focused on the toxicity of the fluorinated compounds over a prolonged incubation period. HEK293T cells treated with compounds 4–7 at a concentration of 25 μ M for 24 hours exhibited viability ranging from 83% to 91%, suggesting that this concentration did not significantly impact cell viability. Subsequently, similarly to other studies, the effect of different pre-treatment times (4 and 20 hours) on protein stabilisation was investigated. Following treatment, the cells were subjected to a thermal pulse at 53 °C. This temperature was identified from the melting curves reported in the manuscript. The results showed that prolonged incubation time shifted the dose-response curve towards higher IC₅₀ values, suggesting that extended incubation times may alter the ligand's intracellular distribution or its competition with other endogenous species (**Figure 3.1**).

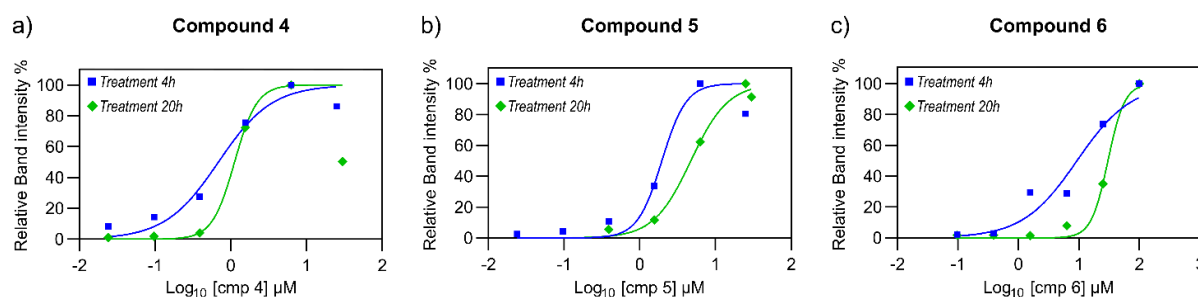


Figure 3.1. Comparison of CA2 dose–response curves obtained after cell pretreatment at different times. Cells were incubated with compounds 4 (a), 5 (b), and 6 (c) for 4 hours (■, blue) or 20 hours (◆, green) and subsequently subjected to heat shock at 53°C.

The influence of denaturation temperature was investigated by treating the cells for four hours and varying the thermal pulse to 53 °C and 55 °C. As expected, the curves shifted towards a higher IC₅₀ as the temperature increased, although the variation was minimal in some cases (**Figure 3.2**). The IC₅₀ values extrapolated from each experiment are in the micromolar range, confirming a moderate interaction under these cellular conditions (**Table 3.1**).

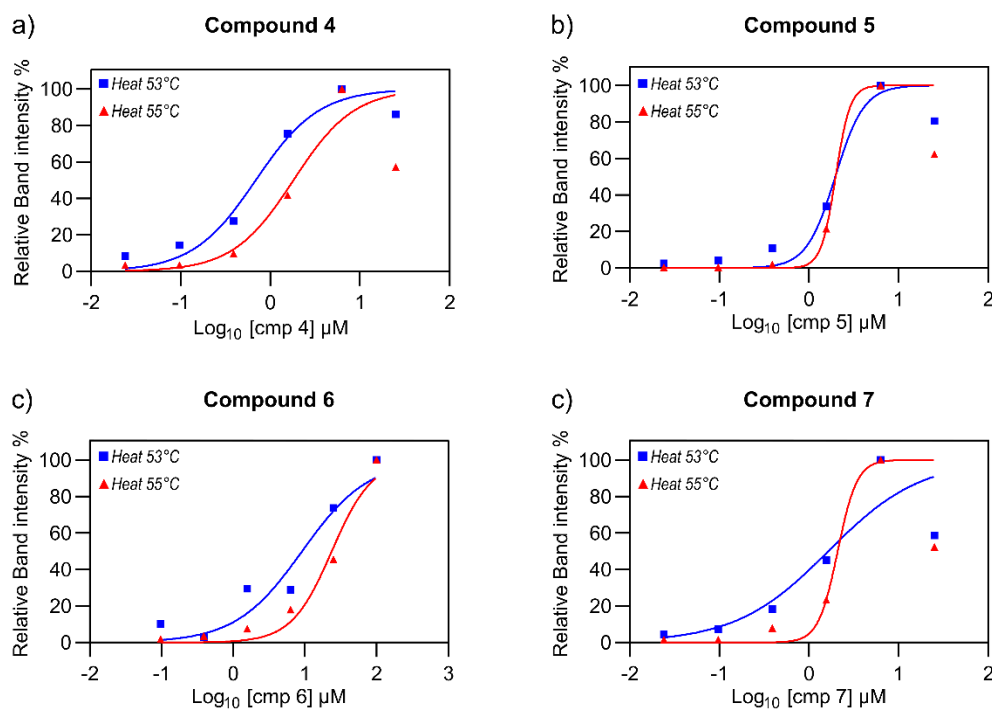


Figure 3.2. Effect of temperature on the dose-response curves of CA2 treated with fluorinated compounds. The cells were pretreated with compounds 4 (a), 5 (b), 6 (c) and 7 (d) for 4 hours, and then subjected to a heat shock of either 53°C (■, blue) or 55°C (▲, red).

Table 3.1. IC₅₀ values obtained from CETSA dose-response curves. Cells were pretreated for 4 h with fluorinated compounds 4–7 and heat shocked at 53 °C, compared with K_I values reported for CA2.

Compound	IC ₅₀ (μM)	K _I (nM) ^a
4	0,6887	1150
5	1,964	203
6	9,148	78.0
7	1,642	57.6

^a In vitro stopped-flow assay (as reported in the manuscript)

For comparison, a dose-response curve was obtained for CA2 treated with acetazolamide (AAZ), a reference ligand with a known K_d in-vitro (12.5 ± 1.0 nM).²⁰ In this case, CETSA returned an IC₅₀ value of 4.9 μM, which is significantly higher than the K_i (**Figure 3.3**). In the scientific literature, Iyer et al. report an IC₅₀ of 0.312 μM for AAZ in-vitro in a system with

CA1 purified at a concentration of 1 μM , while other studies report values between approximately 2.2 and 20.6 μM depending on the experimental conditions and assay method.^{121–}

¹²³ This comparison highlights the variability of IC_{50} values determined in different conditions and underlines that the IC_{50} values obtained in CETSA assay reflects not only the intrinsic binding affinity but also additional factors such as the assay temperature, permeability, target concentration, endogenous competition and non-specific binding.

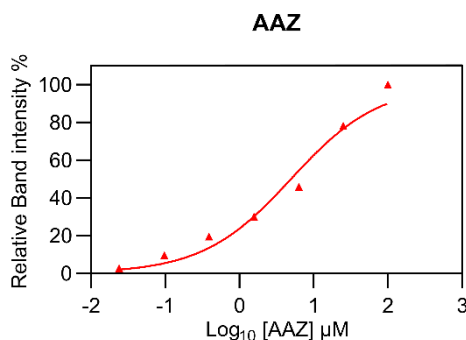


Figure 3.3. Dose-response curve of CA2 treated with acetazolamide (AAZ). Cells were pretreated for 4 hours and subjected to heat shock at 55 °C.

The IC_{50} values obtained after 4 hours of incubation at 53 °C can be used to estimate the apparent affinity ranking of the compounds under these conditions. Compound 4 appears to be the most effective at stabilising CA2, followed by compounds 7 and 5, which have comparable affinities. On the contrary, compound 6 appears to be significantly weaker than the other ligands. Interestingly, this relative affinity ranking is partially consistent with the results of the in-cell ^{19}F NMR study reported in the manuscript. In both cases, compound 6 consistently exhibited the weakest interaction with CA2, whereas compound 4 remained one of the most potent ligands. Meanwhile, the differences observed for compounds 5 and 7 can reasonably be attributed to the divergent experimental conditions. In the in-cell NMR study, cells were treated with higher concentrations of ligands for 1 hour incubation, after which they were detached, washed and pelleted in a closed tube where different equilibria could be established. In the CETSA experiments, however, lower concentrations of the compound were used and the incubations were longer than 4 hours. These variations in experimental conditions and equilibrium dynamics may explain the differences observed in the relative affinity rankings obtained using the two techniques. Finally, we attempted to extend the approach to CA1 and CA13, but HEK293T cells do not express these isoforms at sufficiently high levels for them to

be reliably detected by Western blot. This prevented CETSA experiments from being conducted for these targets.

These preliminary data demonstrate the potential for combining CETSA dose-response curve analysis with in-cell NMR observations to evaluate ligand intracellular engagement. However, it is important to highlight that the CETSA IC_{50} should not be interpreted as an absolute measure of affinity, but rather as a parameter dependent on cellular context and specific experimental conditions.

3.2 Quantitative competition binding of fluorinated ligands by real-time ^{19}F in-cell NMR to assess isoform selectivity in human cells.

Azzurra Costantino^{1,2} | *Letizia Barbieri*^{1,3} | *Simone Giovannuzzi*⁴ |
*Alessio Nocentini*⁴ | *Claudiu T. Supuran*⁴ | *Enrico Luchinat*^{1,2,3}

- ¹ Magnetic Resonance Center - CERM, University of Florence, via Luigi Sacconi, 6, 50019, Sesto Fiorentino, Italy.
- ² Chemistry Department, University of Florence, Via della Lastruccia, 3, 50019, Sesto Fiorentino, Italy.
- ³ Consorzio Interuniversitario Risonanze Magnetiche di Metallo Proteine – CIRMMP, via Luigi Sacconi, 6, 50019, Sesto Fiorentino, Italy.
- ⁴ NEUROFARBA Department, Section of Pharmaceutical and Nutraceutical Sciences, University of Florence, via Ugo Schiff, 6, 50019, Sesto Fiorentino, Italy.

In preparation

3.3.1 Summary

Following on from previous work, competition binding experiments were performed to quantitatively assess the relative affinities of fluorinated benzenesulfonamides towards CA1, CA2 and CA13 isoforms. The fluorinated compounds analysed in this study are the same as those characterised in our previous work, in which we evaluated their binding to different CA isoforms and determined their qualitative affinity ranking. Building on these results, this study employed a bioreactor real-time in-cell ^{19}F NMR spectroscopy to investigate competitive binding among these ligands, ultimately yielding quantitative data relative to the spy ligand.

After verifying the stability of the compounds, compound **6** was selected as the fluorinated spy ligand due to its high cellular uptake and intermediate affinity. Competition binding experiments were then conducted, maintaining the concentration of this spy ligand constant while varying that of the test ligand over time. In-cell ^{19}F NMR spectra were used to monitor the displacement of the spy ligand from the protein and the simultaneous formation of the complex with the test ligand. Analysis of the free ligand signals as a function of the protein–spy complex fraction allowed us to calculate the dissociation constant ratio ($K_{\text{dT}}/K_{\text{dS}}$) between the two ligands. Due to the short repetition times during the experiment, the measured signals and peak areas were corrected post-measurement to account for the effects of T_1 relaxation. Finally, by correlating the fluorinated spy ligand with the non-fluorinated reference inhibitor methazolamide (MZA), we calculate absolute intracellular K_{d} values from the measured K_{d} ratios, enabling quantitative comparison across all tested ligands and isoforms.

The results indicate that all compounds showed good affinity toward the three CA isoforms, except for compound **5** with reduced affinity and compound **7** with enhanced selectivity for CA1. The quantitative results are consistent with the previously established qualitative affinity ranking obtained from closed-tube in-cell NMR experiments, further supporting the reliability of both the closed-tube and real-time bioreactor in-cell NMR approaches. Furthermore, compared to previously NMR bioreactors, the use and monitoring of two fluorinated ligands offers the advantage of avoiding the need to correct the signal intensity related to the progressive depletion of the target, as well as reducing potential biases due to the different solubility or permeability of the compounds.

3.3.2 Attached manuscript [in preparation]

Abstract

Drug development is a risky endeavour with a high failure rate, often caused by the limited ability to predict the efficacy and interactions of candidate drugs in a native cellular environment. In this context, in-cell NMR spectroscopy is a promising tool for assessing drug-target binding directly in living cells, thereby improving the screening and development of new molecules. In this study, we used real-time in-cell ^{19}F NMR spectroscopy in a flow bioreactor to observe competitive binding of fluorinated benzenesulfonamide derivatives to three cytosolic isoforms of carbonic anhydrase. Quantitative measurement of the dissociation constants relative to a spy ligand allowed an accurate ranking the compounds based on their intracellular affinities for each isoform. The use of two fluorinated ligands allowed simultaneous observation of spy ligand displacement and test ligand binding, as well as estimation of the effective ratio of free ligand concentrations under poor solubility conditions. We also show that signal saturation caused by short repetition times, which can significantly impact the analysis, can be easily corrected *a posteriori*. Overall, we show that real-time in-cell ^{19}F NMR spectroscopy can reliably quantify drug–target binding in the cellular environment, paving the way for future applications in drug discovery.

Introduction

Development of new drugs is a long and expensive process, with only a limited number of drug candidates achieving regulatory approval each year. Many promising candidate drugs do not make it through the preclinical tests, and some even fail during clinical trials due to low efficacy, emerging toxic effects, or poor pharmacological properties, thus making drug development a highly risky endeavour.^{1,2} This high attrition rate outlines the need for better predictive tools to better evaluate drug molecules in the early stages of drug discovery process, in a physiologically relevant setting. An important limitation of current preclinical pipelines is that screening and optimization of novel compounds are typically performed *in vitro*, without taking into account the complexity of the cellular milieu. Indeed, the subcellular localization of the target protein, as well as and its interactions with the environment, can greatly influence the pharmacological properties of small molecules. Consequently, drug candidates that appear promising *in vitro* may not actually engage their intracellular target *in vivo*.

In this context, in-cell NMR spectroscopy has the potential to bridge the gap between traditional in vitro studies and in vivo models. When observing intracellular targets, in-cell NMR enables the investigation of macromolecular structure, dynamics, and interactions with cellular partners or exogenous ligands within living cells.^{3–5} This approach can be applied during early-stage drug evaluation, particularly in the optimization phase of lead compounds, when a limited number of candidate compounds require characterization of target engagement under near-physiological conditions.^{6,7} Alternatively, small molecules can be selectively observed as they interact with the target by exploiting the presence of fluorine atoms.^{8–10} ¹⁹F NMR is advantageous for in-cell studies as it is sensitive, reduces spectral complexity, and is less affected by the slow tumbling of the target, as internal motions of the bound ligand lead to favourable spin relaxation properties.^{11–14}

When investigating ligand binding affinity by direct titration, NMR spectroscopy is typically limited to dissociation constants (K_d) in the μ M–mM range. For compounds binding with higher affinity, such as those found in the hit-to-lead optimization phase of drug development, competitive binding can be employed.^{15,16} In this experiment a reference compound, the “spy” ligand, initially bound to the target is progressively displaced by a second “test” ligand. The displacement curve can then be fitted to obtain the K_d of the test ligand. Conducting such experiments on cell pellets sedimented in the NMR tube would require multiple independent samples, due to the limited sample lifetime, thus increasing the overall sample preparation and acquisition time. Alternatively, a flow NMR bioreactor can be employed.^{17–19} In this setup, cells are kept alive and in a metabolically active state through a continuous flow of fresh medium to provide fresh nutrients and remove metabolic byproducts, thereby enabling prolonged monitoring of intracellular processes by NMR spectroscopy. With the NMR bioreactor, competitive binding studies can be performed on a single cell sample by real-time in-cell NMR analysis. This approach has been recently applied to measure ligand binding affinity towards an intracellular target protein, both by protein- and ligand-observed in-cell NMR.^{14,20} In these experiments, a spy ligand is continuously provided to the sample at a constant concentration throughout the experiment, while the test ligand is introduced stepwise at increasing concentrations. In the protein-observed mode, competitive binding is monitored from the signals of the target protein in complex with either the spy or the test ligand, which are in slow exchange.²⁰ In the ligand-observed mode, competitive binding is monitored by the decreasing ¹⁹F signal of the spy ligand in complex with the target.¹⁴ Provided the intracellular affinity of

the spy ligand is known, this approach allows quantitative measurements of intracellular K_d in the nanomolar range.

In this study, we show that the above approach can be used for affinity ranking of fluorinated lead compounds towards different isoforms of the intracellular target. To this aim, we applied real-time in-cell ^{19}F NMR spectroscopy to quantitatively investigate the competition between a set of recently reported fluorinated benzenesulfonamide derivatives targeting carbonic anhydrase (CA).²¹ CA is an ubiquitous metalloenzyme which catalyses the reversible hydration of CO_2 in the organism and is inhibited by benzenesulfonamides, a class of small-molecule inhibitors that coordinate the active site zinc ion. We focused on three cytosolic isoforms, CA1, CA2, and CA13, which share a highly conserved active site. CA1 is primarily expressed in erythrocytes, muscle, and various organs, CA2 is ubiquitously distributed, and CA13 is found mainly in the small intestine, prostate, and colon. Using one of the fluorinated ligands as spy in competition against each of the other fluorinated ligands and correlating the spy ligand with a known reference inhibitor, we determined the intracellular K_d for each ligand relative to the reference inhibitor, thus providing quantitative rankings of the compounds for each CA isoform.

Results and discussion

A series of fluorinated CA inhibitors (**Chart M.1**) previously characterized²¹ were used as ligands for competitive binding targeting different CA isoforms. These compounds shared a benzenesulfonamide scaffold and contain one (**1-6**) or three equivalent (**7**) trifluoromethyl groups, allowing detection by ^{19}F NMR spectroscopy.

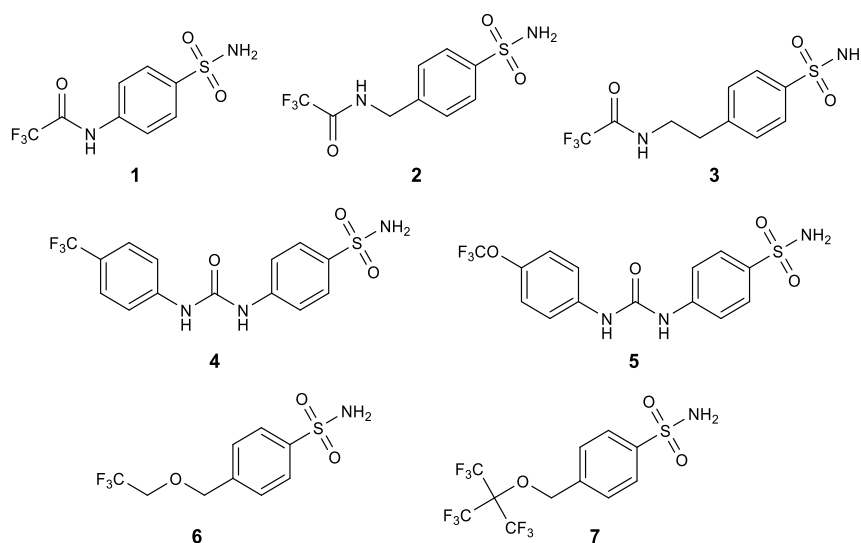


Chart M.1. Structures of the fluorinated benzenesulfonamides derivatives used in this study.

To evaluate the suitability of the ligands for competition binding assays by real-time in-cell NMR, we assessed their stability in DMEM at 37 °C for 72 hours, i.e. the timeframe of the competition binding experiments. The results showed that compounds **4-7** were stable, whereas **2** and **3** exhibited degradation profiles, likely due to the hydrolysis of the trifluoroacetamide moiety (**Figure M.1**). This is consistent with what previously observed for compound **1**, which was hydrolysed in ~ 1 hour.²¹ The slower kinetics of **2-3** makes them suitable for short-time analysis, but precludes analysis by real-time in-cell NMR. Therefore, compounds **1-3** were excluded from further analysis.

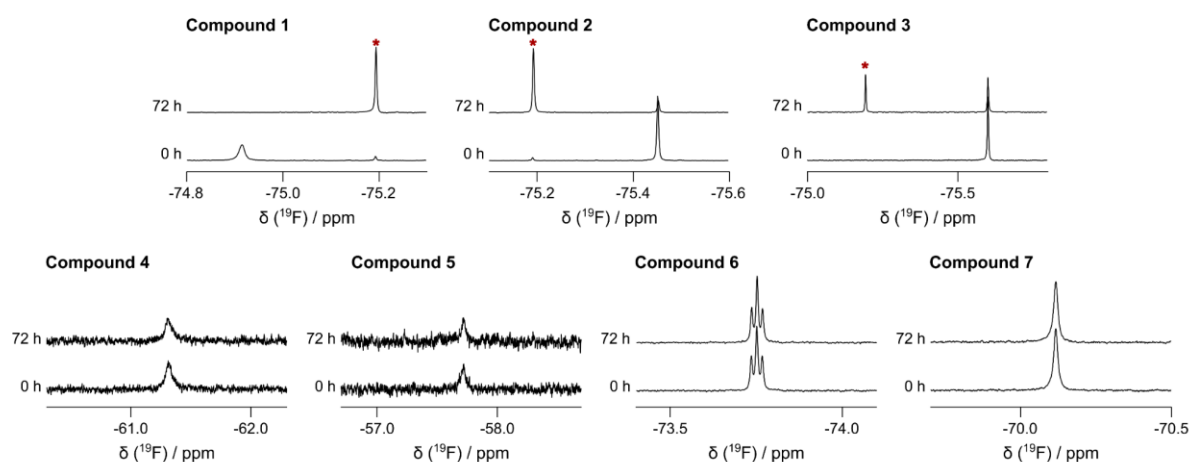


Figure M.1. Stability tests of fluorinated compounds in DMEM at 37°C. ^{19}F NMR spectra of compounds **1-7** dissolved in DMEM (supplemented with 1% antibiotics, 2% FBS, and 3% D_2O), acquired at time zero and after 72 hours at 37°C to assess ligand stability in cell culture medium. The intensities of the original compound peak and the degradation product (indicated by *) were tracked. The same degradation product was observed for compounds **1-3**, thus suggesting that the trifluoroacetamide group present in these three compounds is hydrolysed to produce trifluoroacetate.

CA1, CA2, and CA13 were selected as target CA isoforms for competitive binding experiments with fluorinated ligands **4-7**. Compound **6** was chosen as the spy ligand for real-time in-cell ^{19}F NMR experiments, as it exhibited the fastest cellular uptake, an intermediate estimated affinity in the range of various compounds for both CA2 and CA13 isoforms, and no detectable off-target interactions with other intracellular components.²¹ Compounds **4**, **5**, and **7** were employed as test ligands in the competition assays.

Competitive binding experiments were conducted using real-time ^{19}F NMR on cells expressing CA1, CA2, or CA13, with a constant concentration of compound **6** and gradually increasing the concentration of the test ligand (**Table M-S1**). In each experiment, the displacement of compound **6** was monitored as the test ligand engaged the target CA isoform. This was reflected by a progressive decrease in the ^{19}F NMR signal corresponding to the CA–compound **6** complex (CA:**6**), and by an increase in the signal corresponding to the CA–test ligand complex (CA:Test) (**Figures M.2 and M-S1**). In contrast to a previous real-time in-cell ^{19}F NMR application, in which only the target–spy ligand complex was monitored, here the presence of fluorine on both ligands allowed simultaneous tracking of the two CA–ligand complexes, thereby providing a comprehensive view of the intracellular competitive binding. Notably, the resulting affinity measurements were independent of temporal variations in intracellular protein concentration over time. Furthermore, the signals of the free compounds were monitored in the competition binding experiments, except for compounds **4** and **5** competing with compound **6** for binding to CA1 (**Figure M-S1a, b**), where the signals of free and bound test ligand overlapped and could not be quantified separately. These signals were therefore treated as a single, combined signal.

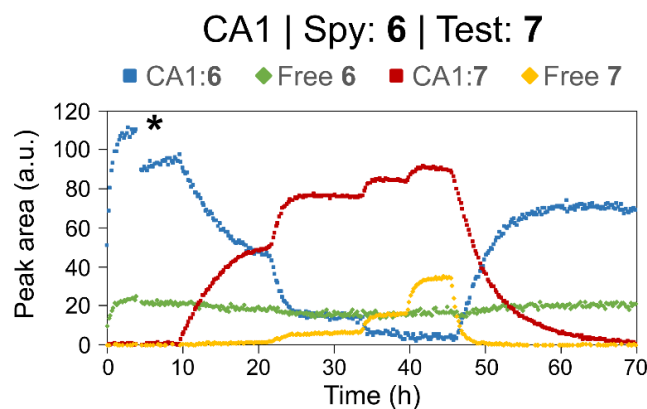


Figure M.2. Time-resolved in-cell ^{19}F NMR competition binding between compounds **6** and **7** targeting CA1 in the NMR bioreactor. The peak areas corresponding to free compound **6** (◆, green), CA1:**6** (■, blue), free test compound **7** (◆, yellow), and CA1:**7** (■, red) are plotted as a function of time. A discontinuity in the peak areas observed after ~4 hours (indicated with *) was caused by a temporary interruption of the flow due to the formation of gas bubbles, which had to be manually removed.

Comparison of the signal intensity of the free spy ligand and the free test ligands revealed discrepancies between the observed ratios and what was expected from their concentration in the reservoirs. Therefore, to obtain a more accurate estimate of the ratio between free test ligand and free spy ligand, the peak areas of the corresponding signals were used. The resulting ratios (**Table M-S2**) deviated significantly from the theoretical values for compounds **4** and **5**, especially at higher concentrations. In contrast, compound **7** showed minimal deviation, with measured ratios closely matching those calculated from the concentrations in the reservoirs. The deviation observed for compounds **4** and **5** is likely attributable to their limited aqueous solubility, which is further supported by the low signal intensity observed in the stability measurements, whereas compounds **6** and **7** at the same concentration were instead fully soluble and resulted in higher signal intensities (**Figure M.1**).

When quantifying populations by signal integration, the partial saturation of the signals caused by a short repetition time can decrease the accuracy of the measurement if the ^{19}F longitudinal relaxation time (T_1) of the two signals differs. To account for signal saturation, intracellular ^{19}F T_1 of compounds **4-7** bound to each CA isoform was measured (**Table M-S3**), and the calculated saturation factors were used to correct the corresponding peak areas, from which the molar fraction of CA bound to compound **6** (CA:**6**) was obtained (see Materials and Methods). For each competition binding experiment, the fraction of CA:**6** as a function of the ratios of free ligand concentrations was fitted with Equation 1 to determine the intracellular K_{dT}/K_{dS} ratio (**Figure M.3**). When the same fitting was performed using the uncorrected peak areas (**Figure M-S2** and **Table M-S4**), some K_{dT}/K_{dS} ratios deviated appreciably (up to 30%) from those obtained from the corrected fitting indicating that saturation effects were indeed not negligible in these experimental conditions.

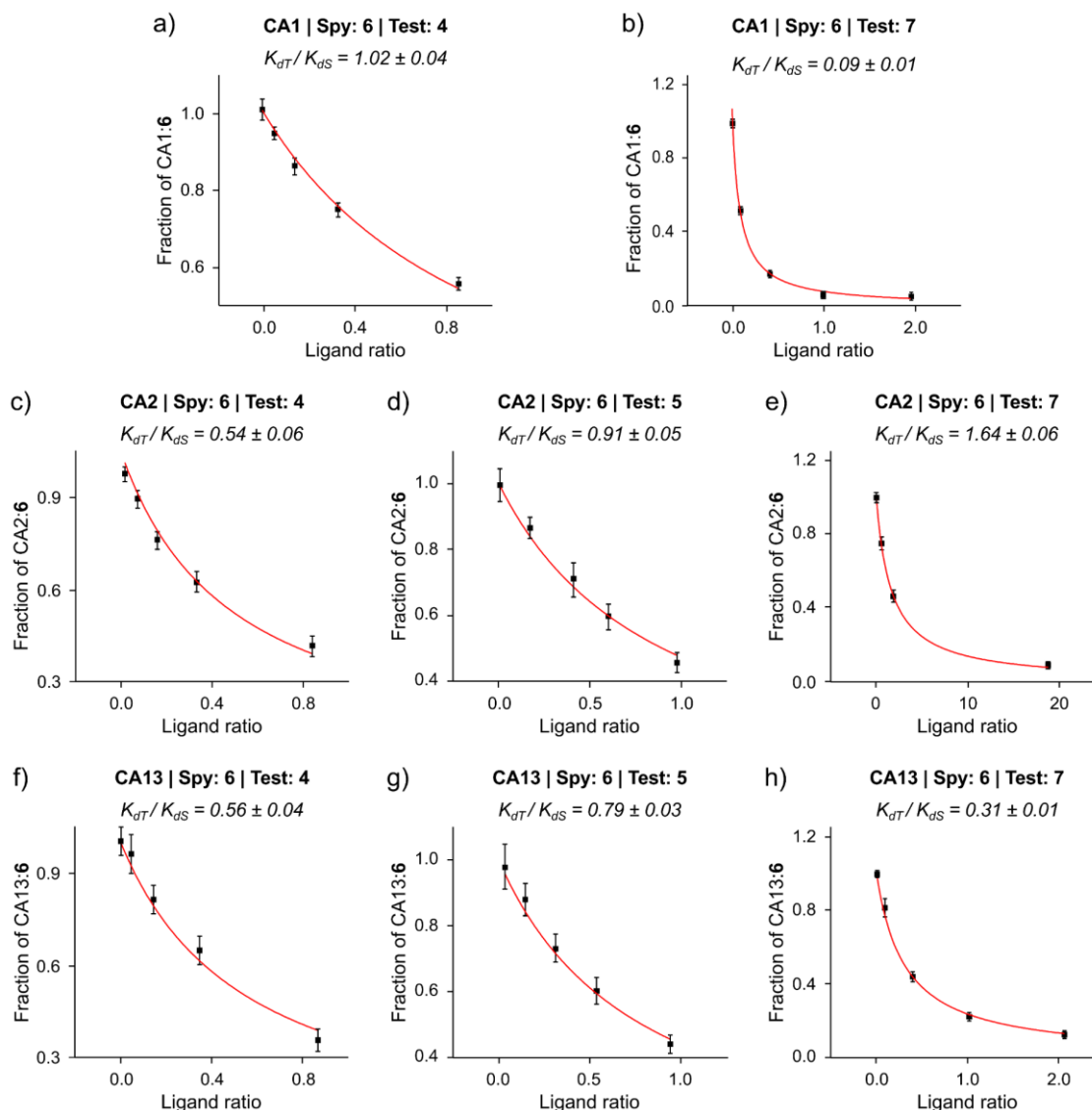


Figure M.3. Fitting of competition binding data corrected for saturation effects for each CA–ligand complex. The plots show the fraction of CA1 (a,b), CA2 (c-e), and CA13 (f-h) bound to compound 6 at the end of each concentration step during the bioreactor runs in competition with compound 4 (a,c,f), 5 (d,g) and 7 (b,e,h). Data are plotted as a function of the ratio between the free concentrations of the test compound and the spy compound. Nonlinear regression fits are shown as red curves. The resulting intracellular K_{dT}/K_{dS} values for each test compound are reported.

Table M.1. Binding affinities relative to compound 6 derived from saturation-corrected real-time in-cell ^{19}F NMR data. For each experiment, the corrected K_{dT}/K_{dS} values are reported.

Compound	K_{dT}/K_{dS} for CA1	K_i (nM) for CA1 ^a	K_{dT}/K_{dS} for CA2	K_i (nM) for CA2 ^a	K_{dT}/K_{dS} for CA13	K_i (nM) for CA13 ^a
4	$>1.02 \pm 0.04$	9.7	0.55 ± 0.06	1150	0.56 ± 0.04	153
5	<i>n.d.</i> ($\gg 1$)	142	0.91 ± 0.05	203	0.79 ± 0.03	82.0
6 (<i>spy</i>)	1	49.3	1	78.0	1	230
7	0.09 ± 0.01	81.6	1.64 ± 0.06	57.6	0.31 ± 0.01	124

^a In vitro stopped-flow assay.²¹

The resulting intracellular K_{dT}/K_{dS} ratios ranged approximately from 0.09 to 1.02 for CA1, from 0.55 to 1.64 for CA2, and from 0.31 to 0.79 for CA13 (**Table M.1**). Notably, in the competition experiment between compounds **5** and **6** targeting CA1 (**Figure M-S1b**), the signals of free compound **5** and CA1:**5** overlapped, with the complex signal appearing poorly resolved or entirely absent. Indeed, the similarity in intensity of free compounds **5** and **6** observed with the same ligands on different isoforms (**Figure M-S1e, h**) supports the conclusion that the formation of CA1:**5** was negligible and, consequently, the K_{dT}/K_{dS} ratio is $\gg 1$. In contrast, in the competition between compounds **6** and **4** on CA1, the signals of free and bound **4** also overlapped, but in this case the total signal intensity of compound **4** in the final step was clearly higher than that of free compound **6** (**Figure M-S1a**), indicating that the overlapping peak corresponded to both free **4** and the CA1:**4**. This overlap results in an overestimation of the CA1:**4** population and thus an underestimation of the corresponding K_{dT}/K_{dS} ratio.

To estimate the K_d values of these novel compounds, additional competition binding experiments were performed between the spy compound **6** and methazolamide, a well-characterized CA inhibitor. Real-time ^{19}F NMR experiments were carried out on cells expressing CA1, CA2, or CA13 (**Figure M-S3** and **Table M-S5**). The resulting K_{dT}/K_{dS} ratios were used to calculate the corresponding intracellular K_d values (**Figure M.4**).

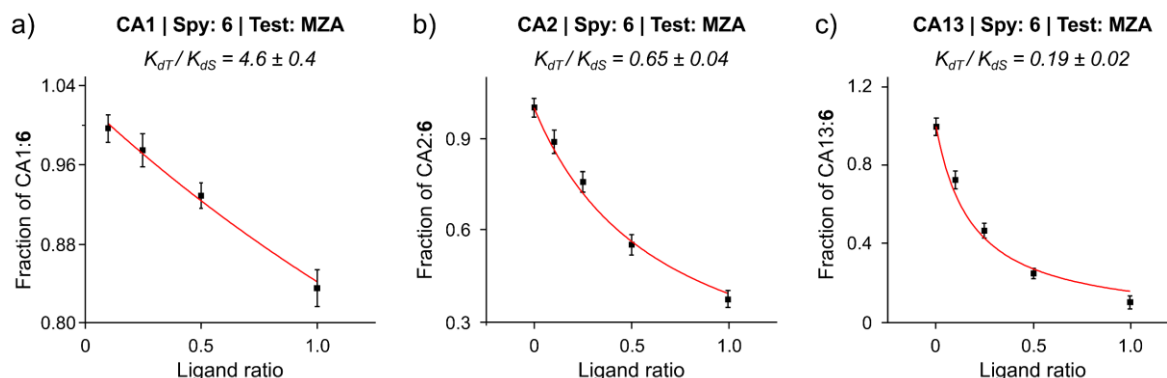


Figure M.4. Fitting of competition binding data between compound **6** and MZA for each CA isoform. The plots show the fraction of CA1 (a), CA2 (b), and CA13 (c) bound to compound **6** at the end of each concentration step during the bioreactor runs in competition with MZA. Data are plotted as a function of the ratio between the concentrations of MZA and the spy compound. Red curves represent nonlinear regression fits used to determine the intracellular K_{dT}/K_{dS} values.

As K_d values for methazolamide are not reported for all isoforms, K_I values determined in vitro (50, 14, and 19 nM for CA1, CA2, and CA13, respectively) were used as reference.²⁶ The calculated dissociation constant (K_{d_calc}) of compound **6** was first determined from its competition with methazolamide, and then used to derive the K_{d_calc} values of the other ligands based on their K_{dT}/K_{dS} ratios (**Table M.2**).

Table M.2. Calculated dissociation constant (K_{d_calc}) values of compounds **4–7** for each CA isoform, derived from the K_d ratios obtained in the real-time in-cell ^{19}F NMR competition binding experiments.

Compound	K_{d_calc} for CA1 (nM)	K_{d_calc} for CA2 (nM)	K_{d_calc} for CA13 (nM)
4	11 ± 2	12 ± 3	57 ± 16
5	n.d.	19 ± 4	80 ± 20
6	11 ± 2	22 ± 3	101 ± 21
7	1.0 ± 0.3	35 ± 7	32 ± 8

Overall, all compounds exhibited good affinity toward the three CA isoforms, except for compound **5**, which showed markedly reduced affinity for CA1. Interestingly, compound **7** displayed higher affinity for CA1, suggesting enhanced selectivity for this isoform compared to CA2 and CA13. Furthermore, the K_{d_calc} values for CA1 and CA13 were in agreement with the qualitative affinity ranking previously reported,²¹ while for CA2 compound **7** appeared weaker than predicted, likely due to the fact that all compounds displayed similar K_{d_calc} values.

Conclusion

In this study, we applied real-time in-cell ^{19}F NMR analysis to quantitatively assess competitive binding among fluorinated benzenesulfonamides targeting three cytosolic CA isoforms in human cells, using a ligand-observed approach. Once a suitable ligand was chosen as spy, competition binding experiments with the other ligands were conducted to determine the intracellular K_d ratio between the test ligand and the spy ligand, thus providing a relative ranking of the fluorinated ligands based on their intracellular binding affinities for each CA isoform. The resulting data indicated that most compounds have overall comparable binding affinities across all isoforms, with the exception of compound **5**, which showed the lowest affinity, and compound **7**, which displayed the highest affinity for the CA1 isoform.

Unlike a previous application of real-time in-cell ^{19}F NMR, in which only one ligand was monitored,¹⁴ competition binding between two fluorinated ligands allows tracking both ligands simultaneously. The observation of both ligand–target complexes allows quantifying the total amount of intracellular over the course of the bioreactor run. Doing so revealed that all three CA isoforms were slowly depleted from the cell sample, which may be caused by the intracellular proteins slowly being metabolised, leaking from the cells, or by a gradual release of cells from the agarose gel, that are then removed by the flow of medium. Accurate analysis is therefore ensured regardless of changes in target protein levels over time. Moreover, the simultaneous detection of the free fluorinated ligands enables the experimental determination of their real concentration ratio in the sample. This is especially relevant in cases where one ligand, such as compound **4** or **5** in this study, exhibits limited aqueous solubility compared to the other. A non-fluorinated test ligand — such as methazolamide — can still be analysed if needed, provided that the compound is sufficiently soluble. We also show that T_1 -related differential signal saturation cannot be ignored when short repetition times are used, but can be easily corrected *a posteriori* by measuring the T_1 of the target-ligand complexes.

Concerning the development of future isoform-selective CA inhibitors, the identification of compound **6** as a suitable spy ligand paves the way for applying this approach to assay other ligands by intracellular competitive binding assays, thus expanding the applicability of the method beyond the current set of fluorinated benzenesulfonamides. Furthermore, since compound **6** forms distinct ^{19}F NMR chemical shifts when bound to different CA isoforms, this could in principle enable simultaneous competition binding experiments across multiple CA isoforms within the same cell sample. Indeed, provided that the ^{19}F signals of the spy ligand in complex with different targets are sufficiently resolved, it should be possible to perform multiplexed assays where two test ligands are evaluated against multiple intracellular targets in parallel. Such approach would offer significant advantages in terms of experimental efficiency and resource optimization.

Experimental Section

Ligand stability assessed by ^{19}F NMR spectroscopy

Fluorinated compounds were dissolved in DMSO to a concentration of 80 mM. To evaluate their stability in cell culture medium, each compound was diluted to a final concentration of 100 μM in high-glucose Dulbecco's Modified Eagle Medium (DMEM, Gibco) supplemented with 2% foetal bovine serum (FBS, Gibco), 1% penicillin/streptomycin (100x solution, Gibco), and 3% D_2O . Sample stability was assessed by ^{19}F NMR spectroscopy immediately after preparation and after 72 hours of incubation at 37 °C. ^{19}F NMR spectra were acquired at 310 K at a 600 MHz Bruker Avance NEO spectrometer equipped with a QCI-F 5-mm Cryoprobe operating at 564.6 MHz. A single 90° pulse was used (zg Bruker pulse program), with a 1.2 s acquisition time, a frequency offset of -66.0 ppm, and a spectral width of 49.2 ppm. Each spectrum was recorded with 32 scans and a 10 s interscan delay, resulting in a total acquisition time of approximately 7 minutes per sample.

Human cell culture and transfection

HEK293T cells (ATCC CRL-3216) were cultured in uncoated 75 cm^2 flasks using high-glucose DMEM, supplemented with 10% FBS, L-glutamine (100x solution, Gibco), and 1% penicillin/streptomycin. Cells were maintained at 37°C in a humidified incubator with 5% CO_2 . For transient expression of carbonic anhydrase isoforms, cells were transfected with pHL-CA1, pHL-CA2, or pHL-CA13 plasmids obtained previously²¹ using polyethylenimine (PEI) at a

DNA:PEI mass ratio of 1:2 (25 µg DNA, 50 µg PEI), according to a previously published protocol.²²

Cell encapsulation in agarose threads

Cells were encapsulated in agarose threads as previously reported.²³ Low-gelling temperature agarose (Sigma–Aldrich) was prepared at 1.5% (w/v) in phosphate-buffered saline (PBS) by heating to 85 °C, sterilized by filtration, and stored at 4°C until use. For cell encapsulation, an aliquot of the agarose solution was melted at 85°C and maintained at 37°C. Cells harvested from a single 75 cm² flask were prewarmed at 37°C and resuspended in 450 µL of agarose solution. The resulting cell-agarose suspension was drawn into a PEEK chromatography tubing (inner diameter, i.d., 0.75 mm) connected to a 1 mL syringe and allowed to cool at room temperature for 2 minutes to solidify. The resulting agarose threads were then transferred into the flow NMR tube, prefilled with PBS.

NMR bioreactor setup and operation

The NMR bioreactor was assembled similarly to previously described.²³ Briefly, the 5 mm flow NMR tube containing the encapsulated cells (**Figure M.5A**) was connected to the tube holder of the flow unit (InsightMR 2.0, Bruker) (**Figure M.5B**). A PEEK tubing inlet (o.d. 1/32", i.d. 0.5 mm) (**Figure M.5C**) was inserted in the sample tube, while a polytetrafluoroethylene (PTFE) tubing outlet (o.d., 1/16", i.d. 1 mm) (**Figure M.5D**) was connected at the top of the tube holder. Compared to the flow unit employed previously (InsightMR 1.0, Bruker), the device used here has a wider exit channel at the top of the flow tube, which may cause sample loss. To prevent this, a 4-mm diameter circular-shaped mesh was cut from a 100-µm mesh cell strainer (Corning), pierced with a needle and slid along the PEEK inlet tubing up to the entrance of exit channel, to confine the agarose gel in the flow tube while allowing liquid and small debris to pass through without clogging. The bioreactor was maintained at 37 °C using a circulating water bath (Julabo) (**Figure M.5J**). The inlet and outlet tubing were connected through a four-way valve (**Figure M.5E**) to the pump (**Figure M.5G**) and to the waste container (**Figure M.5H**). Cells were perfused with unlabelled high-glucose DMEM (Sigma-Aldrich; powder, reconstituted in sterile-filtered Milli-Q water) supplemented with 2% FBS, 10 mM NaHCO₃, 1% penicillin/streptomycin, 3% D₂O, and one or two fluorinated compounds at defined concentration, adjusted to pH 7.4. Using two channels of a multichannel peristaltic pump (RegloICC, Ismatech), the flow from two medium reservoirs (**Figure M.5K**) was merged upstream of the valve using a PEEK tee connector (**Figure M.5F**). The culture media were

stored in 250 mL or 500 mL glass reservoirs and kept at 37 °C in the water bath throughout the experiment. The flow rates of the two channels were varied via the pump software to change the compound concentrations in a stepwise manner, keeping the total flow rate constant at 0.1 mL/min. The compound concentrations and flow profiles used at each step of the experiment are reported in Table S1. Compared to previous configurations, the bioreactor setup was modified to provide constant internal pressure throughout the system in order to minimise gas bubble formation. Specifically, a flow restrictor (**Figure M.5I**) was created by pinching the PTFE tubing at the end of the circuit, before the waste container, and two air-filled balloons were attached to each filter in the bottles containing reconstituted DMEM. These balloons apply a slight positive pressure to the reservoirs, preventing the spontaneous release of dissolved CO₂.

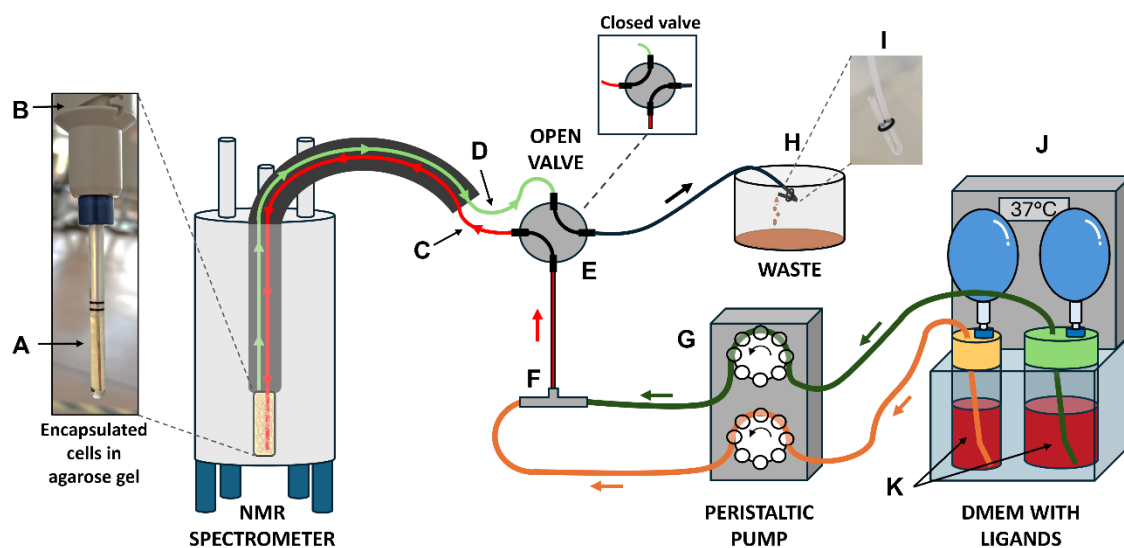


Figure M.5. Experimental setup of the NMR bioreactor.

Time-resolved in-cell ¹⁹F NMR analysis

Time-resolved ¹⁹F NMR spectra were acquired at 310 K at a 600 MHz Bruker Avance NEO spectrometer equipped with a QCI-F 5-mm Cryoprobe operating at 564.6 MHz. The ¹⁹F chemical shift scale was calibrated by setting the signal of trifluoroacetic acid in an external reference sample to −76.55 ppm (with respect to trichlorofluoromethane). A single 90° pulse was employed, followed by FID acquisition (zg Bruker pulse program) with a frequency offset of −54.3 ppm and a spectral window of 50.3 ppm. For time-resolved in cell NMR experiments recorded in the bioreactor, a series of ¹⁹F NMR spectra was acquired with 512 scans and a 1 s recycle delay, yielding a time resolution of 11 minutes per spectrum over a total acquisition

time of up to 66 hours. The resulting spectral series were processed and analysed using the Dynamics Center software (Bruker), where peak integrals were measured as a function of time.

The fraction of intracellular CA bound to each ligand was quantified by measuring the peak area of the corresponding CA–ligand complex at equilibrium, obtained by averaging over a 1-hour period for each step of the run after reaching the plateau values. The peak areas were then corrected to account for signal saturation (*vide infra*). Because 100% of the intracellular CA is bound to either of the two ligands at any moment of the run, the fraction of CA bound to the spy ligand was calculated by dividing the averaged peak area of the CA-spy complex by the sum of the averaged peak areas of both the CA–spy and CA–test complexes when both peaks were visible (bioreactor runs 1–9). In bioreactor runs 10–13, where only the spy-CA complex peak was detected, the total protein concentration was estimated as previously reported,¹⁴ i.e. by fitting the signal intensity between the final point of the first step, where all CA was bound to the spy ligand, and the point of the wash step where the CA-spy complex was fully restored. The obtained total protein concentration at each time point was then used to calculate the fraction of CA bound to the spy ligand throughout the run. For bioreactor runs 1–9, the ratio of the concentrations of free test and free spy ligand in the flow tube was calculated from the peak areas of the corresponding signals at equilibrium, averaged over a 1-hour plateau period for each step. However, in some competition binding experiments, the peak area of the free test ligand was not measurable due to overlap with the corresponding ligand–CA complex peak, such as in bioreactor run 2, where compounds **6** and **5** competed for binding to the CA1 protein. Since the free ligand ratios were found to be consistent across different runs with the same two ligands, the ratios from another run were used whenever signal overlap prevented correct signal integration. Error bars were calculated by standard error propagation from the standard deviations of the corresponding mean peak areas.

Curve fitting

Nonlinear curve fitting was performed using OriginPro 8 (OriginLab) to estimate the dissociation constant of each test ligand relative to that of the spy ligand. The fitted data points correspond to the fraction of intracellular CA bound to the spy ligand (F_S) at a defined ratio of test (T) and spy (S) ligand concentrations. K_d ratios were obtained by fitting the binding curves with the previously described equation:^{20,24}

$$F_S = \frac{1}{1 + \frac{K_{dS}[T]}{K_{dT}[S]}} \quad (1)$$

Where F_S is the fraction of CA isoform bound to S, $[T]/[S]$ is the ratio of the concentrations of free T and S, and K_{dT}/K_{dS} is the ratio of the dissociation constants of T and S.

In the NMR bioreactor setup, direct quantification of intracellular free ligand concentrations cannot be obtained, due to the low filling factor of the cells in the flow tube and, likely, due to exchange broadening from weak interactions with the cellular milieu. However, at equilibrium it is assumed that intracellular concentrations of free ligands are equal to those in the external solution, as both ligands are supplied in excess by the continuously perfused medium and are able to saturate all intracellular binding sites. Therefore, $[S]$ and $[T]$ are taken equal to the concentrations of the ligands in the perfusion medium.

T_1 measurement and saturation factor calculation

Longitudinal relaxation times (T_1) of the CA2-ligand complexes were measured by inversion recovery on HEK293T cells transiently transfected with CA2 and treated with compounds **4**, **5**, **6**, and **7**. A series of ^{19}F NMR spectra was acquired using variable interscan delays of 0.01, 0.1, 0.25, 0.5, 0.75, 1, 2, and 5 seconds. The acquired spectra were analysed using the Dynamics Center software (Bruker), and T_1 values were extracted by fitting the plot of the peak area of the CA2–ligand complex as a function of the delay time. The fitting procedure also provided the corresponding standard error for each T_1 value.

From the measured T_1 values, the saturation factor (SF) was calculated according to equation 2, which describe the ratio between the partially saturated and fully relaxed signal amplitudes, as previously reported.²⁵

$$SF(\theta, T, T_1) = \frac{(1-E_1) \sin \theta}{(1-E_1 \cos \theta)} \quad \text{where} \quad E_1 = \exp\left(-\frac{T}{T_1}\right) \quad (2)$$

Here, θ is the flip angle (90° in this case), T is the repetition time, and T_1 is the longitudinal relaxation time.

Once determined using equation 2, the saturation factor was used as a correction factor to retrieve the fully relaxed (unsaturated) signal amplitudes from the experimentally observed signals, which are partially saturated due to the use of short repetition times:

$$M_0 = M_f \cdot \hat{y} / SF(\theta, T, T_1) \quad (3)$$

Where M_0 denotes the equilibrium magnetization along the static magnetic field, M_f the magnetization vector following a flip during steady state and $\hat{y}$ denotes the unit vector along the y axis.

References

- (1) Sun, D.; Gao, W.; Hu, H.; Zhou, S. Why 90% of Clinical Drug Development Fails and How to Improve It? *Acta Pharm. Sin. B* 2022, 12 (7), 3049–3062. <https://doi.org/10.1016/j.apsb.2022.02.002>.
- (2) Waring, M. J.; Arrowsmith, J.; Leach, A. R.; Leeson, P. D.; Mandrell, S.; Owen, R. M.; Pairaudeau, G.; Pennie, W. D.; Pickett, S. D.; Wang, J.; Wallace, O.; Weir, A. An Analysis of the Attrition of Drug Candidates from Four Major Pharmaceutical Companies. *Nat. Rev. Drug Discov.* 2015, 14 (7), 475–486. <https://doi.org/10.1038/nrd4609>.
- (3) Siegal, G.; Selenko, P. Cells, Drugs and NMR. *J. Magn. Reson.* 2019, 306, 202–212. <https://doi.org/10.1016/j.jmr.2019.07.018>.
- (4) Luchinat, E.; Banci, L. In-Cell NMR: From Target Structure and Dynamics to Drug Screening. *Curr. Opin. Struct. Biol.* 2022, 74, 102374. <https://doi.org/10.1016/j.sbi.2022.102374>.
- (5) Theillet, F.-X.; Luchinat, E. In-Cell NMR: Why and How? *Prog. Nucl. Magn. Reson. Spectrosc.* 2022, 132–133, 1–112. <https://doi.org/10.1016/j.pnmrs.2022.04.002>.
- (6) 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.; Trantirek, L. Monitoring DNA-Ligand Interactions in Living Human Cells Using NMR Spectroscopy. *J. Am. Chem. Soc.* 2019, 141 (34), 13281–13285. <https://doi.org/10.1021/jacs.9b03031>.
- (7) Luchinat, E.; Barbieri, L.; Cremonini, M.; Nocentini, A.; Supuran, C. T.; Banci, L. Drug Screening in Human Cells by NMR Spectroscopy Allows the Early Assessment of Drug Potency. *Angew. Chem. Int. Ed.* 2020, 59 (16), 6535–6539. <https://doi.org/10.1002/anie.201913436>.
- (8) Divakaran, A.; Kirberger, S. E.; Pomerantz, W. C. K. SAR by (Protein-Observed) ^{19}F NMR. *Acc. Chem. Res.* 2019, 52 (12), 3407–3418. <https://doi.org/10.1021/acs.accounts.9b00377>.

-
- (9) Buchholz, C. R.; Pomerantz, W. C. K. ^{19}F NMR Viewed through Two Different Lenses: Ligand-Observed and Protein-Observed ^{19}F NMR Applications for Fragment-Based Drug Discovery. *RSC Chem. Biol.* 2021, 2 (5), 1312–1330. <https://doi.org/10.1039/D1CB00085C>.
- (10) Rüdisser, S. H.; Goldberg, N.; Ebert, M.-O.; Kovacs, H.; Gossert, A. D. Efficient Affinity Ranking of Fluorinated Ligands by ^{19}F NMR: CSAR and FastCSAR. *J. Biomol. NMR* 2020, 74 (10), 579–594. <https://doi.org/10.1007/s10858-020-00325-x>.
- (11) Zhu, W.; Guseman, A. J.; Bhinderwala, F.; Lu, M.; Su, X.-C.; Gronenborn, A. M. Visualizing Proteins in Mammalian Cells by ^{19}F NMR Spectroscopy. *Angew. Chem. Int. Ed.* 2022, 61 (23), e202201097. <https://doi.org/10.1002/anie.202201097>.
- (12) Pham, L. B. T.; Costantino, A.; Barbieri, L.; Calderone, V.; Luchinat, E.; Banci, L. Direct Expression of Fluorinated Proteins in Human Cells for ^{19}F In-Cell NMR Spectroscopy. *J. Am. Chem. Soc.* 2023, 145 (2), 1389–1399. <https://doi.org/10.1021/jacs.2c12086>.
- (13) Chai, Z.; Li, C. In-Cell ^{19}F NMR of Proteins: Recent Progress and Future Opportunities. *Chem. – Eur. J.* 2024, 30 (23), e202303988. <https://doi.org/10.1002/chem.202303988>.
- (14) Luchinat, E.; Barbieri, L.; Davis, B.; Brough, P. A.; Pennestri, M.; Banci, L. Ligand-Based Competition Binding by Real-Time ^{19}F NMR in Human Cells. *J. Med. Chem.* 2024, 67 (2), 1115–1126. <https://doi.org/10.1021/acs.jmedchem.3c01600>.
- (15) Dalvit, C.; Gmür, I.; Rößler, P.; Gossert, A. D. Affinity Measurement of Strong Ligands with NMR Spectroscopy: Limitations and Ways to Overcome Them. *Prog. Nucl. Magn. Reson. Spectrosc.* 2023, 138–139, 52–69. <https://doi.org/10.1016/j.pnmrs.2023.07.001>.
- (16) Dalvit, C.; Flocco, M.; Knapp, S.; Mostardini, M.; Perego, R.; Stockman, B. J.; Veronesi, M.; Varasi, M. High-Throughput NMR-Based Screening with Competition Binding Experiments. *J. Am. Chem. Soc.* 2002, 124 (26), 7702–7709. <https://doi.org/10.1021/ja020174b>.
- (17) Sharaf, N. G.; Barnes, C. O.; Charlton, L. M.; Young, G. B.; Pielak, G. J. A Bioreactor for In-Cell Protein NMR. *J. Magn. Reson.* 2010, 202 (2), 140–146. <https://doi.org/10.1016/j.jmr.2009.10.008>.
- (18) Burz, D. S.; Breindel, L.; Shekhtman, A. Chapter Seventeen - Improved Sensitivity and Resolution of in-Cell NMR Spectra. In *Methods in Enzymology*; Shukla, A. K., Ed.;

- Chemical and Synthetic Biology Approaches To Understand Cellular Functions - Part A; Academic Press, 2019; Vol. 621, pp 305–328. <https://doi.org/10.1016/bs.mie.2019.02.029>.
- (19) Luchinat, E.; Barbieri, L.; Campbell, T. F.; Banci, L. Real-Time Quantitative In-Cell NMR: Ligand Binding and Protein Oxidation Monitored in Human Cells Using Multivariate Curve Resolution. *Anal. Chem.* 2020, 92 (14), 9997–10006. <https://doi.org/10.1021/acs.analchem.0c01677>.
- (20) Luchinat, E.; Barbieri, L.; Cremonini, M.; Pennestri, M.; Nocentini, A.; Supuran, C. T.; Banci, L. Determination of Intracellular Protein–Ligand Binding Affinity by Competition Binding in-Cell NMR. *Acta Crystallogr. Sect. Struct. Biol.* 2021, 77 (10), 1270–1281. <https://doi.org/10.1107/S2059798321009037>.
- (21) Azzurra Costantino; Letizia Barbieri, Simone Giovannuzzi, Alessio Nocentini, Claudio T. Supuran, Mindaugas Raitelaitis, Pär Nordlund, Enrico Luchinat, Lucia Banci. Intracellular Binding of Novel Fluorinated Compounds to Carbonic Anhydrase Isoforms Explored by In-Cell 19F NMR. Submitted. <https://doi.org/10.1021/acs.jmedchem.5c02227>.
- (22) Barbieri, L.; Luchinat, E.; Banci, L. Characterization of Proteins by In-Cell NMR Spectroscopy in Cultured Mammalian Cells. *Nat. Protoc.* 2016, 11 (6), 1101–1111. <https://doi.org/10.1038/nprot.2016.061>.
- (23) Barbieri, L.; Luchinat, E. Monitoring Protein-Ligand Interactions in Human Cells by Real-Time Quantitative In-Cell NMR Using a High Cell Density Bioreactor. *J. Vis. Exp.* 2021, No. 169, 62323. <https://doi.org/10.3791/62323>.
- (24) Dalvit, C.; Parent, A.; Vallée, F.; Mathieu, M.; Rak, A. Fast NMR Methods for Measuring in the Direct and/or Competition Mode the Dissociation Constants of Chemical Fragments Interacting with a Receptor. *ChemMedChem* 2019, 14 (11), 1115–1127. <https://doi.org/10.1002/cmdc.201900152>.
- (25) Spencer, R. G. S.; Ferretti, J. A.; Weiss, G. H. NMR Saturation Factors in the Presence of Chemical Exchange. *J. Magn. Reson.* 1969 1989, 84 (2), 223–235. [https://doi.org/10.1016/0022-2364\(89\)90366-1](https://doi.org/10.1016/0022-2364(89)90366-1).

- (26) Biswas, S.; Aggarwal, M.; Güzel, Ö.; Scozzafava, A.; McKenna, R.; Supuran, C. T. Conformational Variability of Different Sulfonamide Inhibitors with Thienyl-Acetamido Moieties Attributes to Differential Binding in the Active Site of Cytosolic Human Carbonic Anhydrase Isoforms. *Bioorg. Med. Chem.* 2011, 19 (12), 3732–3738. <https://doi.org/10.1016/j.bmc.2011.05.006>.

Supplementary Information

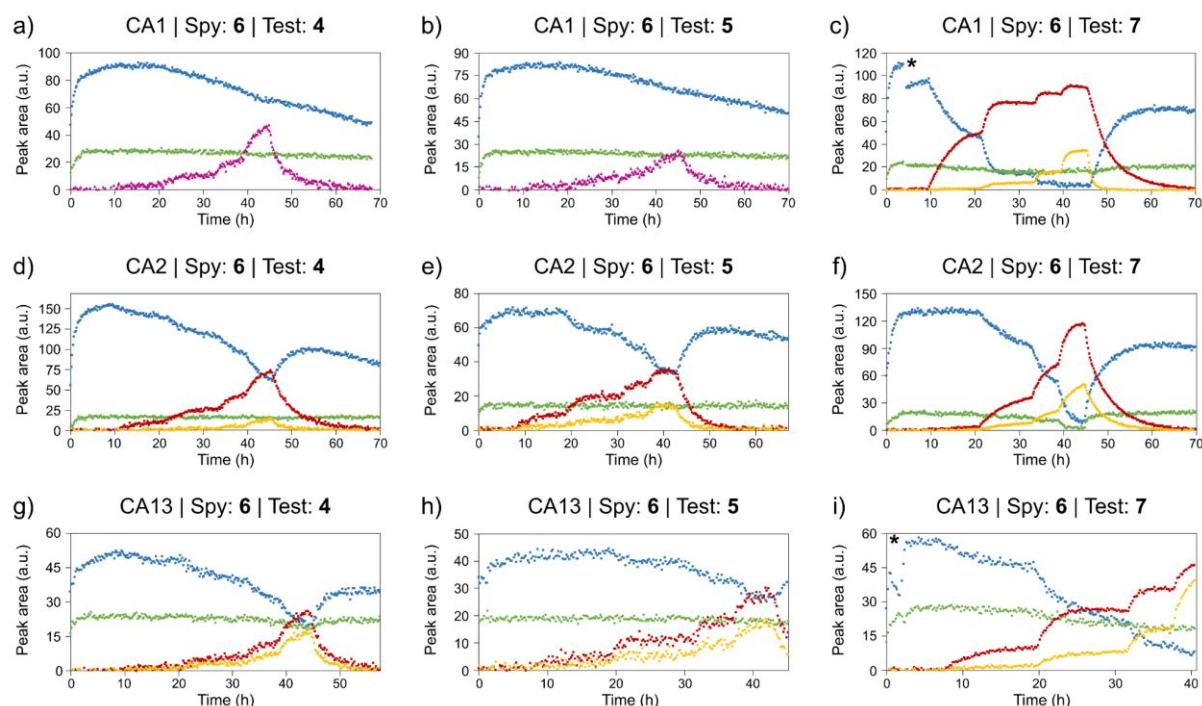


Figure M-S1. Time-resolved in-cell ^{19}F NMR competition binding experiments. Competition assays were performed in the NMR bioreactor with compound **6** as the constant spy ligand and compounds **4** (a,d,g), **5** (b,e,h), and **7** (c,f,i) as test ligands, targeting CA1 (a–c), CA2 (d–f), and CA13 (g–i). Peak areas corresponding to free compound **6** ($\blacklozenge$, green), CA:**6** ($\blacksquare$, blue), free test compound ($\blacklozenge$, yellow), and CA:Test ($\blacksquare$, red) are plotted as a function of time. In bioreactor runs 1 (a) and 2 (b), the free and bound signals of the test compound could not be separated and were therefore treated as a single combined signal ($\blacktriangle$, pink). Theoretical concentrations of compounds **6** and test ligands at each step of the bioreactor runs are reported in Table M-S1. A discontinuity in the peak areas observed for bioreactor runs 3 (c) and 9 (i), indicated with *, was caused by the formation of an air bubble, which required a temporary interruption of the flow to remove it before resuming the experiment.

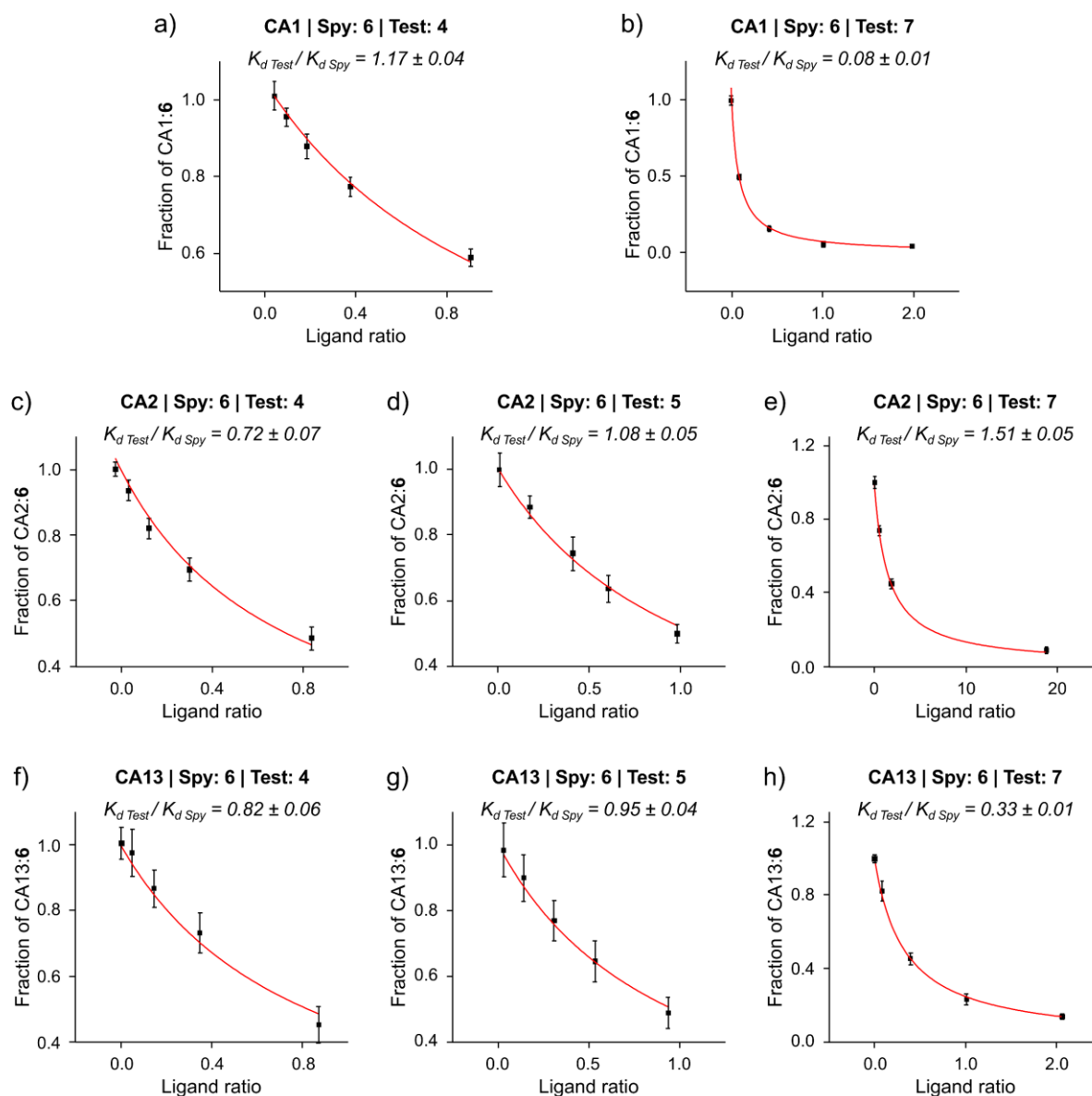


Figure M-S2. Fitting of competition binding data before saturation factor correction. Plots show the fraction of CA1 (a,b), CA2 (c–e), and CA13 (f–h) isoforms bound to compound 6 at the end of each concentration step during the bioreactor runs, plotted as a function of the ratio between the free concentrations of the test compounds 4 (a,c,f), 5 (d,g) and 7 (b,e,h) and the spy compound 6. Nonlinear regression fits are shown as red curves. The resulting intracellular $K_{d\text{T}}/K_{d\text{S}}$ values for each test compound are reported in the corresponding panels.

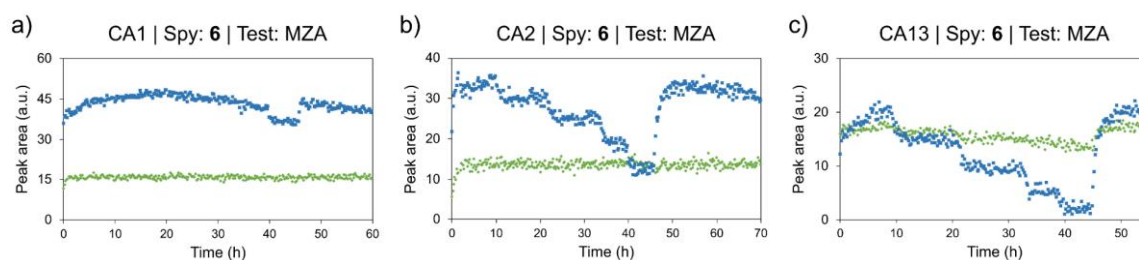


Figure M-S3. Time-resolved in-cell ^{19}F NMR competition binding experiments with MZA. Competition assays were performed in the NMR bioreactor with compound **6** as the constant spy ligand and methazolamide, targeting CA1 (a), CA2 (b), and CA13 (c). Peak areas corresponding to free compound **6** (◆, green) and CA:**6** (■, blue). Concentrations of compounds **6** and test ligands at each step of the bioreactor runs are reported in Table M-S5.

Table M-S1. Durations and ligand concentrations of each step of the bioreactor runs reported in Figure M-S1. For each channel, ligand concentrations in the reservoir and flow rates are also reported.

Bioreactor run 1 (Figure M-S1a) – Target: CA1 – Channel 1: 6 (10 μM); Channel 2: 6 (10 μM) + 4 (20 μM)					
Step	Duration (h)	6 (μM)	4 (μM)	Channel 1 flow rate ($\mu\text{L}/\text{min}$)	Channel 2 flow rate ($\mu\text{L}/\text{min}$)
1	8	10	0	100	0
2	12	10	2	90	10
3	12	10	5	75	25
4	6	10	10	50	50
5	6	10	20	0	100
Bioreactor run 2 (Figure M-S1b) – Target: CA1 – Channel 1: 6 (10 μM); Channel 2: 6 (10 μM) + 5 (20 μM)					
Step	Duration (h)	6 (μM)	5 (μM)	Channel 1 flow rate ($\mu\text{L}/\text{min}$)	Channel 2 flow rate ($\mu\text{L}/\text{min}$)
1	7	10	0	100	0
2	12	10	2	90	10
3	12	10	5	75	25
4	6	10	10	50	50
5	6	10	20	0	100

Bioreactor run 3 (Figure M-S1c) – Target: CA1 – Channel 1: 6 (10 μM); Channel 2: 6 (10 μM) + 7 (20 μM)					
Step	Duration (h)	6 (μ M)	7 (μ M)	Channel 1 flow rate (μ L/min)	Channel 2 flow rate (μ L/min)
1	9	10	0	100	0
2	12	10	2	90	10
3	12	10	5	75	25
4	6	10	10	50	50
5	6	10	20	0	100
Bioreactor run 4 (Figure M-S1d) – Target: CA2 – Channel 1: 6 (10 μM); Channel 2: 6 (10 μM) + 4 (20 μM)					
Step	Duration (h)	6 (μ M)	4 (μ M)	Channel 1 flow rate (μ L/min)	Channel 2 flow rate (μ L/min)
1	9	10	0	100	0
2	12	10	2	90	10
3	12	10	5	75	25
4	6	10	10	50	50
5	6	10	20	0	100
Bioreactor run 5 (Figure M-S1e) – Target: CA2 – Channel 1: 6 (10 μM); Channel 2: 6 (10 μM) + 5 (20 μM)					
Step	Duration (h)	6 (μ M)	5 (μ M)	Channel 1 flow rate (μ L/min)	Channel 2 flow rate (μ L/min)
1	6	10	0	100	0
2	12	10	2	90	10
3	12	10	5	75	25
4	6	10	10	50	50
5	6	10	20	0	100
Bioreactor run 6 (Figure M-S1f) – Target: CA2 – Channel 1: 6 (10 μM); Channel 2: 7 (20 μM)					
Step	Duration (h)	6 (μ M)	7 (μ M)	Channel 1 flow rate (μ L/min)	Channel 2 flow rate (μ L/min)
1	7	10	0	100	0
2	12	9	2	90	10
3	12	7,5	5	75	25
4	6	5	10	50	50
5	6	0	20	0	100

Bioreactor run 7 (Figure M-S1g) – Target: CA13 - Channel 1: 6 (10 μM); Channel 2: 6 (10 μM) + 4 (20 μM)					
Step	Duration (h)	6 (μM)	4 (μM)	Channel 1 flow rate (μL/min)	Channel 2 flow rate (μL/min)
1	9	10	0	100	0
2	12	10	2	90	10
3	12	10	5	75	25
4	6	10	10	50	50
5	6	10	20	0	100
Bioreactor run 8 (Figure M-S1h) – Target: CA13 - Channel 1: 6 (10 μM); Channel 2: 6 (10 μM) + 5 (20 μM)					
Step	Duration (h)	6 (μM)	5 (μM)	Channel 1 flow rate (μL/min)	Channel 2 flow rate (μL/min)
1	8	10	0	100	0
2	12	10	2	90	10
3	12	10	5	75	25
4	6	10	10	50	50
5	6	10	20	0	100
Bioreactor run 9 (Figure M-S1i) – Target: CA13 – Channel 1: 6 (10 μM); Channel 2: 6 (10 μM) + 7 (20 μM)					
Step	Duration (h)	6 (μM)	7 (μM)	Channel 1 flow rate (μL/min)	Channel 2 flow rate (μL/min)
1	7	10	0	100	0
2	12	10	2	90	10
3	12	10	5	75	25
4	6	10	10	50	50
5	6	10	20	0	100

Table M-S2. Free ligand concentration ratios between the test ligand and the spy ligand. The ratios are calculated at the plateau of each concentration step (averaged over a 1-hour interval). The corresponding theoretical ratios, based on the nominal concentrations in the reservoir, are also reported for comparison.

Bioreactor runs 1, 2, and 3 with CA1 target (Figure M-S1a, b, c)				
Step n°	Theoretical ratio [T]/[S]	Calculated [4]/[6] for CA1 *	Calculated [5]/[6] for CA1**	Calculated [7]/[6] for CA1
1	0	-0,01	0,02	-0,01
2	0.2	0,04	0,16	0,08
3	0.5	0,14	0,36	0,41
4	1	0,32	0,57	1,01
5	2	0,85	0,96	1,98
Bioreactor runs 4, 5, and 6 with CA2 target (Figure M-S1d, e, f)				
Step n°	Theoretical ratio [T]/[S]	Calculated [4]/[6] for CA2	Calculated [5]/[6] for CA2	Calculated [7]/[6] for CA2
1	0	-0,02	0,01	0,01
2	0.2	0,04	0,17	0,03
3	0.5	0,12	0,41	0,53
4	1	0,30	0,61	1,84
5	2	0,83	0,98	18,84
Bioreactor runs 7, 8, and 9 with CA1 target (Figure M-S1g, h, i)				
Step n°	Theoretical ratio [T]/[S]	Calculated [4]/[6] for CA13	Calculated [5]/[6] for CA13	Calculated [7]/[6] for CA13
1	0	0	0,03	0,01
2	0.2	0,05	0,14	0,09
3	0.5	0,15	0,31	0,40
4	1	0,35	0,53	1,01
5	2	0,87	0,93	2,06

* Values calculated from the average [4]/[6] ratios for CA2 and CA13.

** Values calculated from the average [5]/[6] ratios for CA2 and CA13.

Table M-S3. Longitudinal relaxation times (T_1) and corresponding saturation factors (SF) for each CA–ligand complex. T_1 values were determined using Equation 1, and the associated saturation factors were calculated using Equation 2. These SF values were subsequently used to correct the peak areas of the CA–ligand complexes prior to nonlinear fitting and determination of the K_{dT}/K_{dS} ratios.

T_1 and SF for CA1-compound complex		
Compound	T_1 (s)	SF
4	1.26 ± 0.13	0.64 ± 0.04
5	1.00 ± 0.06	0.73 ± 0.02
6	0.98 ± 0.09	0.73 ± 0.03
7	0.78 ± 0.02	0.81 ± 0.01
T_1 and SF for CA2-compound complex		
Compound	T_1 (s)	SF
4	1.30 ± 0.15	0.64 ± 0.04
5	1.07 ± 0.09	0.70 ± 0.03
6	0.73 ± 0.02	0.83 ± 0.01
7	0.55 ± 0.01	0.91 ± 0.01
T_1 and SF for CA13-compound complex		
Compound	T_1 (s)	SF
4	1.30 ± 0.12	0.64 ± 0.04
5	0.87 ± 0.04	0.77 ± 0.01
6	0.48 ± 0.03	0.93 ± 0.01
7	0.61 ± 0.03	0.88 ± 0.01

Table M-S4. Affinity Constants for CA1, CA2, and CA13 derived from real-time in-cell ^{19}F NMR Data. For each experiment, the K_{dT}/K_{dS} values are reported.

Compound	K_{dT}/K_{dS} for CA1 by in-cell NMR	K_I (nM) for CA1 ^a	K_{dT}/K_{dS} for CA2 by in-cell NMR	K_I (nM) for CA2 ^a	K_{dT}/K_{dS} for CA13 by in-cell NMR	K_I (nM) for CA13 ^a
4	1.17 ± 0.04	9.7	0.72 ± 0.07	1150	0.82 ± 0.06	153
5	$\gg 1$	142	1.08 ± 0.05	203	0.95 ± 0.04	82.0
6 (spy)	1	49.3	1	78.0	1	230
7	0.08 ± 0.01	81.6	1.51 ± 0.05	57.6	0.33 ± 0.01	124

^a K_I taken from Costantino et al.¹

Table M-S5. Durations and ligand concentrations of each step of the bioreactor runs for competition binding experiments between compound 6 and methazolamide reported in Figure M-S3. For each channel, ligand concentrations in the reservoir and flow rates are also reported.

Bioreactor run 10 (Figure M-S3a) – Target: CA1 – Channel 1: 6 (10 μM); Channel 2: 6 (10 μM) + MZA (10 μM)				
Duration (h)	6 (μM)	MZA (μM)	Channel 1 flow rate ($\mu\text{L}/\text{min}$)	Channel 2 flow rate ($\mu\text{L}/\text{min}$)
8	10	0	100	0
12	10	1	90	10
12	10	2.5	75	25
6	10	5	50	50
6	10	10	0	100
Bioreactor run 11 (Figure M-S3b) – Target: CA2 – Channel 1: 6 (10 μM); Channel 2: 6 (10 μM) + MZA (10 μM)				
Duration (h)	6 (μM)	MZA (μM)	Channel 1 flow rate ($\mu\text{L}/\text{min}$)	Channel 2 flow rate ($\mu\text{L}/\text{min}$)
7	10	0	100	0
12	10	2	90	10
12	10	5	75	25
6	10	10	50	50
6	10	20	0	100

Bioreactor run 12 (Figure M-S3c) – Target: CA13 – Channel 1: 6 (10 μM); Channel 2: 6 (10 μM) + MZA (10 μM)				
Duration (h)	6 (μM)	MZA (μM)	Channel 1 flow rate (μL/min)	Channel 2 flow rate (μL/min)
9	10	0	100	0
12	10	2	90	10
12	10	5	75	25
6	10	10	50	50
6	10	20	0	100

References

- (1) Azzurra Costantino; Letizia Barbieri, Simone Giovannuzzi, Alessio Nocentini, Claudio T. Supuran, Mindaugas Raitelaitis, Pär Nordlund, Enrico Luchinat, Lucia Banci. Intracellular Binding of Novel Fluorinated Compounds to Carbonic Anhydrase Isoforms Explored by In-Cell ^{19}F NMR. *J.Med.Chem* **2025**. <https://doi.org/10.1021/acs.jmedchem.5c02227>.

CHAPTER 4

CONCLUSION AND PERSPECTIVE

In this PhD project, in-cell ^{19}F NMR was implemented as a tool for developing new drug candidates. Two distinct strategies were considered: the target-based approach and the ligand-based approach. The target-based approach focus on controlling the incorporation of aromatic amino acids into proteins expressed directly in human cells, using carbonic anhydrase 2 (CA2), a ubiquitous isoform, as a model protein. It was observed that incorporation of fluorinated amino acids could not be achieved at 100%, and the replacement of natural amino acids with fluorinated counterparts occurred stochastically, independent on the amino acid position along. Previous work combining NMR spectroscopy, mass spectrometry, and X-ray crystallography demonstrated that fluorinated aromatic amino acid incorporation follows a predictable pattern dependent on FAA concentration in the expression medium, with a maximum efficiency of 60–70%. Agreement between NMR- and MS-derived incorporation levels indicates that NMR can reliably estimate FAA incorporation, whereas X-ray crystallography further revealed that the presence of fluorine at specific positions can induce subtle conformational changes, leading to selective exclusion of certain protein forms during crystallization. This incorporation method could be exploited to express fluorinated proteins in human cells under controlled conditions and to investigate their interactions with potential pharmacological drugs. For proteins with multiple fluorinated residues, signal assignment is required, allowing monitoring of chemical shift changes of specific peaks, which can also provide information on the ligand binding site.

On the other hand, the ligand-observed in-cell ^{19}F NMR approach was successfully employed to study the interaction between cytosolic CA isoforms and novel fluorinated benzenesulfonamides. Although this strategy does not provide direct structural information on the target, it offers several advantages, as it enables the investigation of proteins that are otherwise NMR-invisible in cells when using conventional isotopic labelling and ^1H – ^{15}N HSQC experiments, such as the CA7 and CA13 isoforms reported in this work. In this context, the binding of newly synthesised fluorinated benzenesulfonamide-based inhibitors targeting cytosolic CA isoforms was characterised by in-cell ^{19}F NMR spectroscopy, providing a physiologically relevant binding assay compared to classical in-vitro studies. Among the five CA isoforms analysed, CA1, CA2, CA7, and CA13 clearly exhibited binding interactions with the fluorinated compounds, whereas CA3 showed no interaction, as expected due to the steric hindrance in its active site, thus confirming the robustness of the method. Uptake kinetics analysis by real-time NMR further identified the best candidate for use as a “spy” ligand in competitive binding experiments. These experiments between the fluorinated spy ligand and the other fluorinated compounds provided quantitative K_d ratios for the different CA isoforms.

Furthermore, this analysis was extended to the well-known drug methazolamide, which was used to determine first the K_d of the spy ligand and then those of all other compounds, based on the K_d ratios obtained from each competition binding experiment. This allowed the establishment of a quantitative ranking of intracellular binding affinities for the different CA isoforms. Thus, direct spectroscopic observation of drug-target interactions provides a valuable validation tool that complements in-silico predictions and traditional in-vitro studies. In the future, it would be interesting to develop experiments capable of simultaneously monitoring multiple CA isoforms in the same sample. If ^{19}F NMR signals of complexes between the spy ligand and different drug targets were sufficiently distinct and resolved, it might be possible to perform parallel competition binding experiments with a single measurement and evaluate the test ligands against multiple intracellular targets simultaneously. This would expand the applicability of the method, offering advantages in terms of experimental efficiency and resource optimisation and expanding the potential of in-cell ^{19}F NMR in drug discovery.

References

List of references

1. Luchinat E, Banci L. In-cell NMR: a topical review. *IUCrJ*. 2017;4(2):108-118. doi:10.1107/S2052252516020625
2. Reckel S, Hänsel R, Löhr F, Dötsch V. In-cell NMR spectroscopy. *Prog Nucl Magn Reson Spectrosc*. 2007;51(2):91-101. doi:10.1016/j.pnmrs.2007.02.002
3. Rivas G, Minton AP. Macromolecular Crowding In Vitro, In Vivo, and In Between. *Trends Biochem Sci*. 2016;41(11):970-981. doi:10.1016/j.tibs.2016.08.013
4. Serber Z, Corsini L, Durst F, Dötsch V. In-Cell NMR Spectroscopy. In: *Methods in Enzymology*. Vol 394. Nuclear Magnetic Resonance of Biological Macromolecules. Academic Press; 2005:17-41. doi:10.1016/S0076-6879(05)94002-0
5. Theillet FX, Binolfi A, Frembgen-Kesner T, et al. Physicochemical Properties of Cells and Their Effects on Intrinsically Disordered Proteins (IDPs). *Chem Rev*. 2014;114(13):6661-6714. doi:10.1021/cr400695p
6. Theillet FX. In-Cell Structural Biology by NMR: The Benefits of the Atomic Scale. *Chem Rev*. 2022;122(10):9497-9570. doi:10.1021/acs.chemrev.1c00937
7. Li C, Zhao J, Cheng K, et al. Magnetic Resonance Spectroscopy as a Tool for Assessing Macromolecular Structure and Function in Living Cells. *Annu Rev Anal Chem*. 2017;10(Volume 10, 2017):157-182. doi:10.1146/annurev-anchem-061516-045237
8. Ohno A, Inomata K, Tochio H, Shirakawa M. In-Cell NMR Spectroscopy in Protein Chemistry and Drug Discovery. *Curr Top Med Chem*. 2011;11(1):68-73. doi:10.2174/156802611793611878
9. Serber Z, Keatinge-Clay AT, Ledwidge R, Kelly AE, Miller SM, Dötsch V. High-Resolution Macromolecular NMR Spectroscopy Inside Living Cells. *J Am Chem Soc*. 2001;123(10):2446-2447. doi:10.1021/ja0057528
10. Sakai T, Tochio H, Tenno T, et al. In-cell NMR spectroscopy of proteins inside *Xenopus laevis* oocytes. *J Biomol NMR*. 2006;36(3):179-188. doi:10.1007/s10858-006-9079-9

11. Selenko P, Serber Z, Gadea B, Ruderman J, Wagner G. Quantitative NMR analysis of the protein G B1 domain in *Xenopus laevis* egg extracts and intact oocytes. *Proc Natl Acad Sci U S A*. 2006;103(32):11904-11909. doi:10.1073/pnas.0604667103
12. Binolfi A, Limatola A, Verzini S, et al. Intracellular repair of oxidation-damaged α -synuclein fails to target C-terminal modification sites. *Nat Commun*. 2016;7:10251. doi:10.1038/ncomms10251
13. Bertrand K, Reverdatto S, Burz DS, Zitomer R, Shekhtman A. Structure of proteins in eukaryotic compartments. *J Am Chem Soc*. 2012;134(30):12798-12806. doi:10.1021/ja304809s
14. Hamatsu J, O'Donovan D, Tanaka T, et al. High-resolution heteronuclear multidimensional NMR of proteins in living insect cells using a baculovirus protein expression system. *J Am Chem Soc*. 2013;135(5):1688-1691. doi:10.1021/ja310928u
15. Aricescu AR, Lu W, Jones EY. A time- and cost-efficient system for high-level protein production in mammalian cells. *Acta Crystallogr D Biol Crystallogr*. 2006;62(10):1243-1250. doi:10.1107/S0907444906029799
16. Guin D, Gruebele M. Weak Chemical Interactions That Drive Protein Evolution: Crowding, Sticking, and Quinary Structure in Folding and Function. *Chem Rev*. 2019;119(18):10691-10717. doi:10.1021/acs.chemrev.8b00753
17. Kadavath H, Cecilia Prymaczok N, Eichmann C, Riek R, Gerez JA. Multi-Dimensional Structure and Dynamics Landscape of Proteins in Mammalian Cells Revealed by In-Cell NMR. *Angew Chem Int Ed*. 2023;62(4):e202213976. doi:10.1002/anie.202213976
18. Banci L, Barbieri L, Bertini I, et al. Atomic-resolution monitoring of protein maturation in live human cells by NMR. *Nat Chem Biol*. 2013;9(5):297-299. doi:10.1038/nchembio.1202
19. Banci L, Barbieri L, Luchinat E, Secci E. Visualization of Redox-Controlled Protein Fold in Living Cells. *Chem Biol*. 2013;20(6):747-752. doi:10.1016/j.chembiol.2013.05.007
20. Luchinat E, Barbieri L, Cremonini M, Nocentini A, Supuran CT, Banci L. Drug Screening in Human Cells by NMR Spectroscopy Allows the Early Assessment of Drug Potency. *Angew Chem Int Ed*. 2020;59(16):6535-6539. doi:10.1002/anie.201913436
21. Kang C. Applications of In-Cell NMR in Structural Biology and Drug Discovery. *Int J Mol Sci*. 2019;20(1):139. doi:10.3390/ijms20010139
22. Theillet FX, Luchinat E. In-cell NMR: Why and how? *Prog Nucl Magn Reson Spectrosc*. 2022;132-133:1-112. doi:10.1016/j.pnmrs.2022.04.002
23. Luchinat E, Barbieri L, Campbell TF, Banci L. Real-Time Quantitative In-Cell NMR: Ligand Binding and Protein Oxidation Monitored in Human Cells Using Multivariate Curve Resolution. *Anal Chem*. 2020;92(14):9997-10006. doi:10.1021/acs.analchem.0c01677

24. Narayan KS, Moress EA, Chatham JC, Barker PB. ^{31}P NMR of mammalian cells encapsulated in alginate gels utilizing a new phosphate-free perfusion medium. *NMR Biomed.* 1990;3(1):23-26. doi:10.1002/nbm.1940030104
25. Schanda P, Kupče Ě, Brutscher B. SOFAST-HMQC Experiments for Recording Two-dimensional Deteronuclear Correlation Spectra of Proteins within a Few Seconds. *J Biomol NMR.* 2005;33(4):199-211. doi:10.1007/s10858-005-4425-x
26. Luchinat E, Barbieri L, Cremonini M, Banci L. Protein in-cell NMR spectroscopy at 1.2 GHz. *J Biomol NMR.* 2021;75(2-3):97-107. doi:10.1007/s10858-021-00358-w
27. Foster MP, McElroy CA, Amero CD. Solution NMR of Large Molecules and Assemblies. *Biochemistry.* 2007;46(2):331-340. doi:10.1021/bi0621314
28. Rößler P, Ruckstuhl M, Löbbert A, et al. Human Cells for Human Proteins: Isotope Labeling in Mammalian Cells for Functional NMR Studies of Disease-Relevant Proteins. *bioRxiv.* Preprint posted online April 13, 2024:2024.04.09.588766. doi:10.1101/2024.04.09.588766
29. Fuccio C, Luchinat E, Barbieri L, Neri S, Fragai M. Algal autolysate medium to label proteins for NMR in mammalian cells. *J Biomol NMR.* 2016;64(4):275-280. doi:10.1007/s10858-016-0026-0
30. Dutta A, Saxena K, Schwalbe H, Klein-Seetharaman J. Isotope Labeling in Mammalian Cells. *Methods Mol Biol Clifton Nj.* 2012;831:55-69. doi:10.1007/978-1-61779-480-3_4
31. Subedi GP, Roberts ET, Davis AR, Kremer PG, Amster IJ, Barb AW. A comprehensive assessment of selective amino acid ^{15}N -labeling in human embryonic kidney 293 cells for NMR spectroscopy. *J Biomol NMR.* 2024;78(2):125-132. doi:10.1007/s10858-023-00434-3
32. Barbieri L, Luchinat E, Banci L. Characterization of proteins by in-cell NMR spectroscopy in cultured mammalian cells. *Nat Protoc.* 2016;11(6):1101-1111. doi:10.1038/nprot.2016.061
33. Luchinat E, Barbieri L, Cremonini M, et al. Determination of intracellular protein–ligand binding affinity by competition binding in-cell NMR. *Acta Crystallogr Sect Struct Biol.* 2021;77(10):1270-1281. doi:10.1107/S2059798321009037
34. Chai Z, Li C. In-Cell ^{19}F NMR of Proteins: Recent Progress and Future Opportunities. *Chem – Eur J.* 2024;30(23):e202303988. doi:10.1002/chem.202303988
35. Kitevski-LeBlanc JL, Prosser RS. Current applications of ^{19}F NMR to studies of protein structure and dynamics. *Prog Nucl Magn Reson Spectrosc.* 2012;62:1-33. doi:10.1016/j.pnmrs.2011.06.003
36. Li C, Wang GF, Wang Y, et al. Protein ^{19}F NMR in Escherichia coli. *J Am Chem Soc.* 2010;132(1):321-327. doi:10.1021/ja907966n
37. Gronenborn AM. Small, but powerful and attractive: ^{19}F in biomolecular NMR. *Structure.* 2022;30(1):6-14. doi:10.1016/j.str.2021.09.009

38. Yoder NC, Kumar K. Fluorinated amino acids in protein design and engineering. *Chem Soc Rev.* 2002;31(6):335-341. doi:10.1039/b201097f
39. Pham LBT, Costantino A, Barbieri L, Calderone V, Luchinat E, Banci L. Direct Expression of Fluorinated Proteins in Human Cells for ^{19}F In-Cell NMR Spectroscopy. *J Am Chem Soc.* 2023;145(2):1389-1399. doi:10.1021/jacs.2c12086
40. Sharaf NG, Gronenborn AM. ^{19}F -Modified Proteins and ^{19}F -Containing Ligands as Tools in Solution NMR Studies of Protein Interactions. In: Kelman Z, ed. *Methods in Enzymology*. Vol 565. Isotope Labeling of Biomolecules - Labeling Methods. Academic Press; 2015:67-95. doi:10.1016/bs.mie.2015.05.014
41. Crowley PB, Kyne C, Monteith WB. Simple and inexpensive incorporation of ^{19}F -Tryptophan for protein NMR spectroscopy. *Chem Commun.* 2012;48(86):10681. doi:10.1039/c2cc35347d
42. Zhu W, Guseman AJ, Bhinderwala F, Lu M, Su X, Gronenborn AM. Visualizing Proteins in Mammalian Cells by ^{19}F NMR Spectroscopy. *Angew Chem Int Ed.* 2022;61(23):e202201097. doi:10.1002/anie.202201097
43. Luchinat E, Barbieri L, Davis B, Brough PA, Pennestri M, Banci L. Ligand-Based Competition Binding by Real-Time ^{19}F NMR in Human Cells. *J Med Chem.* 2024;67(2):1115-1126. doi:10.1021/acs.jmedchem.3c01600
44. Burz DS, Breindel L, Shekhtman A. Chapter Seventeen - Improved sensitivity and resolution of in-cell NMR spectra. In: Shukla AK, ed. *Methods in Enzymology*. Vol 621. Chemical and Synthetic Biology Approaches To Understand Cellular Functions - Part A. Academic Press; 2019:305-328. doi:10.1016/bs.mie.2019.02.029
45. Boeszoermenyi A, Chhabra S, Dubey A, et al. Aromatic ^{19}F - ^{13}C TROSY: a background-free approach to probe biomolecular structure, function, and dynamics. *Nat Methods.* 2019;16(4):333-340. doi:10.1038/s41592-019-0334-x
46. Luchinat E, Cremonini M, Banci L. Radio Signals from Live Cells: The Coming of Age of In-Cell Solution NMR. *Chem Rev.* 2022;122(10):9267-9306. doi:10.1021/acs.chemrev.1c00790
47. Barbieri L, Luchinat E. Monitoring Protein-Ligand Interactions in Human Cells by Real-Time Quantitative In-Cell NMR using a High Cell Density Bioreactor. *J Vis Exp.* 2021;(169):62323. doi:10.3791/62323
48. Sharaf NG, Barnes CO, Charlton LM, Young GB, Pielak GJ. A bioreactor for in-cell protein NMR. *J Magn Reson.* 2010;202(2):140-146. doi:10.1016/j.jmr.2009.10.008
49. Kubo S, Nishida N, Udagawa Y, Takarada O, Ogino S, Shimada I. A Gel-Encapsulated Bioreactor System for NMR Studies of Protein-Protein Interactions in Living Mammalian Cells. *Angew Chem Int Ed.* 2013;52(4):1208-1211. doi:10.1002/anie.201207243
50. Meehan AJ, Eskey CJ, Domach MM, Koretsky AP. Cultivator for NMR studies of suspended cell cultures. *Biotechnol Bioeng.* 1992;40(11):1359-1366. doi:10.1002/bit.260401110

51. Hartbrich A, Schmitz G, Weuster-Botz D, de Graaf AA, Wandrey C. Development and application of a membrane cyclone reactor for in vivo NMR spectroscopy with high microbial cell densities. *Biotechnol Bioeng.* 1996;51(6):624-635. doi:10.1002/(SICI)1097-0290(19960920)51:6<624::AID-BIT2>3.0.CO;2-J
52. Carvalho J, Alves S, Castro MMCA, et al. Development of a bioreactor system for cytotoxic evaluation of pharmacological compounds in living cells using NMR spectroscopy. *J Pharmacol Toxicol Methods.* 2019;95:70-78. doi:10.1016/j.vascn.2018.11.004
53. Dalvit C, Gmür I, Rößler P, Gossert AD. Affinity measurement of strong ligands with NMR spectroscopy: Limitations and ways to overcome them. *Prog Nucl Magn Reson Spectrosc.* 2023;138-139:52-69. doi:10.1016/j.pnmrs.2023.07.001
54. Sun D, Gao W, Hu H, Zhou S. Why 90% of clinical drug development fails and how to improve it? *Acta Pharm Sin B.* 2022;12(7):3049-3062. doi:10.1016/j.apsb.2022.02.002
55. Kola I. The State of Innovation in Drug Development. *Clin Pharmacol Ther.* 2008;83(2):227-230. doi:10.1038/sj.clpt.6100479
56. Parasrampur DA, Benet LZ, Sharma A. Why Drugs Fail in Late Stages of Development: Case Study Analyses from the Last Decade and Recommendations. *AAPS J.* 2018;20(3):46. doi:10.1208/s12248-018-0204-y
57. Van Vleet TR, Liguori MJ, Lynch I James J, Rao M, Warder S. Screening Strategies and Methods for Better Off-Target Liability Prediction and Identification of Small-Molecule Pharmaceuticals. *SLAS Discov.* 2019;24(1):1-24. doi:10.1177/2472555218799713
58. Buchholz CR, Pomerantz WCK. 19F NMR viewed through two different lenses: ligand-observed and protein-observed 19F NMR applications for fragment-based drug discovery. *RSC Chem Biol.* 2021;2(5):1312-1330. doi:10.1039/D1CB00085C
59. Gossert AD, Jahnke W. NMR in drug discovery: A practical guide to identification and validation of ligands interacting with biological macromolecules. *Prog Nucl Magn Reson Spectrosc.* 2016;97:82-125. doi:10.1016/j.pnmrs.2016.09.001
60. Luchinat E, Banci L. In-cell NMR: recent progresses and future challenges. *Rendiconti Lincei Sci Fis E Nat.* 2023;34(3):653-661. doi:10.1007/s12210-023-01168-y
61. Siegal G, Selenko P. Cells, drugs and NMR. *J Magn Reson.* 2019;306:202-212. doi:10.1016/j.jmr.2019.07.018
62. Inomata K, Ohno A, Tochio H, et al. High-resolution multi-dimensional NMR spectroscopy of proteins in human cells. *Nature.* 2009;458(7234):106-109. doi:10.1038/nature07839
63. Furukawa A, Konuma T, Yanaka S, Sugase K. Quantitative analysis of protein–ligand interactions by NMR. *Prog Nucl Magn Reson Spectrosc.* 2016;96:47-57. doi:10.1016/j.pnmrs.2016.02.002

64. Chandra G, Singh DV, Mahato GK, Patel S. Fluorine-a small magic bullet atom in the drug development: perspective to FDA approved and COVID-19 recommended drugs. *Chem Zvesti*. Published online April 13, 2023;1-22. doi:10.1007/s11696-023-02804-5
65. Shah P, Westwell AD. The role of fluorine in medicinal chemistry: Review Article. *J Enzyme Inhib Med Chem*. 2007;22(5):527-540. doi:10.1080/14756360701425014
66. Nocentini A, Supuran CT. Carbonic anhydrases. In: *Carbonic Anhydrases*. Elsevier; 2019:3-16. doi:10.1016/B978-0-12-816476-1.00001-0
67. Supuran C. Carbonic Anhydrases as Drug Targets - An Overview. *Curr Top Med Chem*. 2007;7(9):825-833. doi:10.2174/156802607780636690
68. Imtaiyaz Hassan Md, Shajee B, Waheed A, Ahmad F, Sly WS. Structure, function and applications of carbonic anhydrase isozymes. *Bioorg Med Chem*. 2013;21(6):1570-1582. doi:10.1016/j.bmc.2012.04.044
69. Supuran CT, Scozzafava A. Carbonic anhydrases as targets for medicinal chemistry. *Bioorg Med Chem*. 2007;15(13):4336-4350. doi:10.1016/j.bmc.2007.04.020
70. Supuran CT. Emerging role of carbonic anhydrase inhibitors. *Clin Sci*. 2021;135(10):1233-1249. doi:10.1042/CS20210040
71. De Simone G, Supuran CT. (In)organic anions as carbonic anhydrase inhibitors. *J Inorg Biochem*. 2012;111:117-129. doi:10.1016/j.jinorgbio.2011.11.017
72. Lomelino C, Supuran C, McKenna R. Non-Classical Inhibition of Carbonic Anhydrase. *Int J Mol Sci*. 2016;17(7):1150. doi:10.3390/ijms17071150
73. Jo J, Hong S, Choi WY, Lee DR. Cell-penetrating peptide (CPP)-conjugated proteins is an efficient tool for manipulation of human mesenchymal stromal cells. *Sci Rep*. 2014;4(1):4378. doi:10.1038/srep04378
74. Räägel H, Säälük P, Pooga M. Peptide-mediated protein delivery—Which pathways are penetrable? *Biochim Biophys Acta BBA - Biomembr*. 2010;1798(12):2240-2248. doi:10.1016/j.bbamem.2010.02.013
75. Ogino S, Kubo S, Umemoto R, Huang S, Nishida N, Shimada I. Observation of NMR signals from proteins introduced into living mammalian cells by reversible membrane permeabilization using a pore-forming toxin, streptolysin O. *J Am Chem Soc*. 2009;131(31):10834-10835. doi:10.1021/ja904407w
76. Theillet FX, Binolfi A, Bekei B, et al. Structural disorder of monomeric α -synuclein persists in mammalian cells. *Nature*. 2016;530(7588):45-50. doi:10.1038/nature16531
77. Sciolino N, Liu A, Breindel L, Burz DS, Sulchek T, Shekhtman A. Microfluidics delivery of DARPP-32 into HeLa cells maintains viability for in-cell NMR spectroscopy. *Commun Biol*. 2022;5(1):451. doi:10.1038/s42003-022-03412-x
78. Mancina F, Patel SD, Rajala MW, et al. Optimization of Protein Production in Mammalian Cells with a Coexpressed Fluorescent Marker. *Structure*. 2004;12(8):1355-1360. doi:10.1016/j.str.2004.06.012

79. Nishida N, Ito Y, Shimada I. *In situ* structural biology using in-cell NMR. *Biochim Biophys Acta BBA - Gen Subj*. 2020;1864(2):129364. doi:10.1016/j.bbagen.2019.05.007
80. Luchinat E, Barbieri L, Banci L. A molecular chaperone activity of CCS restores the maturation of SOD1 fALS mutants. *Sci Rep*. 2017;7(1):17433. doi:10.1038/s41598-017-17815-y
81. Werle Y, Kovermann M. Fluorine Labeling and ¹⁹F NMR Spectroscopy to Study Biological Molecules and Molecular Complexes. *Chem – Eur J*. 2025;31(2):e202402820. doi:10.1002/chem.202402820
82. Gimenez D, Phelan A, Murphy CD, Cobb SL. ¹⁹F NMR as a tool in chemical biology. *Beilstein J Org Chem*. 2021;17:293-318. doi:10.3762/bjoc.17.28
83. Gee CT, Arntson KE, Urick AK, et al. Protein-observed ¹⁹F-NMR for fragment screening, affinity quantification and druggability assessment. *Nat Protoc*. 2016;11(8):1414-1427. doi:10.1038/nprot.2016.079
84. Shehadul Islam M, Aryasomayajula A, Selvaganapathy P. A Review on Macroscale and Microscale Cell Lysis Methods. *Micromachines*. 2017;8(3):83. doi:10.3390/mi8030083
85. Working Principle Of Bead Mill - ELE Mechanical. Accessed August 20, 2025. <https://www.ele-mix.com/news/working-principle-of-bead-mill-24426552.html>
86. Traditional Methods of Cell Lysis - IT. Accessed August 20, 2025. <https://www.thermofisher.com/uk/en/home/life-science/protein-biology/protein-biology-learning-center/protein-biology-resource-library/pierce-protein-methods/traditional-methods-cell-lysis.html>
87. Janson JC, ed. *Protein Purification: Principles, High Resolution Methods, and Applications*. 3rd ed. John Wiley & Sons; 2011.
88. Robert K. Scopes. *Protein Purification: Principle and Practice*. Third Edition. Springer New York
89. Coskun O. Separation Techniques: CHROMATOGRAPHY. *North Clin Istanb*. Published online 2016. doi:10.14744/nci.2016.32757
90. Farheen SR. Membrane Processing Technology: A Review Article. *Int J Eng Res*. 2020;9(11).
91. Banerjee AL, Swanson M, Mallik S, Srivastava DK. Purification of recombinant human carbonic anhydrase-II by metal affinity chromatography without incorporating histidine tags. *Protein Expr Purif*. 2004;37(2):450-454. doi:10.1016/j.pep.2004.06.031
92. Mcpherson A. Introduction to protein crystallization. *Methods*. 2004;34(3):254-265. doi:10.1016/j.ymeth.2004.03.019
93. Stevens RC. High-throughput protein crystallization. 2000;10(5):558-563. doi:10.1016/s0959-440x(00)00131-7

94. Adachi H, Takano K, Morikawa M, Kanaya S, Yoshimura M, Mori Y, Sasaki T. Application of a two-liquid system to sitting-drop vapour-diffusion protein crystallization. *Epub*. Published online December 19, 2002. doi:10.1107/s0907444902019741.
95. Kishwar Hayat Khan S. *Gel Electrophoresis – Principles and Basics*. IntechOpen; 2012. <https://doi.org/10.5772/2205>
96. Wangler MF, Bellen HJ. In Vivo Animal Modeling. In: *Basic Science Methods for Clinical Researchers*. Elsevier; 2017:211-234. doi:10.1016/B978-0-12-803077-6.00012-6
97. Kurien BT, Scofield RH. Western blotting. *Methods*. 2006;38(4):283-293. doi:10.1016/j.ymeth.2005.11.007
98. Taylor SC, Posch A. The Design of a Quantitative Western Blot Experiment. *BioMed Res Int*. 2014;2014(1):361590. doi:10.1155/2014/361590
99. Kurien BT, Scofield RH. Multiple immunoblots after non-electrophoretic bidirectional transfer of a single SDS–PAGE gel with multiple antigens. *J Immunol Methods*. 1997;205(1):91-94. doi:10.1016/S0022-1759(97)00052-5
100. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. doi:10.1073/pnas.76.9.4350
101. Mahmood T, Yang PC. Western Blot: Technique, Theory, and Trouble Shooting. *North Am J Med Sci*. 2012;4(9):429-434. doi:10.4103/1947-2714.100998
102. Marczenko Z, Balcerzak M. Chapter 2 - Principles of Spectrophotometry. In: Marczenko Z, Balcerzak M, eds. *Analytical Spectroscopy Library*. Vol 10. Separation, Preconcentration and Spectrophotometry in Inorganic Analysis. Elsevier; 2000:26-38. doi:10.1016/S0926-4345(00)80066-8
103. Sommer L. *Analytical Absorption Spectrophotometry in the Visible and Ultraviolet: The Principles*. Vol Vol 8. Elsevier; 2012.
104. Rocha FS, Gomes AJ, Lunardi CN, Kaliaguine S, Patience GS. Experimental methods in chemical engineering: Ultraviolet visible spectroscopy—UV-Vis. *Can J Chem Eng*. 2018;96(12):2512-2517. doi:10.1002/cjce.23344
105. Smyth MS. x Ray crystallography. *Mol Pathol*. 2000;53(1):8-14. doi:10.1136/mp.53.1.8
106. Finehout EJ, Lee KH. An introduction to mass spectrometry applications in biological research. *Biochem Mol Biol Educ*. 2004;32(2):93-100. doi:10.1002/bmb.2004.494032020331
107. Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. Electrospray ionization for mass spectrometry of large biomolecules. *Science*. 1989;246(4926):64-71. doi:10.1126/science.2675315
108. Kinter M, Sherman NE. *Protein Sequencing and Identification Using Tandem Mass Spectrometry*. John Wiley & Sons; 2005.

109. Diehl B. Chapter 1 - Principles in NMR Spectroscopy. In: Holzgrabe U, Wawer I, Diehl B, eds. *NMR Spectroscopy in Pharmaceutical Analysis*. Elsevier; 2008:1-41. doi:10.1016/B978-0-444-53173-5.00001-9
110. Nishida N, Zhao Q, Shimada I. Bioreactor in-cell NMR: A powerful tool for observing intracellular biological events. *Curr Opin Struct Biol*. 2025;93:103086. doi:10.1016/j.sbi.2025.103086
111. Cerofolini L, Giuntini S, Barbieri L, et al. Real-Time Insights into Biological Events: In-Cell Processes and Protein-Ligand Interactions. *Biophys J*. 2019;116(2):239-247. doi:10.1016/j.bpj.2018.11.3132
112. Molina DM, Jafari R, Ignatushchenko M, et al. Monitoring Drug Target Engagement in Cells and Tissues Using the Cellular Thermal Shift Assay. *Science*. 2013;341(6141):84-87. doi:10.1126/science.1233606
113. Seashore-Ludlow B, Axelsson H, Almqvist H, Dahlgren B, Jonsson M, Lundbäck T. Quantitative Interpretation of Intracellular Drug Binding and Kinetics Using the Cellular Thermal Shift Assay. *Biochemistry*. 2018;57(48):6715-6725. doi:10.1021/acs.biochem.8b01057
114. Martinez Molina D, Nordlund P. The Cellular Thermal Shift Assay: A Novel Biophysical Assay for In Situ Drug Target Engagement and Mechanistic Biomarker Studies. *Annu Rev Pharmacol Toxicol*. 2016;56(1):141-161. doi:10.1146/annurev-pharmtox-010715-103715
115. Axelsson H, Almqvist H, Seashore-Ludlow B, Lundbäck T. *Screening for Target Engagement Using the Cellular Thermal Shift Assay - CETSA*. Assay Guidance Manual [Internet]; 2016. <https://www.ncbi.nlm.nih.gov/books/NBK374282/>
116. Jafari R, Almqvist H, Axelsson H, et al. The cellular thermal shift assay for evaluating drug target interactions in cells. *Nat Protoc*. 2014;9(9):2100-2122. doi:10.1038/nprot.2014.138
117. Seashore-Ludlow B, Axelsson H, Lundbäck T. Perspective on CETSA Literature: Toward More Quantitative Data Interpretation. *SLAS Discov*. 2020;25(2):118-126. doi:10.1177/2472555219884524
118. Kawatkar A, Schefter M, Hermansson NO, et al. CETSA beyond Soluble Targets: a Broad Application to Multipass Transmembrane Proteins. *ACS Chem Biol*. 2019;14(9):1913-1920. doi:10.1021/acschembio.9b00399
119. Waldron TT, Murphy KP. Stabilization of proteins by ligand binding: application to drug screening and determination of unfolding energetics. *Biochemistry*. 2003;42(17):5058-5064. doi:10.1021/bi034212v
120. Matulis D, Kranz JK, Salemme FR, Todd MJ. Thermodynamic stability of carbonic anhydrase: measurements of binding affinity and stoichiometry using ThermoFluor. *Biochemistry*. 2005;44(13):5258-5266. doi:10.1021/bi048135v
121. Iyer R, Barrese III AA, Parakh S, Parker CN, Tripp BC. Inhibition Profiling of Human Carbonic Anhydrase II by High-Throughput Screening of Structurally Diverse,

- Biologically Active Compounds. *SLAS Discov.* 2006;11(7):782-791. doi:10.1177/1087057106289403
122. Işık A, Acar Çevik U, Celik I, et al. Novel Hydrazide-Hydrazones Bearing a Benzimidazole Ring: Design, Synthesis, and Evaluation of Inhibitor Properties Against CA I and CA II Isozymes. *Chem Biol Drug Des.* 2024;104(6):e70025. doi:10.1111/cbdd.70025
123. Maryam Z, Işık A, Bağcı ER, et al. Synthesis of Thiazole-methylsulfonyl Derivatives, X-ray Study, and Investigation of Their Carbonic Anhydrase Activities: In Vitro and In Silico Potentials. *ACS Omega.* 2025;10(13):13583-13594. doi:10.1021/acsomega.5c00509