UNIVERSITÀ DEGLI STUDI FIRENZE

HR EXCELLENCE IN RESEARCH

DOTTORATO DI RICERCA
INTERNATIONAL DOCTORATE IN STRUCTURAL BIOLOGY

CICLO XXXVIII

COORDINATOR Prof. Roberta Pierattelli

Incorporation of natural and non-natural amino acids in human cells for advanced in-cell NMR labelling schemes

Settore Scientifico Disciplinare CHEM-03/A

PhD student
Dott. Martina Rosati

Supervisor
Prof. Lucia Banci

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

Universiteit Utrecht

FRANKFURT AM MAIN

To my Families,

my strength and inspiration

Table of Contents

Abstract 1

1. Introduction 2

1.1 In-cell nuclear magnetic resonance (NMR) 2

1.1.1 Historical background and evolution 2

1.1.2 Applications 2

1.1.3 Advancements 4

1.1.4 Advanced cellular models for in-cell NMR 5

1.1.5 Future Perspectives 7

1.2 Selective isotopic labelling 7

1.2.1 Selective isotopic labelling with amino acid precursors 8

1.2.2 Selective isotopic labelling of amino acid side chains 10

1.2.3 Selective isotopic labelling of amino acid backbone 11

1.2.4 Selective labelling with fluorinated amino acids 12

2. Complementary methods 14

2.1 Flow cytometry 14

2.1.1 Fluorescence-Activated Cell Sorting (FACS) 15

2.2 Fluorescence microscopy 15

2.2.1 Confocal microscopy 16

3. Towards cost-effective side-chain isotope labelling of proteins expressed in human cells 17

3.1 Summary 17

4. The synthesis of specifically isotope labelled fluorotryptophan and its use in mammalian cell-based protein expression for ^{19}F-NMR applications 34

4.1 Summary 34

5. Selective Labelling Strategies for in-cell NMR studies of intrinsically disordered proteins: the challenging case of BRCA1 60

5.1 Abstract 60

5.2 Introduction 60

5.3 Results and discussion: 62

5.3.1 Expression, stabilization and assessment of experimental conditions: 62

5.3.2 Labelling 64

5.3.3 Co-expression of BRCA1 and MAX 71

5.3.4 BRCA1 monoclonal inducible cell line: 73

5.4 Conclusions and perspectives: 74
5.5 Materials and methods: 75
5.5.1 Expression test 75
5.5.2 Protein stability and cell viability assay 75
5.5.3 Nucleus and cytoplasm separation 75
5.5.4 Western Blot 76
5.5.5 Lysate stabilization 76
5.5.6 ^{15}N uniform labelling 76
5.5.7 ^{15}N, ^{13}C precursor labelling 76
5.5.8 ^{15}N, ^{13}C amino acid labelling 77
5.5.9 Protein coexpression 77
5.5.10 Monoclonal inducible stably transfected cell line generation 77
6. Impact of cell culture complexity on ligand-target interactions determined via in-cell NMR 78
6.1 Abstract 78
6.2 Introduction 78
6.3 Results and discussion 79
6.3.1 In-cell ^{19}F NMR of SLC binding to CA II 79
6.3.2 Ligand observed competition binding to assess K_d in different cell culture models 80
6.1 Conclusions and perspectives 83
6.2 Materials and methods 83
6.2.1 HEK293-TRex$^{TR\text{-}hCAII/Mono}$ sample for in-cell NMR 83
6.2.2 Transfected cell sample for bioreactor 84
6.2.3 Monoclonal cell sample for bioreactor 84
6.2.4 Spheroid sample for bioreactor 84
6.2.5 Bioreactor setup 85
6.2.6 Data analysis 85
7. Conclusions and perspectives 87
8. Bibliography 89

List of Figures

Figure 1.1.2-1 3
Figure 1.1.4-1 6
Figure 1.2.1-1 9
Figure 1.2.4-1 13
Figure 5.3.1-1 62
Figure 5.3.1-2 63
Figure 5.3.1-3 63
Figure 5.3.1-4 64
Figure 5.3.2-1 65
Figure 5.3.2-2 67
Figure 5.3.2-3 67
Figure 5.3.2-4 68
Figure 5.3.2-5 69
Figure 5.3.2-6 70
Figure 5.3.2-7 71
Figure 5.3.3-1 72
Figure 5.3.3-2 73
Figure 5.3.4-1 74
Figure 6.3.1-1 79
Figure 6.3.2-1 80
Figure 6.3.2-2 81
Figure 6.3.2-3 81
Figure 6.3.2-4 82

List of Tables

Table 1 83
Table 2 85

List of Equations

Equation 1 86

List of abbreviations

aaRS – aminoacyl-tRNA Synthetase

AAZ – Acetazolamide

ADP – Adenosine Diphosphate

Ala – Alanine

ALSOFAST HMQC – Alternate SOFAST-HMQC

APC – Allophycocyanin

ASCs – Adult Stem Cells

Asn – Asparagine

Asp – Aspartate

ATP – Adenosine Triphosphate

BEST – Band-Selective Excitation Short-Transient

BRCA1 – Breast Cancer Gene 1

BRCA1 – Breast Cancer Susceptibility Protein 1

CA II – Carbonic Anhydrase isotope II

CPP – Cell-Penetrating Peptides

Cys – Cysteine

DAPI – 4′,6-diamidino-2-phenylindole

DJ-1 – Protein Deglycase 1

DMEM – Dulbecco's Modified Eagle Medium

DNA – Deoxyribonucleic Acid

E.coli – Escherichia Coli

ECL – Enhanced ChemiLuminescenceHEK293T

ETZ – Ethoxzolamide

FACS – Fluorescence-Activated Cell Sorting

FBS – Fetal Bovine Serum

FITC – Fluorescein Isothiocyanate

Flp-In – Flippase recombinase Integration

FM – Fluorescence Microscopy

FSC – Forward Scattering

GFP – Green Fluorescent Protein

Glu – Glutamate

GOI – Gene Of Interest

GSH – Glutathione

HBOC – Hereditary Breast and Ovarian Cancer

HEK293T – Human Embryonic Kidney 293 cells expressing the SV40 large T antigen

HEK293TRex – HEK293T stably transfected with Tet repressor

HEK293TRex$^{TR\text{-}hCAII/Mono}$ – Monoclonal HEK293TRex stably transfected with human CA II gene

His – Histidine

HMQC – Heteronuclear Multiple Quantum Coherence

IDP - Intrinsically Disordered Protein

IDR – Intrinsecally Disordered Region

Ile – Isoleucine

ITR – Inverted Terminal Repeats

Leu – Leucine

MAX – Myc-Associated Factor X

MCF-7 – Michigan Cancer Foundation cell line no. 7

Met – Methionine

MZA – Methazolamide

NMR – Nuclear Magnetic Resonance

nnAA – non-natural Amino Acid.

NOESY – Nuclear Overhauser Effect SpectroscopY

PALB2 – Partner And Localizer of BRCA2

PBS – Phosphate-Buffered Saline

PE – Phycoerythrin

pen/strep – Penicillin-Streptomycin

Phe – Phenylalanine

PI – Propidium Iodide

Pro – Proline

PSCs – Pluripotent Stem Cells

S/N – Signal to Noise ratio

SDS-PAGE – Sodium Dodecyl Sulphate PolyAcrylamide Gel Electrophoresis

SOD1 – Superoxide Dismutase 1

SSC – Side Scattering

STD – Saturation Transfer Difference

Tet – Tetracycline

tRNA – transfer Ribonucleic Acid

trNOESY – Transfer Nuclear Overhauser Effect SpectroscopY

TROSY – Transverse Relaxation-Optimized Spectroscopy

Trp – Tryptophan

Tyr – Tyrosine

Val – Valine

Abstract

This thesis presents significant advancements in developing selective isotopic labelling strategies for in-cell nuclear magnetic resonance (NMR) spectroscopy within human cells, aiming to overcome limitations related to cost, sensitivity, specificity, applicability to complex biological systems and physiological models.

We first established a cost-effective and innovative methodology for selective side-chain isotope labelling in mammalian cells. This method exploits the activity of human cellular transaminases to convert amino acid late precursors (α-keto acids) into their cognate L-amino acids, achieving efficient and residue-specific incorporation of ^{13}C labelled aromatic and aliphatic residues with minimal scrambling. High-quality in-cell and lysate NMR spectra were obtained, and the functional utility was validated by monitoring ligand-induced chemical shift perturbations in live cells expressing carbonic anhydrase II (CA II) treated with different inhibitors.

We also showed the incorporation of a non-natural amino acid, 5-fluorotryptophan selectively labelled with ^{13}C at the C5 position, in proteins expressed in human cells (CA II and superoxide dismutase 1). This strategy allows recording 2D ^{13}C, ^{19}F NMR spectra in which optimal peak dispersion is achieved by exploiting the TROSY effect.

Furthermore, the thesis focused on designing selective labelling schemes for intrinsically disordered regions (IDRs) of proteins, applied to the challenging case of BRCA1. The expression of four fragments was achieved in human cells, including a very long fragment (219-1357) spanning almost the entire IDR of BRCA1. Optimized isotopic labelling schemes were then implemented, utilizing either ^{13}C, ^{15}N uniformly labelled amino acids (L-histidine and L-cysteine) or $^{13}C_{\alpha}$ or $^{13}C'$ labelled α-ketoacid precursors (which are converted to L-leucine and L-phenylalanine), enabling residue-specific observation and greatly reducing spectral crowding in the in-cell and lysate NMR spectra of BRCA1. To facilitate future interaction studies with physiological partners, a co-expression protocol was established, and a monoclonal inducible BRCA1 cell line was developed, enabling the isotopic labelling of only one interacting protein to simplify spectral interpretation.

Finally, the in-cell NMR approach was extended to investigate ligand-target interactions under increasing cell culture complexity. By utilizing a flow NMR bioreactor and time-resolved ^{19}F NMR, the displacement of a fluorinated CA II inhibitor by competition with methazolamide as a reference compound was monitored across four different cellular models: transiently transfected HEK293T cells, stably transfected inducible monoclonal cells expressing CA II (HEK293-TRex$^{TR\text{-}hCAII/Mono}$), HEK293-TRex$^{TR\text{-}hCAII/Mono}$ spheroids, and mixed spheroids (HEK293-TRex$^{TR\text{-}hCAII/Mono}$ mixed with MCF-7 cells). These experiments aimed to detect differences in relative binding affinity amongst the different models.

1. Introduction

1.1 In-cell nuclear magnetic resonance (NMR)

In-cell NMR is a nuclear spectroscopy approach applied to the study of biological macromolecules inside living cells[1–4]. This technique is a powerful tool for investigating the structural and dynamic properties of biological molecules (proteins, nucleic acids, metabolites, and ions) directly within their natural, complex cellular environment, rather than in isolated or artificial conditions. The main aim is to obtain atomic-scale information about specific nuclear isotopes born into the target molecule in the living cells, providing insights into how molecules behave and interact in real time, under physiological conditions[1,2,5–9].

1.1.1 Historical background and evolution

The term "in-cell NMR" was first introduced by Serber and Dötsch in 2001, in pioneering experiments that reported the first in-cell NMR spectra of recombinantly expressed proteins in *E. coli*. In their work, they applied an existing protocol for ^{15}N-labeling and defined key experimental parameters essential for detecting intracellular protein[10]. The first attempt at in-cell NMR on eukaryotic cells was performed on *Xenopus laevis* oocytes in 2006, by purifying ^{15}N-labeled proteins from *E.coli* and injecting them in manually sorted stage-VI oocytes, enabling the observation of proteins and nucleic acids in a more complex cellular setting[11]. By 2009, Inomata et al. achieved the first high-resolution 2D heteronuclear spectra from proteins in living human cells by fusing cell-penetrating peptides to isotopically labelled proteins, while complementary delivery strategies (streptolysin-O permeabilization, hypotonic swelling, and electroporation) were reported in parallel[12–15]. In 2013, direct protein expression for in-cell NMR in human cells (HEK293T cell line) was pioneered by Banci et al. paving the way for studying processes like protein folding, maturation, and post-translational modifications immediately after protein synthesis[16]. From 2013 many advancements have been made in the field of applications with new labelling techniques and technological advancements.

1.1.2 Applications

Unlike in vitro studies, which typically rely on optimized conditions such as buffered aqueous solutions or crystals and often overlook the effects of viscosity, macromolecular crowding (**Figure 1.1.2-1**), and complex intermolecular interactions, in-cell NMR directly probes molecules within the physiological environment of living cells. This unique advantage allows the consideration of key cellular factors such as pH, redox properties, and the presence of natural interaction partners, all of which can profoundly influence biomolecular behaviour[3,4,10,11,17–20].

Indeed, proteins inside cells are subject to a network of transient, weak, and highly dynamic interactions-collectively referred to as the "quinary structure" that extend beyond stable quaternary assemblies[4,6,18,21]. These include electrostatic and hydrophobic contacts, weak interactions with metabolites, RNA, membranes, and other proteins, which, although not forming stable complexes, strongly impact folding, stability, diffusion, and function in vivo[4,6,18,21].

In-cell NMR reveals how the cellular environment shapes protein conformational dynamics and stability, including intrinsically disordered proteins like α-synuclein, which retains its

disordered state in mammalian cells[2–4,8,10,14,18,22]. Remarkably, in-cell NMR has even enabled de novo determination of protein three-dimensional structures, uncovering subtle conformational and dynamic differences compared to in vitro conditions[2–4,8,18,20].

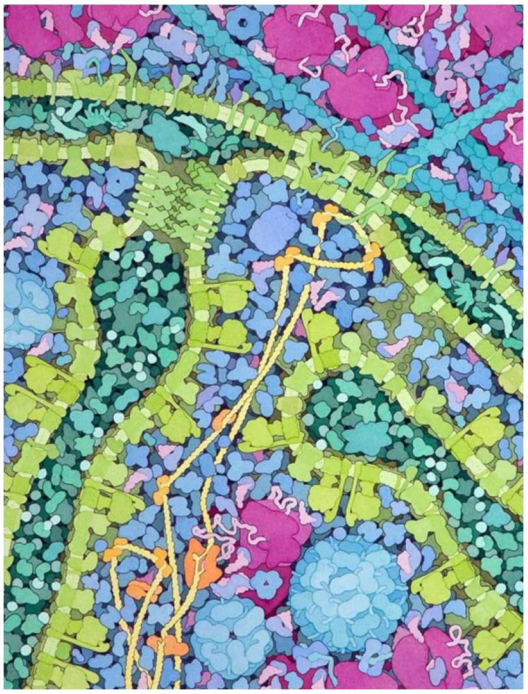


Figure 1.1.2-1
Molecular crowding inside the cell[23].

Importantly, the development of NMR bioreactors has transformed in-cell NMR into a powerful tool for real-time monitoring of molecular processes, spanning from nanosecond-scale dynamics to cellular events unfolding over hours or even days. By maintaining cell viability for up to 72 hours, these systems enable continuous observation of intracellular phenomena such as protein–drug interactions, oxidative modifications, and metabolic changes[1,3,4,9,18,19].

Beyond fundamental insights, in-cell NMR has become a powerful tool for drug discovery, providing information on cell penetration, binding efficiency, and selectivity, data often inaccessible from in vitro studies[9,18–20,24]. Experiments are often carried out through application of specific labelling strategies or sensitive methods such as ^{19}F NMR, ligand-based saturation transfer difference (STD), transfer NOESY (trNOESY) experiments[2–4,4,8–10,17–20,25].

While proteins remain the most extensively studied targets, in-cell NMR is equally valuable for investigating nucleic acids, whose structure and function are strongly modulated by the intracellular milieu[3,4,8,18]. It has been used to monitor conformational changes, stability, ligand binding, protein-RNA interactions, and base-pairing kinetics, and has recently been applied to the study of telomeric DNA G-quadruplexes, structures of significant relevance in anticancer research[3,4,17,18]. Finally, since its earliest applications, NMR has been used for the non-invasive study of small molecules such as metabolites and ions in living systems. The introduction of NMR bioreactors has allowed real-time monitoring of cellular metabolism, enabling the detection and quantification of metabolites including ATP, ADP, and inorganic phosphate, as well as the study of physiologically essential ions such as Na^+, K^+, Mg^{2+}, and Ca^{2+} [4,18,19].

1.1.3 Advancements

While in-cell NMR is a powerful technique for obtaining structural, conformational, and dynamic insights into biomolecules within their physiological environment, it also faces significant challenges. The first among these is the intrinsically low sensitivity of NMR, which requires high intracellular concentrations of the target molecule, typically above 5 μM and often in the range of 50–200 μM for folded proteins, to achieve detectable signals within a viable timeframe[2–4,18]. To address this limitation, researchers have developed optimized strategies to artificially increase intracellular concentrations, either through recombinant overexpression of isotopically labelled proteins within the cells, or through the delivery of purified proteins by methods such as electroporation or the use of cell-penetrating peptides[2–4,10,12,13].

Another critical issue, is that of broad spectral lines and signal overlap[4,8,17,18]. The crowded, viscous, and heterogeneous cellular environment, together with the possibility to form interaction with other cellular components, leads to reduced molecular tumbling, which in time increases spin relaxation rates, and magnetic susceptibility differences, resulting in broad lines and overlapping signals[3,4,6,8,10,17,18,25]. In addition, the finite lifetime of cells under NMR conditions places practical limits on the length of experiments, as nutrient depletion, oxygen starvation, and waste accumulation can quickly compromise viability[1,3,4,17,18]. Finally, the vast background of endogenous cellular signals further complicates detection[2,4,8,17,18].

To address these limitations, researchers have introduced a variety of technical and methodological innovations. The development of bioreactor systems has been particularly impactful, allowing for extended acquisition times and improved signal-to-noise ratio (S/N) by maintaining the cells viable and metabolically active for up to 72 hours. The continuous stream of culture medium not only provides cells always with a flow of fresh nutrients, while removing toxic byproducts, but also allows to control other fundamental parameters like oxygen levels and pH. This has enabled the real-time monitoring of intracellular processes such as protein-ligand binding, oxidation, and cell metabolism[1,4,4,18–20,26–28].

Moreover, advancements in NMR hardware, including the implementation of higher magnetic fields and cryogenic probes, have further increased sensitivity[3,4,18]. In parallel, the development of advanced pulse sequences, such as SOFAST, BEST, and TROSY schemes, along with selective excitation, has reduced acquisition times and improved resolution[4,17,29–32]. Another approach to improve sensitivity and reduce spectral resolution is to apply selective labelling strategies. The incorporation of NMR-active isotopes, such as ^{15}N or ^{13}C is a standard and essential practice in both in vitro and in-cell NMR studies, as it can help filter out background signals from unlabelled cellular components when the target molecule is delivered in the cell[3,4]. However, when proteins are expressed directly in cells, selective labelling of specific amino acids (e.g., methyl-^{13}C aliphatic residues) often in combination with deuteration and background subtraction, lead to reduce spectral complexity and better resolved spectra[33–37]. Similarly, introducing non-natural amino acids containing NMR-active nuclei, such as ^{19}F, provides an additional advantage, as cells naturally lack fluorine, offering excellent sensitivity and a lack of cellular background, though sometimes chemical modifications might affect biological relevance[3,18,38,39]. For immobilized proteins or large complexes, solid-state NMR methods such as Magic Angle Spinning (MAS) can overcome line broadening, caused by slow

tumbling and magnetic susceptibility inhomogeneities, while dynamic nuclear polarization (DNP) provides a substantial gain in sensitivity for cryo-preserved samples[8,18].

When all protein-observed approaches fail, and the protein becomes effectively "invisible" to NMR due to fast relaxation induced by interactions with large cellular components, ligand-observed strategies can be employed[3,4,6,9,18]. In this case, the ligand itself is labelled, most commonly with ^{19}F[4,9,18]. Complementary methods such as STD NMR can reveal whether and where a ligand binds to a protein, while trNOESY provides insights into the bound conformation of the ligand and the nature of its interactions with the target[18].

1.1.4 Advanced cellular models for in-cell NMR

One of the current challenges of in-cell NMR is the use of cellular models that are more physiologically representative and better approximate the native environment in which the processes under investigation actually take place. Lately, three-dimensional (3D) cellular models, are increasingly recognized as advanced systems for in-cell NMR studies due to their ability to better mimic the physiological microenvironment of tissues compared to traditional two-dimensional cultures. Spheroids and organoids are 3D cell culture systems, but they differ significantly in their cellular complexity, formation methods, and applications[40–46].

Spheroids are simplest model, they are spherical aggregates of one or more cell types, typically formed by self-assembly in non-adherent conditions or within microwells. The internal structure of a multicellular spheroid, especially in tumour models, is organized into distinct concentric zones that reflect gradients in nutrient and oxygen availability. Typically, the outer layer (just a few cell layers thick) consists of proliferating cells that have direct access to nutrients and oxygen[47,48]. Beneath this, there is an intermediate "inhibited" region where cells remain viable but are largely non-proliferative due to limited resources. At the core, a necrotic (dead) region often forms as a result of severe nutrient and oxygen deprivation, with cellular material in various stages of disintegration[47,48] (**Figure 1.1.4-1**). This region can expand up to circa 70% of the spheroid in the most extreme cases[47]. The layered architecture closely mimics the microenvironment of avascular tumours, where the centre becomes hypoxic and necrotic as the spheroid grows. Consequently, spheroids are extensively used in cancer research to mimic the tumour microenvironment, allowing for the investigation of tumour growth, invasion, metastasis, drug resistance, and response to therapies in a setting that closely resembles in vivo tumours[41,44–46,49].

By controlling cell-seeding density and growth duration, however, it is possible to generate smaller spheroids, maintaining high cell viability throughout the structure, thereby avoiding the formation of the necrotic core[50]. This approach results in uniform, viable spheroids that better mimic physiological conditions and are especially valuable for regenerative medicine, drug screening, and tissue engineering applications where high cell viability and function are critical.

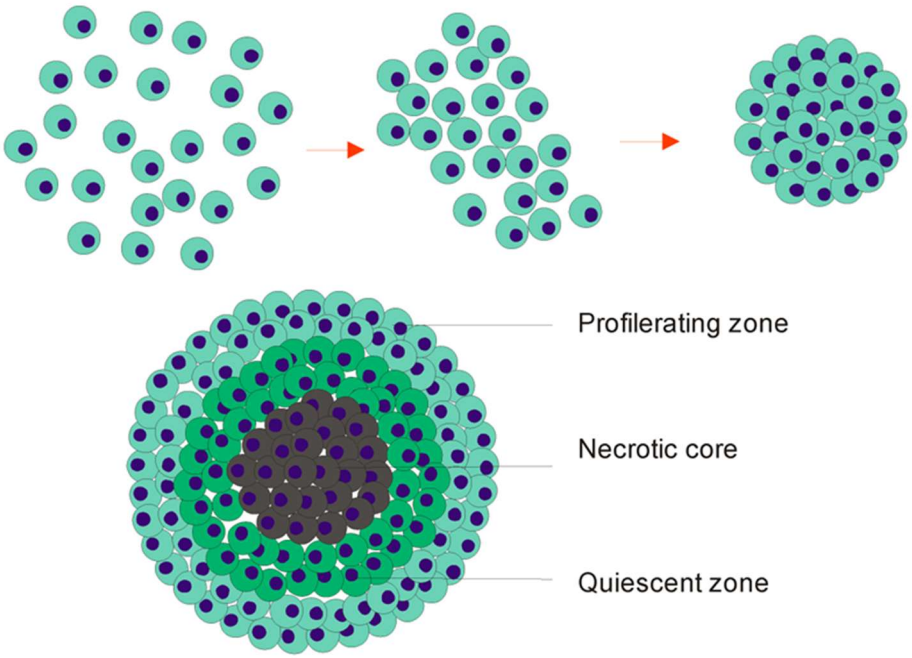


Figure 1.1.4-1
Spheroid structure[51].

Although spheroids remain simpler than fully developed tissues and lack their structural and functional complexity, they recapitulate many essential features of the in vivo microenvironment, including cell–cell and cell–matrix interactions, nutrient and oxygen gradients, and metabolic heterogeneity[43,44]. These properties make them useful not only for cancer biology but also for investigating protein structure, dynamics, metabolism, and ligand binding in physiologically relevant contexts[40,44,46].

In the context of in-cell NMR, spheroids can be embedded in a low-melting agarose matrix, which prevents fusion through self-assembly and thus preserves their size and morphology. This approach makes them highly compatible with bioreactor systems, enabling stable and extended measurements under controlled conditions[52]. Beyond fundamental studies, spheroids are increasingly applied in high-throughput drug screening, personalized medicine, and as alternatives to animal models, offering physiologically relevant yet accessible systems[44,46,49].

Moving along in the complexity scale of the cellular models we encounter organoids, which, in contrast, are more complex 3D structures derived from stem cells or progenitor cells that self-organize into miniaturized, organ-like tissues. Organoids exhibit complex, organ-like architecture, including multiple cell types arranged in spatial patterns that mimic in vivo tissues. They can be derived from pluripotent stem cells (PSCs) or adult stem cells (ASCs), with PSC-derived organoids often displaying higher structural complexity, including epithelial, mesenchymal, and sometimes endothelial components, while ASC-derived organoids more closely resemble adult tissue epithelium[42,53–55]. Importantly, organoids retain genetic stability, the capacity for self-renewal, and the ability to perform organ-specific functions such as secretion, absorption, or electrical activity[53–56].

Despite their remarkable potential, organoids pose significant experimental and analytical challenges due to their size, ranging from tens of micrometres to several millimetres, their dense and heterogeneous internal architecture, and the difficulty of obtaining samples that are reproducible in terms of shape, dimensions, and structural organization of individual

components.[41,42,56,57]. In particular, the reduced sensitivity of NMR signals in such complex samples make it difficult to acquire high-resolution molecular information[40]. While organoids are extremely promising models, further methodological and technological advances will be required before in-cell NMR can fully meet the challenges posed by these systems.

1.1.5 Future Perspectives

Looking ahead, the future of in-cell NMR spectroscopy lies in its continuing technological and methodological evolution. Advances in NMR hardware, such as next-generation electronics, together with the development of more sophisticated bioreactor systems, are expected to significantly expand its applicability by enabling longer measurement times and improved sensitivity[1,4,19]. At the same time, integration with complementary approaches, including cutting-edge cellular imaging techniques such as cryo-electron microscopy and super-resolved fluorescence microscopy, as well as genome engineering tools, will provide a more comprehensive understanding of intracellular systems[3,4,18]. A major goal for the field is to extend in-cell NMR studies beyond single cells to more complex and physiologically relevant systems, including organoids and tissue models, while pushing detection limits to lower concentrations of biomolecules[3,4,18,55]. Ultimately, combining in-cell NMR with these emerging methodologies will offer a powerful, multidimensional view of the structure, dynamics, and interactions of macromolecules within their native environment, with unprecedented spatial and temporal resolution.

1.2 Selective isotopic labelling

Selective labelling is a powerful strategy in nuclear magnetic resonance (NMR) spectroscopy where stable isotopes such as ^{13}C, ^{15}N, ^{2}H or ^{19}F are incorporated only at chosen atomic sites or specific amino acids within a biomolecule[58–64]. Unlike uniform labelling, this targeted enrichment reduces spectral crowding, enhances sensitivity, and simplifies resonance assignment. The approach is essential for studying large proteins, macromolecular complexes, and nucleic acids where uniform labelling produces excessively complex spectra[36,59–63,65–68].

Applying selective isotopic labelling strategies in different cellular hosts yields distinct outcomes, each with specific advantages and limitations.

Bacteria, particularly *E.coli*, are highly efficient and cost-effective platforms for large-scale protein production and amino acid-selective isotopic labelling[58,59,63,67,69,70]. Common strategies include using auxotrophic strains with isotopically labelled amino acids, T7 RNA polymerase with rifampicin to suppress endogenous protein synthesis, or a two-step culture transferring cells from rich to labelled minimal medium to reduce isotope use and growth limitations[59,69–72]. Limitations include lack of eukaryotic post-translational modifications, potential misfolding, and isotope scrambling, which can be mitigated by pathway-specific inhibitors[58–60,63,69,73]. Yeast hosts (*Pichia pastoris, Saccharomyces cerevisiae*) provide a eukaryotic environment suitable for glycosylation and complex folding, support amino acid- and methyl-specific labelling, and allow metabolic glycan labelling, though nonessential amino acids may still experience some scrambling[67,74–76].

Cell-free protein synthesis platforms offer the highest degree of control over selective isotopic labelling[59,77–79]. Because active metabolism is absent, scrambling is essentially eliminated[65,77]. Isotopically enriched amino acids or their precursors can be added directly to the reaction

mixture, generally derived from *E.coli* extracts, allowing precise labelling schemes[65,66,77,80–82]. This includes amino acid-selective incorporation, side-chain methyl labelling, and even site-specific labelling via amber suppression and engineered tRNAs[62,65,66,68,68,78,79,82]. The open nature of cell-free reactions also allows fine-tuning of redox conditions and cofactors[73,77,79]. The main advantages are the elimination of scrambling, flexibility in labelling design, and suitability for proteins that are otherwise difficult to express in living cells. The trade-offs include higher costs and lower yields compared with bacterial systems[65,77–79].

Finally, selective isotopic labelling in mammalian cells has emerged as a powerful approach for studying proteins that require post-translational modifications and complex folding, features often inaccessible in bacterial or yeast expression systems[2,58,60,83]. Recent advancements have made selective labelling strategies more efficient and cost-effective, extending their utility in nuclear magnetic resonance (NMR) and mass spectrometry applications. Amino acid-selective and side-chain-specific protocols have been developed by supplementing mammalian culture media with isotopically labelled amino acids or their precursors[37,58]; for instance, nonadherent Expi293F cells grown in specially formulated media allow high-efficiency incorporation of labels into residues such as phenylalanine, isoleucine, valine, or leucine, with minimal scrambling[58]. This strategy is also compatible with the introduction of fluorinated amino acids for specialized biophysical studies[38,39]. Furthermore, genetic code expansion technologies now permit site-specific incorporation of isotopically labelled or noncanonical amino acids using engineered tRNA/synthetase pairs and amber codon suppression, providing precise control over labelling patterns at defined protein positions for advanced structural and functional analyses[64,84–86]. Despite these advances, mammalian systems remain more complex and costly compared with bacterial or yeast hosts, and label dilution due to scrambling continues to be a significant challenge[87,88]. While certain amino acids, including cysteine, phenylalanine, histidine, lysine, methionine, asparagine, arginine, threonine, tryptophan, and tyrosine, exhibit low scrambling tendencies, others such as alanine, aspartate, glutamate, isoleucine, leucine, and valine are more prone to metabolic redistribution[58,83]. Careful optimization of culture conditions, including tailored media formulations and precursor selection, is therefore essential to minimize scrambling and improve labelling fidelity. Ultimately, the extent of scrambling in mammalian systems depends on the metabolic pathways engaged by the host and the amino acid of interest, underscoring the need for system-specific strategies when designing selective isotopic labelling experiments.

1.2.1 Selective isotopic labelling with amino acid precursors

Selective isotopic labelling with precursors is a useful tool to simplify NMR spectra and reach assignment for NMR spectra which would be too crowded/not resolved in the case of uniformly labelled samples. Its effectiveness depends on both the choice of precursor and the metabolic capabilities of the host organism. It offers a cost-effective alternative to amino acid labelling while enabling equal targeted enrichment of specific isotopically labelled residues, thereby improving spectral resolution and simplifying analysis of large or complex proteins[37,89–91]. The process works by supplying isotope-labelled metabolic precursors, most effectively late-stage intermediates like α-ketoacids, into the growth medium, where they are taken up by the host cell and converted into the corresponding amino acids through the action of endogenous transaminases. These enzymes catalyse the transfer of amino groups (**Figure 1.2.1-1**), enabling

the transformation of the precursor into the desired labelled amino acid, which is then incorporated into newly synthesized proteins.

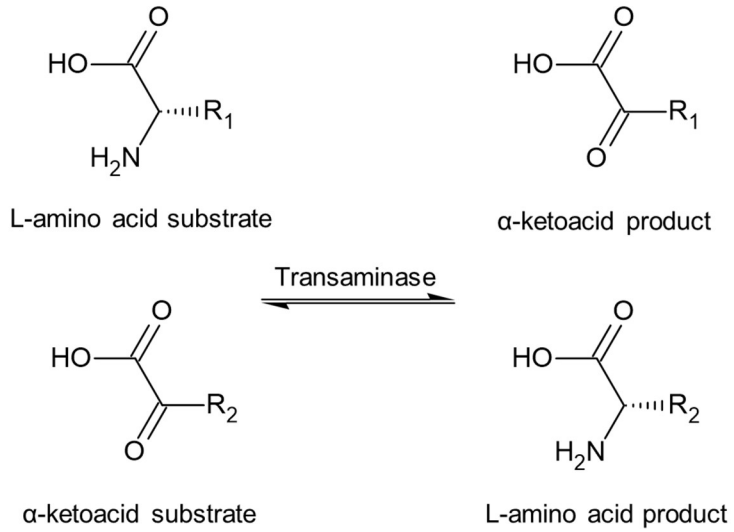


Figure 1.2.1-1
The transaminase enzyme transfers the amino group of an amino donor to an α-ketoacid acceptor thus generating a new amino and a new α-ketoacid products.

In bacterial systems like *E.coli*, this method is well-established: labelled precursors are efficiently converted by highly active but sometimes promiscuous transaminases, allowing for selective and cost-effective labelling of specific residues such as methyl groups in Leu, Val, and Ile, or aromatic rings in Phe and Tyr[66,76,89–91]. The advantages in bacteria include high labelling efficiency, low cost due to the small amounts of precursor required, and minimal metabolic scrambling when late-stage precursors are used, resulting in clear and interpretable NMR spectra. However, challenges can arise from metabolic cross-labelling in shared biosynthetic pathways (e.g., Val/Leu) and difficulties achieving stereo-selectivity and labelling-selectivity for certain side chains unless specific precursors, engineered strains or inhibitors are employed[89,91,91].

In mammalian cells, the approach is more recent but increasingly feasible. Here, a similar principle compared to *E.coli* applies, but the context is more complex: the human genome encodes at least aminotransferases, each with distinct substrate preferences and biological roles spanning amino acid interconversion, nitrogen metabolism, and energy regulation through pathways[92]. When isotope-labelled precursors are supplied in culture media, these enzymes efficiently convert them into target amino acids, enabling selective side-chain labelling, as demonstrated by our group in last year's publication for aromatic and methyl-containing residues[37]. However, this approach also proves efficient when using backbone labelled precursors, as shown in a preliminary work in this thesis (chapter 1). This approach yields minimal scrambling, making the process both cost-effective and straightforward. This is particularly valuable for proteins that require mammalian-specific folding or post-translational modifications. However, mammalian precursor labelling faces unique challenges: uptake efficiency and conversion rate varies by residue, and, furthermore, certain amino acids (Ala, Asp, Asn, Pro, Glu) that are directly linked to essential cellular metabolic pathways, like glycolysis and the TCA cycle, may be difficult to label with α-ketoacid because adding the precursors to the culture medium is likely to cause other amino acids to be unintentionally

labelled in the expressed proteins[37]. Additionally, while metabolic scrambling is generally low, it must be empirically assessed for each system.

The main differences between bacterial and mammalian systems stem from their metabolic flexibility and enzyme repertoire. Bacteria like *E.coli* have robust and versatile transaminase activity, allowing for efficient conversion of a wide range of precursors and enabling highly selective labelling with minimal cost and effort[66,76,89–91]. In contrast, mammalian cells, while capable of similar conversions for some amino acids, are more limited in precursor uptake and enzymatic activity, which can restrict the range of residues that can be selectively labelled and may necessitate more expensive labelled amino acids for certain targets. Thus, while both systems rely on transaminase-mediated conversion of precursors, bacterial hosts offer broader applicability and efficiency, whereas mammalian hosts provide access to proteins with complex modifications but with more constraints on labelling strategies[37].

1.2.2 Selective isotopic labelling of amino acid side chains

Selective isotopic labelling of amino acid side chains, particularly when combined with deuteration, has become a cornerstone of modern NMR spectroscopy, offering critical improvements in spectral resolution and sensitivity for the study of globular proteins and large biomolecular complexes[93–95]. By selectively labelling methyl groups of aliphatic residues (Val, Leu, Ile) or applying selective ^{13}C-labelling strategies to aromatics such as phenylalanine, tyrosine, tryptophan, and histidine, it is possible to drastically reduce spectral overlap and simplify resonance assignments, enabling detailed analysis of protein structure, dynamics, and molecular interactions that are otherwise inaccessible in uniformly labelled or unlabelled samples[37,93,96–98]. Methyl group labelling is especially powerful in combination with methyl TROSY experiments. Methyl TROSY is based on the selection of specific quantum coherences in methyl groups, that experience destructive interference between ^{1}H, ^{1}H and ^{1}H, ^{13}C dipolar interactions. This results in much slower relaxation rates, dramatically improving spectral resolution and signal-to-noise ratio in large, slowly tumbling molecules[99–103]. In practice, the HMQC sequence is already optimal to select the slowly relaxing methyl transitions, maximizing sensitivity and resolution[100–102].

Aromatic side chain labelling, on the other side, is essential for high-resolution structural determination and for probing protein-ligand or protein-protein interfaces, as these residues often play key roles in binding and allosteric regulation[89,104–107].

The integration of deuteration further enhances these benefits: by replacing neighbouring protons with deuterons, dipolar relaxation of the observed ^{1}H is reduced, leading to sharper peaks, longer coherence lifetimes, and greatly improved sensitivity at the labelled sites[102,108,109]. This selective protonation in a deuterated background allows specific signals from methyl groups or aromatics to stand out with high sensitivity, facilitates the acquisition of long-range distance restraints (e.g., methyl-methyl, methyl-NH), and makes it possible to acquire multidimensional NMR experiments such as Heteronuclear Multiple Quantum Coherence (HMQC) and high-dimensional correlation techniques[35,93,109,110].

HMQC is a foundational two-dimensional NMR experiment that correlates signals between protons and heteronuclei (commonly ^{13}C or ^{15}N), enabling detailed analysis of molecular structure and dynamics. HMQC uses a series of radiofrequency pulses to create and transfer

multiple quantum coherences between a proton and a heteronucleus. This process allows the experiment to detect correlations between specific pairs of nuclei, such as ^{1}H, ^{13}C or ^{1}H,^{15}N[111,112]. The experiment records the evolution of the heteronucleus during an indirect time period, but detection is performed on the proton channel, leveraging the higher sensitivity of ^{1}H detection[111,113]. The basic HMQC sequence involves excitation, coherence transfer via scalar (J) coupling, evolution, and detection. Variants and optimizations (e.g., SOFAST-HMQC) improve speed, sensitivity, and resolution[103,112,114,115]. HMQC produces a two-dimensional spectrum where one dimension represents proton chemical shifts and the other the heteronucleus, revealing which protons are directly bonded to which heteronuclei[111,112].

The sensitivity gain, especially in the case of methyl labelling is significant, with a 2.4-fold or more enhancement compared to other labelling schemes. This enables the detection of weak interactions and the possibility to observe protein samples in nanomolar concentration[93,100,116].

Recent methodological advances in the protein production have made these approaches more cost-effective and versatile in both bacterial and mammalian systems, α-ketoacid precursors, or transaminase pathways to maximize labelling efficiency while minimizing scrambling[37]. Biologically, these strategies have enabled high-resolution studies of supramolecular assemblies, membrane proteins, and transient protein-protein or protein-ligand interactions, thereby providing insights that are crucial for understanding molecular mechanisms and advancing drug discovery[30,37,93]. However, despite their success with globular proteins, side chain labelling strategies are less advantageous for intrinsically disordered proteins (IDPs). The lack of stable secondary and tertiary structure in IDPs leads to severe spectral overlap, poor chemical shift dispersion, and broadened signals that largely offset the resolution gains of selective labelling. Moreover, the repetitive nature of IDP sequences complicates resonance assignment, while label scrambling and high experimental costs can further diminish the utility of these approaches in such systems[117–119].

1.2.3 Selective isotopic labelling of amino acid backbone

When side chain labelling is not a viable option it is possible to resort to backbone labelling. This technique is the foundation of structural biology as it has always been used for protein assignment but it's also extremely useful in the investigation of large proteins or proteins with high spectral degeneracy[76,120,121]. In particular two labelling schemes are extremely interesting as they allow great spectral simplification while retaining fundamental structural information: ^{15}N, $^{13}C\alpha$ (alpha carbon) and ^{15}N, $^{13}C'$ (carbonyl). The selective labelling with these schemes applied to single residues has to be carried out in a ^{15}N uniformly labelled background to exploit the whole experimental potential. This can be easily done by supplying an excess of selectively labelled amino acid or cursor in a ^{15}N-labelled culture medium[122] (chapter 1). These labelling strategies enable the use of triple-resonance experiments such as HNCA and HNCO, which are essential for establishing connections between the ^{15}N, ^{13}C amino acids and the subsequent residue in the polypeptide chain.

Specifically, the HNCA experiment transfers phase coherence from the amide proton ($^{1}H_N$) to the amide nitrogen (^{15}N), and then to the alpha carbon ($^{13}C\alpha$) within the same residue and the preceding residue in the protein sequence. Therefore, each cross-peak in the HNCA spectrum links the HN and N of residue i to the Cα of both residue i (intraresidue) and residue i-1 (sequential) via scalar coupling[123–126]. However, when selective isotopic labelling is applied,

only one amino acid will be ^{13}Cα-labelled, thus by observing the HN plane of the 3D spectra we will observe a peak corresponding to the ^{15}N, ^{13}Cα residue and a peak corresponding to the following amino acid in the protein sequence.

Meanwhile, in the HNCO experiment, phase coherence is transferred from the amide proton ($^{1}H_N$) to the amide nitrogen (^{15}N), and then to the carbonyl carbon (^{13}C') of the preceding residue (i-1) via scalar coupling. Consequently, each cross-peak in the HNCO spectrum links the HN and N of residue i to the C' of residue i-1[127,128]. In this case, due to the presence of ^{13}C' just on the selectively labelled amino acid, the HN plane of the HNCO only shows the cross peak of the following amino acid in the protein sequence.

Both labelling strategies are invaluable for interaction studies: they allow for the detection of chemical shift perturbations upon ligand or protein binding, mapping interaction sites with high precision, and supporting combinatorial selective labelling approaches that accelerate assignment and facilitate the identification of binding interfaces[129,130]. Furthermore, by having the background uniformly ^{15}N-labelled, scrambling, that almost always occurs on the nitrogen of the labelled residue, is made completely irrelevant[83,120].

For IDPs, backbone labelling is crucial due to their inherent spectral overlap due to their repetitive and dynamic nature[117–119]. Especially in the case of large proteins, selective backbone labelling of specific residues drastically reduces spectral crowding. This results in greatly simplified spectra that, in turn, allow to better understand local and long-range dynamics, sequence-dependent flexibility, and transient secondary structure in IDPs and monitor interaction-induced changes[67,76,119,120,131,132].

1.2.4 Selective labelling with fluorinated amino acids

The last labelling technique in analysis involves the introduction of non-natural fluorinated amino acids in the protein sequence, which has become a popular tool in NMR-based structural and interaction studies due to the unique properties of the NMR active nucleus and the development of advanced labelling and incorporation strategies. ^{19}F is highly attractive for NMR because it is 100% naturally abundant, highly sensitive, and absent from native biological systems, resulting in background-free spectra and enabling the detection of proteins at low concentrations[74,133–137].

Incorporation of fluorinated amino acids can be achieved through several methods, including genetic code expansion using orthogonal tRNA/aminoacyl-tRNA synthetase (aaRS) pairs, such as pyrolysis-tRNA synthetase/tRNA systems, which have been engineered for high fidelity and efficiency in both *E.coli* and mammalian cells[64,136,138]. These systems rely on the introduction of a stop codon (e.g., the amber stop codon TAG) into the gene of interest (GOI) at the position corresponding to the amino acid targeted for substitution with its non-natural counterpart. Within the host cell, an engineered orthogonal aaRS is co-expressed; this enzyme specifically recognizes the matching orthogonal tRNA and charges it with the desired non-natural amino acid (nnAA). When the ribosome encounters the reassigned stop codon during translation, the orthogonal tRNA delivers the nnAA[139]. These systems allow for site-specific labelling, including the use of various fluorinated phenylalanine analogues, and can be multiplexed for the simultaneous incorporation of multiple non-natural amino acids[136,140]. In mammalian cells, direct expression and medium-switch strategies have enabled efficient incorporation of

fluorinated amino acids, with incorporation efficiencies up to 60% and minimal impact on protein structure, as confirmed by mass spectrometry and X-ray crystallography[39,74,133]. The stochastic nature of incorporation in mammalian systems allows for predictive control over labelling levels[133].

Fluorinated amino acids provide several advantages in NMR studies: they offer high sensitivity to local environmental changes, enabling the detection of conformational heterogeneity, protein folding, and ligand binding events that may be invisible to traditional ^{1}H or ^{15}N NMR[74,134,137,140]. Their site-specific incorporation reduces spectral overlap, simplifying analysis of large or complex proteins[133,138]. Advanced NMR techniques, such as 2D ^{13}C, ^{19}F correlation spectra, allow for site-specific assignments, long-range distance measurements, and enhanced spectral resolution, especially when using selectively fluorinated and isotopically labelled amino acids[39]. These methods have been successfully applied to study protein-ligand and protein-protein interactions, as well as to probe spatial proximity and conformational dynamics in both bacterial and mammalian systems[39,74,137,140,141].

Regarding toxicity, the effects of fluorinated amino acids vary by structure and host. In *E.coli*, some fluorinated amino acids (e.g., 3-fluoroalanine) are highly toxic and can only be incorporated using cell-free systems, while others are tolerated and incorporated efficiently[142]. In mammalian cells, toxicity is less well characterized and must be evaluated for each ^{19}F nnAA and every protein but studies indicate that incorporation of many fluorinated aromatic residues is achievable in globular proteins (such as carbonic anhydrase isoform II (CA II) (**Figure 1.2.4-1**) and superoxide dismutase 1 (SOD1)) without disruption of protein structure or function at controlled concentrations[39,74,133].

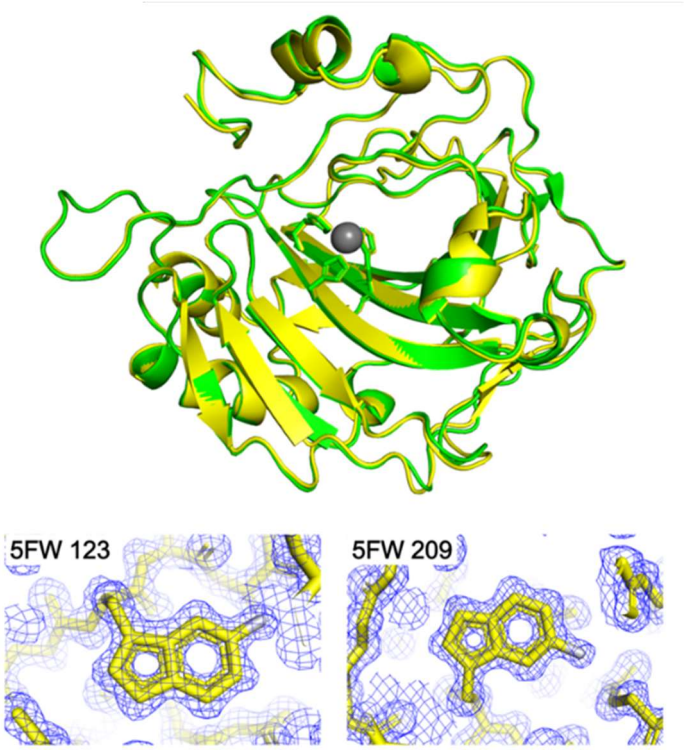


Figure 1.2.4-1
Structure of carbonic anhydrase isoform II with 5F tryptophan[133].

2. Complementary methods

2.1 Flow cytometry

Flow cytometry is a sophisticated technique for the rapid, multi-parametric analysis of individual cells or particles flowing in suspension through a measuring device. Its operation fundamentally relies on the light scattering and fluorescence emission that occur when a light source, typically a laser beam, strikes the moving cells[143,144].

A flow cytometer is comprised of three core systems: fluidics, optics, and electronics[143–146]. The fluidics system is crucial for precisely transporting cells to the laser interrogation point. This is achieved through hydrodynamic focusing, where a sample containing suspended cells is injected into a flow chamber, surrounded by a sheath fluid (often phosphate-buffered saline, PBS). Due to pressure differences, cells align in a single line as they pass through the laser beam, ensuring uniform illumination of each cell. The injection rate can be adjusted depending on the purpose of the analysis. High flow rates are generally used for qualitative measurements while slow rates for higher resolution applications. Maintaining laminar flow and preventing air bubbles or debris is critical for proper operation[143–147].

The optical system is divided into excitation and collection components. Lasers (e.g., 488 nm blue, 405 nm violet, 640 nm red) energize electrons in fluorochromes, causing them to produce light[143,145]. When the laser strikes a cell, light is scattered in two main ways: Forward Scatter (FSC), collected along the same axis as the laser, is proportional to the cell's size or surface area, and Side Scatter (SSC), collected at approximately 90 degrees, is proportional to the cell's granularity or internal complexity. Both FSC and SSC are valuable for differentiating cell types in heterogeneous populations. Additionally, cells can be stained with fluorescent dyes or antibodies conjugated to fluorochromes, which bind to specific cellular molecules[143–145]. When excited by the laser, these fluorochromes absorb and then emit light at a longer wavelength as they return to their ground state. The difference between excitation and emission maxima is known as the Stokes shift[143–146]. Common fluorochromes include fluorescein isothiocyanate (FITC), phycoerythrin (PE), allophycocyanin (APC), and propidium iodide (PI)[143–145].

The collection optics (lenses, mirrors, filters) gather the scattered and fluorescent light, using dichroic filters to transmit or reflect light based on wavelength, and bandpass filters to allow only a narrow range of specific wavelengths to reach the detectors, ensuring independent detection of each fluorochrome[143,145]. Recent advancements include spectral flow cytometers, which capture the full emission spectrum of all fluorophores and use algorithms to distinguish them, reducing the need for traditional compensation, and imaging flow cytometers, which combine flow cytometry with fluorescence microscopy to capture detailed cell morphology and marker distribution[147,148].

Finally, the electronic network converts light signals into proportional electrical currents using photodetectors like photodiodes for strong FSC signals and photomultiplier tubes for weaker SSC and fluorescence signals[143,145]. These electrical currents are linearly or logarithmically amplified and converted from analog to digital data by analog-to-digital converters for computer processing. The digital data is then displayed as plots or histograms, allowing researchers to analyse, interpret, and "gate" specific cell populations for further study[143,144,146,147].

2.1.1 Fluorescence-Activated Cell Sorting (FACS)

Fluorescence Activated Cell Sorters (FACS) are a specialized type of flow cytometer that provide a method to physically separate cells into subpopulations based on their cytometric analysis[143,144,149].

This is achieved by generating regular and uniformly sized droplets from the cell stream containing the already analysed cells, charging said droplets via passing through a charging electrode, and then deflecting them into collection vessels using high-voltage electrostatic plates[143,147,149,150].

Cells can be sorted in various mode, however, a key parameter to consider is the drop delay, which is the time interval between the cell's detection by the laser and the moment the droplet containing that cell breaks off and is charged[147]. The most commonly used sorting logics are purity mode, where if the exact location of the cell cannot be confirmed the droplet is discarded to minimize the presence of non-target cell, enrichment mode, where if the target cell is present, the droplet is collected regardless of other cells, and finally single-cell mode, where only droplets containing a single cell are collected[145,147,149].

There are primarily two types of cell sorters: traditional electrostatic sorters and microfluidic-based sorters. Electrostatic sorters are further divided into "jet-in-air" systems, where cells are ejected from a nozzle into the open air for laser interrogation, and "hybrid cuvette" systems, where cells are first analysed in a quartz cuvette and then ejected into the open air for droplet formation. Hybrid cuvette systems often offer better fluorescence sensitivity[145,149].

Microfluidic-based cell sorters are an innovation that utilize chips with microchannels and mechanical or air pressure deflection mechanisms[147,149]. These systems offer several advantages: they use disposable chips (preventing cross-contamination), are closed systems (eliminating aerosols), and apply lower pressures, which enhances cell viability. However, they are generally slower than traditional electrostatic sorters (<1,000 cells per second vs. >10,000 cells per second) and typically collect fewer populations[147,149]. Despite their slower speed, microfluidic systems are gaining traction for applications requiring high cell viability and sterility, especially for therapeutic cell sorting[147].

2.2 Fluorescence microscopy

Fluorescence microscopy (FM) is a fundamental biomedical imaging tool and a critical technique to understand biological processes at the cellular level. It provides high-resolution images of molecular contrast in living samples and is widely used to observe discrete subcellular structures and processes. This technique allows for the visualization and analysis of complex dynamic events within cells, organelles, and their sub-components[151–153]. Fluorescence microscopy works by exploiting the phenomenon of fluorescence, where certain atoms or molecules, called fluorophores, absorb electromagnetic radiation at a specific wavelength and then emit it at lower-energy and longer wavelength[151,153,154].

In its basic setup, a fluorescence microscope typically consists of an illumination source, such as a xenon or mercury lamp or a laser, a magnifying lens (objective), and an image acquisition device[153,155]. A popular configuration is epi-illumination FM, where the objective lens serves a dual role in both illuminating the sample and collecting the emitted fluorescence[153,154]. To separate the excitation light from the weaker fluorescence signal, a dichroic beam splitter is

used, reflecting the shorter excitation wavelengths towards the sample and transmitting the longer emission wavelengths towards the detectors[154]. Additional filters, such as excitation and emission band-pass filters, further improve the signal-to-noise ratio[153,154]. The fluorescence signal is then recorded by a camera, such as a charge-coupled device or scientific complementary metal-oxide-semiconductor camera[155,156]. This selective filtering process allows the fluorophore-labelled structures to appear brightly against a dark background, providing significantly enhanced contrast compared to other imaging methods[153–155]. The field has evolved from visualizing autofluorescent specimens to using fluorescent stains, labelled antibodies, and genetically encoded fluorescent proteins like GFP, enabling targeted live-cell imaging[152,153].

2.2.1 Confocal microscopy

Confocal microscopy is an advanced form of fluorescence microscopy that achieves higher contrast, and spatial resolution allows to visualize subcellular details beyond conventional wide-field systems[152,157]. Invented by Minsky in 1955, its fundamental purpose is to provide optical sectioning, enabling the recording of three-dimensional (3D) images that are impossible with wide-field microscopy[152–154,157]. This is achieved by using a focused laser beam for excitation, raster scanning the sample with galvanometric mirrors, and incorporating a pinhole in the detection pathway. The pinhole spatially filters out-of-focus and stray light, ensuring that only fluorescence originating from the tightly focused focal plane reaches the single-photon detector[151,153,154,154,157]. This process, which often involves descanning the emitted light back through the scanning system, significantly enhances image contrast and resolving power, making labelled structures appear brightly against a dark background[153,154,157]. While confocal microscopy can achieve a twofold higher lateral resolution than wide-field microscopy with an infinitely small pinhole, this comes at the cost of reduced light detection efficiency and signal-to-noise ratio[155]. The sequential, point-by-point scanning process makes image acquisition inherently slower compared to wide-field methods, although it limits light exposure only to the scanned area, which can reduce overall photobleaching in certain contexts[154,157].

3. Towards cost-effective side-chain isotope labelling of proteins expressed in human cells

3.1 Summary

In this work, we introduced an innovative and economical methodology to selectively label protein side chains with stable isotopes in human cells. We demonstrated that selective isotopic labelling can be achieved in mammalian cells by supplementing the culture medium with amino acid late precursors (α-keto acids) instead of the corresponding amino acids. Human transaminases are, in fact, capable of converting these precursors into the cognate amino acids, allowing efficient incorporation into proteins.

We showed that transfecting HEK293T cells with a pHL-sec plasmid encoding carbonic anhydrase II (CA II) in the presence of a custom-made DMEM medium, where an amino acid was replaced by a two-fold molar concentration of its α-keto acid precursor, led to protein expression levels comparable to those obtained in standard commercial DMEM. Protein expression levels in the presence of precursors of seven amino acids (Tyr, Phe, Leu, Val, Met, Ile, His) was assessed by overexpression of CA II in HEK293T cells, followed by lysis through freeze-thaw cycles and subsequent SDS-PAGE analysis.

After confirming that protein expression level was maintained in the presence of precursors, we tested two different side chain labelling strategies to demonstrate specific incorporation without scrambling. First, we employed ^{1}H, ^{13}C labelling of aromatic amino acids (Phe, Tyr) in a deuterated background; second, we implemented ^{1}H, ^{13}C methyl labelling of aliphatic residues (Val, Leu, Met). Incorporation was validated by transfecting HEK293T cells with two proteins, CA II and DJ-1, using custom DMEM lacking the target amino acid but supplemented with the corresponding α-keto acid precursor. The resulting cell samples were analysed by in-cell NMR, then recovered, lysed, and subjected to the same NMR experiments in the lysate state.

All NMR experiments were carried out at 950 MHz and 37 °C using SOFAST and XL-ALSOFAST HMQC pulse sequences. In both in-cell and lysate conditions, the spectra confirmed efficient and residue-specific incorporation of labelled side chains without evidence of isotopic scrambling.

We also proved expression and incorporation of two precursors in the same sample by overexpressing DJ-1 in the presence of both Val and Leu precursors.

Finally, to assess the functional utility of this labelling strategy, CA II-expressing cells were treated with known inhibitors, acetazolamide (AAZ), methazolamide (MZA), and ethoxzolamide (ETZ). The resulting chemical shift perturbations observed in the NMR spectra confirmed that this selective side chain labelling approach enables sensitive detection of ligand binding events directly in a cellular environment.

Journal of Biomolecular NMR (2024) 78:237–247
https://doi.org/10.1007/s10858-024-00447-6

ARTICLE

Towards cost-effective side-chain isotope labelling of proteins expressed in human cells

Martina Rosati[1] · Letizia Barbieri[1,2] · Matus Hlavac[3] · Sarah Kratzwald[3] · Roman J. Lichtenecker[3,4] · Robert Konrat[3,5] · Enrico Luchinat[1,2,6] · Lucia Banci[1,2,6]

Received: 23 February 2024 / Accepted: 17 July 2024 / Published online: 22 August 2024

Abstract

Side chain isotope labelling is a powerful tool to study protein structure and interactions by NMR spectroscopy. $^{1}H,^{13}C$ labelling of side-chain methyl groups in a deuterated background allows studying large molecules, while side-chain aromatic groups are highly sensitive to the interaction with ligands, drugs, and other proteins. In *E. coli*, side chain labelling is performed by substituting amino acids with isotope-labelled precursors. However, proteins that can only be produced in mammalian cells require expensive isotope-labelled amino acids. Here we provide a simple and cost-effective method to label side chains in mammalian cells, which exploits the reversible reaction catalyzed by endogenous transaminases to convert isotope-labelled α-ketoacid precursors. We show by in-cell and in-lysate NMR spectroscopy that replacing an amino acid in the medium with its cognate precursor is sufficient to achieve selective labelling without scrambling, and how this approach allows monitoring conformational changes such as those arising from ligand binding.

Keywords Protein labelling · Alpha-ketoacid · In-cell NMR · HEK293T cells · Mammalian expression · Transaminases

Introduction

Interactions underpin all biological processes. They occur between proteins, between proteins and their biological ligands and inhibitors, or between proteins and synthetic molecules. NMR spectroscopy is ideally suited to investigate such interactions at atomic resolution, in physiologically relevant settings, e.g. in intact cells and lysates (Theillet and Luchinat 2022). The various types of interactions involve specific types of amino acids. Therefore, it is essential to develop novel, efficient, and cost-effective technologies to examine interactions by looking at the side chains of the specific amino acids relevant for molecular interactions (Schörghuber et al. 2018). Furthermore, structural characterization of large proteins by NMR requires residue-specific labelling in order to improve spectral resolution and sensitivity, and to reduce signal overlap. Two side chain labelling types are particularly useful to investigate protein structure, dynamics and interactions: $^{1}H,^{13}C$ labelling of aromatic residues and methyl-$^{1}H,^{13}C$ labelling of aliphatic residues. Aromatic residues are often involved in the interaction with other proteins, ligands or drugs, and take part in enzymatic activities (Serrano et al. 1991; Schörghuber et al. 2018; Ninković et al. 2020). On the other hand, labelled methyl groups of aliphatic residues are highly useful when investigating slow-tumbling macromolecules, thanks to their high intrinsic mobility and three-fold proton multiplicity, which result in a greatly enhanced sensitivity. Methyl labelling allows for the recording of NMR experiments of large proteins that would otherwise feature exceedingly broadened signals due to fast transverse relaxation (Kerfah et al. 2015). In cells, a similar

✉ Enrico Luchinat
eluchinat@cerm.unifi.it

✉ Lucia Banci
banci@cerm.unifi.it

[1] CERM — Magnetic Resonance Center, Università degli Studi di Firenze, Sesto Fiorentino, Italy

[2] Consorzio Interuniversitario Risonanze Magnetiche di Metallo Proteine — CIRMMP, Sesto Fiorentino, Italy

[3] MAG-LAB GmbH, Vienna, Austria

[4] Institute of Organic Chemistry, University of Vienna, Vienna, Austria

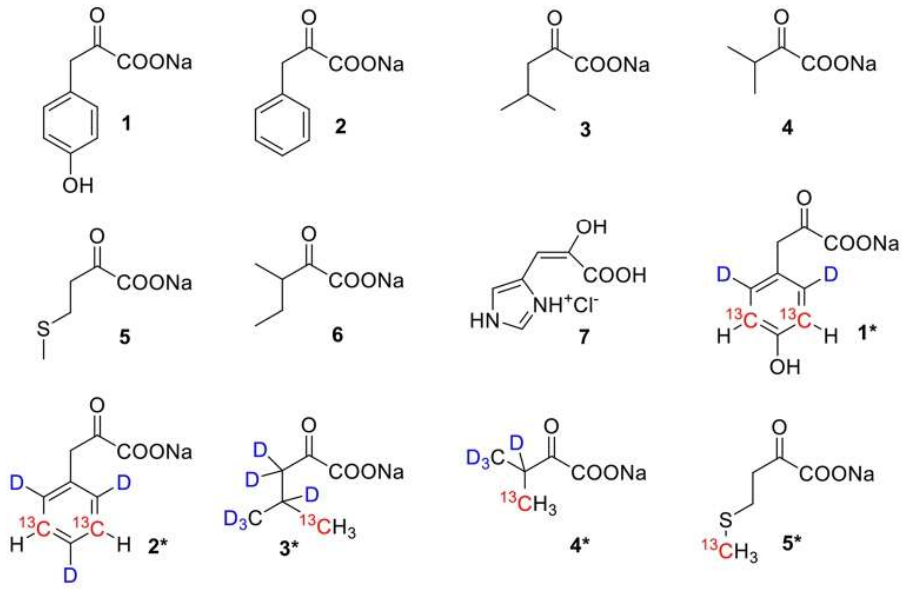
[5] Department of Structural and Computational Biology, Max Perutz Laboratories, University of Vienna, Vienna, Austria

[6] Dipartimento di Chimica, Università degli Studi di Firenze, Sesto Fiorentino, Italy

scenario is also encountered when studying small globular proteins which interact with large, abundant cellular components (Barbieri et al. 2015; Burz et al. 2019). In both cases, the presence of a deuterated side chain background further increases the sensitivity by decreasing the ^{1}H spin relaxation rates. Moreover, sparse labelling schemes greatly simplify the spectra of large proteins by reducing spectral overlap.

In *E. coli*, side chain labelling is typically performed by substituting amino acids with the cognate labelled α-ketoacid precursors to the growth medium, as these compounds can be synthesized from commercially available isotope sources without the need of creating chiral carbons by elaborate organic synthesis. Isotope-labelled α-ketoacids are then effectively incorporated into the host organism's biosynthetic pathways and converted into the respective amino acids by bacterial transaminases, which substitute the keto group with an amino group in vivo (Goto et al. 1999; Lichtenecker et al. 2004, 2013a, c). Although this method is extremely effective, it has never been applied to those proteins, which require production in mammalian cells. Indeed, previous reports on selective isotope labelling in mammalian cells had to make use of synthetically more demanding and thus more expensive isotope-labelled amino acids as additives to substitute the unlabelled target residues (Werner et al. 2008; Banci et al. 2013). In mammals, transaminases reversibly catalyze the first catabolic reaction in the degradation of amino acids, transferring their amino group to a ketoacid substrate, usually α-ketoglutarate, in order to obtain the corresponding α-ketoacid and glutamate (Caligiore et al. 2022). As some transaminases are known to be expressed inside mammalian tissues (Caligiore et al. 2022; Neinast et al. 2019; Toyokawa et al. 2021; Sivaraman and Kirsch 2006; Mehere et al. 2010; Cooper 2004), we reasoned that supplementing expression media with α-ketoacids should result in the biosynthesis of the corresponding amino acids in cultured mammalian cells, thus allowing the expression of proteins with amino acid-specific labelling schemes at a lower cost.

In this work, we show that HEK293T cells supplied with custom-made culture media containing an α-ketoacid precursor in place of the corresponding amino acid result in a high-yield transient expression of the desired proteins. These experiments were performed with the unlabelled α-ketoacid precursors of tyrosine (p-hydroxyphenylpyruvate **1**), phenylalanine (phenylpyruvate **2**), leucine (α-ketoisocaproate **3**), valine (α-ketoisovalerate **4**), methionine (4-methylthio-α-ketobutanoic acid **5**), isoleucine (3-methyl-α-ketovalerate **6**) and histidine (4-(2-carboxy-2-hydroxyvinyl)-*1H*-imidazolium chloride **7**). Selective side-chain labelling protocols were demonstrated on two model proteins, carbonic anhydrase II (CA II) and the Parkinson-related protein deglycase 1 (DJ-1), by recording fast 2D $^{1}H,^{13}C$ NMR spectra of both intact cells and cell lysates. The unlabelled precursors, as well as the corresponding heavy isotope containing isotopologues (compounds **1***-**5***) are shown in scheme 1 and have been synthesized using previously reported protocols



Scheme 1 Amino acid precursors used in this study. p-Hydroxyphenylpyruvate **1**; Phenylpyruvate **2**; α-Ketoisocaproate **3**; α-Ketoisovalerate **4**; 4-Methylthio-α-ketobutanoate **5**; 3-Methyl-α-ketovalerate **6**; 4-(2-Carboxy-2-hydroxyvinyl)-*1 H*-imidazolium chloride **7**; ([3,5-$^{13}C_2$, 2,6-$^{2}H_2$] p-Hydroxyphenyl)pyruvate **1***; ([3,5-$^{13}C_2$, 2,4,6-$^{2}H_3$] Phenyl)pyruvate **2***; [5-^{13} C, 3,3,4,5,5,5-$^{2}H_6$] α-Ketoisocaproate **3***; [4-^{13} C, 3, 4,4,4-$^{2}H_4$] α-Ketoisovalerate **4***; 4-[^{13}C] Methylthio-α-ketobutanoate **5***

(Lichtenecker et al. 2004, 2013a; Fischer et al. 2007; Lichtenecker 2014).

Materials and methods

Plasmid design

Following the procedures outlined in previous studies, DNA sequences encoding for carbonic anhydrase 2 (CA II, NP_000058.1), deglycase protein 1 (DJ-1, NP_009193.2), and empty vector (pHL-empty) were cloned in the pHLsec vector, deprived of the secretion sequence (Aricescu et al. 2006).

Custom medium preparation

The composition of the custom-made medium was made accordingly to that of commercial high-glucose Dulbecco's modified Eagle's medium (DMEM, Sigma D6546), excluding the amino acid(s) to be labelled or to be replaced with precursors. The components were weighed on an analytical scale and dissolved in ultrapure water. The solution was left stirring overnight at 4 °C and the pH was brought to 7.4. The medium was then filtered under a laminar flow hood with a 0.22 nm filter and stored at -80 °C in 50 mL aliquots. The labelled amino acids or precursors were supplemented to the medium from concentrated stock solutions at the time of transfection.

Expression tests

HEK293T cells (ATCC CRL3216) were cultured in T25 flasks with DMEM (Life Technologies). At 90-95% confluence the cells were transiently transfected with the vectors in a custom-made medium where one amino acid was substituted with the corresponding precursor (MAG-LAB) in different concentrations with respect to the concentration of the amino acid (aa: precursor 1:1, 1:2, 1:5). The transfection was carried out following a previously reported protocol (Barbieri et al. 2016), in which vector and polyethylenimine (PEI, Sigma-Aldrich) were mixed in a 1:2 ratio (8.3 µg of DNA: 16.7 µg of PEI). After transfection the cells were incubated for 48 h at 37 °C, 5% CO_2, with custom-made DMEM supplemented with 2% (v/v) fetal bovine serum (FBS, Life Technologies) and 100 µg/mL penicillin – streptomycin (Life Technologies), respectively. It was previously shown that CA II and DJ-1 after 48 h of transient expression are correctly folded and do not undergo degradation (Barbieri et al. 2018; Luchinat et al. 2020a). The cells were then harvested in 150 µL of phosphate buffered saline (PBS, Life Technologies), lysed by freeze and thaw and the soluble part of the lysate was collected after centrifugation. The expression level of CA II was evaluated by Coomassie-stained SDS-polyacrylamide gel electrophoresis (SDS-PAGE). The loading sample, consisting in 7 µL of a solution made by diluting the soluble lysate 1:3 in 100mM Tris-HCl pH 6.8, 12.5% (v/v) glycerol, 2% (w/v) sodium dodecyl sulfate (SDS), 0.1% (w/v) bromophenol blue, and 50 mM dithiothreitol (DTT), was loaded onto pre-cast gels (Bio-Rad) and run at 200 V for 30 min. The gels were then stained with Coomassie Blue dye (ProBlue Safe Stain, Giotto Biotech).

Protein quantification

CA II concentration in the lysates from each condition was quantified by densitometric analysis on Coomassie-stained SDS-PAGE. Three dilutions of the lysates of the expression tests of CA II with two doses of each precursor, and with commercial DMEM were loaded onto a pre-cast gel as previously described and compared to a linear regression standard curve made of four samples of purified CA II at increasing concentrations (1, 2, 5, 10 µM). The gel was stained with Coomassie Blue dye, imaged by ChemiDoc XRS (Bio-Rad), and analyzed with Image J. Densitometry was used to quantify the protein expression based on the standard curve.

NMR sample preparation

HEK293T cells cultured in T75 flasks with DMEM were transiently transfected in custom-made medium with one amino acid substituted by a 2x dose of the cognate labelled precursor. Cell transfection was performed following the aforementioned protocol scaled up for T75 flasks (25 µg of DNA: 50 µg of PEI). Protein expression was carried out for 48 h in custom-made DMEM as above. Cells were harvested by trypsinization and suspended in 180 µL of NMR buffer (DMEM, 70 mM HEPES, and 20% (v/v) D_2O), transferred in a 3 mm Shigemi NMR tube, gently pelleted and inserted in the NMR instrument for analysis. At the end of the NMR experiment the cells were collected and lysed in PBS buffer by freeze and thaw as described before. The soluble part of the lysate was supplemented with 10% D_2O, transferred in a 3 mm NMR tube and inserted in the NMR instrument for analysis. For samples supplied with two precursors at the same time, the experiment was carried out with the same protocol using a custom-made medium containing both precursors and lacking the corresponding amino acids.

NMR spectra acquisition and analysis

All NMR spectra were recorded at 310 K on a 950 MHz ^1H Avance III (Bruker) with pulse sequences, acquisition

Table 1 NMR acquisition parameters

Precursor		1*	2*	3*	4*	5*	1*+2*	3*+4*
Pulse sequence		SOFAST HMQC	SOFAST HMQC	XL ALSO-FAST HMQC	ALSOFAST HMQC	XL ALSO-FAST HMQC	SOFAST HMQC	XL ALSO-FAST HMQC
Duration (minutes)		44	44	61	46	59	44	61
^{1}H spectral width (ppm)		17.5	17.5	13	16	13	17.5	13
^{13}C spectral width (ppm)		13	13	30	25	20	16	30
^{13}C frequency offset (ppm)		117.6	130.6	25.1	22.6	17.6	124.6	25.1
^{13}C FID size	**in-cells**	64	64	128	128	128	80	152
	lysate	96	96	256	256	256	116	256
Dummy scans		64	64	64	64	64	64	64
Scans	**in-cells**	96	96	64	64	64	80	56
	lysate	64	64	32	32	32	56	32
Interscan delay (ms)		350	350	350	250	350	350	350

Table 2 Labelled precursors employed in this study

Precursor	Extended name	Labelling scheme
1*	[$^{13}C_2$, $^{2}H_2$] Hydroxyphenylpyruvate	Aromatic ^{13}C-^{1}H, deuterated ring
2*	[$^{13}C_2$, $^{2}H_3$] Phenylpyruvate	Aromatic ^{13}C-^{1}H, deuterated ring
3*	[^{13}C, $^{2}H_6$] α-Ketoisocaproate	Methyl $^{13}CH_3$, deuterated side chain
4*	[^{13}C, $^{2}H_4$] α-Ketoisovalerate	Methyl $^{13}CH_3$, deuterated side chain
5*	[^{13}C] Methylthio- α-ketobutanoate	Methyl $^{13}CH_3$

windows and parameters specifically selected for each labelled residue (Table 1). Aromatic-labelled samples were analyzed with 2D ^{1}H-^{13}C SOFAST HMQC spectra (Schanda and Brutscher 2005; Sathyamoorthy et al. 2014), while methyl-labelled samples were analyzed with either 2D ^{1}H-^{13}C ALSOFAST or 2D ^{1}H-^{13}C XL-ALSOFAST HMQC spectra (Mueller 2008; Rößler et al. 2020). The experiments were performed with the same spectral window for both in-cell and lysate samples; for lysates we increased the size and decreased the number of scans per increment to increase the FID resolution while keeping constant the duration of the experiment (Table 1). Peak counting was performed automatically in Topspin and manually edited to account for partial peak overlap and to exclude peaks arising from the cellular background. CA II concentration in the cell lysates analyzed by NMR was estimated by comparing the signal intensity in the imino region of 1D ^{1}H NMR spectra, which is free from cellular background signals, with those of a reference cell lysate sample containing CA II recorded with identical acquisition parameters. The concentration of CA II in the reference lysate sample (240 μM) was determined by titration with the CA inhibitor methazolamide (MZA), previously shown to bind CA II in the slow exchange regime (Luchinat et al. 2020a).

CA II-ligand interaction

CA II was overexpressed in presence of compound **4*** and treated with inhibitors to validate the labelling method for the study of interactions. After transfection and incubation for 48 h the cells were treated with one of the following inhibitors: acetazolamide (AAZ) 50 μM for 2 h, MZA 10 μM for 1 h, ethoxzolamide (ETZ) 10 μM for 1 h. After treatment, cells were collected and the samples were prepared and analyzed as reported above.

Results

Protein expression with α-ketoacid precursors

Custom-made media were prepared, in which a single α-ketoacid precursor was supplemented in place of the corresponding amino acid. Tyrosine precursor **1**, Phenylalanine precursor **2**, Leucine precursor **3**, Valine precursor **4**, Methionine precursor **5**, Isoleucine precursor **6** and Histidine precursor **7** were tested (Scheme 1). For subsequent NMR analysis, isotope-labelled precursors were employed (**1*-5***, Scheme 1 and Table 2).

Protein expression efficiency in media supplemented with each precursor was tested by evaluating the expression level of CA II by SDS-PAGE. Each expression test was carried out with increasing amount of precursor (1x, 2x, and 5x doses relative to the concentration of the corresponding amino acid in the commercial DMEM) and compared to the expression in commercial DMEM (Fig. 2). For all precursors tested, the 2x dose was sufficient to achieve expression levels comparable to those obtained with the corresponding

amino acid. Lower levels were obtained with both 1x dose and 5x dose of precursor, the latter being most likely a result of cytotoxic effects due to the high concentration of precursor (Fig. 1). Quantitative analysis revealed that in the lysates from cells treated with DMEM (either commercial or custom-made) CA II concentration was ~100 µM, whereas with 1x, 2x and 5x doses of precursor CA II was ~35–70 µM, 70–100 µM and 45–85 µM, respectively (Fig. S1).

In *E. coli*, α-kesoisovalerate **4**, acts as a precursor for both valine and leucine (Lichtenecker et al. 2013b). However, the pathway for the synthesis of leucine from α-ketoisovalerate is known to be inactive in mammalian cells (Neinast et al. 2019). Nevertheless, protein expression was tested in media lacking only valine, only leucine, or both. As expected, expression was only achieved in the medium lacking valine, confirming that α-ketoisovalerate only works as a valine precursor in mammalian cells (Fig. 1B).

Incorporation of isotopically labelled precursors

For a subset of the above precursors (Table 2), protein side chain-specific isotopic labelling was evaluated through NMR spectroscopy in both intact cells and cell lysates. To this aim, two model globular proteins, CA II and DJ-1, were transiently overexpressed in presence of isotope-labelled precursors at the optimal doses reported above. To correctly

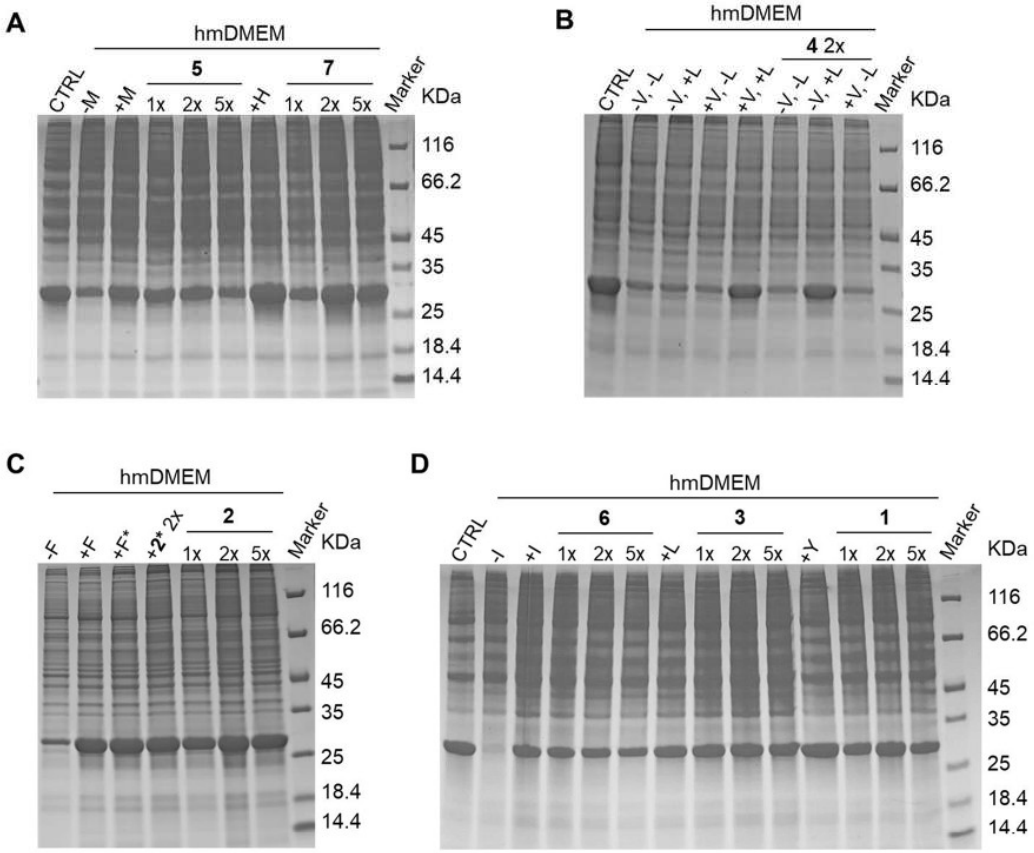


Fig. 1 Protein expression tests. (**A**) HEK293T overexpressing CAII, incubated with commercial DMEM, custom-made DMEM w/o Met, custom-made complete DMEM, custom-made DMEM w/ Methionine precursor **5** 1x, custom-made DMEM w/ **5** 2x, custom-made DMEM w/ **5** 5x, custom-made complete DMEM, custom-made DMEM w/ Histidine precursor **7** 1x, custom-made DMEM w/ **7** 2x, custom-made DMEM w/ **7** 5x, protein marker. (**B**) HEK293T overexpressing CAII, incubated with commercial DMEM, custom-made DMEM w/o Val and Leu, custom-made DMEM w/o Val, custom-made DMEM w/o Leu, custom-made complete DMEM, custom-made DMEM w/ Valine precursor **4** 2x w/o Val and Leu, custom-made DMEM w/ **4** 2x w/o Val, custom-made DMEM w/ **4** 2x w/o Leu, protein marker. (**C**) HEK293T overexpressing CAII, incubated with custom-made DMEM w/o Phe, custom-made complete DMEM, custom-made DMEM w/ Phenylalanine precursor **2**, custom-made DMEM w/ labelled Phenylalanine precursor **2***, custom-made DMEM w/ **2** 1x, custom-made DMEM w/ **2** 2x, custom-made DMEM w/ **2** 5x, protein marker. (**D**) HEK293T overexpressing CAII, incubated with commercial DMEM, custom-made DMEM w/o Ile, custom-made complete DMEM, custom-made DMEM w/ Isoleucine precursor **6** 1x, custom-made DMEM w/ **6** 2x, custom-made DMEM w/ **6** 5x, custom-made complete DMEM, custom-made DMEM w/ Tyrosine precursor **1** 1x, custom-made DMEM w/ **1** 2x, custom-made DMEM w/ **1** 5x, protein marker

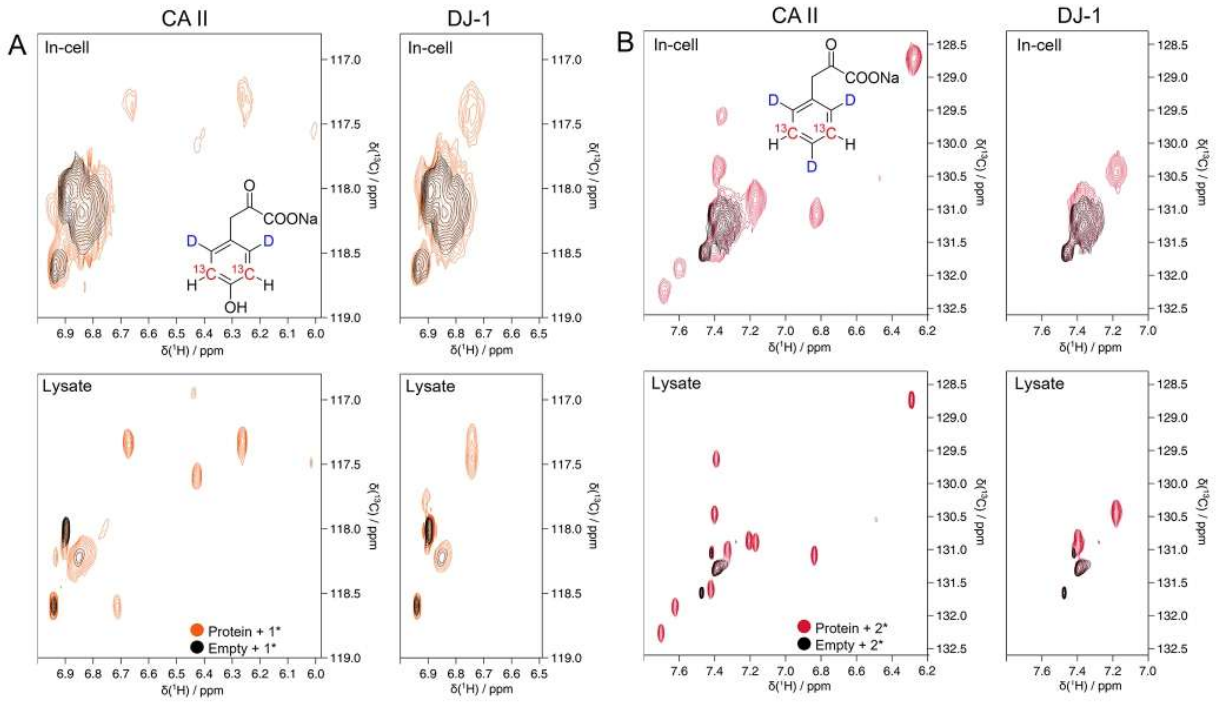


Fig. 2 NMR spectra of tyrosine and phenylalanine-labelled proteins. **(A)** Top: In-cell spectra on HEK293T cells overexpressing CA II (left) and DJ-1 (right), incubated with **1*** (orange); HEK293T cells transfected with empty vector, incubated with **1*** (black). Bottom: In-solution spectra on HEK293T lysate. Lysate of cells overexpressing CA II (~20 μM, left) and DJ-1 (right), incubated with **1*** (orange); lysate of cells transfected with empty vector, incubated with **1*** (black). **(B)** Top: In-cell NMR spectra on HEK293T cells overexpressing CA II (left) and DJ-1 (right), incubated with **2*** (red); HEK293T cells transfected with empty vector, incubated with **2*** (black). Bottom: NMR spectra on HEK293T lysate. Lysate of cells overexpressing CA II (~40 μM, left) and DJ-1 (right), incubated with **2*** (red); lysate of cells transfected with empty vector, incubated with **2*** (black)

Table 3 Observed vs. expected peaks detected in the NMR spectra of cell lysates. [a] Excluding N-terminal methionine

Precursor	Corresponding amino acid	Observed/expected peaks	
		CA II	DJ-1
1*	Tyr	5/8	3/3
2*	Phe	10/12	2/3
3*	Leu	50/52	35/36
4*	Val	34/34	36/38
5*	Met	1/1[a]	4/4[a]

identify the signals arising from the isotope-labelled overexpressed proteins, the spectra were compared with those obtained from control cells transfected with an empty vector and incubated in the same labelled media.

Aromatic labelling

We tested the incorporation of tyrosine and phenylalanine precursors **1*** and **2*** via in-cell and in-lysate NMR (Fig. 2) and compared the spectra of proteins obtained from a precursor-containing medium to those obtained from a medium containing the corresponding side-chain labelled amino acids, which have been synthetized according to literature (Young et al. 2021) (Supp. Fig. S2). The ensuing in-cell NMR spectra revealed well-defined peaks for the side chain-labelled amino acids. The same peaks were better resolved in the lysate NMR spectra. Peak counting revealed a lower number of peaks than expected (Table 3). Notably, however, the same number of signals was also observed in the spectra obtained using the cognate amino acids (Supp. Fig. S2), indicating that the missing peaks do not arise from incomplete incorporation, but are likely caused by signal overlap and/or exchange broadening. Taken together, these results confirm that the α-ketoacids are transformed in the corresponding amino acids and subsequently integrated in the protein sequence at the correct positions.

Methyl labelling

We performed in-cell and in-lysate NMR experiments on protein samples methyl-^{13}C labelled at valine, leucine, and methionine residues using compounds **3***, **4*** and **5***. The resulting in-cell NMR spectra showed well-resolved peaks corresponding to the side chain-labelled amino acids, which

were even better resolved in the lysate NMR spectra, confirming the incorporation in the protein sequence (Fig. 3). In the case of valine and leucine precursors, where the methyl groups are not stereospecifically labelled (resulting in a racemic mixture), the number of expected peaks corresponds to twice as many residues in the protein sequence, due to the labelling of both prochiral methyl groups (Fig. 3A, B). Despite the relatively low protein concentrations, ranging between 25 and 50 µM (Supp. Fig. S3), most of the expected methyl peaks were observed in the lysate spectra (Table 3), thanks to the high sensitivity and spectral resolution attained by selective methyl labelling. To assess the contribution of natural abundance ^{13}C in the spectra of selectively labelled samples, a lysate from cells expressing CA II in unlabelled medium was compared to a lysate containing leucine-labelled CA II. No signals from unlabelled protein were detected within the duration of the experiment, indicating that ^{13}C natural abundance does not interfere with the detection of methyl-labelled proteins (Supp. Fig. S4).

Combined precursor incorporation

We then combined multiple precursors in the same culture medium to assess whether the expression and incorporation levels remained unchanged. We chose valine and leucine precursors, which share the same transaminase enzyme, the branched-chain amino acid aminotransferase (BCAT) (Neinast et al. 2019; Toyokawa et al. 2021). This test would thus provide further information on enzymatic overload. Therefore, a custom culture medium containing Val and Leu precursor was made and the NMR experiments was performed overexpressing DJ-1. The resulting in-cell and lysate spectra show a good DJ-1 expression level, similar to the one of DJ-1 with the precursors used separately, and an efficient incorporation of both labelled amino acids (Fig. 4).

Application of the method

To show an example of a possible application of this labelling technique, we chose a set of well-characterized sulfonamide-derived CA inhibitors: Acetazolamide (AAZ) and methazolamide (MZA) which are currently in use as drugs to treat glaucoma, and ethoxzolamide (ETZ) which is an inhibitor of CAs in proximal renal tubules, widely used as a diuretic. All three compounds were previously shown to tightly bind CA II by in vitro and in-cell NMR (Luchinat et al. 2020a, b). We selected the valine precursor **4*** as Val residues are present in the CA II ligand binding site. Indeed, from the spectra comparison it is possible to appreciate the change in the shifts of those valine residues involved in the interaction or experiencing an environment variation due to the proximity of the ligand (Fig. 5).

Discussion and conclusions

Exclusive isotope labelling of backbone positions is often not sufficient to unfold the full potential of NMR spectroscopy and perform in-depth protein characterization. Special isotope patterns in the side chains are required to investigate interaction surfaces via chemical shift perturbation or study side-chain dynamics in spin relaxation experiments. However, selective isotope labelling is often compromised by high costs for the heavy isotope-containing amino acids needed. Metabolic precursors such as α-ketoacids can be efficiently synthesized from low-cost heavy isotope-containing starting compounds; the corresponding protocols have been summarized in various reviewing articles (Schörghuber et al. 2018; Kerfah et al. 2015; Tugarinov et al. 2006; Rowlinson et al. 2022; Lichtenecker et al. 2015). Many of these precursor isotopologues are even commercially available today. Our results show that using precursors to introduce side chain-labelled amino acids is not limited to *E. coli* expression but can be transferred to mammalian cell hosts. This is especially useful when bacterial or yeast hosts are not viable options to produce a mammalian protein, which is usually the case for large proteins or proteins that undergo post-translational modifications. We furthermore demonstrated that the efficacy of the method is guaranteed even in the case of precursor combinations, granting an amino acid-selective labelling on more than one amino acid, thus expanding the possible applications of this labelling technique. The successful expression of proteins in the presence of the precursors employed in this work is allowed by the presence of transaminases in cultured mammalian cells: the reversible transamination of Ile, Leu, and Val is catalyzed by branched chain amino acid amino transferase (BCAT), Tyr and Phe are substrates of tyrosine aminotransferase (TAT) and glutamic-oxaloacetic transaminase (GOT), the transamination of Met is catalyzed by glutamine transaminase K (GTK), and His is a substrate of both glutamine transaminase L (GTL) and serine-pyruvate transaminase (SPT) (Caligiore et al. 2022; Neinast et al. 2019; Toyokawa et al. 2021; Sivaraman and Kirsch 2006; Mehere et al. 2010; Cooper 2004). Concerning the general applicability of the method, our results suggest that α-ketoacid precursors of other amino acids could be employed, provided that they are recognized by transaminases expressed in the cell culture. GTK and TAT are reported to catalyze the transamination of Trp with high efficiency, whereas a similar activity for Lys and Arg is not reported (Caligiore et al. 2022).

Furthermore, side chain isotope labelling may not be straightforward for some amino acids (Ala, Asp, Asn, Pro, Glu) that are directly connected to core cellular metabolic pathways, such as glycolysis and the TCA cycle. In such cases, adding the corresponding α-ketoacid precursors to

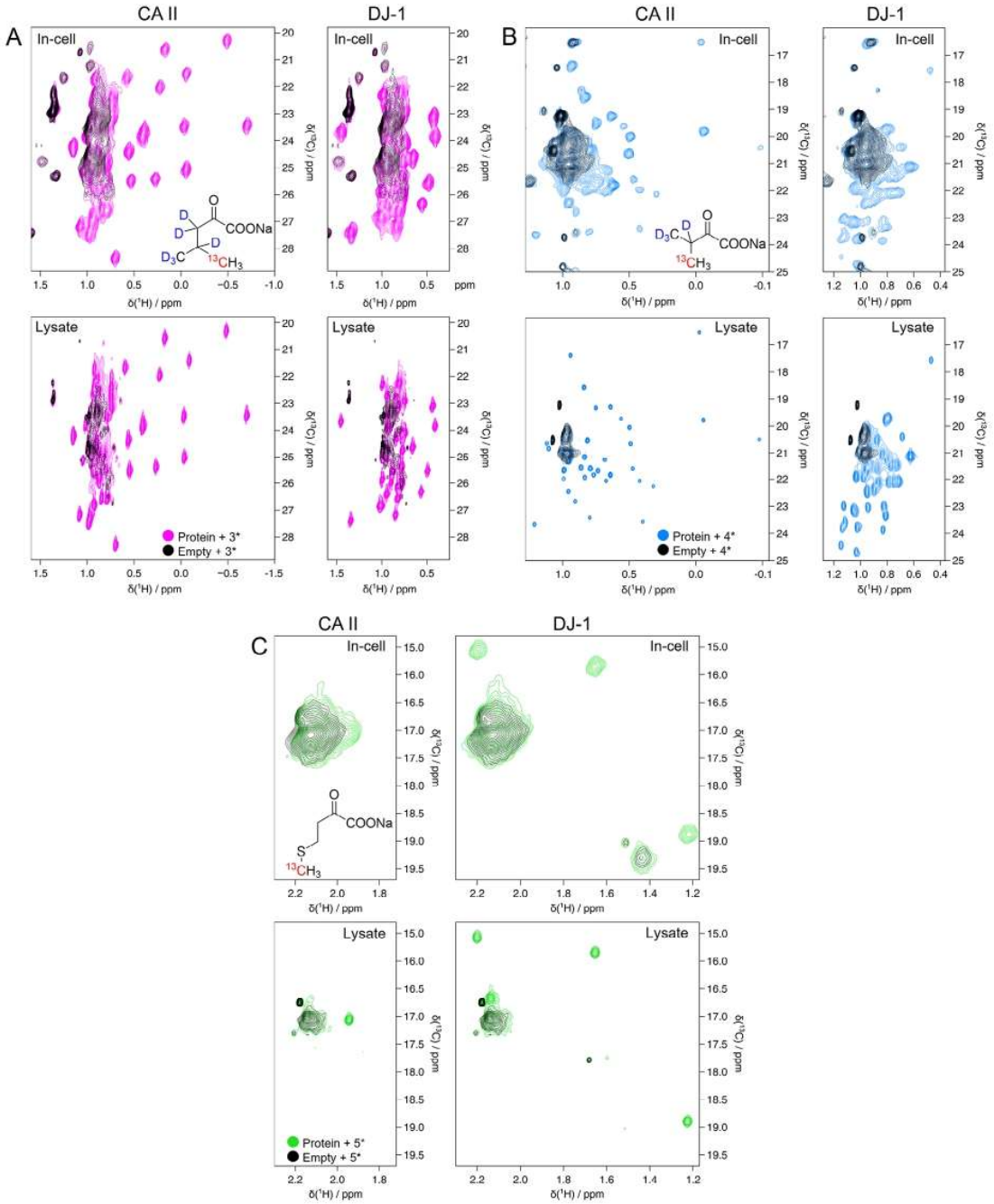


Fig. 3 NMR spectra of valine, leucine and methionine-labelled proteins. (**A**) Top: In-cell NMR spectra on HEK293T cells overexpressing CA II (left) and DJ-1 (right), incubated with **3*** (magenta); HEK293T cells overexpressing empty vector, incubated with **3*** (black). Bottom: NMR spectra on HEK293T lysate. Lysate of cells overexpressing CA II (~50 µM, left) and DJ-1 (right), incubated with **3*** (magenta); lysate of cells transfected with empty vector, incubated with **3*** (black). (**B**) Top: In-cell NMR spectra on HEK293T cells overexpressing CA II (left) and DJ-1 (right), incubated with **4*** (light blue); HEK293T cells transfected with empty vector, incubated with **4*** (black). Bottom: NMR spectra on HEK293T lysate. Lysate of cells overexpressing CA II (~25 µM, left) and DJ-1 (right), incubated with **4*** (light blue); lysate of cells transfected with empty vector, incubated with **4*** (black). (**C**) Top: In-cell NMR spectra on HEK293T cells overexpressing CA II (left) and DJ-1 (right), incubated with **5*** (green); HEK293T cells transfected with empty vector, incubated with **5*** (black). Bottom: NMR spectra on HEK293T lysate. Lysate of cells overexpressing CA II (~25 µM, left) and DJ-1 (right), incubated with **5*** (green); lysate of cells transfected with empty vector, incubated with **5*** (black)

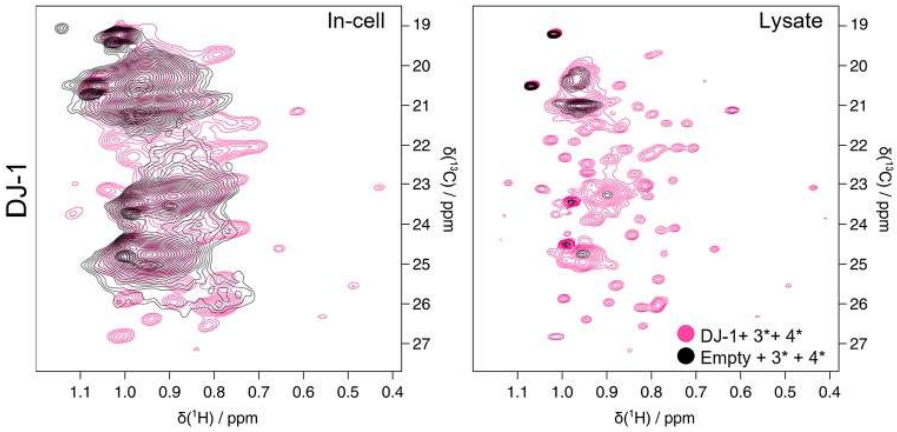


Fig. 4 Simultaneous labelling with valine and leucine. Left: In-cell spectra on HEK293T overexpressing DJ-1, incubated with **3*** and **4*** (magenta); HEK293T cells overexpressing empty vector, incubated with **3*** and **4*** (black). Right: In-solution spectra on HEK293T lysate overexpressing DJ-1, incubated with **3*** and **4*** (magenta); lysate of cells overexpressing empty vector, incubated with **3*** and **4*** (black)

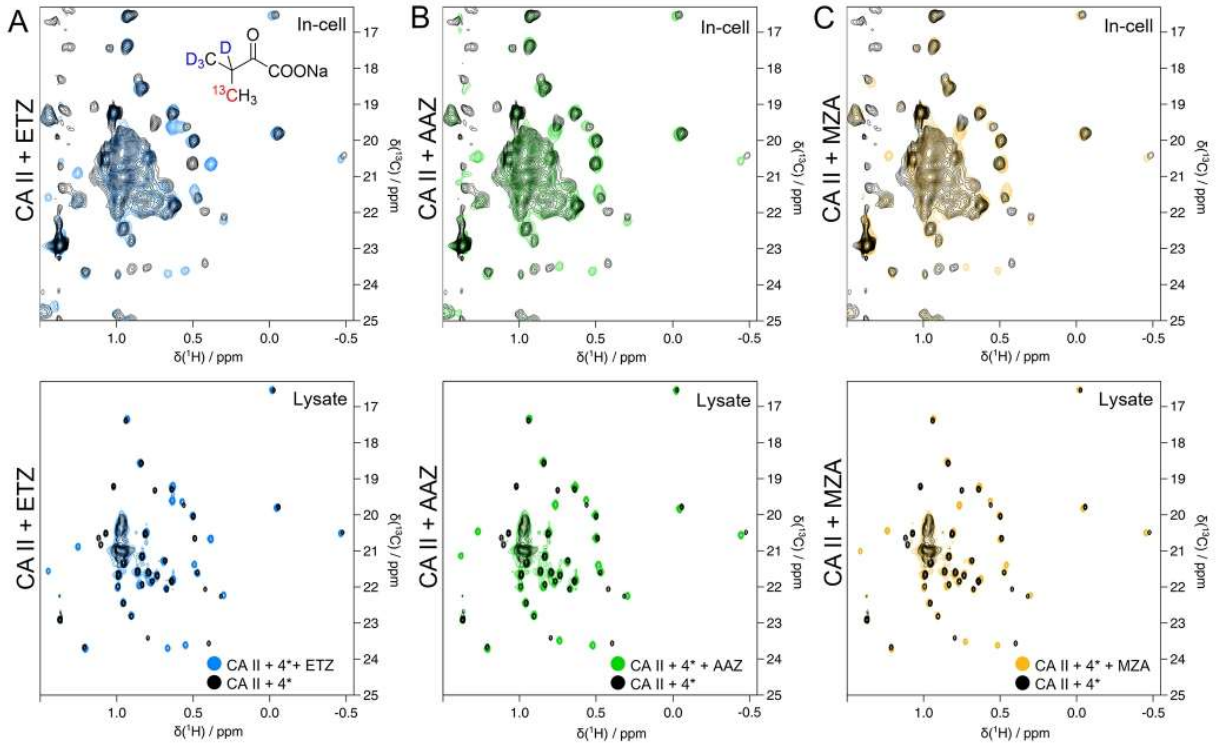


Fig. 5 Ligand binding to CA II. (**A**) Top: In-cell NMR spectra on HEK 293T overexpressing CA II, incubated with **4*** and treated with ETZ (light blue); In-cell NMR spectra on HEK 293T overexpressing CA II, incubated with **4*** (black) Bottom: NMR spectra on HEK293T lysate overexpressing CA II, incubated with **4*** and treated with ETZ (light blue); In-solution spectra on HEK293T lysate overexpressing CA II, incubated with **4*** (black). (**B**) Top: In-cell NMR spectra on HEK 293T overexpressing CA II, incubated with **4*** and treated with AAZ (green); In-cell NMR spectra on HEK 293T overexpressing CA II, incubated with **4*** (black). Bottom: NMR spectra on HEK293T lysate overexpressing CA II, incubated with **4*** and treated with AAZ (green); In-solution spectra on HEK293T lysate overexpressing CA II, incubated with **4*** (black). (**C**) Top: In-cell NMR spectra on HEK 293T overexpressing CA II, incubated with **4*** and treated with MZA (yellow); In-cell NMR spectra on HEK 293T overexpressing CA II, incubated with **4*** (black). Bottom: NMR spectra on HEK293T lysate overexpressing CA II, incubated with **4*** and treated with MZA (yellow); In-solution spectra on HEK293T lysate overexpressing CA II, incubated with **4*** (black)

the culture medium will likely result in the unwanted labelling of other amino acids in the expressed proteins. Notably, however, the above amino acids are not predominant in protein-ligand and protein-protein interfaces, and therefore their labeling would not be the first choice for studying such phenomena.

As a first example, we demonstrated how this labelling technique results in well-resolved methyl resonances in the NMR spectra of two globular proteins, the 29 kDa CA II and the 40 kDa DJ-1 homodimer. Addition of selected known ligands led to substantial perturbation of the CA II chemical shifts, which could be recorded in-lysate, as well as in-cell. This observation implies that combinations of precursors and mammalian expression systems are generally applicable to investigate protein-ligand- and, in principle, protein-protein interactions. Overall, we believe that introducing α-ketoacid precursors to mammalian cell-based protein expression will greatly facilitate the application of NMR to study structure and function of challenging human proteins in a purified state, in cell lysates, or even in intact cells.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10858-024-00447-6.

Author contributions R.K., E.L. and Lu.B. conceived the work; M.R., Le.B. R.J.L., R.K., E.L. and Lu.B. designed the experiments; M.H., S.K. and R.J.L. synthesized the precursor compounds; M.R. and Le.B. prepared the NMR samples; M.R. and E.L. collected, processed and analyzed the NMR data; M.R. and E.L. drafted the manuscript; M.R., Le.B., R.J.L., R.K., E.L. and Lu.B. contributed to the manuscript; Lu.B. secured funding.

Funding Open access funding provided by Università degli Studi di Firenze within the CRUI-CARE Agreement. This work was supported by Instruct-ERIC, a Landmark ESFRI project, and specifically by the CERM/CIRMMP Italian Instruct Centre, and by Fragment-Screen, grant agreement No. 101094131, funded by the Horizon Europe program of the European Commission. The authors acknowledge the support by the Italian Ministry for University and Research (MUR) FOE funding to CERM/CIRMMP. E.L. contract is funded by MUR through the Dipartimento di Eccellenza 2023–2027 (DICUS 2.0) funding to the Department of Chemistry "Ugo Schiff" of the University of Florence. Open access funding provided by Università degli Studi di Firenze within the CRUI-CARE Agreement.

Data availability The raw NMR data will be made available upon reasonable request.

Declarations

Competing interests R.J.L. and R.K. are shareholders in the company Mag-Lab, which is mentioned in the affiliations. R.J.L., M.H. and S.K. are employed at Mag-Lab.

References

Aricescu AR, Lu W, Jones EY (2006) A time- and cost-efficient system for high-level protein production in mammalian cells. Acta Crystallogr D Biol Crystallogr 62:1243–1250. https://doi.org/10.1107/S0907444906029799

Banci L, Barbieri L, Bertini I et al (2013) Atomic-resolution monitoring of protein maturation in live human cells by NMR. Nat Chem Biol 9:297–299. https://doi.org/10.1038/nchembio.1202

Barbieri L, Luchinat E, Banci L (2015) Protein interaction patterns in different cellular environments are revealed by in-cell NMR. Sci Rep 5:14456. https://doi.org/10.1038/srep14456

Barbieri L, Luchinat E, Banci L (2016) Characterization of proteins by in-cell NMR spectroscopy in cultured mammalian cells. Nat Protoc 11:1101–1111. https://doi.org/10.1038/nprot.2016.061

Barbieri L, Luchinat E, Banci L (2018) Intracellular metal binding and redox behavior of human DJ-1. J Biol Inorg Chem JBIC Publ Soc Biol Inorg Chem 23:61–69. https://doi.org/10.1007/s00775-017-1509-5

Burz DS, Breindel L, Shekhtman A (2019) Improved sensitivity and resolution of in-cell NMR spectra. Methods Enzymol 621:305–328. https://doi.org/10.1016/bs.mie.2019.02.029

Caligiore F, Zangelmi E, Vetro C et al (2022) Human cytosolic transaminases: side activities and patterns of discrimination towards physiologically available alternative substrates. Cell Mol Life Sci 79:421. https://doi.org/10.1007/s00018-022-04439-3

Cooper AJL (2004) The role of glutamine transaminase K (GTK) in sulfur and α-keto acid metabolism in the brain, and in the possible bioactivation of neurotoxicants. Neurochem Int 44:557–577. https://doi.org/10.1016/j.neuint.2003.12.002

Fischer M, Kloiber K, Häusler J et al (2007) Synthesis of a 13 C-Methyl-group-labeled methionine precursor as a useful Tool for simplifying protein structural analysis by NMR spectroscopy. ChemBioChem 8:610–612. https://doi.org/10.1002/cbic.200600551

Goto NK, Gardner KH, Mueller GA et al (1999) A robust and cost-effective method for the production of Val, Leu, Ile (delta 1) methyl-protonated 15 N-, 13 C-, 2H-labeled proteins. J Biomol NMR 13:369–374. https://doi.org/10.1023/a:1008393201236

Kerfah R, Plevin MJ, Sounier R et al (2015) Methyl-specific isotopic labeling: a molecular tool box for solution NMR studies of large proteins. Curr Opin Struct Biol 32:113–122. https://doi.org/10.1016/j.sbi.2015.03.009

Lichtenecker RJ (2014) Synthesis of aromatic 13 C/2H-α-ketoacid precursors to be used in selective phenylalanine and tyrosine protein labelling. Org Biomol Chem 12:7551–7560. https://doi.org/10.1039/C4OB01129E

Lichtenecker R, Ludwiczek ML, Schmid W, Konrat R (2004) Simplification of protein NOESY Spectra using Bioorganic Precursor synthesis and NMR spectral editing. J Am Chem Soc 126:5348–5349. https://doi.org/10.1021/ja049679n

Lichtenecker RJ, Coudevylle N, Konrat R, Schmid W (2013a) Selective isotope labelling of leucine residues by using α-Ketoacid

precursor compounds. ChemBioChem 14:818–821. https://doi.org/10.1002/cbic.201200737
Lichtenecker RJ, Weinhäupl K, Reuther L et al (2013b) Independent valine and leucine isotope labeling in Escherichia coli protein overexpression systems. J Biomol NMR 57:205–209. https://doi.org/10.1007/s10858-013-9786-y
Lichtenecker RJ, Weinhäupl K, Schmid W, Konrat R (2013c) α-Ketoacids as precursors for phenylalanine and tyrosine labelling in cell-based protein overexpression. J Biomol NMR 57:327–331. https://doi.org/10.1007/s10858-013-9796-9
Lichtenecker RJ, Schörghuber J, Bisaccia M (2015) Synthesis of metabolic amino acid precursors: tools for selective isotope labeling in cell-based protein overexpression. Synlett 26:2611–2616. https://doi.org/10.1055/s-0034-1378878
Luchinat E, Barbieri L, Cremonini M et al (2020a) Drug screening in human cells by NMR spectroscopy allows the Early Assessment of Drug Potency. Angew Chem Int Ed Engl 59:6535–6539. https://doi.org/10.1002/anie.201913436
Luchinat E, Barbieri L, Cremonini M et al (2020b) Intracellular Binding/Unbinding Kinetics of approved drugs to Carbonic anhydrase II observed by in-cell NMR. ACS Chem Biol 15:2792–2800. https://doi.org/10.1021/acschembio.0c00590
Mehere P, Han Q, Lemkul JA et al (2010) Tyrosine aminotransferase: biochemical and structural properties and molecular dynamics simulations. Protein Cell 1:1023–1032. https://doi.org/10.1007/s13238-010-0128-5
Mueller L (2008) Alternate HMQC experiments for recording HN and HC-correlation spectra in proteins at high throughput. J Biomol NMR 42:129–137. https://doi.org/10.1007/s10858-008-9270-2
Neinast M, Murashige D, Arany Z (2019) Branched chain amino acids. Annu Rev Physiol 81:139–164. https://doi.org/10.1146/annurev-physiol-020518-114455
Ninković DB, Blagojević Filipović JP, Hall MB et al (2020) What is special about aromatic–aromatic interactions? Significant attraction at large horizontal displacement. ACS Cent Sci 6:420–425. https://doi.org/10.1021/acscentsci.0c00005
Rößler P, Mathieu D, Gossert AD (2020) Enabling NMR studies of high Molecular Weight systems without the need for Deuteration: the XL-ALSOFAST experiment with delayed decoupling. Angew Chem Int Ed 59:19329–19337. https://doi.org/10.1002/anie.202007715
Rowlinson B, Crublet E, Kerfah R, Plevin MJ (2022) Specific isotopic labelling and reverse labelling for protein NMR spectroscopy: using metabolic precursors in sample preparation. Biochem Soc Trans 50:1555–1567. https://doi.org/10.1042/BST20210586
Sathyamoorthy B, Lee J, Kimsey I et al (2014) Development and application of aromatic [(13)C, (1)H] SOFAST-HMQC NMR experiment for nucleic acids. J Biomol NMR 60:77–83. https://doi.org/10.1007/s10858-014-9856-9
Schanda P, Brutscher B (2005) Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds. J Am Chem Soc 127:8014–8015. https://doi.org/10.1021/ja051306e
Schörghuber J, Geist L, Platzer G et al (2018) Late metabolic precursors for selective aromatic residue labeling. J Biomol NMR 71:129–140. https://doi.org/10.1007/s10858-018-0188-z
Serrano L, Bycroft M, Fersht AR (1991) Aromatic-aromatic interactions and protein stability. J Mol Biol 218:465–475. https://doi.org/10.1016/0022-2836(91)90725-L
Sivaraman S, Kirsch JF (2006) The narrow substrate specificity of human tyrosine aminotransferase – the enzyme deficient in tyrosinemia type II. FEBS J 273:1920–1929. https://doi.org/10.1111/j.1742-4658.2006.05202.x
Theillet F-X, Luchinat E (2022) In-cell NMR: why and how? Prog Nucl Magn Reson Spectrosc. 132–133:1–112. https://doi.org/10.1016/j.pnmrs.2022.04.002
Toyokawa Y, Koonthongkaew J, Takagi H (2021) An overview of branched-chain amino acid aminotransferases: functional differences between mitochondrial and cytosolic isozymes in yeast and human. Appl Microbiol Biotechnol 105:8059–8072. https://doi.org/10.1007/s00253-021-11612-4
Tugarinov V, Kanelis V, Kay LE (2006) Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy. Nat Protoc 1:749–754. https://doi.org/10.1038/nprot.2006.101
Werner K, Richter C, Klein-Seetharaman J, Schwalbe H (2008) Isotope labeling of mammalian GPCRs in HEK293 cells and characterization of the C-terminus of bovine rhodopsin by high resolution liquid NMR spectroscopy. J Biomol NMR 40:49–53. https://doi.org/10.1007/s10858-007-9205-3
Young BM, Rossi P, Slavish PJ et al (2021) Synthesis of isotopically labeled, spin-isolated tyrosine and phenylalanine for protein NMR applications. Org Lett 23:6288–6292. https://doi.org/10.1021/acs.orglett.1c02084

Supporting Information

Towards cost-effective side-chain isotope labelling of proteins expressed in human cells

Martina Rosati[1], Letizia Barbieri[1,2], Matus Hlavac[3], Sarah Kratzwald[3], Roman J. Lichtenecker[3,4], Robert Konrat[3,5], Enrico Luchinat[1,2,6,*], Lucia Banci[1,2,6,*]

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

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

[3] MAG-LAB GmbH, Vienna, Austria.

[4] Institute of Organic Chemistry, University of Vienna, Vienna, Austria.

[5] Department of Structural and Computational Biology, Max Perutz Laboratories, University of Vienna, Vienna, Austria.

[6] Dipartimento di Chimica, Università degli Studi di Firenze, Sesto Fiorentino, Italy.

* Enrico Luchinat: eluchinat@cerm.unifi.it; Lucia Banci: banci@cerm.unifi.it

Supplementary Figures S1-S4

Supplementary Figures

	CA II concentration (μM)		
precursor	**1x**	**2x**	**3x**
pTyr	56	83	45
pPhe	47	86	59
pHis	34	99	85
pLeu	55	74	65
pIle	57	72	67
pMet	68	78	53
pVal		91	

cDMEM	106, 96, 98
hDMEM	89, 103

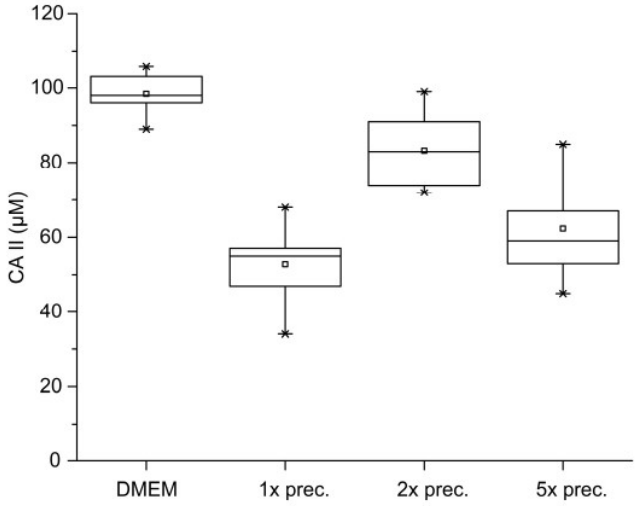


Fig. S1 Left: CA II concentration in the lysates obtained from cell samples treated with different doses of each precursor. Right: box plot showing the range of CA II concentrations obtained in DMEM (either commercial or custom made) and in custom medium containing 1x, 2x and 5x doses of precursor.

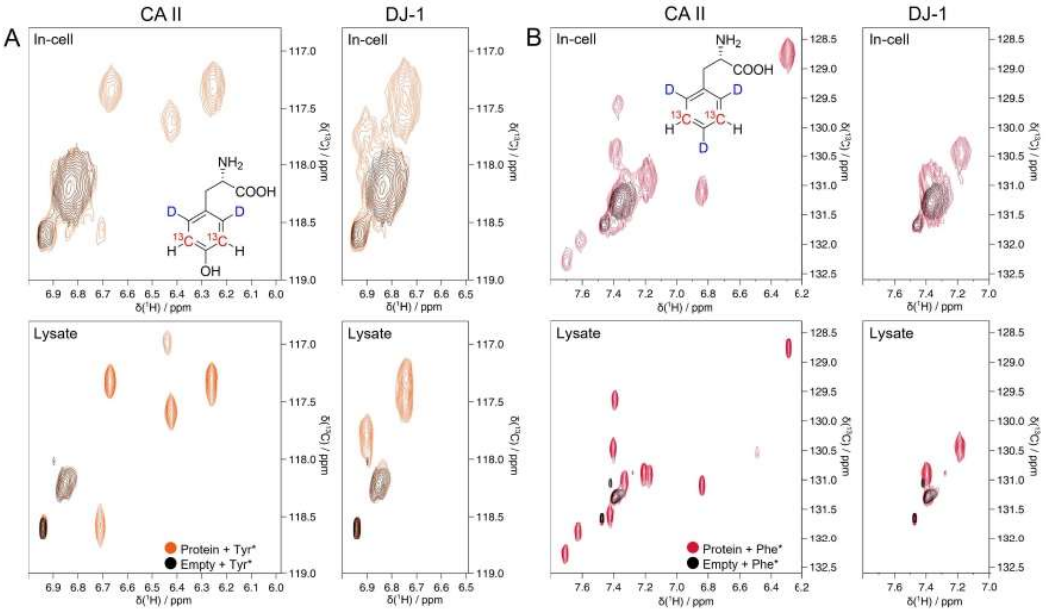


Fig. S2 Phe* and Tyr* NMR spectra. A) Top: In-cell NMR spectra on HEK293T cells overexpressing CA II (left) and DJ-1 (right), incubated with Tyr* (orange); HEK293T cells transfected with empty vector, incubated with Tyr* (black). Bottom: NMR spectra on HEK293T lysate. Lysate of cells overexpressing CA II (~60 µM, left) and DJ-1 (right), incubated with Tyr* (orange); lysate of cells transfected with empty vector, incubated with Tyr* (black). B) Top: In-cell NMR spectra on HEK293T cells overexpressing CA II (left) and DJ-1 (right), incubated with Phe* (red); HEK293T cells transfected with empty vector, incubated with Phe* (black). Bottom: NMR spectra on HEK293T lysate. Lysate of cells overexpressing CA II (~60 µM, left) and DJ-1 (right), incubated with Phe* (red); lysate of cells transfected with empty vector, incubated with Phe* (black).

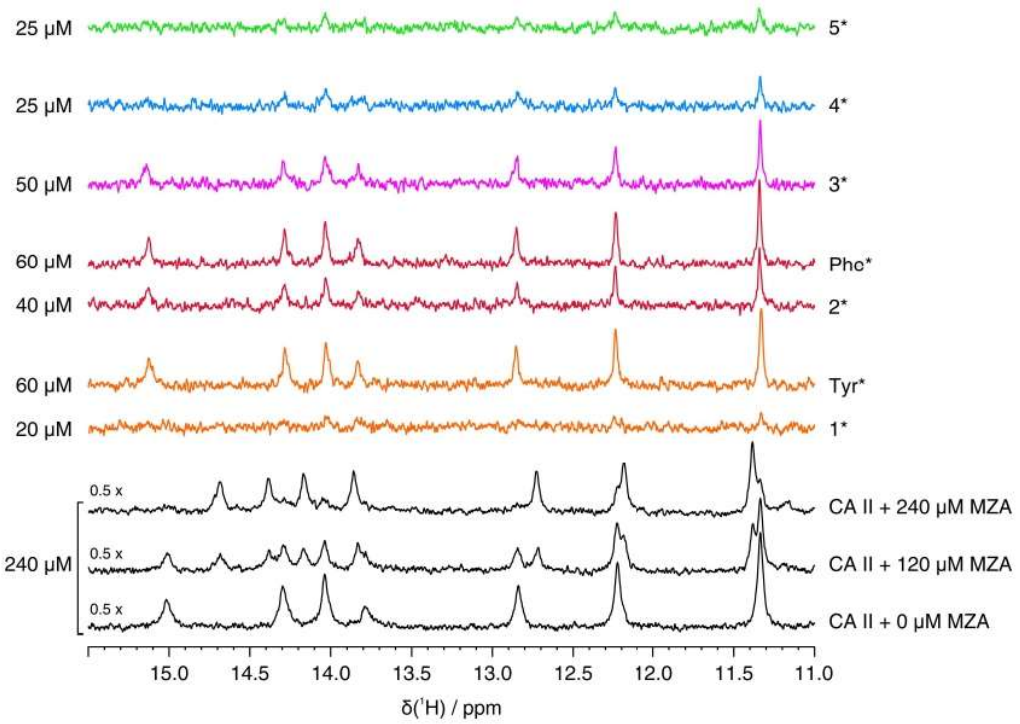


Fig. S3 CA II quantification in the cell lysates by ^{1}H NMR signal integration. From bottom to top: reference cell lysate containing 240 µM CA II (black) titrated with methazolamide (MZA); lysates from cells incubated with **1*** and Tyr* (orange); lysates from cells incubated with **2*** and Phe* (red); lysates from cells incubated with **3*** (magenta), **4*** (blue) and **5*** (green). Estimated CA II concentrations are reported for each sample.

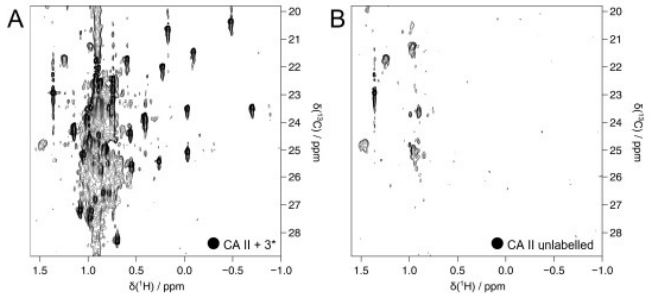


Fig. S4 A) NMR spectrum of a lysate from cells overexpressing leucine-labelled CA II; B) NMR spectrum of a lysate from cells overexpressing unlabelled CA II, recorded with the same duration of (A).

4. The synthesis of specifically isotope labelled fluorotryptophan and its use in mammalian cell-based protein expression for ^{19}F-NMR applications

4.1 Summary

In this work, we exploited a cost-effective protocol for the synthesis of deuterated 5-fluorotryptophan selectively labelled with ^{13}C at the C5 position, developed by Roman J. Lichtenecker's group at Mag-Lab, in collaboration with the University of Vienna, for the the incorporation of this fluorinated amino acid into two proteins: carbonic anhydrase isoform II (CA II) and superoxide dismutase 1 (SOD1).

Protein expression was carried out in HEK293T cells transiently transfected with a PEI:plasmid DNA mixture at a 2:1 ratio in standard commercial DMEM. After 24 hours of incubation at 37°C and 5% CO_2, the culture medium was replaced with a custom-made DMEM in which L-tryptophan was substituted with an equimolar amount of 5-^{13}C, ^{19}F-Trp. The cells were further incubated for 24 hours before harvesting and lysing by freeze-thaw cycles in liquid nitrogen.

The successful incorporation of the fluorinated tryptophan was assessed by ^{13}C, ^{19}F TROSY NMR experiments performed on the cell lysates. The spectra clearly demonstrated the advantage of exploiting the direct spin-spin coupling between the isotopically labelled fluorine and carbon nuclei. High resolution and excellent peak dispersion were observed, particularly in the CA II overexpressing lysate, where four of the seven tryptophan residues, which strongly overlap in conventional 1D ^{19}F spectra, appear well resolved in the 2D ^{13}C, ^{19}F correlation spectra.

ChemComm

COMMUNICATION

View Article Online
View Journal | View Issue

Cite this: *Chem. Commun.*, 2024, **60**, 14188

Received 16th September 2024,
Accepted 1st November 2024

DOI: 10.1039/d4cc04789c

rsc.li/chemcomm

The synthesis of specifically isotope labelled fluorotryptophan and its use in mammalian cell-based protein expression for ^{19}F-NMR applications†

Giorgia Toscano, ‡[abc] Martina Rosati, ‡[d] Letizia Barbieri, [de] Katharina Maier,[a] Lucia Banci, [def] Enrico Luchinat, [def] Robert Konrat [bgh] and Roman J. Lichtenecker *[abh]

^{19}F nuclei serve as versatile sensors for detecting protein interactions and dynamics in biomolecular NMR spectroscopy. Although various methods have been developed to incorporate fluorine-containing aromatic residues into proteins using *E. coli* or cell-free expression techniques, similar approaches for protein production in mammalian cell lines remain limited. Here, we present a cost-effective synthetic route to obtain selectively deuterated, carbon-13 labeled fluorotryptophan and demonstrate its use in introducing ^{19}F–^{13}C spin pairs into carbonic anhydrase 2 and superoxide dismutase, following an expression protocol utilizing HEK cells.

NMR spectroscopy is a powerful tool for studying the conformational properties and interactions of biomolecules under conditions that closely mimic the natural cellular environment.[1] The challenges associated with NMR spectroscopy, particularly in the case of large molecular weight samples, such as limited resolution and sensitivity, have been mitigated by advancements in NMR hardware (higher magnetic fields, cryogenic probes),[2] improvements in experimental techniques (multidimensional NMR sequences, transverse relaxation optimized spectroscopy TROSY),[3] and various selective isotope labeling strategies. In the context of these strategies, NMR-active nuclei such as carbon-13 or nitrogen-15 are only introduced in certain residue types, while protons are selectively replaced by deuteration.[4] Proteins with such defined isotope patterns exhibit clear and unambiguous signals in simplified NMR spectra, which allow to follow more easily the changes in the chemical environment during interactions with binding partners. The resulting chemical shift changes can be analyzed to probe binding pockets and explore large interaction surfaces.[5] NMR studies of largely unstructured, very large or even multi-component biomolecular complexes may require the introduction of bioorthogonal NMR active nuclei, such as fluorine. ^{19}F offers excellent NMR properties, like a high sensitivity (similar to ^{1}H), and a wide chemical shift range (~400 ppm compared to ~15 ppm for ^{1}H). Since fluorine is largely absent in biological systems, ^{19}F-NMR serves as a powerful tool to analyze proteins in complex biological matrices, such as in-cell- or in-lysate NMR spectroscopy.[6] However, introducing non-proteinogenic fluorine-containing amino acids might significantly alter a protein's conformation and function. Changing hydrogen for fluorine does not result in a drastic steric change, but being the most electronegative element in the periodic table fluorine polarizes bonds, changes pK_a values and affects the sensitive networks of hydrogen bond interaction.[7]

Aryl-fluorinated aromatic residues have been shown to have minimal impact on the function of the target proteins in various applications.[7] Additionally, fluorophenylalanines, fluorotyrosines, and fluorotryptophans can be easily incorporated by substituting the natural amino acids with their fluorinated counterparts in *E. coli*-based expression media.[8] This can be achieved by either using auxotrophic host organisms or adding small molecules as inhibitors of the relevant biosynthetic pathways. A notable example of enhancing the uptake efficiency of non-natural amino acids through inhibition is the addition of *N*-(phosphonomethyl)glycine (glyphosate®) to block the shikimate pathway.[9] In the case of fluorotryptophans

[a] *Institute of Organic Chemistry, University of Vienna, Währinger Str. 38, 1090-Vienna, Austria. E-mail: roman.lichtenecker@univie.ac.at*
[b] *Christian Doppler Laboratory for High-Content Structural Biology and Biotechnology, Max Perutz Laboratories, University of Vienna, Campus Vienna Biocenter 5, 1030 Vienna, Austria*
[c] *Vienna Doctoral School in Chemistry (DoSChem), University of Vienna, Währingerstraße 42, 1090 Vienna, Austria*
[d] *CERM Magnetic Resonance Center, Università degli Studi di Firenze, Sesto Fiorentino, Italy*
[e] *Consorzio Interuniversitario Risonanze Magnetiche di Metallo Proteine CIRMMP, Sesto Fiorentino, Italy*
[f] *Dipartimento di Chimica, Università degli Studi di Firenze, Sesto Fiorentino, Italy*
[g] *Department of Structural and Computational Biology, Max Perutz Laboratories, University of Vienna, Dr-Bohr-Gasse 9, 1030-Vienna, Austria*
[h] *Mag-Lab, Karl-Farkas-Gasse 22, 1030 Vienna, Austria*

† Electronic supplementary information (ESI) available: Experimental details concerning organic synthesis, protein expression and NMR spectroscopy. See DOI: https://doi.org/10.1039/d4cc04789c

‡ These authors contributed equally.

(fTrp), even metabolic precursors like fluoroindoles are accepted by overexpressing bacterial microorganisms. Alternatively, in-vivo incorporation can be achieved by site-selective fluorination using *F*-amino acid/tRNA/aminoacyl tRNA synthetase combinations orthogonal to the translation machinery of the host organism.[10] This extended genetic encoding strategy reduces the number of fluorine atoms within the target protein sequence, helping to preserve its natural properties. This approach was recently implemented for the site-selective incorporation of various fTrp isomers.[10b]

For many large proteins, especially for those modified by post-translational modifications, protein production in prokaryotic organisms is not an option. Consequently, protocols for mammalian cell expression have been developed and implemented.[11] Recently, Costantino *et al.* reported on a robust method to incorporate differently fluorinated aromatic amino acids by substituting the target residue with its corresponding fluorinated derivative in HEK cell cultures, transiently transfected with the corresponding gene of interest.[12] Building on these findings, we aimed to synthesize a fluorinated tryptophan with a selective $^{13}C/^{2}H$ isotope pattern (compound **1** in Fig. 1) to be used in mammalian cell overexpression systems.

5-Fluorotryptophan **1** (hereafter referred to as fTrp) contains a $^{13}C–^{19}F$ spin pair, which enables the additional resolution of overlapping fluorine resonances in the carbon-13 dimension, particularly in large proteins with multiple tryptophan residues. Moreover, Boeszoermenyi *et al.* reported the observation of ^{13}C signals of remarkable narrow linewidths by exploiting the $^{19}F–^{13}C$ TROSY effect in large proteins.[13] Corresponding synthetic procedures to incorporate such spin pairs have been reported for fluorinated tyrosine, -phenylalanine and -tryptophan.[14] The authors of the latter used $^{13}C/^{19}F/^{2}H$ indoles as precursors in *E. coli* overexpression – a compound that is not applicable in mammalian-based systems. Therefore, we wanted to close this methodological gap by introducing the corresponding fluorinated amino acid **1**. FTrp is especially interesting in this regard, because Trp is generally less abundant compared to Phe and Tyr meaning a fluorinated Trp therefore alters the protein's properties to a lesser extent than fluorinated versions of the other aromatic residues. Additionally, Trp is frequently found at interaction sites, where the large, apolar indole side chain plays a crucial part in intermolecular interaction, contributing to π–π, cation–π or CH–π forces.[15] We started our synthetic route with commercially available [2-^{13}C] acetone **2**, which we used to assemble the aryl ring by condensation with sodium nitromalonaldehyde as reported in earlier work (Fig. 2).[16] The resulting [^{13}C] nitrophenol **3** was selectively deuterated ortho to the OH-moiety under acidic conditions at high temperature (→**4**). In the next step, dehydroxyfluorination (**4** → **5**) using phenofluoromix® **11**, as described by the group of Ritter, was performed.[17] This reaction works best with electron deficient aryl rings, which makes nitrophenol an ideal substrate for this conversion. The electron density of the aromatic ring is drastically changed within the scope of the next reaction, which is the reduction of the nitro-group to the amine **6**. This conversion activates the protons *ortho* to the NH_2-group for the exchange with deuterium at elevated temperatures. While reviewing the literature for effective methods to construct the indole system from aniline derivatives and to introduce the asymmetric carbon atom of the target compound in a straightforward and efficient manner, we discovered a method developed by James Cook's group involving alkyne **12**.[18] This compound can be easily synthesized from the classic chiral Val/Gly Schöllkopf auxiliary. To implement this synthetic approach, we successfully prepared the isotope-labeled fluoro-*ortho*-iodoaniline **8** using iodine, achieving a satisfactory yield. With compounds **8** and **12** in hand we could apply the palladium catalyzed heteroannulation reaction to prepare the indole-bislactim ether **9**. This compound was hydrolyzed to the amino acid ester with concomitant cleavage of the silyl group.[18] Final saponification of the ester gave the target compound **1** with high deuteration grades in the aromatic positions (95% for ε-, 96% for ζ- and 93% for η-position, respectively, according to NMR analysis – see spectral characterization in the ESI†). In total, the target compound was prepared in nine consecutive steps with an overall yield of 19%. Notably, the costs for chemicals and reagents are low, at less than €3 per mg of the final compound, with potential for scale-up to gram quantities. The compound was then applied to the expression of two different model proteins using HEK293T cells, namely the 29.3 kDa carbonic anhydrase II (CA2), containing 7 tryptophans and superoxide dismutase (SOD1), with a molecular weight of 32 kDa, containing a single tryptophan. Both proteins were expressed following methods outlined in previous studies, assuming approximately 70% incorporation rates based on these findings; detailed experimental procedures are provided in the ESI.†[12] In short, cells were transiently transfected with

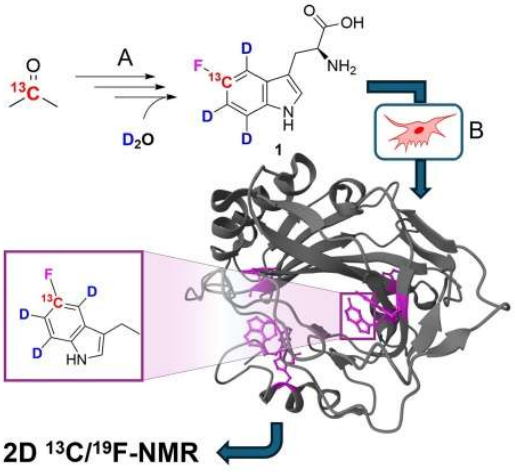


Fig. 1 General strategy of selective $^{19}F/^{13}C/^{2}H$ tryptophan labelling using mammalian cell lines. (A) Synthesis of labelled fTrp **1** by multistep organic synthesis from simple, commercially available isotope sources. (B) Expression of the target protein in mammalian cells by replacing the natural amino acid Trp with compound **1** in the corresponding media. The protein structure shows fTrp-containing carbonic anhydrase 2 (CA2; PDB structure: 8P6U) with fTrp shown in purple.

linewidths is still under investigation but is expected to improve transverse relaxation in the ^{13}C evolution period and remove $^{2}J_{CH}$ couplings. Moreover, compound **1** is not limited to mammalian overexpression systems but can also be used in *E. coli*, yeast, cell-free and even genetic encoding methods. Additional applications, particularly involving in-cell NMR, are currently part of our ongoing research.

This work was supported by the Christian Doppler Laboratory for High-Content Structural Biology and Biotechnology (to G. T.), by Instruct-ERIC and specifically the CERM/CIRMMP Italian Instruct Centre, by Fragment-Screen, No. 101094131, funded by the Horizon Europe program of the EC (to L. B.), and by ITACA. SB "Potentiating the Italian Capacity for Structural Biology Services in Instruct-ERIC" (IR0000009) within the call MUR 3264/2021 PNRR M4/C2/L3.1.1, funded by the European Union–Next Generation EU.

Data availability

The data supporting this article have been included as part of the ESI.†

Conflicts of interest

RJL and RK are shareholders of the company Mag-Lab, which is mentioned in the affiliations. There are no other conflicts to declare.

Notes and references

1 (*a*) I. R. Kleckner and M. P. Foster, *Biochim. Biophys. Acta*, 2011, **1814**(8), 942; (*b*) A. Bax and G. M. Clore, *J. Magn. Reson.*, 2019, **306**, 187; (*c*) J. A. Purslow, B. Khatiwada, M. J. Bayro and V. Venditti, *Front. Mol. Biosci.*, 2020, **7**, 9; (*d*) Y. Hu, K. Cheng, L. He, X. Zhang, B. Jiang, L. Jiang, C. Li, G. Wang, Y. Yang and M. Liu, *Anal. Chem.*, 2021, **93**, 1866; (*e*) E. Luchinat, M. Cremonini and L. Banci, *Chem. Rev.*, 2022, **122**(10), 9267.

2 (*a*) E. Moser, E. Laistler, F. Schmitt and G. Kontaxis, *Front. Phys.*, 2017, **5**, 33, DOI: 10.3389/fphy.2017.00033; (*b*) H. Kovacs, D. Moskau and M. Spraul, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2005, **2–3**, 131.

3 (*a*) M. Salzmann, K. Pervushin, G. Wider, H. Senn and K. Wüthrich, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 13585; (*b*) M. Sattler, J. Schleucher and C. Griesinger, *Prog. Nucl. Magn. Reson. Spectrosc.*, 1999, **34**(2), 93; (*c*) C. Fernández and G. Wider, *Curr. Opin. Struct. Biol.*, 2003, **13**(5), 570.

4 (*a*) M. Kainosho, T. Torizawa, Y. Iwashita, T. Terauchi, A. M. Ono and P. Güntert, *Nature*, 2006, **440**, 52; (*b*) S.-Y. Ohki and M. Kainosho, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2008, **53**(4), 208; (*c*) D. Lacabanne, B. H. Meier and A. Böckmann, *J. Biomol. NMR*, 2018, **71**, 141; (*d*) J. Schörghuber, L. Geist, G. Platzer, M. Feichtinger, M. Bisaccia, L. Scheibelberger, F. Weber, R. Konrat and R. J. Lichtenecker, *J. Biomol. NMR*, 2018, **71**(3), 129; (*e*) B. Rowlinson, E. Crublet, R. Kerfah and M. J. Plevin, *Biochem. Soc. Trans.*, 2022, **50**(6), 1555.

5 (*a*) M. P. Williamson, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2013, **73**, 1; (*b*) W. Becker, K. C. Bhattiprolu, N. Gubensäk and K. Zangger, *Chem. Phys. Chem.*, 2018, **19**(8), 895.

6 (*a*) J. L. Kitevski-LeBlanc and R. S. Prosser, *Prog. Nucl. Magn. Reason. Spectrosc.*, 2012, **62**, 1; (*b*) H. Chen, S. Viel, F. Ziarelli and L. Peng, *Chem. Soc. Rev.*, 2013, **42**(20), 7971; (*c*) K. E. Arntson and W. C. K. Pomerantz, *J. Med. Chem.*, 2016, **59**(11), 5158; (*d*) C. T. Gee, K. E. Arntson, A. K. Urick, N. K. Mishra, L. M. L. Hawk, A. J. Wisniewski and W. C. K. Pomerantz, *Nat. Protoc.*, 2016, **11**(8), 1414; (*e*) A. Divakaran, S. E. Kirberger and W. C. K. Pomerantz, *Acc. Chem. Res.*, 2019, **52**(12), 3407; (*f*) D. Gimenez, A. Phelan, C. D. Murphy and S. L. Cobb, *Beilstein J. Org. Chem.*, 2021, **17**(1), 293; (*g*) A. M. Gronenborn, *Structure*, 2022, **30**(1), 6; (*h*) W. Zhu, A. J. Guseman, F. Bhinderwala, M. Lu, X.-C. Su and A. M. Grunenborn, *Angew. Chem., Int. Ed.*, 2022, **61**(23), e202201097.

7 (*a*) M. Salwiczek, E. K. Nyakatura, U. I. M. Gerling, S. Ye and B. Koksch, *Chem. Soc. Rev.*, 2012, **41**, 2135; (*b*) H. Welte, T. Zhou, X. Mihajlenko, O. Mayans and M. Kovermann, *Sci. Rep.*, 2020, **10**, 2640; (*c*) D. T. Yang, A. M. Gronenborn and L. T. Chong, *J. Phys. Chem. A*, 2022, **126**, 2286.

8 (*a*) C. Li, G.-F. Wang, Y. Wang, R. Creager-Allen, E. A. Lutz, H. Scronce, K. M. Slade, R. A. S. Ruf, R. A. Mehl and G. J. Pielak, *J. Am. Chem. Soc.*, 2010, **132**(1), 321; (*b*) P. B. Crowley, C. Kyne and W. B. Monteith, *Chem. Commun.*, 2012, **48**(86), 10681; (*c*) M. G. Lete, A. Franconetti, S. Bertuzzi, S. Delgado, M. Azkargorta, F. Elortza, O. Millet, G. Jiménez-Osés, A. Arda and J. Jiménez-Barbero, *Chem. – Eur. J.*, 2023, **29**(5), e202202208.

9 H.-W. Kim, J. A. Perez, S. J. Ferguson and I. D. Campbell, *FEBS Lett.*, 1990, **1**(2), 34.

10 (*a*) J. C. Jackson, J. T. Hammill and R. A. Mehl, *J. Am. Chem. Soc.*, 2007, **129**(5), 1160; (*b*) H. Qianzhu, E. H. Abdelkader, G. Otting and T. Huber, *J. Am. Chem. Soc.*, 2024, **146**(19), 13641.

11 (*a*) L. Barbieri, E. Luchinat and L. Banci, *Nat. Protoc.*, 2016, **11**(6), 1101; (*b*) A. R. Aricescu, W. Lu and E. Y. Jones, *Acta. Cryst. D*, 2006, **62**(10), 1243.

12 (*a*) L. B. T. Pham, A. Costantino, L. Barbieri, V. Calderone, E. Luchinat and L. Banci, *J. Am. Chem. Soc.*, 2023, **145**(2), 1389; (*b*) A. Costantino, L. B. T. Pham, L. Barberi, V. Calderone, G. Ben-Nissan, M. Sharon, L. Banci and E. Luchinat, *Protein Sci.*, 2024, **33**(3), e4910.

13 (*a*) A. Boeszoermenyi, S. Chhabra, A. Dubey, D. L. Radeva, N. T. Burdzhiev, C. D. Chanev, O. I. Petrov, V. M. Gelev, M. Zhang, C. Anklin, H. Kovacs, G. Wagner, I. Kuprov, K. Takeuchi and H. Arthanari, *Nat. Methods*, 2019, **16**, 333; (*b*) A. Boeszoermenyi, B. Ogórek, A. Jain, H. Arthanari and G. Wagner, *J. Biomol. NMR*, 2020, **74**, 365.

14 (*a*) A. Maleckis, I. D. Herath and G. Otting, *Org. Biomol. Chem.*, 2021, **19**, 5133; (*b*) G. Toscano, J. Holzinger, B. Nagl, G. Kontaxis, H. Kählig, R. Konrat and R. J. Lichtenecker, *J. Biomol. NMR*, 2024, **78**, 139.

15 J. Shao, B. P. Kuiper, A.-M. W. H. Thunnissen, R. H. Cool, L. Zhou, C. Huang, B. W. Dijkstra and J. Broos, *J. Am. Chem. Soc.*, 2022, **144**(30), 13815.

16 R. J. Lichtenecker, *Org. Biomol. Chem.*, 2014, **38**(12), 7551.

17 T. Fujimoto and T. Ritter, *Org. Lett.*, 2015, **17**(3), 544.

18 (*a*) C. Ma, S. Yu, X. He, X. Liu and J. M. Cook, *Tetrahedron Lett.*, 2000, **41**(16), 2781; (*b*) C. Ma, X. Liu, X. Li, J. Flippen-Anderson, S. Yu and J. M. Cook, *J. Org. Chem.*, 2001, **66**(13), 4525.

19 E. Luchinat, L. Barbieri, M. Cremonini, A. Nocentini, C. T. Supuran and L. Banci, *Angew. Chem., Int. Ed.*, 2020, **59**(16), 6535.

20 S. Lindskog, *Pharmacol. Ther.*, 1997, **74**(1), 1.

21 S. Kumar, S. Rulhania, S. Jaswal and V. Monga, *Eur. J. Med. Chem.*, 2021, **209**, 112923.

22 M. Berdyński, P. Miszta, K. Safranow, P. M. Andersen, M. Morita, S. Filipek, C. Żekanowski and M. Kuźma-Kozakiewicz, *Sci. Rep.*, 2022, **12**, 103.

23 (*a*) R. Mikulski, J. F. Domsic, G. Ling, C. Tu, A. H. Robbins, D. N. Silverman and R. McKenna, *Arch. Biochem. Biophys.*, 2011, **516**(2), 97; (*b*) E. Pokrishevsky, L. McAlary, N. E. Farrawell, B. Zhao, M. Sher, J. J. Yerbury and N. R. Cashman, *Sci. Rep.*, 2018, **8**, 15590.

Supporting Information

The synthesis of specifically isotope labelled Fluorotryptophan and its use in mammalian cell-based protein expression for ^{19}F-NMR applications

Table of Contents

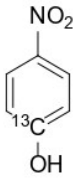
1. Synthesis of compound **1**
    1.1. General Information
    1.2. Synthetic Protocols and Spectral Data
    1.3. NMR Spectra
2. Expression of [^{13}C, $^{2}H_3$]-^{19}F Trp CA2 and SOD1
3. Purification of [^{13}C, $^{2}H_3$]-^{19}F Trp CA2
4. NMR experiments
5. Supplementary figures
6. References

1. Synthesis of compound **1**

1.1. General Information

NMR spectra were recorded on Bruker BioSpin AV NEO 400. The chemical shifts are reported in parts per million (ppm) on the delta (δ) scale. The solvent peak was used as a reference value, for ^{1}H NMR: $CDCl_3$ = 7.27 ppm, $(CD_3)_2SO$ = 2.50 ppm, D_2O = 4.79 ppm; for ^{13}C NMR: $CDCl_3$ = 77.16 ppm, (CD3)2SO = 39.52 ppm. The coupling data are reported as follows: s = singlet; d = doublet; t = triplet; q = quartet; m = multiplet. All the ^{13}C spectra are proton decoupled. High resolution mass spectra were collected on an Orbitrap Exploris 120 (Thermo Fisher Scientific, Bremen, Germany) and on timsTOF fleX ESI/MALDI dual source - Qq-TOF mass spectrometer (Bruker Daltonics, Bremen, Germany) in the positive- and/or negative ion mode. The sum formulas of the detected ions were determined using Bruker Compass DataAnalysis 5.3 based on the mass accuracy ($\Delta m/z \leq 5$ ppm) and isotopic pattern matching (SmartFormula algorithm). Dry solvents were purchased from ThermoFisher Scientific. Analytical TLC was performed on pre-coated (25 mm) silica gel 60 with fluorescent indicator plates. Visualization was done under UV (254 nm) and by staining with ninhydrin or phosphomolybdic acid staining solutions with subsequent heating using a heat-gun. Column chromatography was done using Silica 60 (0.04-0.063) from Merk. All reactions were performed in flame-dried glassware under a positive pressure of argon with magnetic stirring unless noted otherwise. All reagents were purchased from commercial sources and used without further purification, unless otherwise noted. [2- ^{13}C] acetone was purchased from Sigma Aldrich and stored into at – 25°C. D_2O was purchased from Eurisotop and stored at 25 °C. 7.6 N DCl was purchased by Eurisotop. Yields refer to isolated compounds. Microwave reactions were performed in a Biotage® Initiator+. Microwave reaction vials, caps with septum and stir bars were also purchased from Biotage.

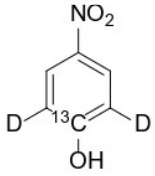
1.2. Synthetic Protocols and Spectral Data



[1- ^{13}C] 4-nitrophenol[1] **(3):** Sodium nitromalonaldehyde (3.25 g, 23.37 mmol) was weighed into a 500 mL round-bottom flask and dissolved in 200 mL of H_2O. The mixture was placed in an ice bath, and a magnetic stir bar was added. 1 g of [2-^{13}C] acetone (16.93 mmol, 1.24 mL) was removed from the freezer at -25°C and promptly added to the mixture at 0°C using a syringe. While stirring the mixture at 0°C, an aqueous NaOH solution (4.4 g, 0.11 mol in 20 mL) was slowly added via a dropping funnel. After the addition was complete, the flask was tightly closed with a glass stopper and stirred for 6 days at 4°C. The resulting dark brown solution was cooled to 0°C, and 52 mL of 6 M HCl were slowly added. The mixture was then transferred to a separating funnel, and 50 mL of ethyl acetate were added. The phases were separated, and the aqueous phase was extracted again 5 times with 30 mL of ethyl acetate each. The combined organic phases were dried over $MgSO_4$, filtered, and evaporated under reduced pressure using a rotavapor, yielding a brown oil. The crude product was purified using a silica gel chromatography column eluted with heptane–ethyl acetate (7:3). The reaction yielded 1.47 g (63%) of [1-^{13}C] 4-nitrophenol **3** as a yellow solid.

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.19 (t, *J* = 9.4 Hz, 2H, CH arom.), 6.93 (dd, *J* = 9.2, 2.4 Hz, 2H, CH aromatic), 5.49 (d, *J* = 3.4 Hz, 1H, OH).

^{13}C NMR (101 MHz, $CDCl_3$) δ 161.11 (s, ^{13}C-OH), 126.43 (s), 142.03 (s), 115.79 (d, *J* = 65.9 Hz).

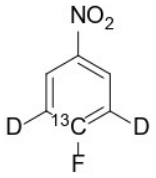


[1-^{13}C, 2,6-$^2H_2$] 4-nitrophenol[1] **(4):** A 2-5 mL microwave reaction vial was loaded with 200 mg of [1-^{13}C] 4-nitrophenol **3** (1.44 mmol), 1 mL of D_2O, and 1 mL of DCl 7.6 N. The mixture was irradiated for 1.5 h at 170 °C in a Biotage® Initiator+ and then cooled to RT. The solvent was evaporated under reduced pressure, and the resulting solid was purified via silica column chromatography eluted with ethyl acetate-heptane (6:4). The product **4** was isolated in 98% yield as a yellow solid. ^{1}H-NMR spectroscopy analysis indicated 97.5% of deuterium incorporation at positions 2 and 6.

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.19 (d, *J* = 9.6 Hz, 2H), 6.92 (dd, *J* = 9.6, 2.5 Hz, 2.5% residual CH position 3 and 5).

^{13}C NMR (101 MHz, $CDCl_3$) δ 161.18 (s, ^{13}C-OH), 126.30 (s).

HRMS calcd for $C_5{}^{13}CH_2D_2NO_3$ $[M - H]^-$ 141.03557, found 141.0353.

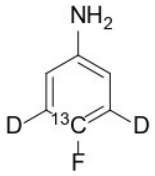


[4-^{13}C, 3,5-$^{2}H_2$]4-fluoronitrobenzene[2] (5): To a 50 mL two-neck round-bottom flask, 1.52 g (9.99 mmol) of CsF and 0.68 g (1.55 mmol) of N,N′-1,3-Bis(2,6-diisopropylphenyl)-2-chloro imidazolium chloride were added and stirred at 140 °C under vacuum for 3 hours. The solid mixture was then allowed to cool to room temperature, and 142 mg (0.99 mmol) of [1-^{13}C, 2,6-$^{2}H_2$]4-nitrophenol **4** were added. The flask was equipped with a reflux condenser and a septum, and placed under argon atmosphere. Next, 7 mL of dry toluene were added, and the mixture was stirred at 110 °C for 24 hours. Upon completion, the reaction mixture was allowed to cool to room temperature and filtered through Celite, eluting with dichloromethane. The dichloromethane was removed under reduced pressure (>500 mbar, 40 °C). The residual toluene was evaporated at a pressure above 90 mbar at 40 °C after forming an azeotrope with methanol, using a rotary evaporator (product bp = 205 °C, lit.). The resulting crude brown oil was further purified using column chromatography with heptane/ethyl acetate (20:1) as the eluent. This step resulted in the isolation of 110 mg (76%) of [4-^{13}C, 3,5-$^{2}H_2$]4-fluoronitrobenzene **5** as a yellow oil, which solidifies below 20 °C. ^{1}H-NMR spectroscopy analysis indicated a deuteration level of 96% at both positions 2 and 6.

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.29 (dd, J = 10.7, 4.6 Hz, 2H), 7.25 – 7.19 (m, 4% residual CH position 3 and 5).

^{13}C NMR (101 MHz, $CDCl_3$) δ 166.36 (d, J = 258.0 Hz, ^{13}C-F), 126.36 (d, J = 10.0 Hz, CH arom.), 117.49 – 115.06 (m, C-D arom.)

HRMS (m/z): calcd for $C_5{}^{13}CH_2D_2FNO_2$ 144.0385, found 144.0377.

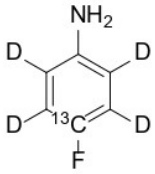


4- ^{13}C, 3,5-$^{2}H_2$]4-fluoroaniline[3] (6): 111 mg (0.77 mmol) of substrate **5** were dissolved in 4 mL of dry methanol in a flame dried Schlenk flask and 40 mg (0.38 mmol) of 10% Pd/C catalyst were added. The flask was purged with argon and then exposed to H_2 using a hydrogen balloon. The mixture was stirred under H_2 overnight and then filtered through a Celite pad, eluting with dichloromethane. The solvent was removed at reduced pressure (P > 300 mbar, bp product = 188 °C) and the resulting residue was dissolved in 10 mL of dichloromethane. The solution was washed with 10 mL of 1 M aqueous NaOH and 10 mL of brine. The organic phase was dried over $MgSO_4$, and the solvent evaporated, resulting in 108 mg (95%) of [4-^{13}C, 3,5-$^{2}H_2$]4-fluoroaniline **6** as a brown oil. ^{1}H-NMR spectroscopy analysis indicated a deuteration level of 96% at both positions 2 and 6.

^{1}H NMR (400 MHz, $CDCl_3$) δ 6.90 – 6.82 (m, 4% residual CH position 3 and 5), 6.63 (dd, *J* = 10.5, 4.2 Hz, 2H, CH arom.), 3.54 (bs, 2H, NH_2).

^{13}C NMR (101 MHz, $CDCl_3$) δ 156.55 (d, *J* = 235.7 Hz, ^{13}C-F).

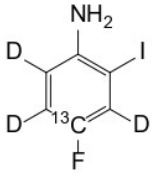
HRMS: calcd for $C_5{}^{13}CH_4D_2FN$ $[M + H]^+$ 115.07161, found 115.0714.



[4- ^{13}C, 2,3,5,6-$^{2}H_4$]4-fluoroanilinium chloride (7): In a 2-5 mL microwave reaction vial, 114 mg of [4-^{13}C, 3,5-$^{2}H_2$] 4-fluoroaniline **6** (1 mmol) was mixed with 1 mL of D_2O and 1 mL of DCl 7.6 N. The vial was irradiated for 1.5 h at 180 °C. After cooling to RT, the solvent is evaporated under vacuum. The resulting solid was then taken up in 10 mL of 1 M NaOH and extracted with dichloromethane (5 × 20 mL). The combined organic phases were dried over $MgSO_4$ and the solvent was evaporated, yielding 110 mg (95%) of [4-^{13}C, 2,3,5,6-$^{2}H_4$]4-fluoroaniline **7** as a pale brown oil. ^{1}H-NMR spectroscopy indicated almost quantitative aryl deuteration with 97.5 % deuteration for each of the four positions.

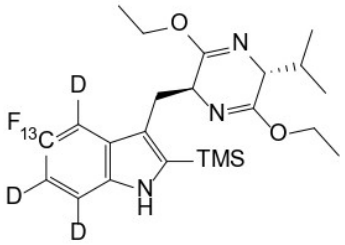
^{1}H NMR (400 MHz, $CDCl_3$) δ 6.86 (m, 2.5% residual CH position 3 and 5.), 6.64 (m, 2.5% residual CH position 2 and 6), 3.53 (bs, 2H, NH_2).

^{13}C NMR (101 MHz, $CDCl_3$) δ 156.43 (d, *J* = 235.7 Hz, ^{13}C-F).



[4- ^{13}C, 1,3,5-^{2}H$_3$]4-fluoro-2-iodoaniline[4] (8): 581 mg of compound **7** (5.23 mmol), 1.33 g of I_2 (5.23 mmol), and 879 mg of $NaHCO_3$ were added to a brown screw-cap vial. To this, 6 mL of H_2O and 4 mL of dichloromethane were added, and the mixture was stirred at room temperature for 30 hours. After this time, 413 mg (2.61 mmol) of sodium thiosulfate were added, and the mixture was stirred for an additional 15 minutes. The phases were then separated, and the aqueous phase was extracted three times with 10 mL portions of dichloromethane. The combined organic phases were washed with 20 mL of brine, dried over $MgSO_4$, and evaporated under vacuum. The product **8** was isolated as a yellow oil by short-path distillation at 120°C and 0 mbar, yielding 1.087 g (87.6%).

^{1}H NMR ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.42 – 7.34 (m, 2% residual CH position 3), 6.94 – 6.86 (m, 5% residual CH position 5), 6.73 – 6.67 (m, 5% residual CH position 6), 3.95 (bs, 2H).
^{13}C NMR (101 MHz, $CDCl_3$) δ 155.36 (d, J = 240.8 Hz, ^{13}C-F).
HRMS calcd for $C_5{}^{13}CH_2D_3FIN$ $[M + H]^+$ 241.9745; found 241.9743.

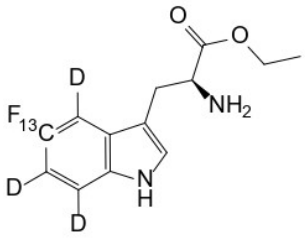


(2R,5S)-3,6-Diethoxy-2-isopropyl-5-[5-^{13}C, 4,6,7-^{2}H$_3$, 5-fluoro-2-(trimethylsilyl)-3-indolyl]methyl-2,5 -dihydro pyrazine[5] (9): In a 100 mL round-bottom flask equipped with a stirring bar were placed compound **8** (161 mg, 0.67 mmol), (2R,5S)-3,6-diethoxy-2-iso-propyl-5-[3-(trimethylsilyl)prop-2-ynyl]-2,5-dihydropyrazine **12** (237 mg, 0.73 mmol), palladium(II) acetate (6 mg, 0.027 mmol), lithium chloride (28.3 mg, 0.67 mmol), sodium carbonate (141 mg, 1.33 mmol), and DMF (10 mL). The reaction mixture was degassed and then heated at 100°C under argon until the starting compound **8** was no longer detected on analysis by TLC (30 hours). The DMF was removed under reduced pressure, and the residue was taken up in dichloromethane (50 mL). The suspension that resulted was allowed to pass through a Celite pad to remove insoluble solids. The solution was concentrated in vacuum, and the product was purified by flash chromatography (silica gel, hexane/EtOAc 98:2) to afford 197 mg of compound **9** as an oil in 70% yield.

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.90 (s, 1H), 7.44 – 7.38 (m, 2% residual CH position 4), 7.25 – 7.21 (m, 10% residual CH position 6), 6.92 – 6.87 (m, 2% residual CH position 7), 4.29 – 3.98 (m, 5H), 3.88 (t, *J* = 3.4 Hz, 1H), 3.52 (dd, *J* = 14.3, 3.3 Hz, 1H), 2.82 (dd, *J* = 14.3, 9.7 Hz, 1H), 2.33 – 2.21 (m, 1H), 1.32 (t, *J* = 7.1 Hz, 3H), 1.22 (t, *J* = 7.1 Hz, 3H), 1.03 (d, *J* = 6.8 Hz, 3H), 0.68 (d, *J* = 6.8 Hz, 3H), 0.42 (s, 9H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 163.45 (s), 163.01 (s), 157.45 (d. ^{13}C-F, J = 233.49 Hz), 156.37 (s), 136.30 (s), 60.75 (s), 58.78 (s), 31.84 (d, *J* = 19.8 Hz), 19.19 (s), 16.72 (s), 14.47 (s), 14.39 (s) - 0.56 (s).

HRMS calcd. for $C_{22}{}^{13}CH_{31}D_3FN_3O_2Si$ $[M + H]^+$ 436.2699, found 436.2696.

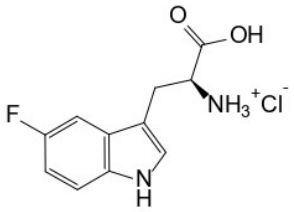


[5- ^{13}C, 4,6,7-$^{2}H_3$]-5-Fluoro-L-tryptophan ethyl ester[5] (10): To a solution of optically pure compound **9** (235 mg, 0.53 mmol) in THF (8 mL) at 0°C was slowly added an aqueous solution of 6 N HCl (7 mL). The mixture was allowed to warm to room temperature and stirred for 4 h. Ice (10 g) was added to the solution, and the pH of the reaction mixture was adjusted to 8 with aqueous NH_4OH (concentrated) at 0°C. The mixture was then extracted with dichloromethane (3 × 50 mL). The combined organic layers were dried over $MgSO_4$, and the solvent was removed under reduced pressure. The residue which resulted was purified by flash chromatography (silicagel, EtOAc) to afford 108 mg of compound **10** in 79% yield.

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.07 (s, 1H), 7.13 (d, *J* = 2.0 Hz, 1H), 4.21 – 4.13 (m, 2H), 3.79 (m, 1H), 3.22 (dd, *J* = 14.4, 5.0 Hz, 1H), 3.04 (dd, *J* = 14.5, 7.4 Hz, 1H), 1.59 (s, 2H), 1.26 (t, *J* = 7.1 Hz, 3H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 175.47 (s), 157.97 (d, *J* = 234.9 Hz), 124.77 (s), 111.85 (s), 61.14 (s), 55.08 (s), 30.84 (s), 14.32 (s).

HRMS calcd for $C_{12}{}^{13}CH_{12}D_3FN_2O_2$ $[M + H]^+$ 255.1412, found 255.1413.

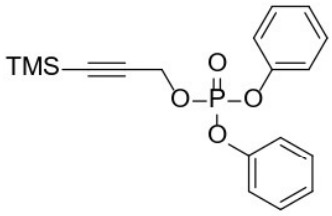


[5- ^{13}C, 4,6,7-$^{2}H_3$]-5-Fluoro-L-tryptophan[5] (1): Into a 25 mL round-bottom flask were added compound **10** (131 mg, 0.51 mmol), aqueous 1 N NaOH (1.0 mL, 1 mmol), and ethanol (1.5 mL). The solution was heated to 50°C for 2 h. Analysis by TLC (silica gel) indicated that all the starting material had disappeared. Most of the EtOH was removed under reduced pressure. A piece of ice was dropped into the flask and the mixture was brought to pH 6-7 with an aqueous solution of 2 N HCl at 0°C. The water in the solution that resulted was removed under reduced pressure until a white precipitate appeared. The total volume of solution was reduced to 0.5 mL. Cold water (1 mL) was then added, and the precipitate that formed was collected by vacuum filtration, washed with cold water (3 × 1 mL), and dried to provide compound **1** (54 mg) in 94.6% yield.

^{1}H NMR (400 MHz, D_2O) δ 7.46 (m, 7% residual CH position 7), 7.36 (s, 1H), 7.07 – 7.02 (m, 4% residual CH position 4 and 6), 4.30 – 4.18 (m, 1H), 3.46 (dd, *J* = 15.4, 5.2 Hz, 1H), 3.41 – 3.32 (dd, *J* = 15.5, 7.38 Hz, 1H).

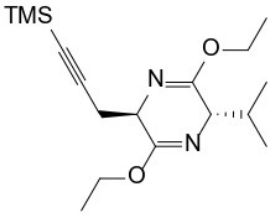
^{13}C NMR (101 MHz, MeOD) δ 157.72 (d, *J* = 233.0 Hz), 126.18 (s), 52.97 (s), 26.04 (s).

HRMS calcd for $C_{10}{}^{13}CH_8D_3FN_2O_2$ $[M + H]^+$ 227.10992, found 227.1095.



3-(Trimethylsilyl)-2-propyn-1-yl-P,P-diphenylphosphinate[5] (13): To a solution of 3 -(trimethylsilyl)prop-2-yn-1-ol (1.4 g, 11.7 mmol) and diphenyl chlorophosphate (3.52 g, 13.1 mmol) in diethyl ether (30 mL) at -50°C was added powdered potassium hydroxide (4.18 g, 74 mmol). The mixture that resulted was allowed to warm to 0°C over 20 min and stirred at 0°C for 12 h. The reaction mixture was poured into ice-water (100 mL), and the ether layer was separated. The aqueous layer was extracted with ether (2 × 50 mL). The combined ether layer was washed with water until the water washing was no longer alkaline to pH paper. The ether solution was dried ($MgSO_4$), and the solvent was removed under reduced pressure. The colorless residue was chromatographed on silica gel (hexane/EtOAc 8:1) to provide pure **13** (3.6 g) in 92% yield.

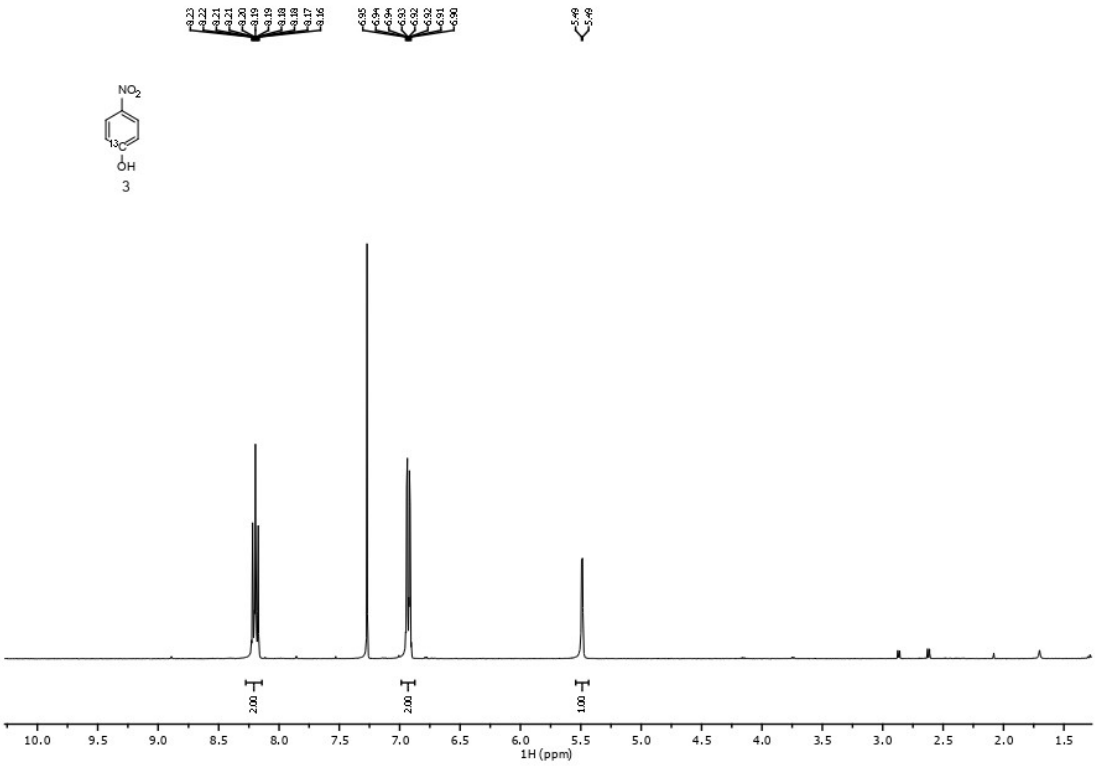
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.38 – 7.19 (m, 10H), 4.86 (d, *J* = 10.7 Hz, 2H), 0.16 (s, 9H).

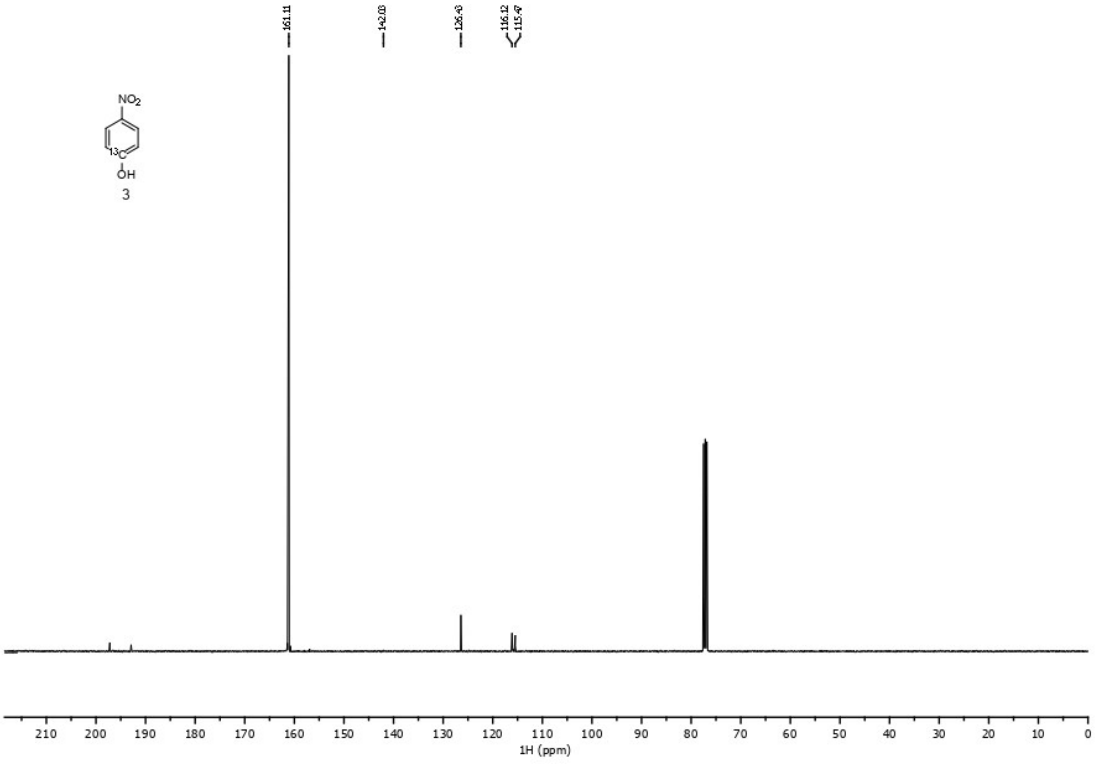


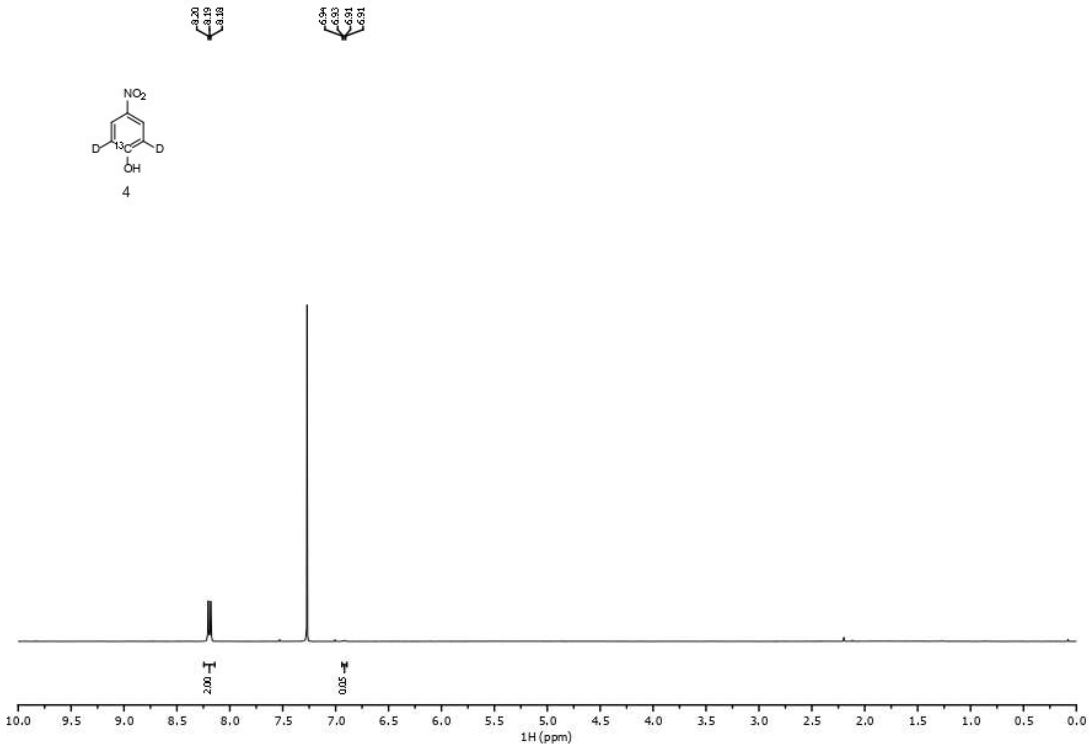
(2R,5S)-3,6-Diethoxy-2-isopropyl-5-[3-(trimethylsilyl)-prop-2-ynyl]-2,5-dihydropyrazine[5] (12): To a solution of (2R)-3,6-diethoxy-2-isopropyl-2,5-dihydropyrazine (0.827 g, 3.9 mmol) in dry THF (25 mL) under nitrogen was syringed n-BuLi (2.5 M, 1.72 mL, 4.3 mmol) dropwise at -78 °C. This solution was stirred at -78 °C for 30 min. To this solution was added slowly a solution of diphenyl 3-(trimethylsilyl)prop- 2-ynyl phosphate **13** (1.4 g, 39 mmol) in dry THF (20 mL) that was cooled to -78 °C. After the reaction mixture was allowed to stir for 6 h at -78 °C, it was allowed to slowly warm to room temperature. The solution was then quenched with the addition of water (2 mL). The THF was removed under reduced pressure, and the residue was partitioned between water (20 mL) and diethyl ether (60 mL). The organic layer was separated, and the aqueous layer was extracted with ether (3 × 30 mL). The combined organic layers were washed with brine and dried ($MgSO_4$). After removal of solvent under reduced pressure, the residue was purified by chromatography (silica gel, hexane/EtOAc 98:2) to afford **12** as an oil (1.0 g) in 80% yield.

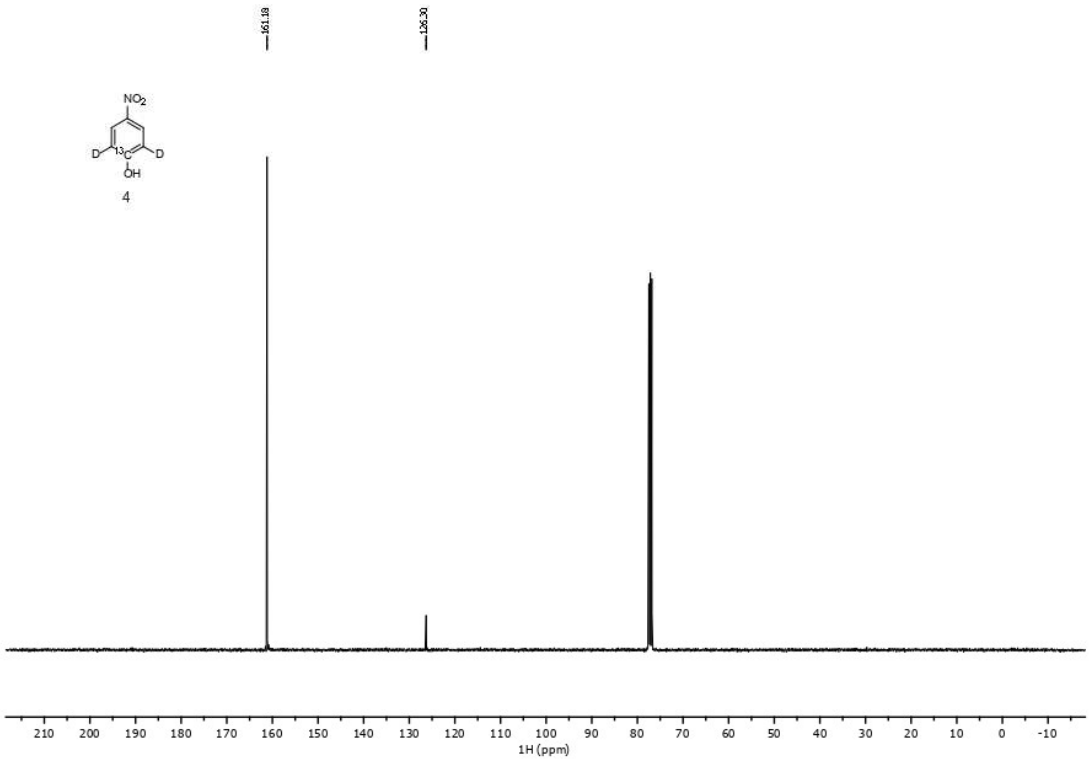
^{1}H NMR (400 MHz, $CDCl_3$) δ 4.30 – 4.06 (m, 5H), 3.98 (t, *J* = 3.3 Hz, 1H), 2.74 (ddd, *J* = 42.2, 16.6, 4.5 Hz, 2H), 2.30 (dtd, *J* = 13.7, 6.8, 3.2 Hz, 1H), 1.30 (td, *J* = 7.1, 2.8 Hz, 6H), 1.06 (d, *J* = 6.9 Hz, 3H), 0.71 (d, *J* = 6.8 Hz, 3H), 0.11 (s, 9H).

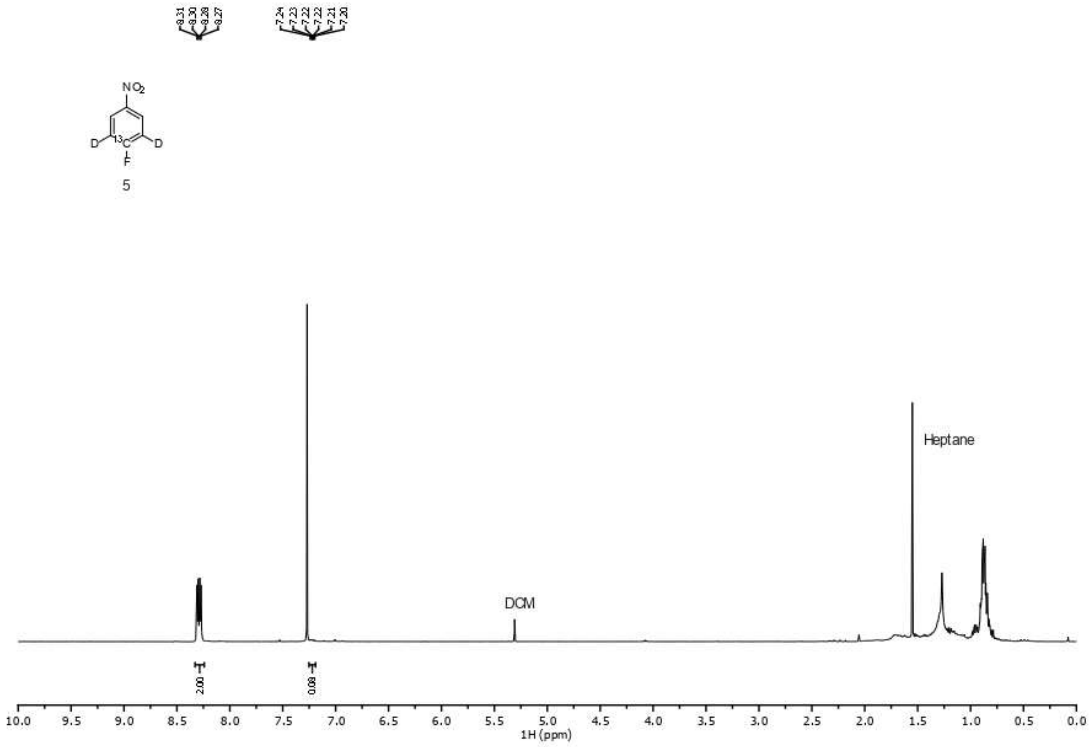
1.3. NMR spectra

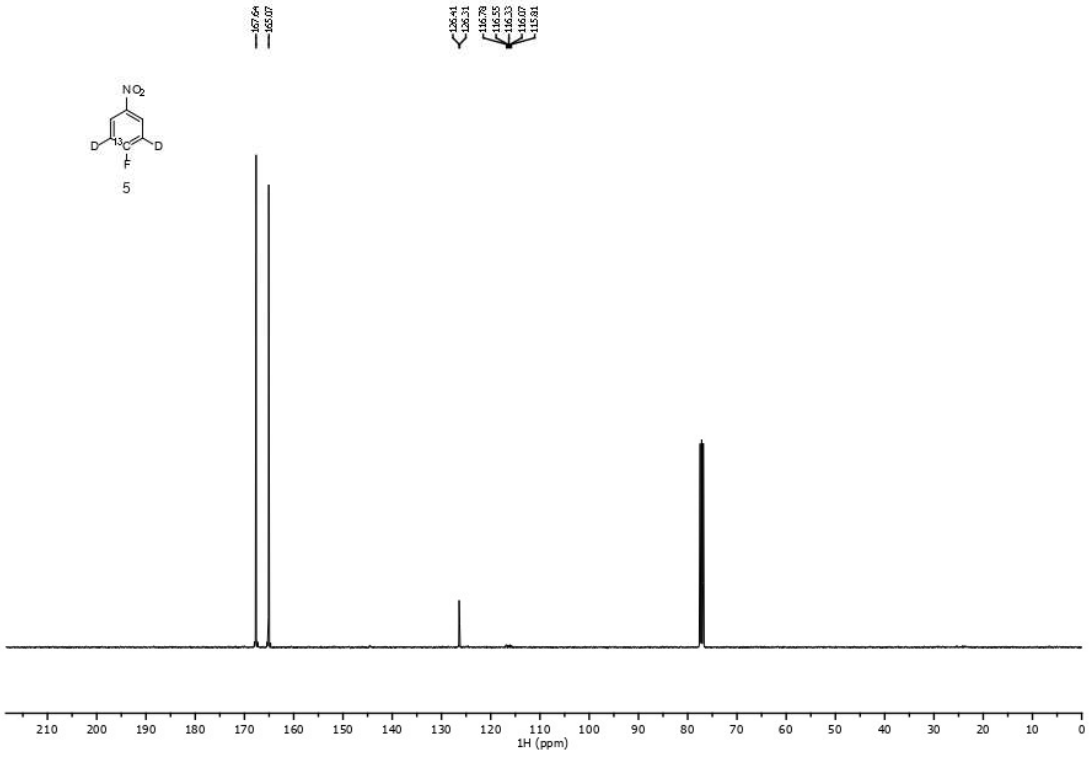


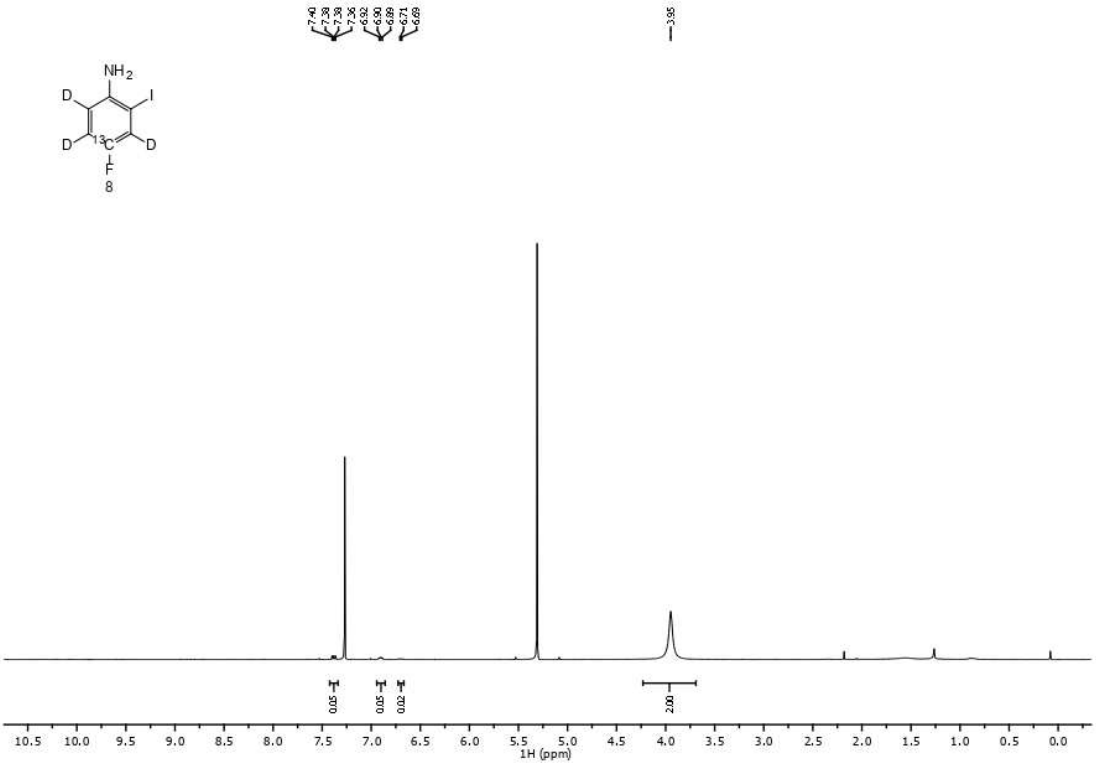


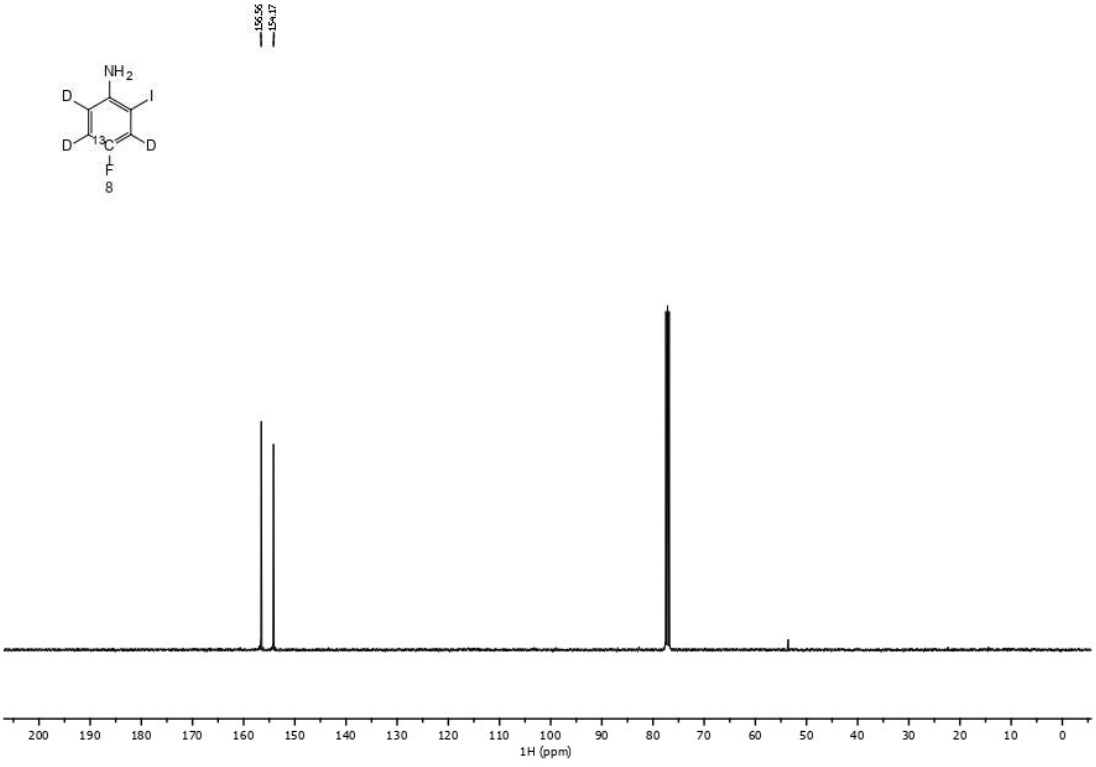

8.20
8.19
8.18
6.94
6.93
6.91
6.91
NO2
D
13C
D
OH
4
2.00
0.05
10.0 9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0
1H (ppm)

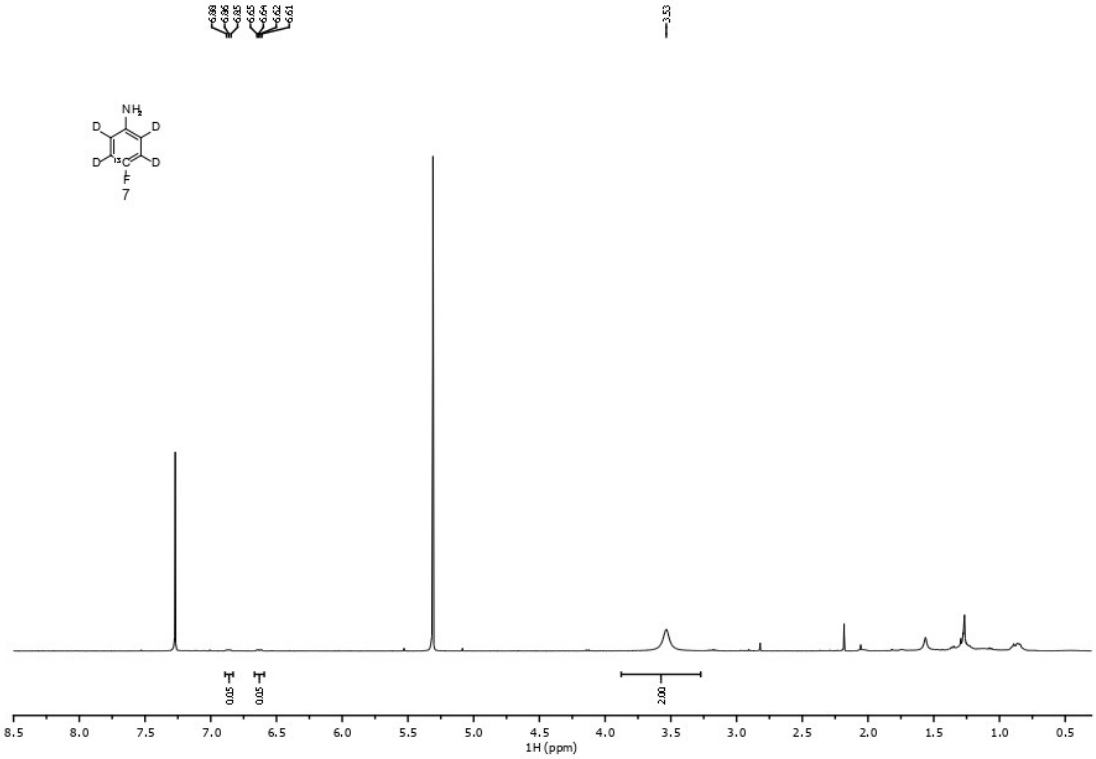

161.18
126.30
NO2
D
13C
D
OH
4
210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
1H (ppm)

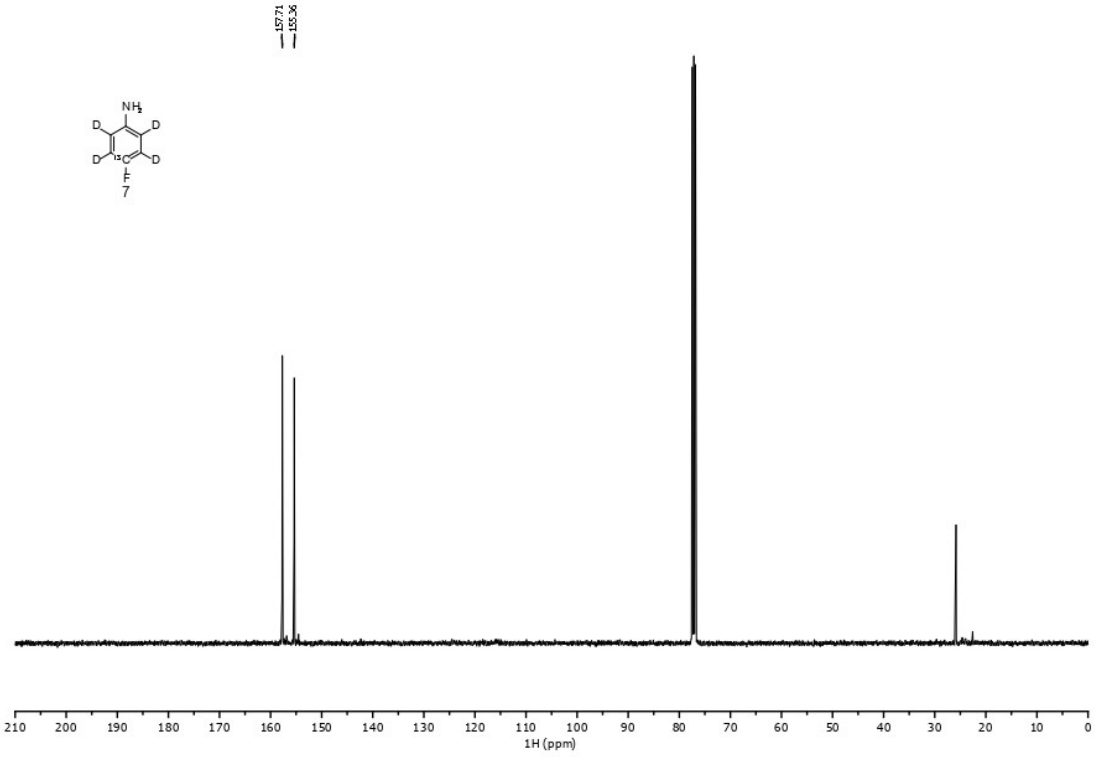

9.31
9.30
9.28
9.27
7.24
7.23
7.22
7.22
7.21
7.20
NO2
D
13C
D
F
5
Heptane
DCM
2.00
0.98
10.0
9.5
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
1H (ppm)

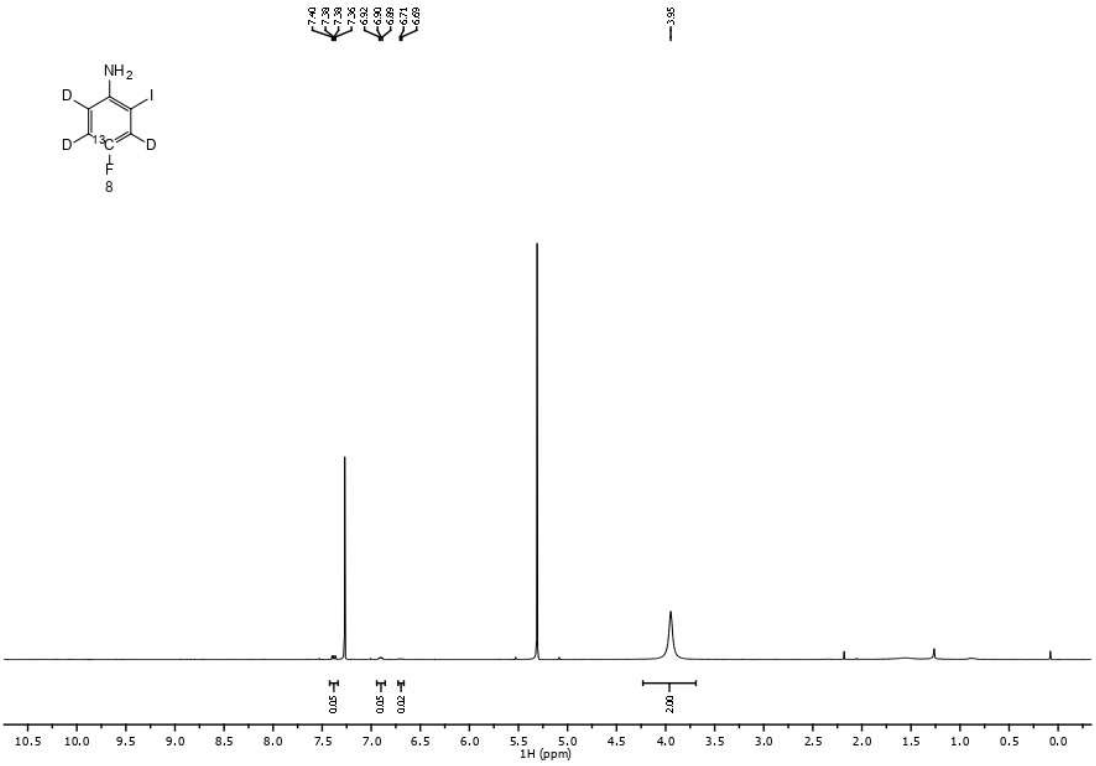

167.64
165.07
126.41
126.31
116.78
116.55
116.33
116.07
115.81
NO2
D
13C
D
F
5
210
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
1H (ppm)

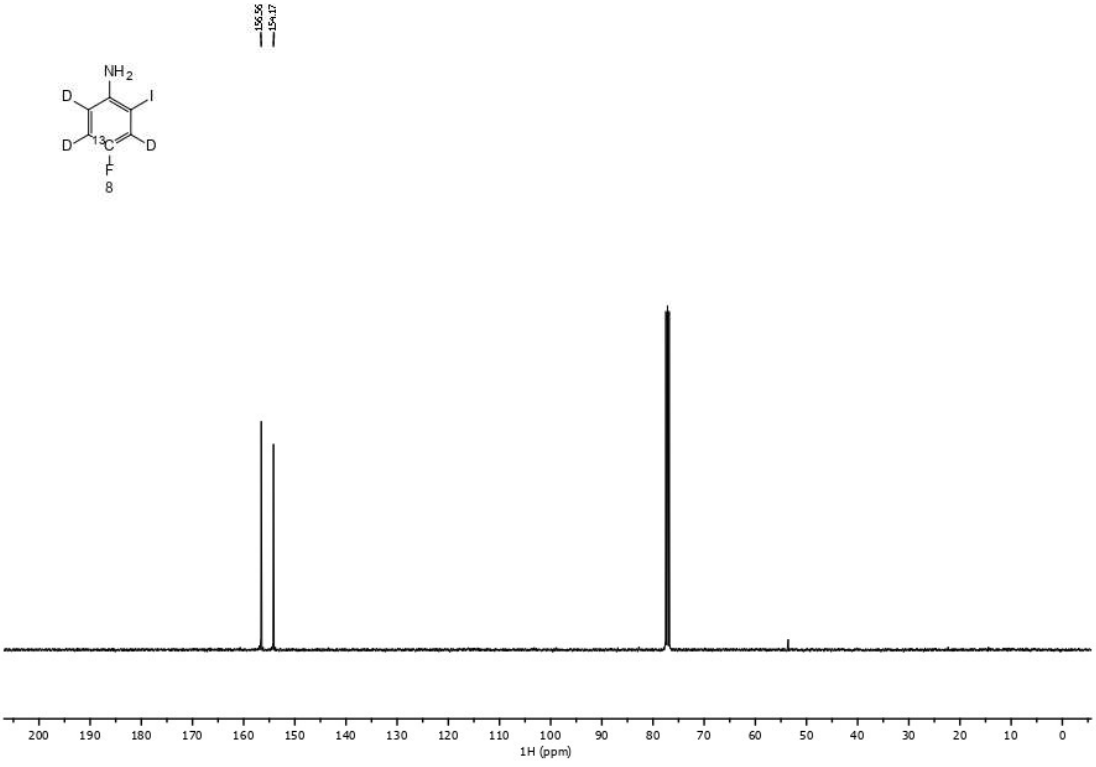

7.40
7.39
7.38
7.36
6.92
6.90
6.89
6.71
6.69
3.95
NH2
D
I
D
13C
D
F
8
0.05
0.05
0.02
2.00
10.5 10.0 9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0
1H (ppm)

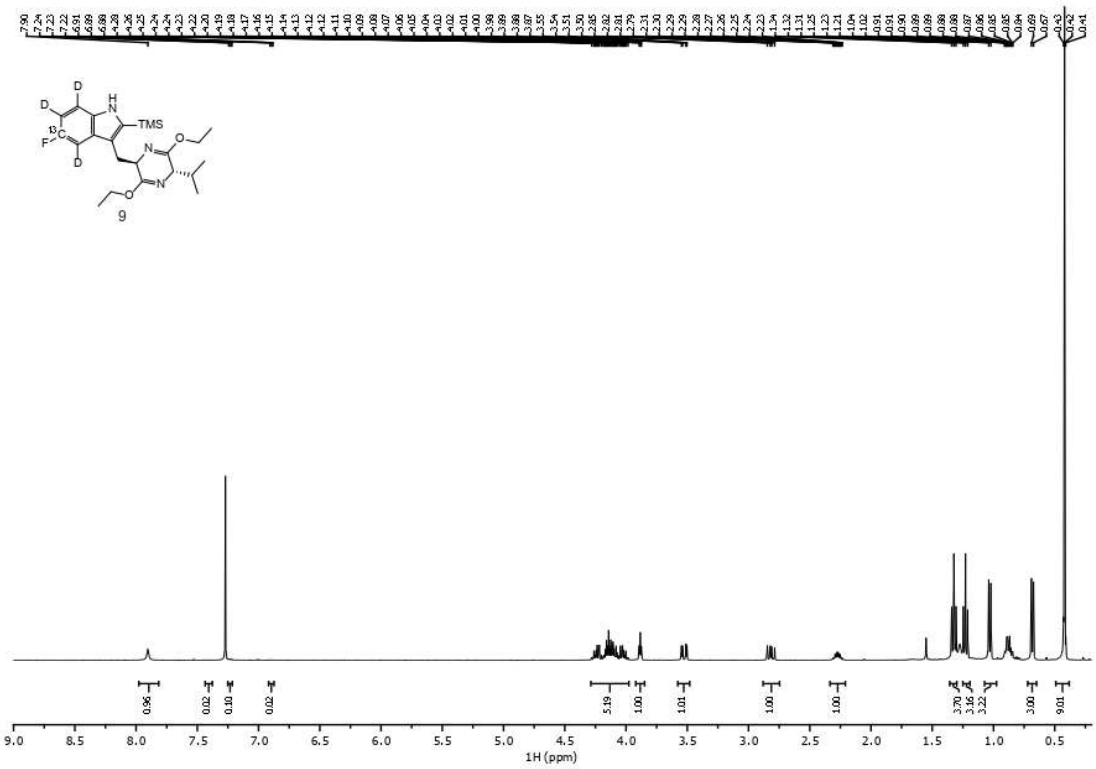

156.56
154.17
NH2
D
I
D
13C
D
F
8
200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0
1H (ppm)

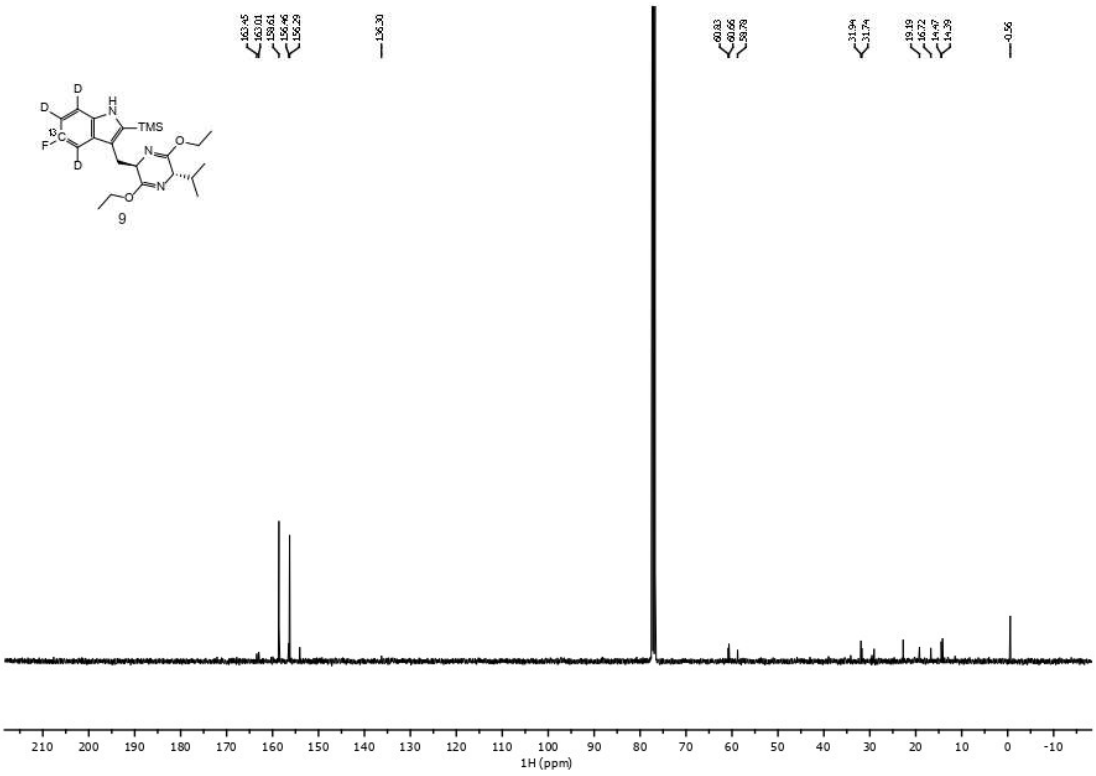

6.88
6.86
6.85
6.65
6.64
6.62
6.61
3.53
NH2
D
D
D
D
13C
F
7
0.05
0.05
2.00
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
1H (ppm)

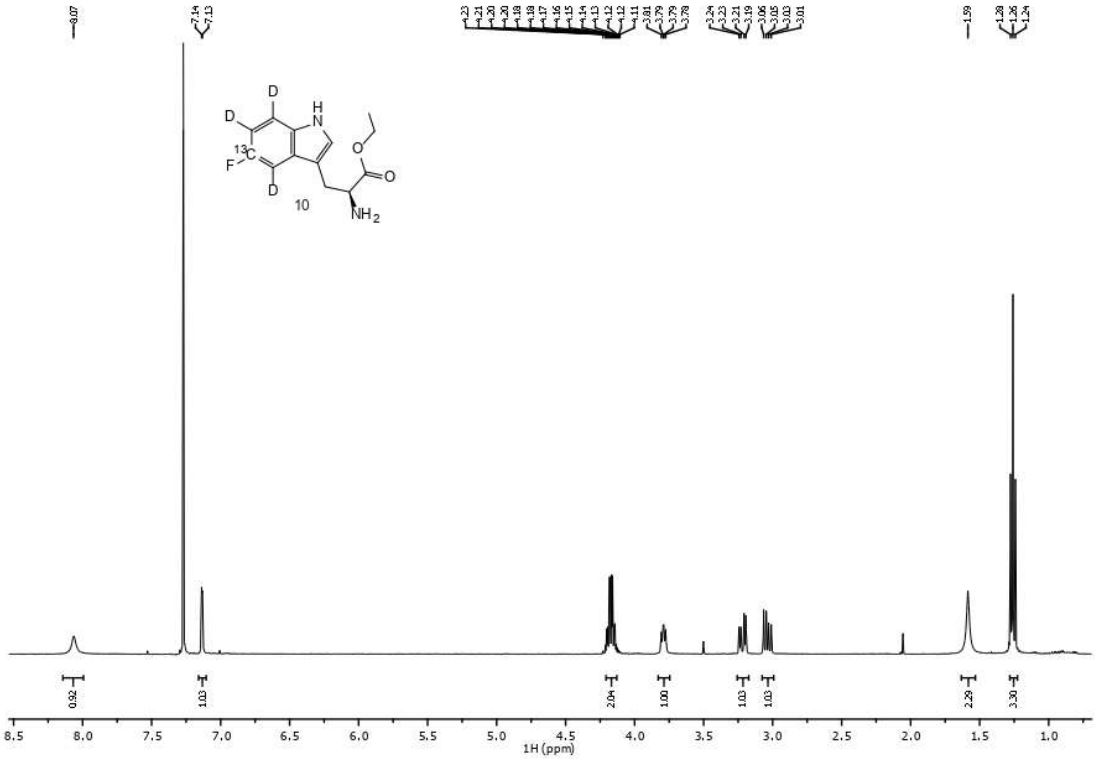

157.71
155.36
NH2
D
D
D
D
13C
F
7
210
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
1H (ppm)

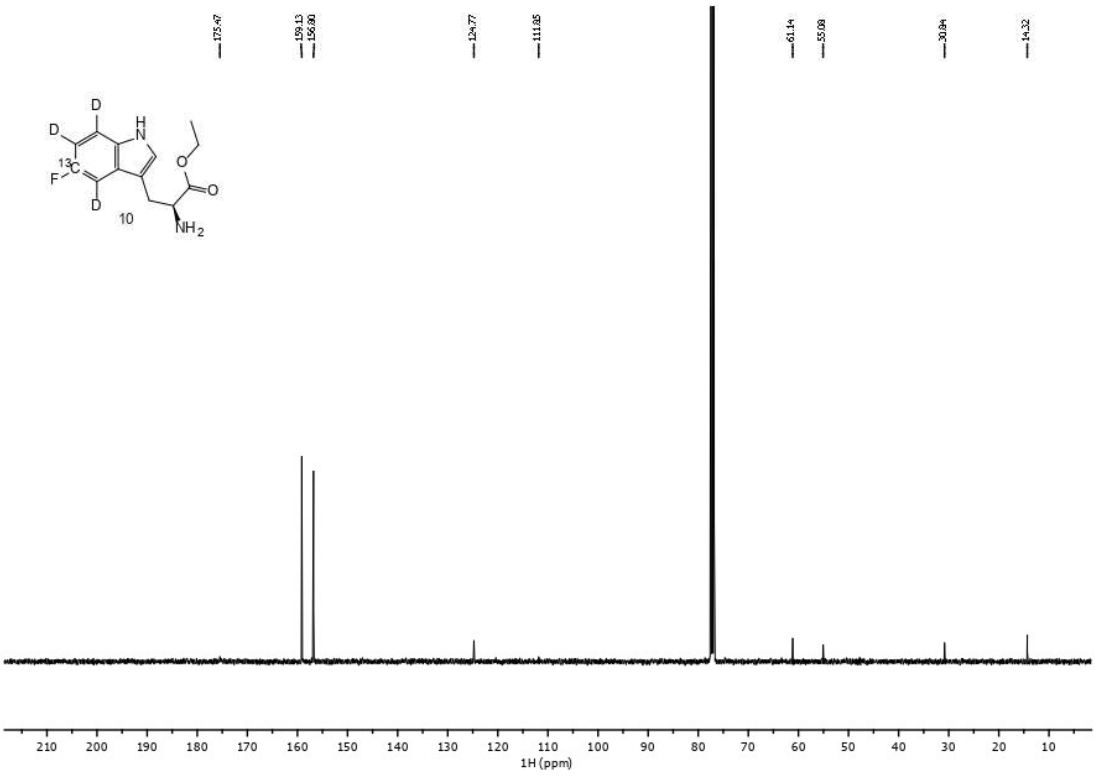

7.40
7.39
7.38
7.36
6.92
6.90
6.89
6.71
6.69
3.95
NH2
D
I
D
13C
D
F
8
0.05
0.05
0.02
2.00
10.5 10.0 9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0
1H (ppm)

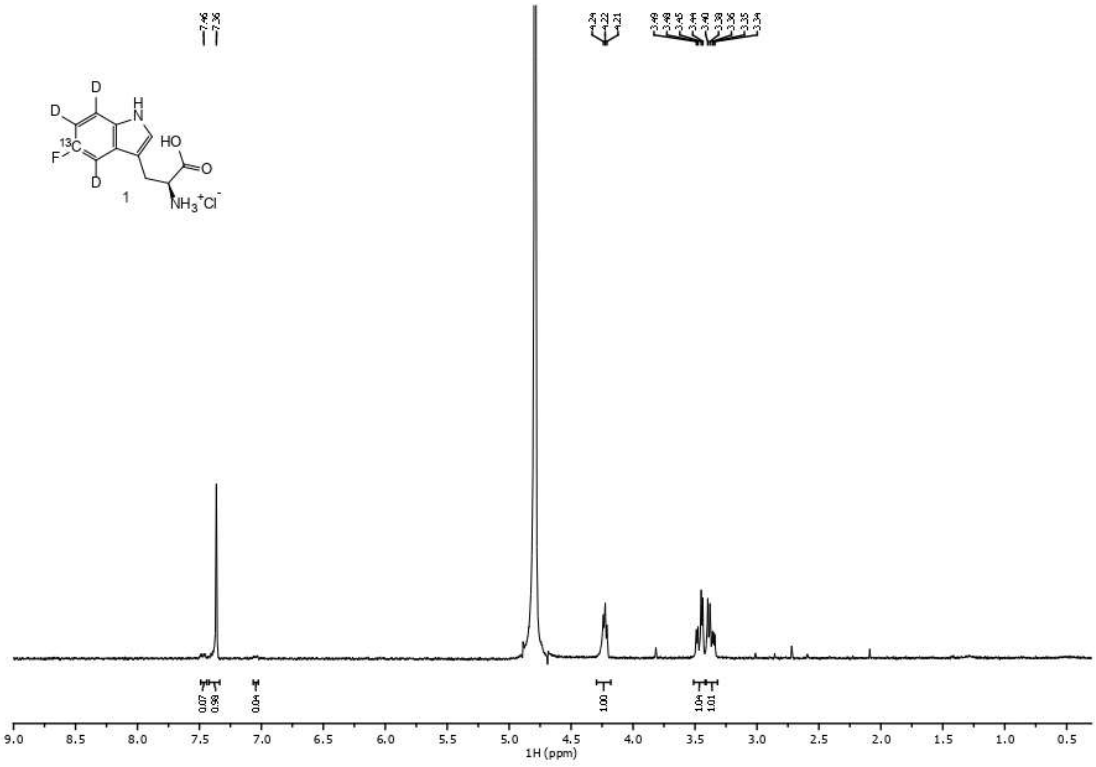

156.56
154.17
NH2
D
I
D
13C
D
F
8
200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0
1H (ppm)

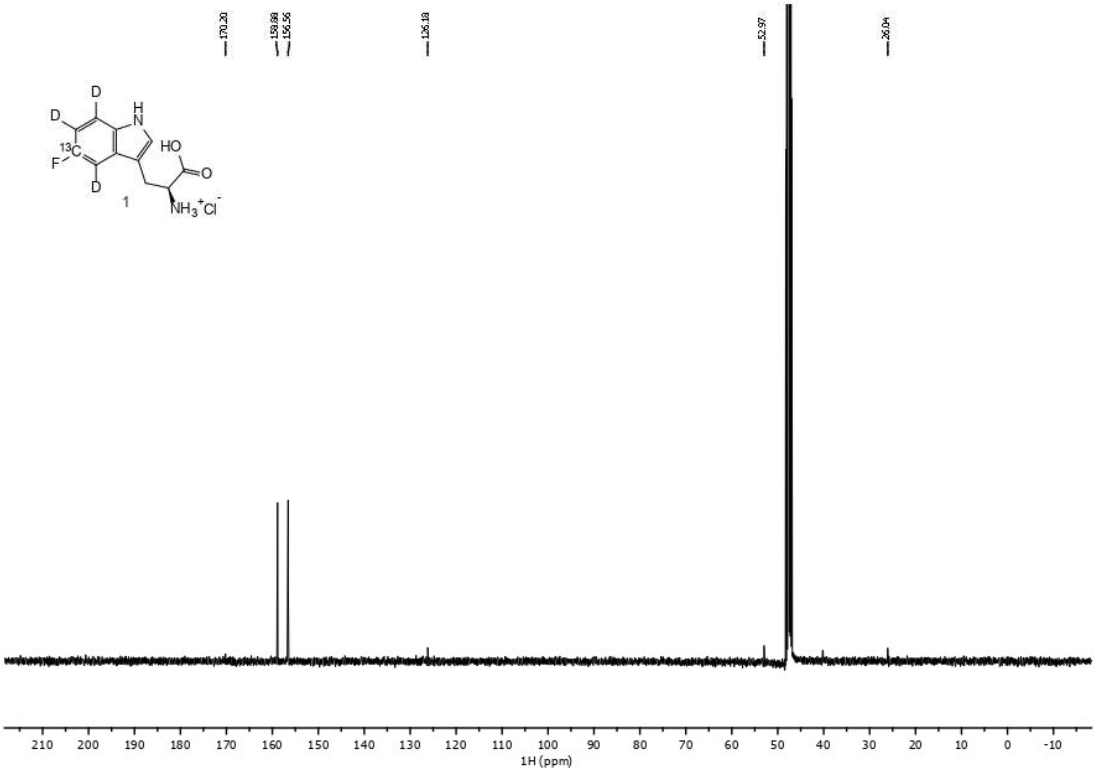

TMS
9
1H (ppm)

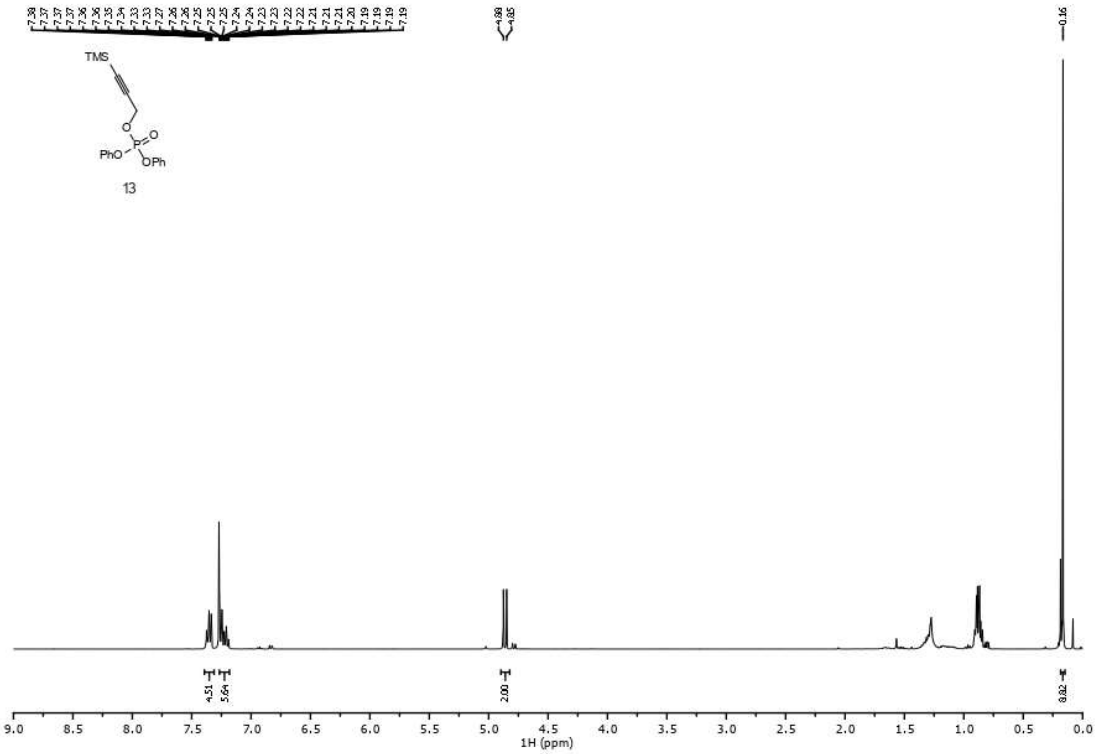

TMS
9
1H (ppm)

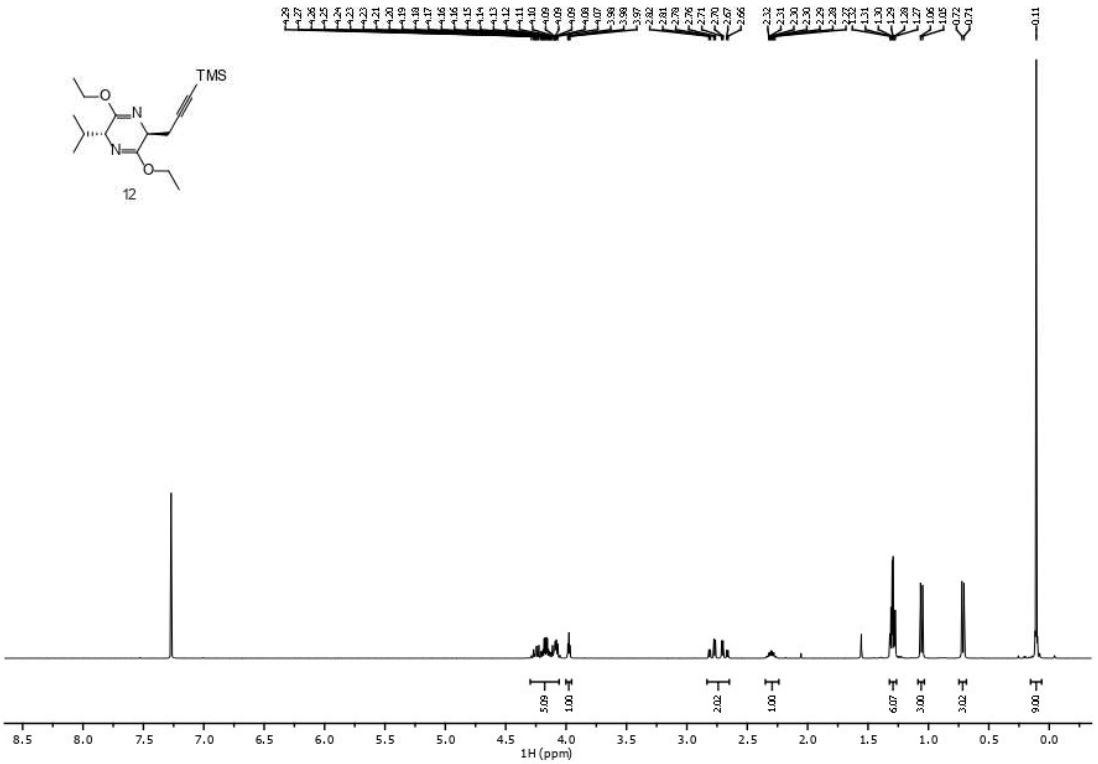

D
H
N
D
13C
F
D
O
O
10
NH2
0.92
1.03
2.04
1.00
1.03
1.03
2.29
3.30
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
1H (ppm)


175.47
159.13
156.93
124.77
111.85
61.14
55.08
30.94
14.32
D
H
N
D
13C
F
D
O
O
10
NH2
210
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
1H (ppm)


D
D
H
N
13C
F
HO
O
D
1
NH3+Cl-
7.46
7.36
4.24
4.22
4.21
3.49
3.48
3.45
3.44
3.40
3.38
3.36
3.35
3.34
0.97
0.98
0.94
1.00
1.04
1.01
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
1H (ppm)


D
D
H
N
13C
F
HO
O
D
1
NH3+Cl-
170.20
159.89
156.56
126.18
52.97
26.04
210
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
-10
1H (ppm)


TMS
PhO
OPh
13
1H (ppm)


TMS
12
1H (ppm)

2. Expression of [^{13}C, $^{2}H_3$]-^{19}F Trp CA2 and SOD1

Plasmids containing the coding sequences for carbonic anhydrase 2 (CA2, NP_000058.1) and superoxide dismutase 1 (SOD1, NP_000445.1) were previously obtained from the pHLsec vector by removal of its secretion sequence.[6] The plasmids were introduced in HEK293T cells (ATCC CTRL-3216) by transient transfection.[7] Cells were cultured in T75 flasks in Dulbecco's modified eagle medium (DMEM, Gibco). HEK293T were transfected with a mixture of 50 µg of polyethylenimine (Sigma Aldrich) and 25 µg of plasmid (2:1 ratio) after reaching 90-95% confluence. The transfected cells were cultured in DMEM with 2% v/v of fetal bovine serum (Life Technologies), 2 mM of glutamine (Life Technologies) and 100 µg/mL of penicillin/streptomycin (Life Technologies), in a sterile incubator at 37°C, 5% CO_2. After 24 h of incubation, a medium switch was performed with a custom-made DMEM, prepared according to the recipe of commercial DMEM but lacking L-tryptophan, and to which [^{13}C, $^{2}H_3$]-^{19}F Trp was added at a final concentration of 80 µM. The transfected cells were harvested after 48 h of protein expression, resuspended in 150 µL of phosphate saline buffer (Life Technology), lysed by freezing and thawing cycles in liquid nitrogen, and centrifuged at 16000 g for 1 h at 4°C. The supernatant containing the soluble components was collected, supplemented with 10% D_2O, and transferred into a 3 mm NMR tube.

3. Purification of [^{13}C, $^{2}H_3$]-^{19}F Trp CA2

[^{13}C, $^{2}H_3$]-^{19}F Trp CA2 was purified by affinity chromatography following an existing strategy.[8] HEK293T cells were used for protein expression as described above. The soluble lysates from three T75 flasks of transfected cells were collected, pooled, and diluted in 3 mL of binding buffer (20 mM Tris buffer, pH 8). The solution was then loaded onto a 1 mL NiNTA column, previously equilibrated with binding buffer. After three washes with 1 mL of binding buffer, the protein was eluted in 3 mL steps of buffer with increasing concentration of imidazole. An SDS-PAGE of the solution samples confirmed that CA2 was eluted at high purity with 20 mM of imidazole. The buffer of the purified ^{13}C-^{19}F Trp CA2 was exchanged to 500 µL of 10 mM HEPES pH 6.8. The protein sample was supplemented with 10% D_2O, reaching a final concentration of 180 µM, and transferred in a 5 mm NMR tube.

4. NMR experiments

^{13}C-^{19}F TROSY NMR spectra were recorded on a 14.1 T (600 MHz ^{1}H, 564.6 MHz ^{19}F) Bruker Avance NEO spectrometer equipped with a QCI-F cryoprobe. The ^{13}C-start, ^{19}F-detect TROSY pulse program from Boeszoermenyi et al. was employed.[9] Lysate samples were analyzed at 310 K with a series of ~3.75 h-long spectra recorded with 16 scans and 512 increments each, acquisition times of 69 ms (^{19}F) and 68 ms (^{13}C), and an interscan delay of 1.5 s. The spectra were processed in TopSpin 4.4.0 (Bruker) with sine squared window function and inspected to exclude protein degradation occurring over time. The first eight spectra were summed together. Purified [^{13}C, $^{2}H_3$]-^{19}F Trp CA2 was analyzed at 298 K. A ~9.75 h-long ^{13}C-^{19}F TROSY spectrum was recorded with 32 scans and 512 increments, acquisition times of 69 ms (^{19}F) and 68 ms (^{13}C), and an interscan delay of 2 s, and processed as above. The chemical shifts were calibrated at each temperature with an external reference, trifluoroacetic acid for ^{19}F (set at -76.55 ppm), DSS for ^{13}C (indirectly referenced from ^{1}H).

5. Supplementary figures

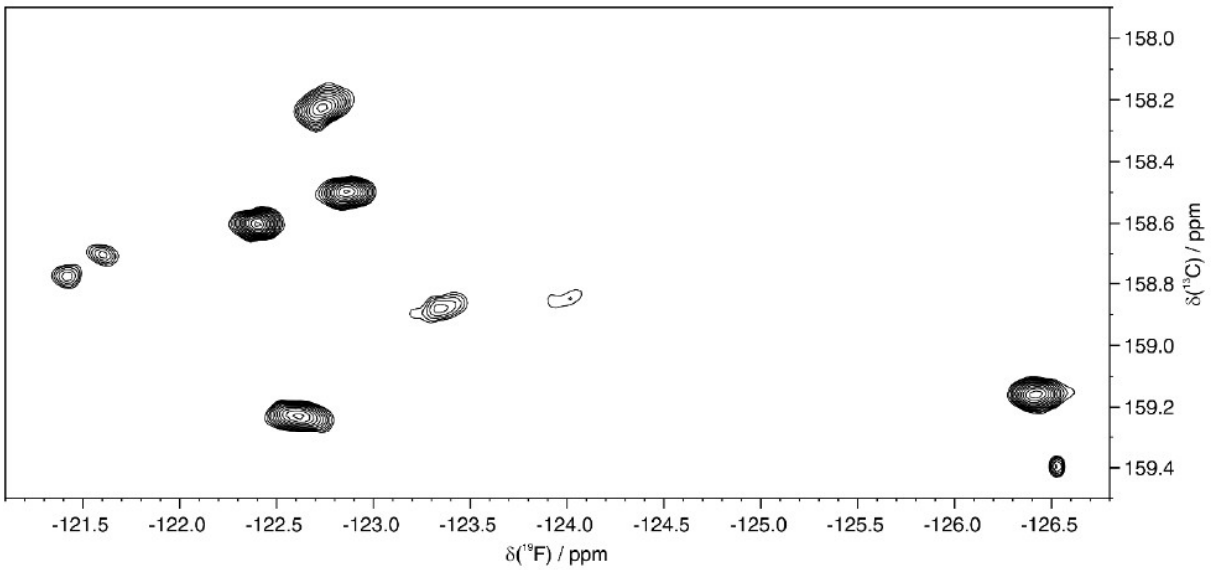


Figure S1: ^{13}C-^{19}F TROSY spectrum of purified [^{13}C, $^{2}H_3$]-^{19}F Trp CA2. The signal splitting at -121.5 ppm is attributed to incomplete fluorotryptophan incorporation. The sharp signal at -126.5 ppm for ^{19}F and 159.4 ppm for ^{13}C is likely due to residual free fluorotryptophan that was initially bound to the protein and subsequently released into the solution.

6. References

1) R. J. Lichtenecker, *Org. Biomol. Chem*., 2014, **12**, 7551.
2) T. Fujimoto T. Ritter, *Org. Lett.,* 2015, **17**, 3, 544.
3) G. Toscano, J. Holzinger, B. Nagl, G. Kontaxis, H. Kählig, R. Konrat, R. J. Lichtenecker, *J Biomol* NMR, 2024, https://doi.org/10.1007/s10858-024-00440-z.
4) L. Gurskaya, L. Politanskaya, J. Wang, P. Ilyina, A. Volobueva, V. Zarubaev, *J. Fluor. Chem.,* 2024, **274**, 110240.
5) C. Ma, X. Liu, X. Li, J. Filippen-Anderson, S. Yu, J. M. Cook, *J. Org. Chem*, 2001, **66**, 4525.
6) a) L. Banci, L. Barbieri, I. Bertini, E. Luchinat, E. Secci, Y. Zhao, A. R. Aricescu, *Nat. Chem. Biol.*, 2013, **9**, 297; b) E. Luchinat, L. Barbieri, M. Cremonini, A. Nocentini, C. T. Supuran, L. Banci, *Angew. Chem. Int. Ed.*, 2020, **59**, 16, 6535.
7) L. Barbieri, E. Luchinat, L. Banci, *Nat. Protoc.*, 2016, **11**, 1001.
8) A. L. Banerjee, M. Swanson, S. Mallik, D. K. Srivastava, *Protein Expr. Purif.*, 2004, **37**, 2, 450.
9) A. Boeszoermenyi, S. Chhabra, A. Dubey, D. L. Radeva, N. T. Burdzhiev, C. D. Chanev, O. I. Petrov, V. M. Gelev, M. Zhang, C. Anklin, H. Kovacs, G. Wagner, I. Kuprov, K. Takeuchi, H. Arthanari, *Nature Methods*, 2019, **16**, 333.

5. Selective Labelling Strategies for in-cell NMR studies of intrinsically disordered proteins: the challenging case of BRCA1

5.1 Abstract

Isotopic labelling is a fundamental tool in NMR spectroscopy for structural characterization of proteins, yet studying intrinsically disordered proteins (IDPs) remains challenging due to low chemical shift dispersion, spectral overlap, and dynamic interactions. Here, we report optimized strategies for selective ^{15}N and ^{13}C isotopic labelling of BRCA1 fragments encompassing nearly its entire intrinsically disordered region (IDR), expressed in human HEK293T cells. Using a combination of amino acid-specific labelling and ^{13}C-labeled α-ketoacid precursors, we recorded simplified HMQC, HNCA, and HNCO spectra for both in-cell and in-lysate samples, enabling residue-specific observation while mitigating spectral crowding. We achieved selective labelling of leucine, histidine, cysteine, and phenylalanine residues, demonstrating efficient incorporation and favourable signal resolution, particularly in lysates. Furthermore, we established co-expression protocols for BRCA1 and its partner MAX and developed a monoclonal inducible BRCA1 cell line, allowing selective isotopic labelling of individual proteins to facilitate analysis of interactions in intact cells. Our results demonstrate that these selective labelling strategies, coupled with optimized experimental conditions and lysate stabilization protocols, provide a viable framework for NMR investigation of large IDPs and their interactions.

5.2 Introduction

Isotopic labelling represents a cornerstone of modern NMR spectroscopy and is an essential tool for structural studies of proteins both in vitro and in cellular environments. While uniform ^{15}N and ^{13}C labelling has long been the standard, its application to intrinsically disordered proteins (IDPs) is often limited as the lack of a stable three-dimensional structure in IDPs leads to poor chemical shift dispersion, especially in the proton dimension, resulting in severe spectral overlap, especially for long disordered regions, when employing uniform isotopic labelling. This feature is further complicated by the presence of repetitive sequences common in IDPs, which further reduce peak separation and make conventional ^{1}H, ^{15}N HSQC experiments less effective[117–119].

When studying interactions between IDPs and their partners using NMR, additional challenges arise. For example, the inherent conformational flexibility of IDPs leads to highly dynamic and heterogeneous complexes, which can cause exchange broadening, leading to weak or invisible NMR signals for the interacting regions, especially when partial folding or transient binding occurs[117,158]. Moreover, IDP complexes often display a range of binding modes, from folding-upon-binding to highly dynamic "fuzzy" complexes, making it difficult to capture a single structural state and complicating resonance assignment and interpretation[158–161]. Finally, the low chemical shift dispersion typical of IDPs increases spectral crowding, while the presence of multiple, weak, and transient interactions can hide specific binding events or lead to ambiguous results[117,162–164]. However, amongst the techniques employed in structural biology, NMR still remains, the most suited to investigate IDPs, as their intrinsic mobility makes X-ray crystallography and high-resolution cryo-electron microscopy non-viable.

One effective strategy to address the long-standing challenge of studying intrinsically disordered proteins is the use of selective isotopic labelling combined with heteronuclear detection schemes. The relevance of such methodological approach is particularly evident when studying proteins with extensive disordered regions that mediate critical biological functions. A prime example is the breast cancer susceptibility protein BRCA1. This protein is a large, multi-domain tumour suppressor protein essential for maintaining genomic stability, with critical roles in DNA repair, transcriptional regulation, and cell cycle control. Structurally, BRCA1 contains a well-characterized N-terminal RING (residues 1-103) and two C-terminal BCRT domains (residues 1646-1863), both of which are globular and mediate specific protein-protein interactions. However, the central region (residues 104-1645), is predominantly intrinsically disordered, lacking stable secondary or tertiary structure under physiological conditions[165–168]. This intrinsically disordered region (IDR) is characterized by high flexibility and a large hydrodynamic radius[165–168]. Functionally, the IDR of BRCA1 serves as a dynamic scaffold, enabling the protein to interact with a diverse array of partners, including DNA, and many gene expression-regulating proteins such as p53, MYC, MAX and PALB2, often with affinities in the low micromolar range[166,168,169]. The flexibility of the IDR allows BRCA1 to mediate multiple, sometimes simultaneous, protein-protein and protein-DNA interactions, which is crucial for integrating cellular signals in the DNA damage response pathway[166–168,170].

Many functions of BRCA1 involve the interaction of the IDR: firstly DNA repairing via formation of "fuzzy" complexes with DNA[170], then facilitating histone ubiquitylation and chromatin remodelling, necessary for DNA repair and cell survival, via recruitment of BRCA1/BARD1 complexes to chromatin and sites of DNA damage[171], and finally modulating transcription of proto-oncogenic (Myc) and tumour suppressing (p53) genes[168,169]. Another transcriptional regulation pathway involves the formation of the BRCA1-Myc and BRCA1-Myc-MAX complexes that are responsible for regulation of MYC-driven cell proliferation, cell growth and differentiation[169,172–174]. Recently a new interaction has been discovered amongst BRCA1 and MAX that does not involve the participation of Myc. The BRCA1-MAX complex, particularly with the monomeric form of MAX, may modulate the availability of MAX for Myc-MAX heterodimer formation, indirectly influencing Myc-dependent gene expression and cellular proliferation[169]. The biological pathways arising from this interaction remain unclear and need to be further investigated.

In this preliminary work, we optimized protocols for selective isotopic labelling of four BRCA1 fragments, covering nearly the entire IDR, expressed in human HEK293T cells. The labelling strategies combined amino acid-selective ^{15}N, ^{13}C labelling schemes using both labelled amino acids and ^{13}C-labelled α-ketoacid precursors, resulting in uniformly ^{15}N, ^{13}C-labelled histidines and cysteines, $^{13}C_{\alpha}$-labelled leucines, and $^{13}C'$-labelled phenylalanines. These approaches enabled simplified spectral readouts of HMQC, HNCA and HNCO experiments for in-cell and in-lysate samples, paving the way for interaction investigations. Furthermore, we demonstrated the feasibility of co-expressing BRCA1 with its partner MAX in mammalian cells, and we established a monoclonal inducible stably transfected BRCA1 cell line to allow selective labelling of individual proteins in co-expression experiments, to further alleviate the spectral crowding issue.

5.3 Results and discussion:

5.3.1 Expression, stabilization and assessment of experimental conditions:

The first major challenge in this project was the expression of the BRCA1 protein. We approached this by expressing four different fragments (**Figure 5.3.1-1**), all located within the intrinsically disordered region of the protein: BRCA1 219-504, BRCA1 495-685, BRCA1 675-1357, and BRCA1 219-1357. Among these fragments, the first one (219-504) has already been assigned[168] and its role in interacting with the Myc-associated factor X (MAX) has been documented[169] and the last one (219-1357) with its 1139 amino acids-long sequence, spans almost the entire IDR of BRCA1. Due to its length, the fourth fragment represents a significant challenge, both in terms of expression yield and spectral complexity. The four constructs were expressed in mammalian cells using a protocol previously established by Barbieri et al.[7], which involves the transient transfection of HEK293T cells with a mixture of plasmid DNA (pHL-sec) and polyethyleneimine (PEI). All four fragments were successfully overexpressed, as shown by the corresponding bands observed on the SDS-PAGE gel (**Figure 5.3.1-2**). To our knowledge, BRCA1 219-1357 represents the largest BRCA1 fragment ever expressed to date.

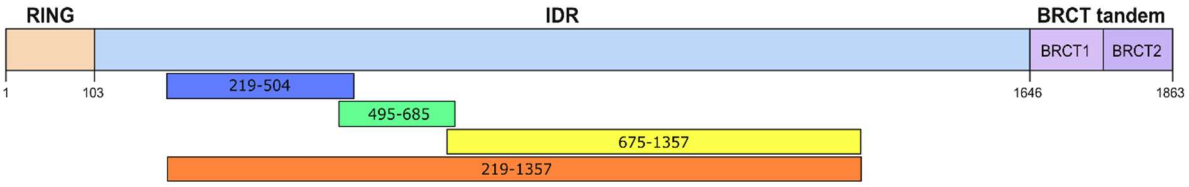


Figure 5.3.1-1
Schematic representation of BRCA1 sequence: *structural domains represented in pink, light blue and purple, expressed fragments represented in blue, green, yellow and orange.*

We investigated these constructs via NMR both in cells, which allows preservation of the physiological environment as well as all native protein interactions, and in cell lysates. The latter proved to be a valuable alternative to purification, as it simplifies the spectra compared to in-cell experiments while avoiding the need for scale-up, which in mammalian systems would be costly and labour-intensive. The experimental conditions for in-cell NMR investigations were then optimized to maximise cell viability and protein integrity throughout the time required for NMR analysis. In order to do this we simulated an in-cell NMR experiment, by incubating HEK293T cells overexpressing the BRCA1 219-504 fragment in a 3 mm Shigemi NMR tube for up to 4 hours at different temperatures (t 4, 15, 20, 25, and 37 °C) to assess cell viability (Figure 5.3.1-2; C) and protein integrity (Figure 5.3.1-3; A, B) during the in-cell NMR experiment. A second critical parameter was the determination of the temperature that yielded the highest spectral quality. To this end, we conducted a temperature scan from 5 to 37°C, by acquiring 1H15N HMQC spectra on a 900MHz magnet, on a 15N uniformly labelled BRCA1 219-504 in-cell sample (**Figure 5.3.1-4**).

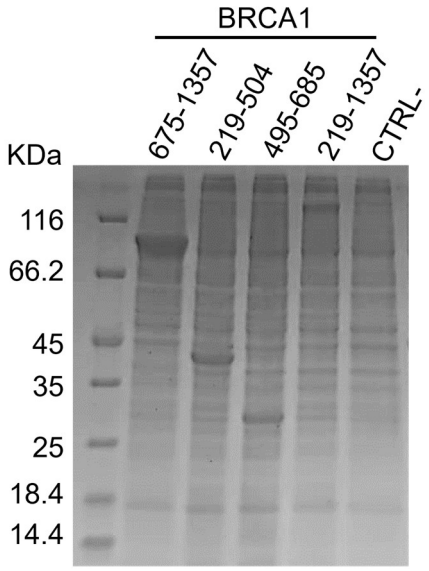


Figure 5.3.1-2
Expression of BRCA1 fragments: *HEK293T transfected with a pHL-sec plasmid containing a sequence coding for BRCA1 675-1357, BRCA1 219-504, BRCA1 495-685, BRCA1 219-1357, a non-coding pHL-sec (empty pHL-sec) as negative control (CTRL-).*

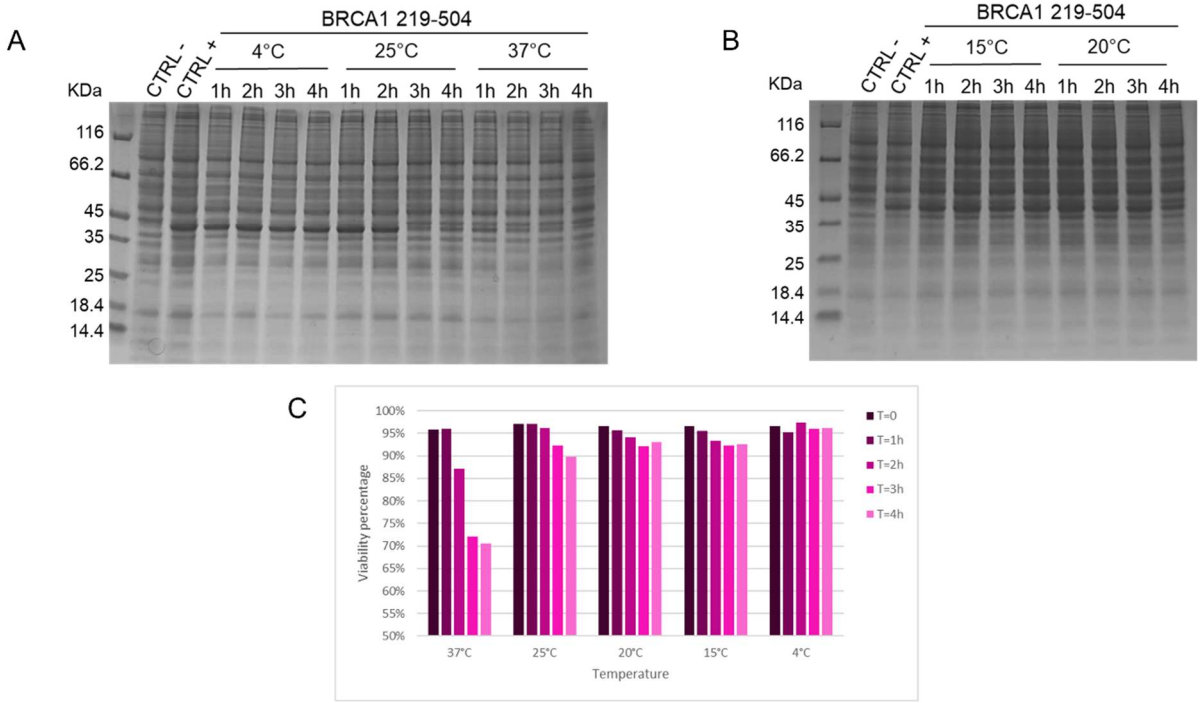


Figure 5.3.1-3
Cell viability and protein integrity assay: *(**A, B**) SDS- PAGE of lysates of HEK293T cells transfected with pHL-sec coding for BRCA1 219-504, incubated in a 3mm Shigemi tube for 1, 2, 3, 4 hours at 4, 15, 20, 25, 37 °C, in the negative control (CTRL-) HEK293T were transfected with empty pHL-sec, in the positive control (CTRL+) HEK293T were transfected with pHL-sec coding for BRCA1 219-504 and immediately lysed without incubation. (**C**) Viability percentages of HEK293T cells transfected with pHL-sec coding for BRCA1 219-504 before incubation in a 3mm Shigemi (T=0) and after incubation for 1h (T=1h), 2h (T=2h), 3h (T=3h) and 4h (T=4h) at 4, 15, 20, 25, 37°C.*

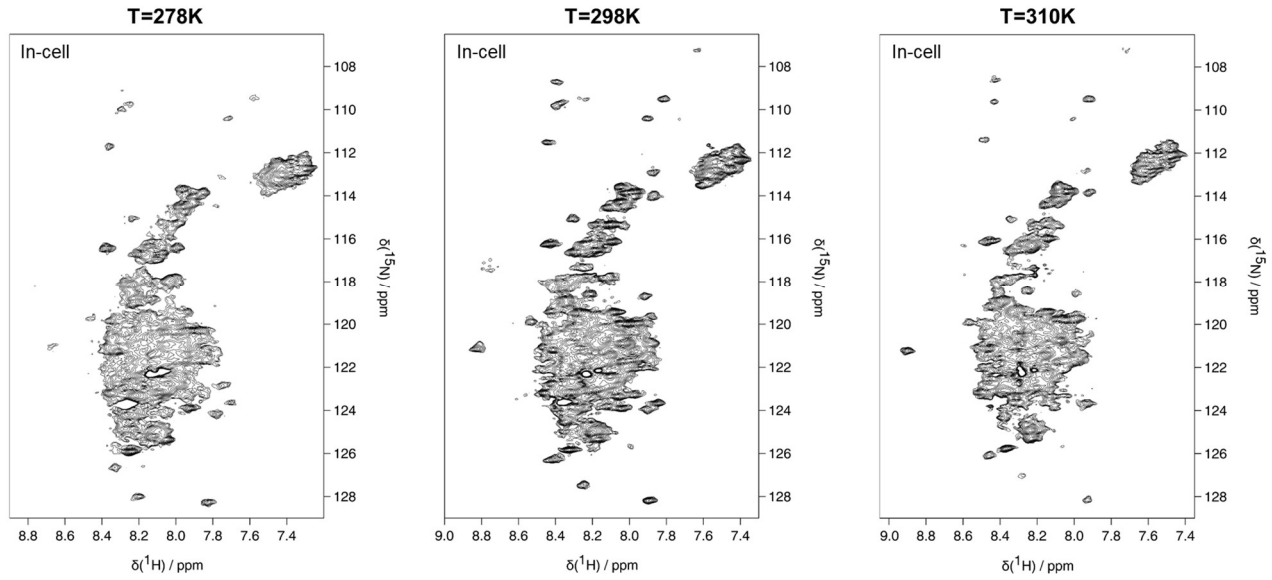


Figure 5.3.1-4
In-cell temperature scan on BRCA1 219-504: *In-cell ^{1}H, ^{15}N HMQC of HEK293T cells overexpressing BRCA1 219-504 at 278, 298 and310 K with background subtraction.*

The analysis of in-cell spectral quality and protein stability at different temperatures revealed two opposing trends. While increasing the temperature generally enhanced spectral quality primarily due to the faster protein dynamics, at too high temperatures the increased exchange rate of the amide protons and the accelerated protein degradation led to a decrease of signal intensity. Conversely, lowering the temperature improved protein stability over time but compromised spectral resolution, possibly due to the stabilization of weak interactions with the cellular environment. Based on these observations, analysis at 20 °C offered the best compromise between cell viability (up to 4 hours), (**Figure 5.3.1-3**, **C**), protein integrity (up to 3 hours), (**Figure 5.3.1-3**, **B**), and spectral quality (**Figure 5.3.1-4**).

For lysate experiments, we also maintained the temperature at 20 °C and supplemented the samples with a protease inhibitor to further improve protein stability over time. Additionally, we implemented a lysate "clean-up" protocol to prepare for future, more complex experiments that will require longer acquisition times and therefore increased protein stability. This protocol involves boiling the diluted lysate at 80 °C for 20 minutes, followed by filtration using a Centricon centrifugal filter with a 30 kDa molecular weight cut-off. These steps are commonly used in purification protocols for intrinsically disordered proteins and help to remove larger cellular components such as structured proteins, enzymes, and organelles[175–177].

5.3.2 Labelling

NMR spectra of long IDPs are typically challenging to interpret due to the large number of resonances and the inherently low chemical shift dispersion of their peaks[117–119]. To overcome this limitation, we employed various amino acid-selective labelling strategies aimed at simplifying spectral analysis. For optimal peak separation in very long constructs, uniform ^{15}N,^{13}C labelling of the protein backbone is not ideal, in addition to being prohibitively expensive when expressing proteins in mammalian cells. Therefore, we sought to establish amino acid- and backbone-selective ^{15}N,^{13}C labelling. However, selective isotope labelling of single backbone atoms for certain amino acids can still be extremely expensive or, in some

cases, not easily achieved due to lack of commercially available compounds. Additionally, ^{15}N scrambling can hinder the selective labelling of specific residues, rendering backbone-specific labelling impossible[83,178,179]. To address these challenges, and as previously demonstrated (**chapter 0**), protein expression in mammalian cells can be performed using the corresponding α-ketoacid analogues of the desired amino acids. These α-ketoacids, also referred to as precursors, can be added directly to the cell culture medium. Human cellular transaminases then convert the precursors into their corresponding amino acids, offering a cost-effective method for introducing ^{13}C-labeled residues to the expressed protein[37]. However, since α-ketoacids lack the amide nitrogen, they can only provide ^{13}C labelling.

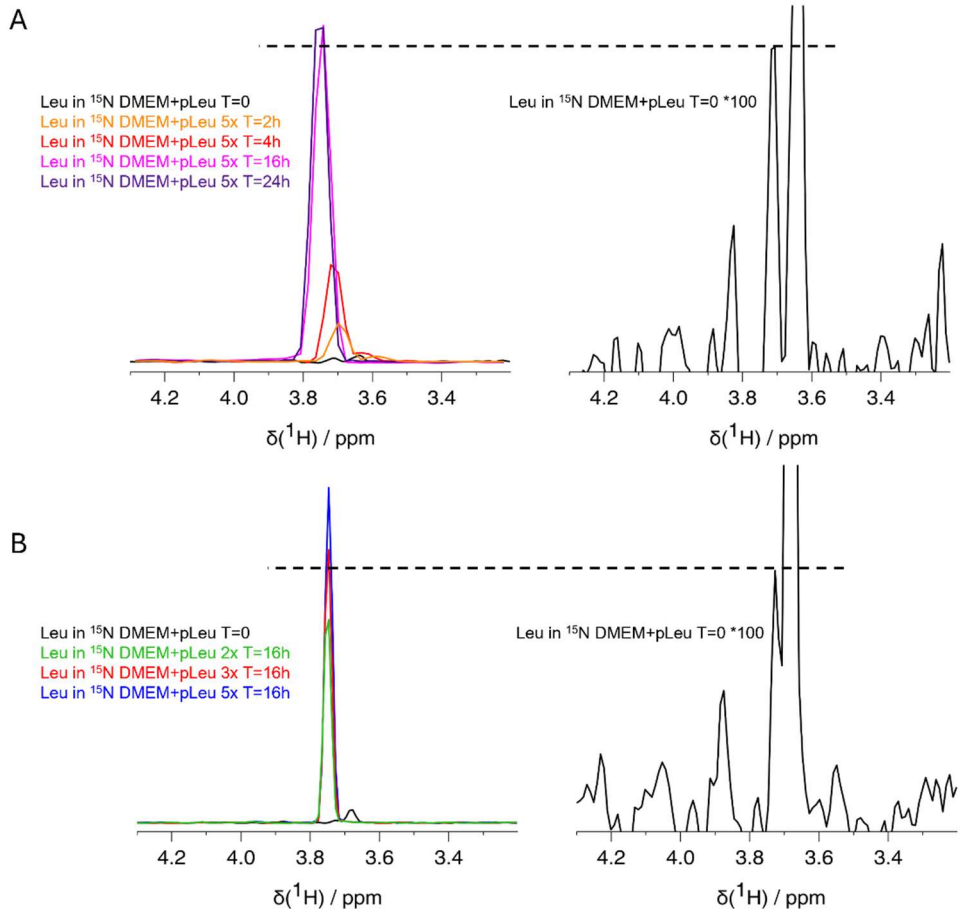


Figure 5.3.2-1
Evaluation of the conversion level of $^{13}C_{\alpha}$-pLeu into ^{15}N, $^{13}C_{\alpha}$ Leu: *(**A**) Left: ^{1}H plane of ^{1}H ^{13}C HSQC acquired on ^{15}N-labelled DMEM samples supplemented with 5x $^{13}C_{\alpha}$-pLeu and incubated on HEK293T cells for different time intervals (0, 2, 4, 16, and 24 hours). Right: ^{1}H plane of ^{1}H ^{13}C HSQC acquired on ^{15}N-labelled DMEM, detecting carbon signals from unlabelled Leu present in the medium and scaled to account for natural abundance. (**B**) Left: ^{1}H plane of ^{1}H ^{13}C HSQC acquired on ^{15}N-labelled DMEM samples supplemented with different concentrations of $^{13}C_{\alpha}$-pLeu (2x, 3x, 5x) and incubated on HEK293T cells for 16 hours. Right: ^{1}H plane of ^{1}H ^{13}C HSQC acquired on ^{15}N-labelled DMEM, detecting carbon signals from unlabelled Leu present in the medium and scaled to account for natural abundance.*

To achieve double labelling (^{15}N,^{13}C), the precursors must be supplied alongside a commercial uniformly ^{15}N labelled culture medium, which already contains the ^{15}N-labelled form of the amino acid of interest. We then developed a protocol to obtain a ^{15}N-labelled medium containing the double labelled amino acid of interest, that involves the incubation of non-transfected HEK293T cells with a uniformly ^{15}N-labelled medium, supplied with the ^{13}C-labelled precursor. These cells are exploited to synthetize the ^{15}N,^{13}C amino acid. Once the

medium is enriched in the doubly labelled amino acid it is collected and provided to the transfected cells for protein expression. To increase the incorporation of the double labelled amino acid over the ^{15}N labelled counterpart already present in the medium, an excess of precursor is provided in the first step. We tested different incubation times and precursor concentration with a $^{13}C_{\alpha}$-labelled leucine precursor, and we found that suppling a 5-fold molar excess of the labelled α-ketoacid (with respect to the leucine already present in the medium), and incubating the medium on non-transfected cells overnight (~16 h), yielded a culture medium with a concentration of ^{15}N,$^{13}C_{\alpha}$-Leu comparable to that of ^{15}N-Leu (**Figure 5.3.2-1)**. The selective labelling of BRCA1 was then tested using these conditions.

$^{13}C_{\alpha}$ leucine precursor labelling: in disordered proteins, leucine residues are often found in regions which interact with protein partners, making them a useful NMR reporter of such interactions[180,181]. Sequence analysis of the BRCA1 219-1357 fragment revealed the presence of 90 leucine residues (**Figure 5.3.2-2**). To pursue selective labelling on Leu residues, we employed the previously mentioned $^{13}C_{\alpha}$ leucine precursor and prepared the NMR samples following the Barbieri et al protocol[7], implementing the ^{15}N culture medium incubation described in the previous paragraph. We then recorded 2D ^{1}H-^{15}N planes of a standard HNCA experiment on both cell and lysate samples overexpressing BRCA1 219-504 and on lysate samples overexpressing BRCA1 219-1357 (**Figure 5.3.2-3**). In these spectra, the $^{13}C_{\alpha}$ is used as a filter to detect only Leu residues and the following ones in the polypeptide chain (Leu+1). The obtained HNCA spectra contained several peaks, which were compatible with the reported assignments of Leu and Leu+1 residues. As observed from the BRCA1 219-504 spectra (**Figure 5.3.2-3**; **A, B**), many of these signals were only detected in the lysate spectrum, suggesting that the ones missing in the in-cell spectrum are broadened due to interactions with cellular components. Moreover, leucine-leucine (Leu-Leu) pairs are also present in the sequence, distributed throughout all the sequence (**Figure 5.3.2-2**). Selective observation of these pairs would result in a highly simplified NMR spectrum, especially critical when analysing the long BRCA1 constructs, where only ten peaks would be expected. Therefore, 2D planes of a double-quantum filtered HNCA (DQ-HNCA), which would only detect the amide peak of the second Leu in a Leu–Leu pair, was also recorded. The resulting spectra showed extremely weak, sometimes undetectable, Leu–Leu signals (**Figure 5.3.2-3**), likely as a result of the incomplete incorporation of the double labelled amino acid into the protein. Indeed, a ~40% incorporation efficiency of ^{13}C-labelled leucine was estimated from the standard HNCA spectra, which would result in a ~16% probability of two adjacent leucine residues being labelled. Such a low probability was particularly evident in the in-cell spectra, where no Leu-Leu peaks could be detected. Therefore, while the $^{13}C\alpha$ leucine precursor labelling protocol is effective for single-residue labelling, it does not provide sufficient incorporation levels for DQ-HNCA experiments. In principle, detection of Leu-Leu signals may be achieved at a higher cost by supplementing an excess ^{15}N,^{13}C-L-leucine to the ^{15}N-medium, resulting in higher incorporation efficiency. Nonetheless, the detection of Leu and Leu+1 in standard HNCA spectra was deemed highly promising. In particular, the lysates spectra, especially on the smaller construct shows excellent signal intensity and peak dispersion, indicating the viability of leucine-selective labelling for interaction studies.

DSAKKAACEFSETDVTNTEHHQPSNNDLNTTEKRAAERHPEKYQGSSVSNLHVEPCGTNTHASSLQHENSSLLLTKDRMNVEKAEFCNKSKQPGLARSQHNRWAGSKETCN
DRRTPSTEKKVDLNADPLCERKEWNKQKLPCSENPRDTEDVPWITLNSSIQKVNEWFSRSDELLGSDDSHDGESESNAKVADVLDVLNEVDEYSGSSEKIDLLASDPHEALIC
KSERVHSKSVESNIEDKIFGKTYRKKASLPNLSHVTENLIIGAFVTEPQIIQERPLTNKLKRKRRPTSGLHPEDFIKKADLAVQKTPEMINQGTNQTEQNGQVMNITNSGHENKTK
GDSIQNEKNPNPIESLEKESAFKTKAEPISSSISNMELELNIHNSKAPKKNRLRRKSSTRHIHALELVVSRNLSPPNCTELQIDSCSSSEEIKKKKYNQMPVRHSRNLQLMEGKEP
ATGAKKSNKPNEQTSKRHDSDTFPELKLTNAPGSFTKCSNTSELKEFVNPSLPREEKEEKLETVKVSNNAEDPKDLMLSGERVLQTERSVESSSISLVPGTDYGTQESISLLEVS
TLGKAKTEPNKCVSQCAAFENPKGLIHGCSKDNRNDTEGFKYPLGHEVNHSRETSIEMEESELDAQYLQNTFKVSKRQSFAPFSNPGNAEEECATFSAHSGSLKKQSPKVTFE
CEQKEENQGKNESNIKPVQTVNITAGFPVVGQKDKPVDNAKCSIKGGSRFCLSSQFRGNETGLITPNKHGLLQNPYRIPPLFPIKSFVKTKCKKNLLEENFEEHSMSPEREMGN
ENIPSTVSTISRNNIRENVFKEASSSNINEVGSSTNEVGSSINEIGSSDENIQAELGRNRGPKLNAMLRLGVLQPEVYKQSLPGSNCKHPEIKKQEYEEVVQTVNTDFSPYLISDN
LEQPMGSSHASQVCSETPDDLLDDGEIKEDTSFAENDIKESSAVFSKSVQKGELSRSPSPFTHTHLAQGYRRGAKKLESSEENLSSEDEELPCFQHLLFGKVNNIPSQSTRHST
VATECLSKNTEENLLSLKNSLNDCSNQVILAKASQEHHLSEETKCSASLFSSQCSELEDLTANTNTQDPFLIGSSKQMRHQSESQGVGLSDKELVSDDEERGTGLEENNQE

Figure 5.3.2-2
Leucines in BRCA1 219-1357 sequence: *BRCA1 219-504 protein sequence with highlighted leucines (yellow) and leucines pairs (magenta).*

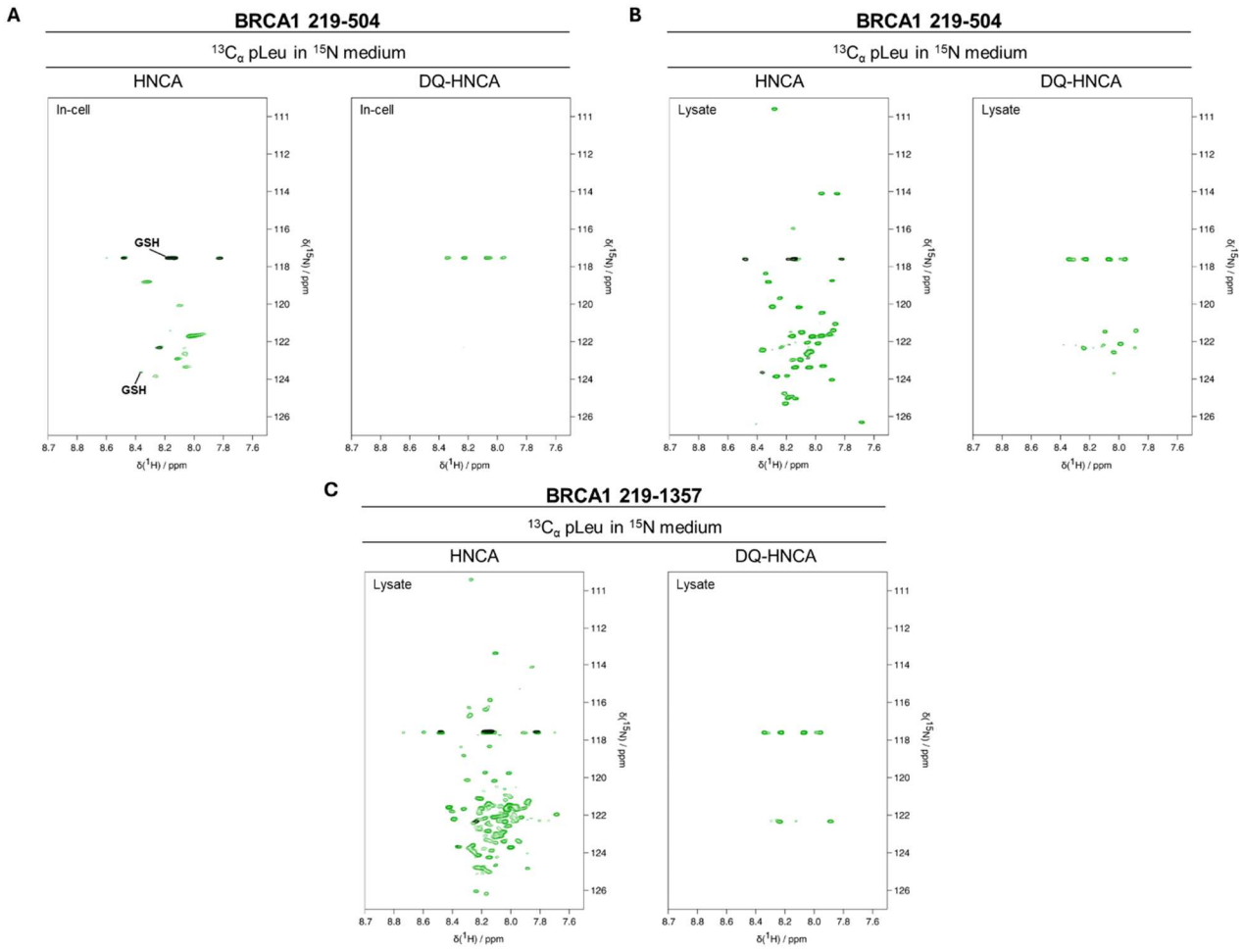


Figure 5.3.2-3
Leucine precursor labelling: ***(A)*** *Left: in-cell HNCA on HEK293T transfected with BRCA1 219-504 (green) and empty pHL-sec (black) in presence of ^{15}N uniformly labelled culture medium supplied with $^{13}C_\alpha$ leucine precursor 5x (green); Right: in-cell double quantum filtered HNCA on HEK293T transfected with BRCA1 219-504 (green) in presence of ^{15}N uniformly labelled culture medium supplied with $^{13}C_\alpha$ leucine precursor 5x.* ***(B)*** *Left: HNCA on lysate of HEK293T transfected with BRCA1 219-504 (green) and empty pHL-sec (black) in presence of ^{15}N uniformly labelled culture medium supplied with $^{13}C_\alpha$ leucine precursor 5x; Right: double quantum filtered HNCA on lysate of HEK293T transfected with BRCA1 219-504 (green) in presence of ^{15}N uniformly labelled culture medium supplied with $^{13}C_\alpha$ leucine precursor 5x.* ***(B)*** *Left: HNCA on lysate of HEK293T transfected with BRCA1 219-1357 (green) and empty pHL-sec (black) in presence of ^{15}N uniformly labelled culture medium supplied with $^{13}C_\alpha$ leucine precursor 5x; Right: double quantum filtered HNCA on lysate of HEK293T transfected with BRCA1 219-1357 (green) in presence of ^{15}N uniformly labelled culture medium supplied with $^{13}C_\alpha$ leucine precursor 5x.*

^{15}N,^{13}C histidine and cysteine labelling: We then extended the amino acid-selective labelling strategy to histidine and cysteine residues. In this case, ^{15}N,^{13}C uniformly labelled L-histidine and L-cysteine amino acids were employed. Two distinct labelling schemes were applied. In the first scheme, BRCA1 was expressed in a custom-made unlabelled DMEM medium lacking the amino acid of interest, which was then supplemented with its ^{15}N,^{13}C-labelled counterpart.

This strategy allowed the acquisition of ^{15}N SOFAST-HMQC spectra in which only signals from the labelled amino acids present in the BRCA1 sequence were observed. Conversely, in the second protocol, expression was carried out in a commercial uniformly labelled ^{15}N medium, which was supplemented with the ^{15}N,^{13}C-labelled amino acid. Under these conditions, all protein residues incorporate ^{15}N, but only the supplemented amino acid is ^{15}N, ^{13}C-labelled. Supplementing the medium with a 5-fold molar excess of the double-labelled amino acid ensures a ratio of single to double-labelled amino acid of 1:5, yielding an expected incorporation efficiency of the double-labelled amino acid of ~80%. This labelling scheme allows for the acquisition of 2D ^{1}H-^{15}N planes of HNCA, where both the double-labelled residue (i) and its immediate neighbour (i+1) are observed, as well as 2D ^{1}H-^{15}N planes of HNCO, in which only the residue (i+1) is observed. Given the results obtained with the Leu precursor, we focussed on lysate samples for testing these labelling schemes.

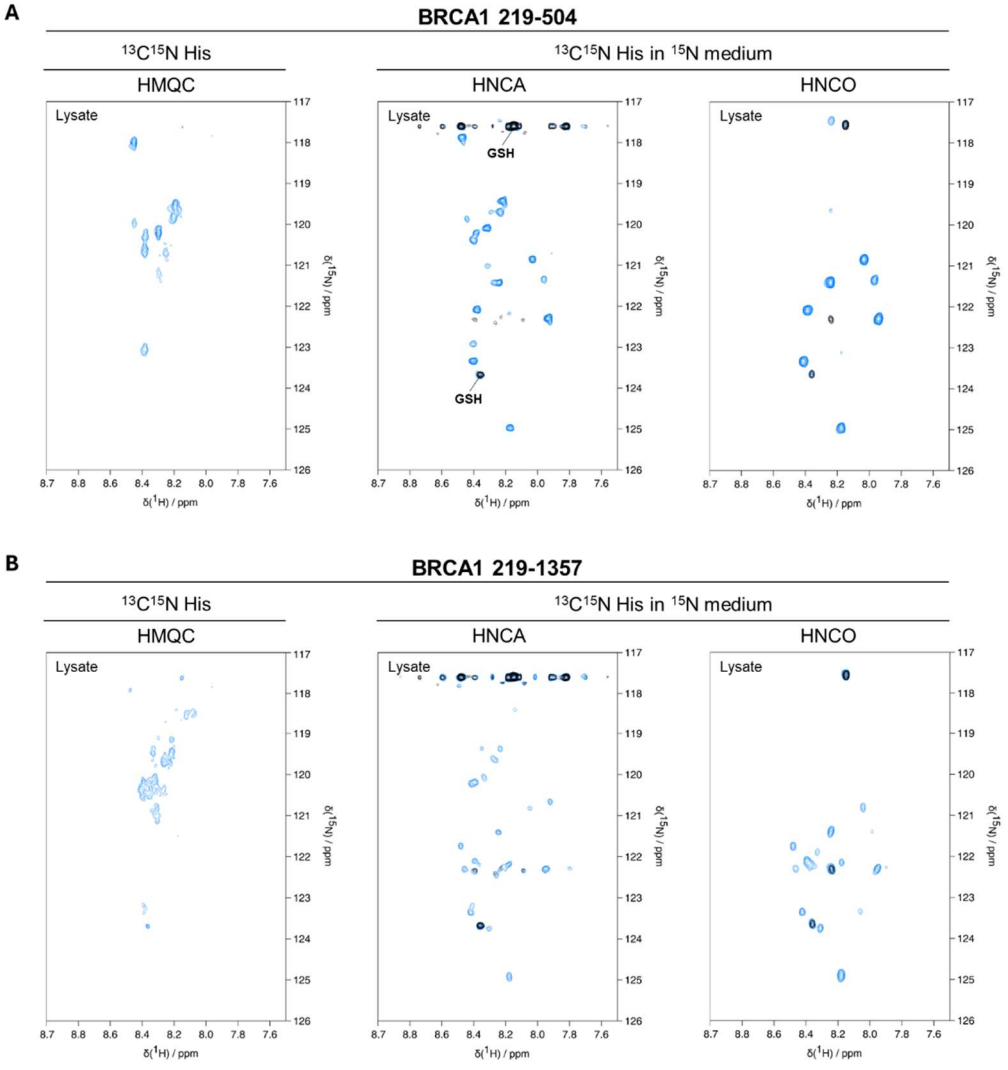


Figure 5.3.2-4
Histidine labelling: *(**A**) Left: HMQC on lysate of HEK293T transfected with BRCA1 219-504 (blue) and empty pHL-sec (black) in presence of custom made DMEM containing ^{15}N, ^{13}C histidine; Middle: HNCA on lysate of HEK293T transfected with BRCA1 219-504 (blue) and empty pHL-sec (black) in presence of ^{15}N uniformly labelled culture medium supplied with ^{15}N, ^{13}C histidine 5x. Right: HNCO on lysate of HEK293T transfected with BRCA1 219-504 (blue) and empty pHL-sec (black) in presence of ^{15}N uniformly labelled culture medium supplied with ^{15}N, ^{13}C histidine 5x. (**B**) Left: HMQC on lysate of HEK293T transfected with BRCA1 219-1357 (blue) and empty pHL-sec (black) in presence of custom made DMEM containing ^{15}N, ^{13}C histidine; Middle: HNCA on lysate of HEK293T transfected with BRCA1 219-1357 (blue) and empty pHL-sec (black) in presence of ^{15}N uniformly labelled culture medium supplied with ^{15}N, ^{13}C histidine 5x. Right: HNCO on lysate of HEK293T*

transfected with BRCA1 219-1357 (blue) and empty pHL-sec (black) in presence of ^{15}N uniformly labelled culture medium supplied with ^{15}N, ^{13}C histidine 5x.

Examination of the spectra of BRCA1 labelled with ^{15}N, ^{13}C histidine, clearly indicates that amino acid incorporation was successful, leading to improved peak dispersion (**Figure 5.3.2-4**). In particular, the HNCO spectrum of the short fragment (BRCA1 219-504) displays intense, well-resolved signals, with 8 out of the 10 expected peaks being observed (**Figure 5.3.2-4**; **A**). Indeed, the fragment sequence contains 11 histidines, one of which is followed by a proline residue.

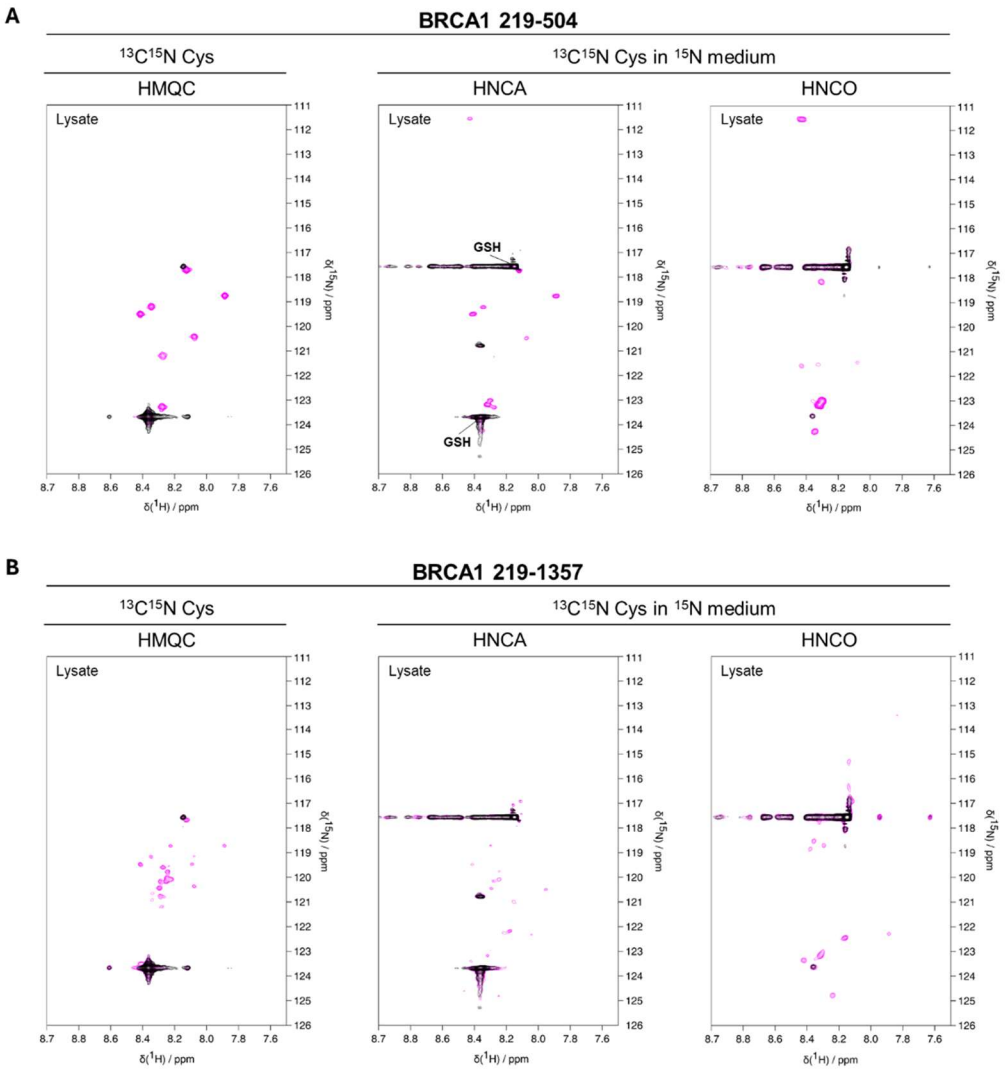


Figure 5.3.2-5
Cysteine labelling: *(**A**) Left: HMQC on lysate of HEK293T transfected with BRCA1 219-504 (magenta) and empty pHL-sec (black) in presence of custom made DMEM containing ^{15}N, ^{13}C cysteine; Middle: HNCA on lysate of HEK293T transfected with BRCA1 219-504 (magenta) and empty pHL-sec (black) in presence of ^{15}N uniformly labelled culture medium supplied with ^{15}N, ^{13}C cysteine 5x. Right: HNCO on lysate of HEK293T transfected with BRCA1 219-504 (magenta) and empty pHL-sec (black) in presence of ^{15}N uniformly labelled culture medium supplied with ^{15}N, ^{13}C cysteine 5x. (**B**) Left: HMQC on lysate of HEK293T transfected with BRCA1 219-1357 (magenta) and empty pHL-sec (black) in presence of custom made DMEM containing ^{15}N, ^{13}C cysteine; Middle: HNCA on lysate of HEK293T transfected with BRCA1 219-1357 (magenta) and empty pHL-sec (black) in presence of ^{15}N uniformly labelled culture medium supplied with ^{15}N, ^{13}C cysteine 5x. Right: HNCO on lysate of HEK293T transfected with BRCA1 219-1357 (magenta) and empty pHL-sec (black) in presence of ^{15}N uniformly labelled culture medium supplied with ^{15}N, ^{13}C cysteine 5x.*

Labelling with ^{15}N, ^{13}C cysteine also resulted in efficient incorporation and spectral simplification. In this case, the most well-resolved spectrum is the HMQC, in the short fragment, signals corresponding to all seven cysteine residues present in its sequence can be

observed (**Figure 5.3.2-5**). A drawback of cysteine labelling in cell lysates, however, is the occurrence of two very strong signals from glutathione (GSH), an abundant intracellular tripeptide containing cysteine, which may obscure protein peaks if located within the same spectral region.

Simultaneous labelling with $^{13}C_{\alpha}$ leucine precursor and $^{13}C'$ phenylalanine precursor: As a final labelling strategy, we combined two precursors with different labelling schemes: a $^{13}C'$-labelled phenylalanine precursor together with the previously tested $^{13}C_{\alpha}$-labelled leucine precursor. Labelling was carried out in ^{15}N uniformly enriched medium to label all the amino acids, including leucine and phenylalanine, on the amide nitrogen. To this end, we employed the previously described strategy of incubating ^{15}N medium supplemented with an excess of precursors on cells overnight, prior to using the same medium for transfection. This approach increases the concentration of doubly labelled amino acids in the medium and favours their incorporation over the corresponding ^{15}N-only species.

To assess the efficiency of precursor conversion under these conditions, we incubated the medium overnight and subsequently recorded 1D ^{13}C-detected spectra both on the unincubated medium (containing only precursor) and on the medium after incubation (**Figure 5.3.2-6**; **A**). To determine whether the presence of pLeu interfered with the conversion of pPhe, we also tested media containing both precursors in excess (**Figure 5.3.2-6**; **B**). From these spectra, we concluded that in the absence of pLeu approximately 80% of pPhe was converted into the amino acid, whereas in the presence of pLeu the conversion of pPhe efficiency decreased to 40%, nevertheless yielding a favourable ^{15}N, $^{13}C'$ Phe:^{15}N Phe ratio of 2:1. Conversely, by recording ^{1}H-^{13}C HSQC spectra we observed that the presence of pPhe in the culture medium did not affect the conversion rate of pLeu.

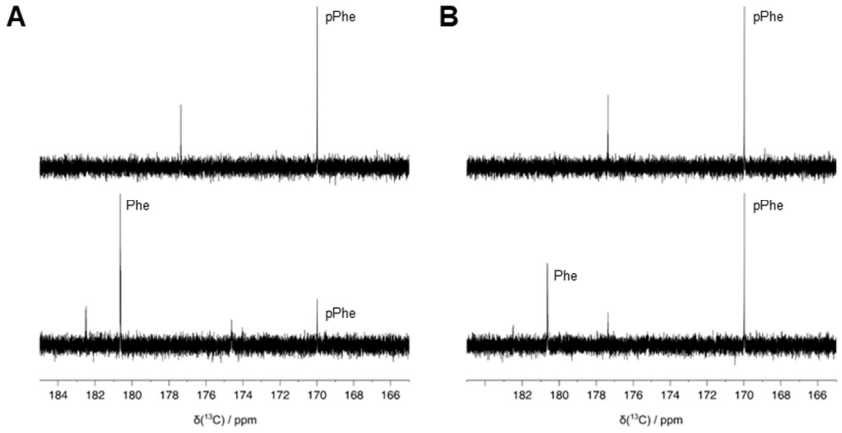


Figure 5.3.2-6
Evaluation of the conversion level of $^{13}C'$-pPhe into ^{15}N, $^{13}C'$ Phe: *(**A**) Top: ^{13}C detected 1D spectrum acquired on ^{15}N-labelled DMEM samples supplemented with 5x $^{13}C'$pPhe; Bottom: ^{13}C detected 1D spectrum acquired on ^{15}N-labelled DMEM samples supplemented with 5x $^{13}C'$-pPhe and incubated on HEK293T cells 16 hours. (**B**) Top: ^{13}C detected 1D spectrum acquired on ^{15}N-labelled DMEM samples supplemented with 5x $^{13}C'$pPhe and 5x $^{13}C_{\alpha}$pLeu ; Bottom: ^{13}C detected 1D spectrum acquired on ^{15}N-labelled DMEM samples supplemented with 5x $^{13}C'$pPhe and 5x $^{13}C_{\alpha}$pLeu and incubated on HEK293T cells 16 hours.*

We then applied this labelling strategy on a BRCA1 219-504 lysate sample, on which both HNCA and HNCO spectra were recorded, thereby demonstrating that simplified spectra

representative of two different residues can be obtained from the same sample. While the HNCA spectrum was essentially superimposable to the one obtained with the pLeu single labelling (see **Figure 5.3.2-3**), the HNCO spectrum displayed only two of the five expected peaks (**Figure 5.3.2-7**). This outcome may be attributed to local conformational or solvent exchange phenomena causing broadening and loss of some of the expected signals.

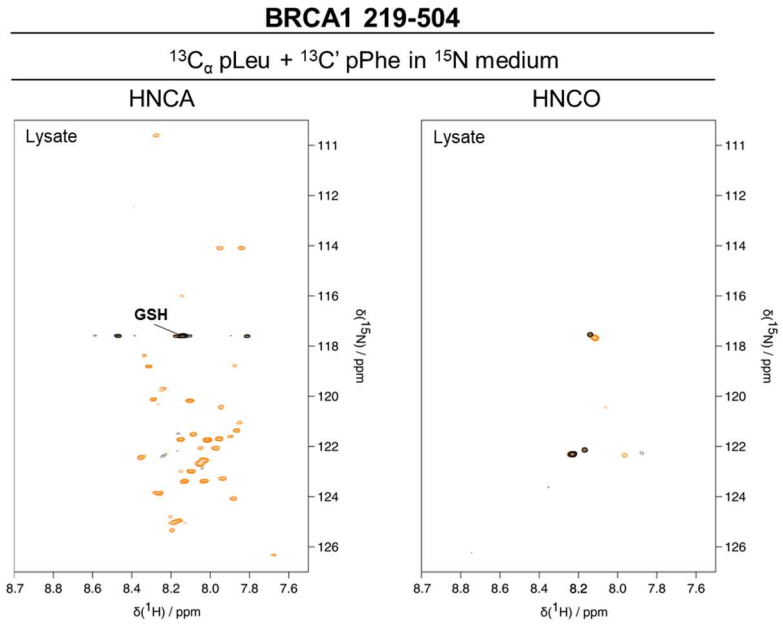


Figure 5.3.2-7
Simultaneous labelling with leucine and phenylalanine precursors: *(A) Left: HNCA on lysate of HEK293T transfected with BRCA1 219-504 (orange) and empty pHL-sec (black) in presence of ^{15}N uniformly labelled culture medium supplied with $^{13}C_\alpha$ leucine precursor 5x and $^{13}C'$ phenylalanine precursor 5x. Right: HNCO on lysate of HEK293T transfected with BRCA1 219-504 (orange) and empty pHL-sec (black) in presence of ^{15}N uniformly labelled culture medium supplied with $^{13}C_\alpha$ leucine precursor 5x and $^{13}C'$ phenylalanine precursor 5x.*

Overall, all labelling schemes tested yielded favourable outcomes in terms of efficiency of incorporation and spectral simplification. The somewhat limited resolution observed in certain spectra can be ascribed not only to the intrinsic sensitivity of the experimental setup but, more critically, to the protein concentration in the sample. This limitation is particularly pronounced for the longer BRCA1 fragment (219-1357), where the concentration is significantly lower than that of BRCA1 (219-504), and below those typically employed in NMR studies. In such instance, prolonged acquisition times will prove decisive for enhancing spectral quality. Consequently, future interaction studies will incorporate the previously described "clean-up" protocol, which has demonstrated excellent performance in preserving protein stability in the lysate for several days.

Finally, we note that the two signals arising from the amide groups of GSH, were detected in nearly all spectra, even in the absence of cysteine-specific labelling, suggesting that a residual level of ^{13}C scrambling invariably occurs in all labelling strategies (**Figure 5.3.2-3**, **Figure 5.3.2-4**, **Figure 5.3.2-5**, **Figure 5.3.2-7**).

5.3.3 Co-expression of BRCA1 and MAX

One of the aims of this project is to investigate whether the newly discovered interaction between BRCA1 and MAX occurs in human cells. To this aim, it's necessary to co-express both

proteins within a single cell sample. Two MAX constructs were tested: one corresponding to the full-length protein, and the other comprising only the BRCA1-interacting region (MAX fragment, residues 22–110). SDS-PAGE analysis revealed that, in both cases, the proteins were absent from the soluble fraction of the non-denaturing lysate, suggesting that the protein, which is expected to be localized in the nucleus, remains trapped in the insoluble fraction due to incomplete lysis of the nuclei and/or interactions with nuclear components, leading to its precipitation during the non-denaturing lysis process. To test this hypothesis, we performed separation of the nuclear and cytoplasmic compartments, which were then denatured and fully solubilized in RIPA buffer, followed by western blot analysis. The results revealed that, while MAX fragment was mainly localized in the nucleus, full-length MAX was localized in both the nuclear and the cytoplasmic fractions, where it is also interacting with non-soluble cytoplasmic components which are lost in non-denaturing lysis conditions (**Figure 5.3.3-1)**.

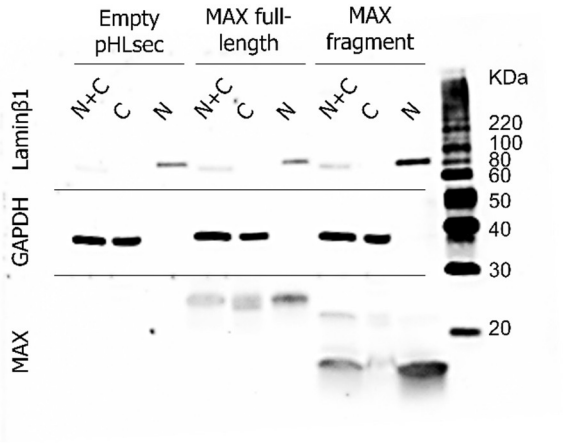


Figure 5.3.3-1
MAX expression: *western blot to assess nuclear and cytoplasmic localization of MAX full length and MAX fragment in HEK293T lysates.*

After confirming expression and localization of both MAX constructs, we proceeded to test the co-expression of full-length MAX with BRCA1 219-504. We transfected HEK293T cells with varying plasmid DNA ratios of the two proteins to identify the condition that yielded comparable expression levels. Once the optimal combination ratio was determined, we performed the in-cell NMR experiment for the co-expression of the two proteins in ^{15}N uniformly labelled medium, at 950 MHz and 20°C. The resulting in-cell NMR spectra showed signals corresponding to both proteins. Strikingly, several signals of MAX (both alone and together with BRCA1) are clearly detected in intact cells, despite the fact that the protein is extensively involved in interactions that prevent its solubilization in non-denaturing conditions. However, the significant spectral crowding prevented proper analysis, suggesting that an amino acid selective labelling such as those described above will be beneficial to detect any chemical shift perturbations that would indicate the occurrence of an interaction (**Figure 5.3.3-2**).

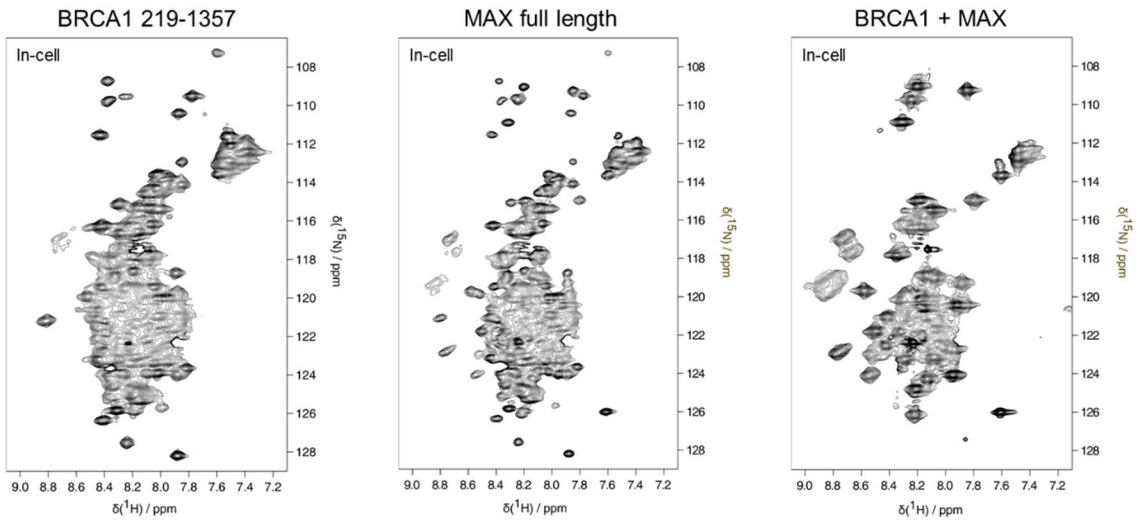


Figure 5.3.3-2
BRCA1 219-504 and MAX coexpression: *In-cell HMQC spectra of HEK293T spectra overexpressing BRCA1 219-504, MAX full length, BRCA1 219-504 and MAX full length simultaneously.*

5.3.4 BRCA1 monoclonal inducible cell line:

In addition to employing selective amino acid labelling strategies, we sought out a method to label only one of the two interacting proteins in order to address the issue of spectral crowding in in-cell NMR experiments involving both BRCA1 and MAX. To this aim, we adopted the PiggyBac transposon system, which enables the generation of stably transfected and inducible cell lines able to reach expression levels suitable for in-cell NMR detection, as previously demonstrated[52]. This method involves co-transfecting HEK293T-REx, a HEK293T stably transfected with the tetracycline repressor (Tet-repressor), with two plasmids: one encoding the transposase enzyme, and a second containing the GOI under the control of a tetracycline-responsive promoter (Tet-promoter), flanked by two inverted terminal repeat (ITR) sequences. Upon expression, the transposase identifies the ITRs and excises the GOI-containing fragment from the plasmid. The resulting fragment, containing the operator and the GOI, associates with the transposase to generate the transpososome complex, which in turn targets specific TTAA sites within the cell's genome, characterized by open chromatin and high expression, and randomly integrates the GOI in multiple copies. The GOI also contains the sequence encoding for GFP to facilitate assessment of successful gene integration. This method ensures high copy inducible expression of proteins. We applied this protocol to generate a stably transfected monoclonal HEK293T-REx cell line expressing BRCA1 219-1357. Achieving monoclonality ensures uniform protein expression across all cells, thus reducing sample heterogeneity during in-cell NMR analysis. High expression levels and monoclonality were confirmed by evaluating GFP expression using fluorescence-activated cell sorting (FACS) and immunofluorescence analysis (**Figure 5.3.4-1**). This newly established inducible monoclonal cell line expressing BRCA1 219-1357 allows the independent control of the induction and transfection times and, as a result, will enable in-cell NMR analysis of samples in which only one of the interacting proteins is isotopically labelled, significantly decreasing spectral complexity and facilitating the study of their interaction.

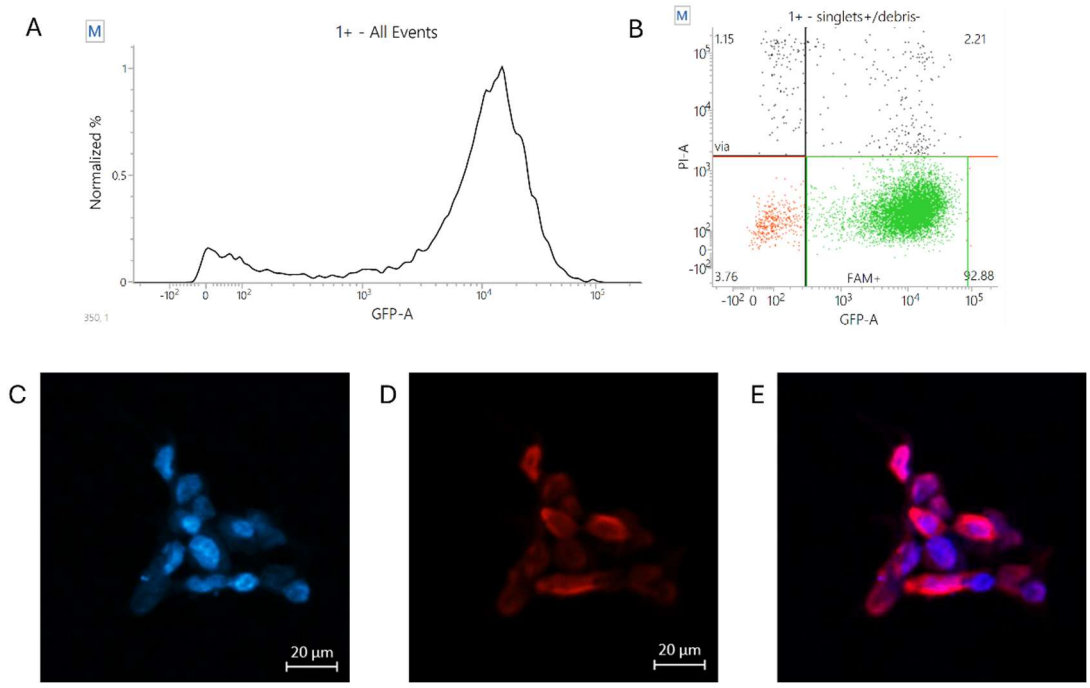


Figure 5.3.4-1
Stably transfected inducible BRCA1 219-1357 monoclonal cell line: *(**A**) Flow cytometry analysis of GFP fluorescence intensity from induced monoclonal HEK293T-Rex stably transfected with BRCA1 219-1357 (**B**) Flow cytometry analysis of GFP and PI fluorescence intensity from induced monoclonal HEK293T-Rex stably transfected with BRCA1 219-1357. (**C-E**) immunofluorescence analysis of fixed induced monoclonal HEK293T-Rex stably transfected with BRCA1 219-1357. (**C**) DAPI stained nuclei, (**D**) BRCA1 219-1357 stained with Alexa Fluor 594-conjugated secondary antibody, (**E**) Merge of C and D.*

5.4 Conclusions and perspectives:

In summary, this work establishes a set of selective labelling strategies for in-cell and in-lysate NMR investigations of intrinsically disordered proteins aimed to uncover interactions, with BRCA1 serving as a biologically relevant, challenging case study. By optimizing experimental conditions and implementing amino acid-specific labelling protocols, we demonstrated viable approaches to overcome the inherent difficulties of investigating IDPs via NMR. Specifically, we showed that optimized ^{15}N, ^{13}C amino acid-specific isotopic labelling schemes allow resolving spectral crowding in IDPs, which suffer from low chemical shift dispersion and overlapping peaks. Moreover, when labelling with an L-amino acid is not feasible, due to unavailability on the market, extremely high cost, or the occurrence of ^{15}N scrambling, we demonstrated that labelling can be achieved using ^{13}C-labelled α-ketoacid precursors. This approach offers a useful alternative to L-amino acid labelling, despite suffering from limited incorporation rate, which makes it unsuitable for low-sensitivity experiments. We also highlighted the importance of adjusting sample preparation and NMR acquisition parameters which, together with using high-field spectrometers, can significantly improve spectral quality and resolution for large or full-length IDPs. Importantly, we achieved expression of a very large fragment of the BRCA1 protein, spanning almost the entire IDR. This is crucial to our aim of investigating interactions, since studying an intact protein segment, rather than isolated fragments, better preserves the physiological complexity of its binding to multiple targets. This is particularly relevant for IDPs, which are known to exhibit a range of binding modes, from folding-upon-binding to the formation of "fuzzy" complexes: missing parts of the sequence can lead to altered or incomplete interaction patterns. Finally, the development of a monoclonal inducible BRCA1 cell line enables the isotope labelling of one protein at a time, further simplifying the interpretation of the in-cell NMR data for the study of interactions. Although

many challenges remain, this preliminary work paves the way for the investigation of BRCA1 in intact cells and its interaction with partners like MAX, ultimately contributing to a broader understanding of IDP biology.

5.5 Materials and methods:

5.5.1 Expression test

HEK293T cells were seeded in T25 flasks and transfected at 90–95% confluency using a polyethyleneimine (PEI)/pHL-sec plasmid mixture at a 2:1 ratio, following the transfection protocol described by Barbieri et al.[7]. As a negative control, an additional flask was transfected with a pHL-sec plasmid containing a non-coding sequence (empty vector). Transfected cells were incubated at 37 °C with 5% CO_2 for 48 hours under sterile conditions in DMEM supplemented with 2% (v/v) foetal bovine serum (FBS) and 100 µg/mL penicillin-streptomycin (Pen/Strep). After incubation, cells were harvested and pelleted by centrifugation. For protein extraction, cell pellets were resuspended in 150 µL of either PBS (for non-denaturing lysis) and lysed by one freeze–thaw cycle in liquid nitrogen, or in 150 µL of RIPA buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS) and incubated on ice for 30 minutes. Lysates were subsequently centrifuged at 14,000 g for 1 hour at 4 °C, and the supernatants were collected and stored at -20 °C. Protein expression levels were assessed via SDS-polyacrylamide gel electrophoresis (SDS-PAGE), followed by Coomassie Brilliant Blue staining.

5.5.2 Protein stability and cell viability assay

Cell viability and protein expression were evaluated at 4, 15, 20, 25, and 37°C, after 1, 2, 3, and 4 hours of incubation in a 3 mm Shigemi NMR tube. For each temperature condition, four T25 flasks of HEK293T cells were seeded and transfected as described above. After 48 hours of incubation at 37 °C and 5% CO_2, cells were harvested, pelleted, and resuspended in 180 µL of NMR buffer (DMEM supplemented with 70 mM HEPES and 20% (v/v) D_2O). A 10 µL aliquot of the cell suspension was taken at t = 0 to assess initial cell viability. Cells were then transferred into 3 mm Shigemi tubes and incubated at the designated temperatures. After 1, 2, 3, and 4 hours, one sample was retrieved for each time point. A 10 µL aliquot was taken from each for viability assessment, after which the cells were lysed by a freeze–thaw cycle using liquid nitrogen. Cell viability was evaluated by staining of the collected 10 µL aliquots of cell suspension with trypan blue (1:1) and counting viable cells using a Luna II automated. Protein integrity over time was assessed by SDS-PAGE analysis of the lysates.

5.5.3 Nucleus and cytoplasm separation

HEK293T cells were plated in T25 flasks and transfected as described above. After 48 hours of incubation, cells were harvested, pelleted, and resuspended in 2 volumes (relative to pellet volume) of cold homogenization buffer (10 mM Tris-HCl pH 8.0, 5 mM $MgCl_2$, 2 mM DTT, 0.1% Triton X-100, 0.25 M sucrose). The suspension was transferred to a glass Dounce homogenizer and homogenized with 20 strokes using the loose pestle (A) followed by 40 strokes with the tight pestle (B). The homogenate was collected and centrifuged at 500 g for 5 minutes at 4 °C. The supernatant (cytoplasmic fraction) was transferred to a fresh tube. The pellet was washed three times with cold homogenization buffer, then resuspended in 150 µL of

RIPA buffer, incubated on ice for 30 minutes, and centrifuged at 14,000 g for 1 hour at 4 °C. The resulting supernatant (nuclear fraction) was collected in a new tube for further analysis.

5.5.4 Western Blot

Samples were lysed in RIPA buffer. Protein extracts were separated by SDS-PAGE and transferred onto nitrocellulose membranes using a Bio-Rad Trans-Blot Turbo semi-dry transfer system. Membranes were blocked in 6% (w/v) non-fat dry milk in tris buffered saline-Tween (TBS-T) buffer (20 mM Tris-HCl, 150 mM NaCl, 0.1% Tween-20) for 1 hour at room temperature, then cut as needed and incubated overnight at 4 °C with the appropriate primary antibodies diluted in 2% (w/v) non-fat dry milk in TBS-T (anti-FLAG 1:1000, anti-GAPDH 1:2000, and anti-Lamin β1 1:1000). Following three washes with TBS-T, membranes were incubated for 1 hour at room temperature with HRP-conjugated secondary antibodies (1:80,000 in TBS-T). After three additional washes, signal detection was performed using enhanced chemiluminescence (ECL) reagents (Euroclone, Cat# P0131506), and membranes were imaged with a ChemiDoc MP imaging system (Bio-Rad).

5.5.5 Lysate stabilization

HEK293T cells were plated in two T75 flasks and transfected with BRCA1 219-1357. After 48 hours, cells were harvested and pelleted. Each pellet was lysed in 150 µL of PBS supplemented with protease inhibitor (PI) cocktail (Roche, Cat. No. 05892791001) via freezing and thawing cycles in liquid nitrogen, followed by centrifugation at 14,000 g for 1 hour at 4 °C. The two supernatants were pooled and diluted with PBS + PI to a final volume of 2 mL. The diluted lysate was incubated at 80 °C for 20 minutes, then centrifuged again at 14,000 g for 1 hour at 4 °C. The lysate was loaded into a 15 mL Centricon filter (30 kDa cut-off) and washed three times with PBS + PI buffer. The washed lysate was then concentrated to a final volume of ~300 µL using a 0.5 mL Centricon filter (30 kDa cut-off).

5.5.6 ^{15}N uniform labelling

HEK293T were plated in T75 flasks then transfected as previously described. The transfection was performed in ^{15}N uniformly labelled DMEM with 2% (v/v) FBS and 100 µg/mL pen/strep. The transfected cells were incubated at 37°C, 5% CO_2 for 48 hours in sterile conditions, harvested, pelleted, resuspended in 180 µL of NMR buffer (DMEM, 70 mM HEPES, and 20% (v/v) D_2O), and then transferred to a 3 mm Shigemi tube. The in-cell NMR experiment consisted of a ^{1}H, ^{15}N SOFAST-HMQC recorded at a 950 MHz Bruker Avance NEO spectrometer equipped with a TXI cryogenic probe.

5.5.7 ^{15}N, ^{13}C precursor labelling

A T75 flask of HEK293T cells at 90% confluence was supplied with 15 mL of ^{15}N uniformly labelled medium, supplemented with 2% FBS (v/v), 100 µg/mL pen/strep and 5x molar excess of ^{13}C labelled precursor (4 mM ^{13}Cα Leu precursor, 2 mM ^{13}C' Phe precursor) and incubated overnight at 37°C, 5% CO_2. After incubation, the medium was collected and used to transfect a second HEK293T T75 flask. The transfected flasks were incubated at 37°C, 5% CO_2 for 48h. Following incubation, the cells were harvested pelleted and either resuspended in 150 µL of NMR medium for in-cell NMR experiments or lysed in 150 µL of PBS+PI via freezing and thawing cycles in liquid nitrogen and supplemented with 10% D_2O for in-lysate NMR.

5.5.8 ^{15}N, ^{13}C amino acid labelling

To perform SOFAST-HMQC experiments HEK293T were plated in T75 flasks then transfected as previously described using a custom made DMEM lacking the amino acid to be labelled, supplemented with the uniformly labelled ^{15}N,^{13}C residue at a concentration equivalent to that of the unlabelled amino acid in commercial DMEM (i.e., 2.7 mM His, 0.5 mM Cys), besides 2% (v/v) FBS and 100 µg/mL pen/strep. To perform HNCA and HNCO experiments, HEK293T cells were transfected using a ^{15}N uniformly labelled medium supplied with 5x molar excess of labelled amino acid (i.e., 13.5 mM His, 2.5 mM Cys) along with 2% (v/v) FBS and 100 µg/mL pen/strep. The transfected flasks were incubated at 37°C, 5% CO_2 for 48h. After incubation the cells were harvested pelleted lysed in 150 µL of PBS+PI via freezing and thawing cycles in liquid nitrogen for in-lysate NMR. The lysates were supplemented with 10% D_2O and transferred to a 3mm NMR tube.

5.5.9 Protein coexpression

HEK293T cells were transfected with various ratios of plasmids encoding BRCA1 219-504 and MAX full-length, as follows:

- BRCA1 219-504 150% + MAX full length 50%
- BRCA1 219-504 50% + MAX full length 50%
- BRCA1 219-504 25% + MAX full length 75%
- BRCA1 219-504 10% + MAX full length 90%
- BRCA1 219-504 50% + MAX full length 150%

Protein expression tests via SDS-PAGE determined that the best ratio for comparable expression levels of both protein is BRCA1 219-504 150% + MAX full length 50%. While this condition uses a higher-than-standard DNA:PEI ratio, it was shown not to compromise cell viability or transfection efficiency.

5.5.10 Monoclonal inducible stably transfected cell line generation

We produced a monoclonal inducible stably transfected cell line, expressing BRCA1 219-1357 by using the PiggyBac Tetracycline inducible expression developed previously[52]. HEK293T-Rex Flp-In cells were co-transfected with two plasmids, one encoding for BRCA1 219-1357 with a FLAG tag and GFP, flanked by Inverted Terminal Repeats regions (ITR) and the other encoding for the Piggy-Bac transposase protein. Cells were incubated at 37°C, 5% CO_2 for 48 hours in DMEM supplemented with 10% (v/v) FBS, 1X Zell Shield, 5 µg/mL blasticidine. After 24 hours, transfected cells were selected with 3 µg/mL puromycin for 36 hours to ensure genomic integration of the gene of interest (GOI). Protein expression was then induced using 3 µg/mL tetracycline for 48 hours. Expression of GFP was analysed via flow cytometry to assess successful integration and induction. To achieve monoclonality, the top 5% of GFP-expressing cells were isolated by cell sorting via FACS and individually plated into wells of a 96-well plate. Clonal populations were expanded and further evaluated for GFP expression via flow cytometry after re-induction to confirm monoclonality. The best performing clones were selected based on GFP fluorescence intensity and uniformity. BRCA1 219-1357 expression was further confirmed by immunofluorescence analysis following fixation and staining of cells with an anti-FLAG primary antibody (1:2000) and a fluorescent Alexa Fluor 594-conjugated secondary antibody (1:1000) and 4′,6-diamidino-2-phenylindole (DAPI).

6. Impact of cell culture complexity on ligand-target interactions determined via in-cell NMR

6.1 Abstract

Accurate chaacterization of ligand-target interactions in physiologically relevant environments is essential for predicting in vivo efficacy. Conventional in vitro assays often fail to capture critical factors such as limited membrane permeability, off-target binding, and protein conformational changes within the cellular milieu. Three-dimensional (3D) cell cultures, particularly spheroids, provide a more representative intermediate model between monolayer systems and animal models, while remaining compatible with high-resolution structural techniques such as in-cell NMR. In this study, we investigated how cell culture complexity influences ligand binding to human carbonic anhydrase II (CA II) using methazolamide (MZA) and SLC-0111 (SLC) as ligands. Ligand-observed time-resolved ^{19}F in-cell NMR experiments performed with an optimized bioreactor setup enabled the monitoring of SLC displacement by MZA and the determination of relative in-cell K_d values across three models of increasing complexity: transiently transfected HEK293T cells, HEK293-TRex$^{TR\text{-}hCAII/Mono}$ spheroids, and mixed spheroids combining HEK293-TRex$^{TR\text{-}hCAII/Mono}$ and MCF-7 cells. While the results suggest a possible modulation of SLC affinity for CA II with increasing cell culture complexity, the high experimental uncertainty prevents drawing definitive conclusions at this stage.

6.2 Introduction

Investigating drug-cellular partner interactions in cellular models is of utmost importance, as traditional in vitro drug development often overlooks the complexity of the cellular environment[9,18,19,24]. Critical factors such as poor cell penetration, off-target binding, or conformational changes of the target protein within the cellular milieu can significantly alter drug efficacy compared to standard in vitro assays[4,20,24]. Indeed, even when certain drugs demonstrate high efficacy in cells, their performance in vivo often fails to meet expectations. Furthermore, the complexity of animal models introduces additional variables that single-cell systems cannot effectively represent. Intermediate models that simulate more complex environments and better recapitulate tissue architecture are therefore essential. One notable advancement in this area is the development of three-dimensional (3D) cell cultures, particularly spheroids[41,44,49]. Spheroids represent an incremental level of complexity compared to traditional monolayer cultures. They are spherical aggregates of cells, typically formed via self-assembly under non-adherent conditions or within microwells[41,43,52]. Their internal organization is arranged into distinct concentric zones reflecting gradients of nutrient and oxygen availability. The outer layer consists of proliferating cells, while the inner layers become quiescent and eventually a necrotic core is formed when the spheroid size prevents sufficient oxygen and nutrient penetration [40,41,44–46]. This structural organization makes spheroids excellent models for tissues or tumours. Unlike more complex organoids, spheroids present crucial advantages: they can be easily grown in large number, and provide good size homogeneity within the sample, which is important when analysing a large ensemble via in-cell NMR. This approach is particularly powerful for studying cellular processes, as it provides atomic-resolution insights into the structure, dynamics, and kinetics of ligand-target interactions directly within living cells[1,3,4,16]. In-cell NMR, therefore, enables early assessment

of drug potency and reliable prediction of in vivo activity by evaluating both membrane permeability and intracellular binding selectivity[19,20,24].

To demonstrate how the cellular model can influence drug studies, we used a pharmacologically relevant target: human carbonic anhydrase II (CA II). CA II is one of 15 isoforms of a ubiquitous metalloenzyme that catalyses the reversible reaction between CO_2 and H_2O to produce HCO_3^- and H^+. CAs play crucial roles in various physiological processes and are implicated in numerous pathological states such as glaucoma, epilepsy, and cancer[182–185]. As ligands, we selected two well-established inhibitors: methazolamide (MZA), an approved drug for glaucoma, and SLC-0111 (SLC), a recently developed inhibitor, currently in Phase II-b clinical trials as an anticancer agent (**Figure 6.3.2-1**)[186]. The fluorine atom present in SLC allowed us to perform ligand observed competition binding through ^{19}F in-cell NMR, using an optimized bioreactor setup, using SLC as reference ligand and MZA as its competitor.

6.3 Results and discussion

6.3.1 In-cell ^{19}F NMR of SLC binding to CA II

The fluorine atom present in the structure of SLC makes it an excellent probe for ligand-observed monitoring of protein-ligand binding in live cells, enabling the use of sensitive and background-free 1D ^{19}F NMR spectra. Notably, three distinct peaks were detected in the in-cell ^{19}F NMR spectrum of a sample of HEK293-TRex$^{TR\text{-}hCAII/Mono}$ cells expressing CA II and treated with SLC, two of which were also present in a negative control sample of uninduced cells (**Figure 6.3.1-1**).

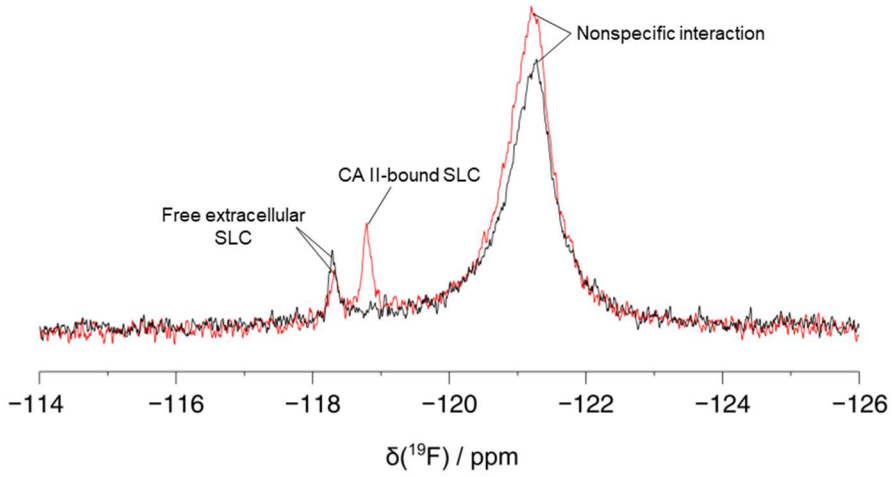


Figure 6.3.1-1
^{19}F in-cell NMR spectra of HEK293-TRex$^{TR\text{-}hCAII/Mono}$ treated with SLC-0111 100µM, with induced CA II expression (red) and without induced CA II expression (black).

The peak at -118.8 ppm absent in the uninduced cells corresponds to the drug bound to CA II, while the smaller peak at -118.3 ppm present in both samples arises from the free drug outside the cells. The third, strong signal at -121.2 ppm is also present in the negative controls and corresponds to a nonspecific interaction of the drug with other abundant cellular component(s) (**Figure 6.3.1-1**).

Thanks to the high sensitivity of the ^{19}F chemical shift to its chemical surroundings, all three forms of the drug are clearly resolved in the spectrum, allowing quantitative analysis of the drug-target interaction.

6.3.2 Ligand observed competition binding to assess K_d in different cell culture models

A competition binding experiment was therefore performed, where the concentration of SLC was kept constant while the concentration of the second, higher-affinity drug (MZA) was gradually increased (**Figure 6.3.2-1**).

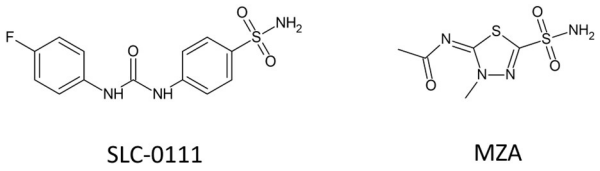


Figure 6.3.2-1
Chemical structure of CA II ligands SLC-0111 and methazolamide (MZA)

By monitoring the decrease of intensity of the ^{19}F signal of the SLC:CA II complex, the progressive displacement of SLC was quantified and used to determine the relative in-cell K_d of SLC with respect to MZA. An optimized NMR bioreactor setup was employed that enables in-cell experiments for up to 72 hours while maintaining high cell viability. This is accomplished through the continuous flow of fresh culture medium, regulated by a programmable multichannel peristaltic pump[1,19]. Using this system, we conducted 60-hour experiments divided into five 12-hour steps. At each step, the concentration of SLC was kept constant, while the concentration of MZA in the flowing medium was gradually increased (**Figure 6.3.2-2**). The step duration was chosen to ensure that the equilibrium between free and bound ligand within the cells was reached by the end of each step.

This bioreactor experiment was repeated on three cellular models of increasing complexity to evaluate whether the K_d of SLC remained constant or varied depending on the model used. We began with the simplest model, HEK293T cells transiently transfected with CA II. This reference system, already tested in previous studies, allowed us to perform an initial assessment of the difference between the drug's K_d measured in vitro and in living cells[187]. Then the experiment was reproduced using a monoclonal, stably transfected, and inducible cell line expressing CA II (HEK293-TRex$^{TR\text{-}hCAII/Mono}$). The next step involved using a HEK293-TRex$^{TR\text{-}hCAII/Mono}$ to generate spheroids. These spheroids reached a diameter of approximately 350 µM, large enough to develop two concentric zones: an inner, non-proliferating quiescent layer, and an outer, proliferating layer, but without the formation of a necrotic core. Notably, in HEK293-TRex$^{TR\text{-}hCAII/Mono}$, the expression level of CA II is significantly lower than in transiently transfected HEK293T cells, thus providing the further advantage of approaching protein concentrations closer to endogenous levels. Finally, the third and most complex model consisted of mixed spheroids. These were produced by combining spheroid pellets from HEK293-TRex$^{TR\text{-}hCAII/Mono}$ cells with pellets from a second cell line, the Michigan Cancer Foundation cell line no. 7 (MCF-7), which exhibit an epithelial-like morphology[52].

The displacement of SLC by MZA during each step was quantified by integrating the corresponding signal in the time-resolved 1D ^{19}F NMR series to obtain the fraction of CA II

bound to SLC at each time point (**Figure 6.3.2-3**). The averaged plateau values at each step (**Figure 6.3.2-4**) were fitted with **Equation 1** to retrieve the K_d of SLC relative to MZA. By taking a previously determined K_d value of 37 nM for MZA[187], the intracellular K_d of SLC was obtained for each cell culture model.

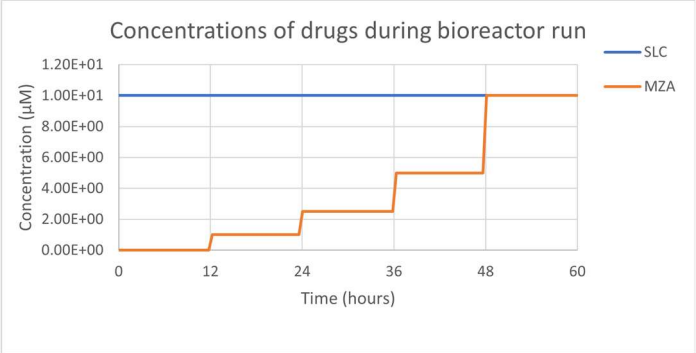


Figure 6.3.2-2
Concentration profile of the two ligands, SLC-0111 and MZA, during the five 12-hours steps of the bioreactor run.

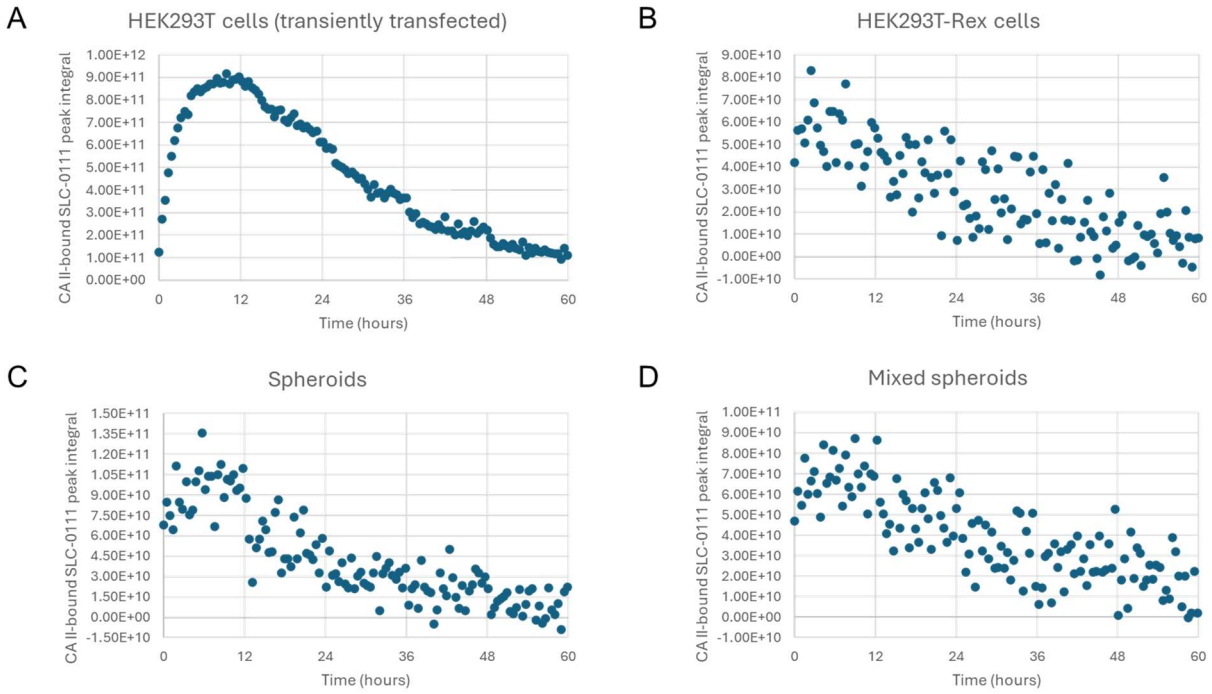


Figure 6.3.2-3
Concentration profiles of CA II–bound fraction of SLC-0111 obtained by MCR-ALS in (A) transiently transfected HEK293T cells, (B) HEK293-TRex$^{TR\text{-}hCAII/Mono}$ cells (C) HEK293-TRex$^{TR\text{-}hCAII/Mono}$ spheroids and (D) mixed spheroids (HEK293-TRex$^{TR\text{-}hCAII/Mono}$ spheroids with MCF-7 cells) .

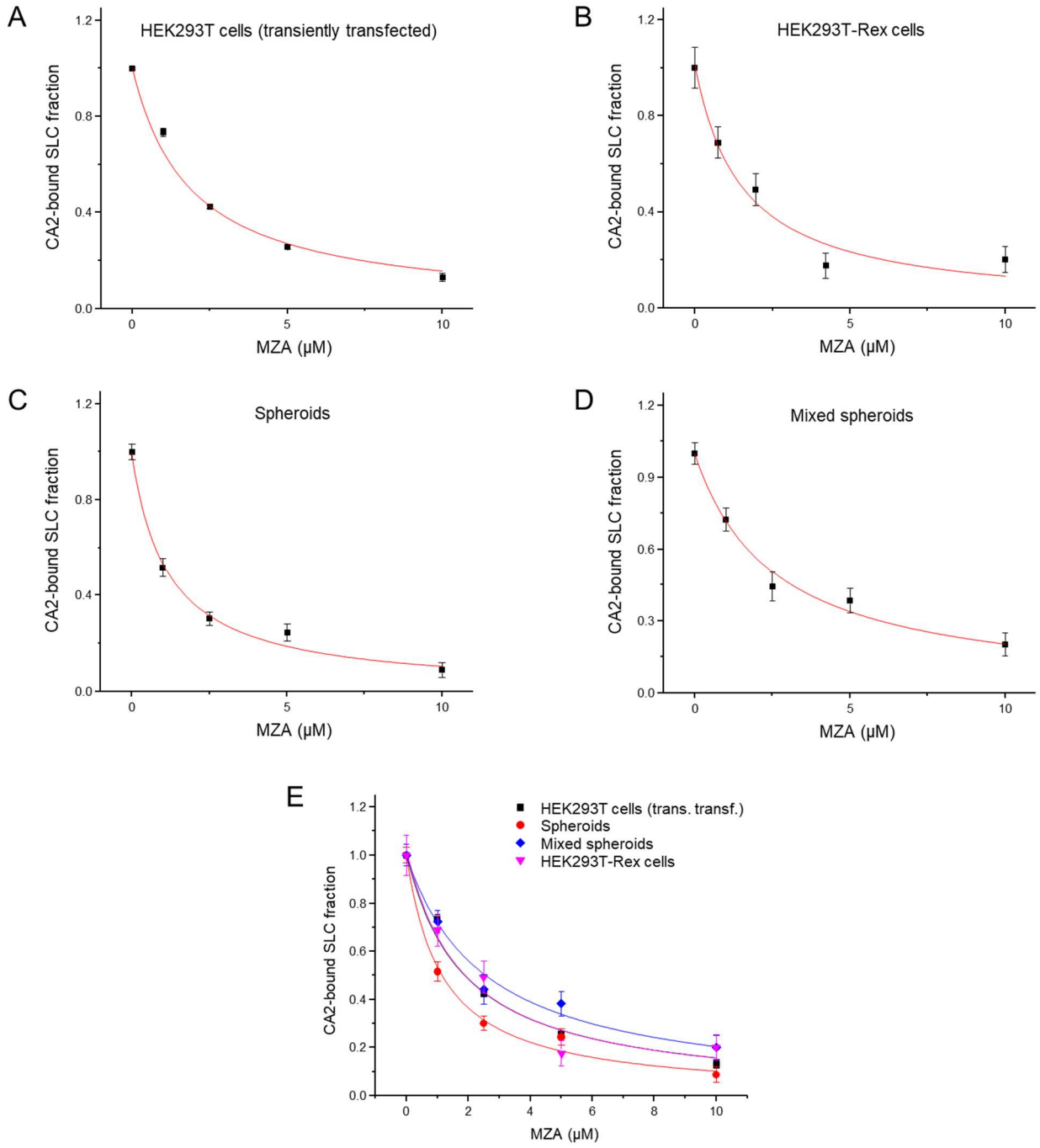


Figure 6.3.2-4
CA II-bound fractions of SLC-0111 obtained from the plateau values after each step of bioreactor run plotted as a function of MZA concentration in (A) transiently transfected HEK293T cells, (B) HEK293-TRex$^{TR\text{-}hCAII/Mono}$ cells (C) HEK293-TRex$^{TR\text{-}hCAII/Mono}$spheroids and (D) mixed spheroids (HEK293-TRex$^{TR\text{-}hCAII/Mono}$ spheroids with MCF-7 cells). Binding curves from nonlinear fitting are shown as dashed lines.(E) Comparison of CA II-bound fractions of SLC-0111and binding curves in the different cellular models.

The obtained K_d values and error values obtained from the fitting (**Table 1**) suggest a significant alteration in the affinity of SLC for CA II depending on the cellular model. However, the extremely low signal intensity for monoclonal cells, spheroids and hybrid spheroids, which results in an unfavourable signal-to-noise (S/N) ratio in the spectra (**Figure 6.3.2-3** and **Figure 6.3.2-4**), suggests that additional replicates of the experiments are necessary to draw this conclusion with reasonable certainty. Such low S/N ratio is attributable to the low concentration of CA II in the HEK293-TRex$^{TR\text{-}hCAII/Mono}$ cells. In addition, as observed with the transiently transfected HEK293T cells (**Figure 6.3.2-4**), the equilibrium between the concentrations of

bound drugs may not be fully reached at the end of each step, thus introducing an additional source of error.

Table 1
SLC-0111 K_d in different cellular models

Cellular model	K_d SLC/ K_d MZA	K_d SLC (M)
HEK293T transiently transfected	5.4 ± 0.4	$2.0 \cdot 10^{-7} \pm 2 \cdot 10^{-8}$
HEK293-TRex$^{TR\text{-}hCAII/Mono}$	4.0 ± 0.7	$1.5 \cdot 10^{-7} \pm 3 \cdot 10^{-8}$
HEK293-TRex$^{TR\text{-}hCAII/Mono}$ spheroids	8.8 ± 1.0	$3.2 \cdot 10^{-7} \pm 4 \cdot 10^{-8}$
HEK293-TRex$^{TR\text{-}hCAII/Mono}$ and MCF-7 mixed spheroids	3.9 ± 0.5	$1.4 \cdot 10^{-7} \pm 2 \cdot 10^{-8}$

6.1 Conclusions and perspectives

In this work, we demonstrated that ligand-observed ^{19}F in-cell NMR experiments can be applied to study ligand-target interactions even when the protein of interest is present at very low intracellular concentrations, such as those attained in stable inducible human cell lines. Under these conditions, protein-observed experiments would not be practical, making the ligand-based approach particularly valuable. Using SLC-0111 as a fluorinated probe and methazolamide as a competitor, we monitored relative in-cell K_d values for CA II across cellular models of increasing complexity, including transiently transfected HEK293T cells, HEK293-TRex$^{TR\text{-}hCAII/Mono}$ cells, HEK293-TRex$^{TR\text{-}hCAII/Mono}$ spheroids, and mixed spheroids produced by combining HEK293-TRex$^{TR\text{-}hCAII/Mono}$ spheroids with MCF-7 cells. Although the different dissociation constants obtained suggest that cellular complexity may modulate ligand affinity, the high experimental uncertainty derived from the low expression levels of the target protein in HEK293-TRex$^{TR\text{-}hCAII/Mono}$ currently prevents definitive conclusions. While it is not possible to increase protein expression in monoclonal cells to reduce this measurement-intrinsic error, other strategies can be adopted to increase the reliability of the result, such as repeating bioreactor runs to perform a statistically significant analysis of K_d changes. In addition, the duration of bioreactor steps can be further optimized to ensure complete equilibration between free and bound ligand at the end of each step. The latter is especially true for cells expressing higher levels of protein, where slow drug diffusion rates through the cell membrane may become a bottleneck in the real-time measurement.

6.2 Materials and methods

6.2.1 HEK293-TRex$^{TR\text{-}hCAII/Mono}$ sample for in-cell NMR

HEK293-TRex$^{TR\text{-}hCAII/Mono}$ cells[52], cells from a monoclonal stably transfected inducible cell line overexpressing human CA II and GFP (at a lower concentration with respect to that of CA II) obtained and characterized previously[52] were seeded in a T75 flask and cultured in DMEM supplemented with 10% (v/v) FBS, 100 µg/mL penicillin-streptomycin, 5 µg/mL blasticidine, 3 µg/mL puromycin. When cells reached 70–80% confluence, protein expression was induced by addition of 3 µg/mL tetracycline. Cultures were then incubated for 48 h at 37 °C and 5% CO_2 under sterile conditions. Following induction, cells were treated with 100 µM SLC-0111 for 1 h, harvested by trypsinization, centrifuged at 1000 rpm for 5 min, resuspended in 150 µL of NMR buffer (DMEM, 70 mM HEPES, and 20% (v/v) D_2O) and transferred into a 3mm Shigemi tube. Ligand-observed in-cell NMR experiments were carried out at 310 K on a 600 MHz Bruker Avance NEO spectrometer.

6.2.2 Transfected cell sample for bioreactor

HEK293T cells were seeded in a T75 flask and transfected at 90–95% confluency using a mixture of polyethyleneimine (PEI)/pHL-sec plasmid encoding for CA II in a ratio of 2:1, following the protocol reported by Barbieri et al.[7]. Transfected cells were incubated at 37 °C with 5% CO_2 for 48 hours under sterile conditions in DMEM supplemented with 2% (v/v) FBS and 100 µg/mL Pen/Strep. After incubation, cells were harvested by trypsinization, pelleted by centrifugation at 800g for 5 minutes, then resuspended in 400 µL of sterile low-melting agarose in PBS, melted at 85°C and kept in solution at 37°C. The suspension was transferred using a syringe in a 0.75 mm (i.d.) PEEK tubing, let cool down for 2 minutes at room temperature to allow hydrogel formation, and transferred in the 5 mm flow NMR tube following the protocol reported by Barbieri and Luchinat[1].

6.2.3 Monoclonal cell sample for bioreactor

A T75 flask of HEK293-TRex$^{TR\text{-}hCAII/Mono}$, cells cultivated in DMEM supplemented with 10% (v/v) FBS, 100 µg/mL penicillin-streptomycin, 5 µg/mL blasticidine, 3 µg/mL puromycin at 80-90% confluence was split 1:2 in two new T75 flasks. CA II expression was induced by treatment with tetracycline (final concentration in well 3 µg/mL). Induced cells were incubated at 37 °C with 5% CO_2 for 48 hours under sterile conditions. Cell harvesting and hydrogel preparation was conducted as previously described for the transfected cell sample.

6.2.4 Spheroid sample for bioreactor

Spheroid culture plates were prepared by agarose stamping. A 1.5% (w/v) solution of ultrapure agarose was prepared in PBS, melted until fully dissolved, and maintained at 80 °C. The solution was sterilized by filtration through a 0.22 µm filter under a laminar flow hood and dispensed into 6-well plates (3 mL per well). Immediately after dispensing, optimized sterile stamps (kindly provided by Lukas Trantirek's group) were applied onto the liquid agarose to create wells for spheroid growth, avoiding bubble formation. Agarose was allowed to solidify at room temperature before stamp removal. The wells were filled with 1.5 mL of complete culture medium and incubated at 37 °C for 24 h prior to cell seeding.

HEK293-TRex$^{TR\text{-}hCAII/Mono}$ cells were cultivated in DMEM supplemented with 10% (v/v) FBS, 100 µg/mL penicillin-streptomycin, 5 µg/mL blasticidine, 3 µg/mL puromycin, harvested from confluent T75 flasks by trypsinization, resuspended in complete medium, and counted using a Burker chamber. For spheroid seeding, 3.7×10^7 cells were collected, centrifuged (1000 rpm, 5 min), and resuspended in 9.5 mL of complete medium. The cell suspension was filtered through a 40 µm cell strainer (DNase/RNase-free) and 1.5 mL of the filtered suspension was added to each well of an agarose stamped 6 well plate containing 1.5 mL of fresh medium, resulting in a final volume of 3 mL per well. Immediately after plating cells were supplemented with $ZnSO_4$ (final concentration in well 10 µM) and expression of CA II was induced with tetracycline (final concentration in well 3 µg/mL). Plates were the incubated at 37 °C with 5% CO_2 for 48 hours under sterile conditions, during which spheroid formation was monitored. Three plates were prepared for one NMR experiment.

After 48 hours protein expression was assessed by checking intensity of GFP signal through fluorescence microscopy, then culture medium was gently aspirated, and spheroids were extracted from the plates by gently pipetting fresh medium on top of the agarose disk. The

spheroid suspension was collected into Falcon tubes and centrifuged at 1000 rpm for 5 min. A 300 µL volume of pelleted spheroids were resuspended in 200 µL of 1.5% (w/v) sterile low-melting agarose solution in PBS kept at 37°C. The agarose filament was prepared as described above and inserted into the 5 mm flow NMR tube.

For mixed spheroids, the HEK293-TRex$^{TR\text{-}hCAII/Mono}$ induced spheroids pellet was mixed with a MCF-7 pellet in a ratio of 1:2 before the agarose suspension preparation, followed by the same procedure as above.

6.2.5 Bioreactor setup

The bioreactor medium was prepared by dissolving 7,37 g of DMEM powder in 530 mL of a solution of ultrapure H_2O, 10 mM $NaHCO_3$ and 2% D_2O. pH was adjusted to 7.4 by adding HCl and the solution was sterilized by filtration. The medium was then supplemented with 2% FBS, 1% Pen/Strep, and 10 µM SLC-0111 before being divided into two sterile bottles; bottle A, containing 360 mL of the prepared medium, and bottle B, containing 190 mL of the same medium further supplemented with 10 µM MZA.

The bioreactor flow unit was assembled according to the setup described by Luchinat et al.[187] and connected to a programmable peristaltic pump equipped with two independent channels. A tee junction allowed the mixing of flows from the two channels (A and B), which were connected respectively to bottle A and bottle B, both maintained at 37 °C in a water bath. During the bioreactor run, pump operation was controlled via a PC using a pre-programmed protocol, with the combined flow rate of the two channels kept constant at 0.1 mL/min (**Table 2**).

Table 2
Steps, timings, ligand concentrations and flow rates of bioreactor runs.

Step	Timing (hours)	SLC-0111 (µM)	MZA (µM)	Channel A flow rate (µL/min)	Channel B flow rate (µL/min)
1	12	10	0	100	0
2	12	10	1	90	10
3	12	10	2.5	75	25
4	12	10	5	50	50
5	12	10	10	0	100

A series of time-resolved 1D ^{19}F NMR spectra were acquired at 310 K with a 600 MHz Bruker Avance NEO spectrometer over the course of 60 hours, with an acquisition time of 294 ms, interscan delay of 1 s, 1280 scans for an experiment time of 28 minutes per spectrum.

6.2.6 Data analysis

The time-resolved in-cell NMR spectra were processed in Topspin (Bruker) and analysed using Dynamic Center (Bruker). The signal intensities at each step of the run were obtained by averaging the last 13 time points of each step and were normalized by the averaged signal intensity at the end of the first step to obtain the molar fraction of CA II bound to SLC. The obtained values were fitted as a function of the concentration of MZA with the nonlinear fitting function in OriginPro 8 (OriginLab) to retrieve the K_d of the tested ligands from the K_d of the reference ligand.

For in-cell NMR competition binding experiments in the bioreactor, where the free ligand concentration is constant and known, the K_d of the test ligand was retrieved from **Equation 1**.

In this equation F_I is the reference ligand-bound CAII fraction, K_{dI} is the dissociation constant for the reference ligand, K_{dL} is the dissociation constant for the tested ligand, *[I]* is the concentration of the free reference ligand and *[L]* is the concentration of the free tested ligand, under the approximation that the concentration of free ligand is equal to the total concentration of ligand, and that both are in excess with respect to the target.

Equation 1

$$F_I = \frac{1}{1 + \frac{K_{dI}[L]}{K_{dL}[I]}}$$

7. Conclusions and perspectives

This thesis demonstrates substantial progress in the development of selective isotopic labelling strategies for in-cell NMR spectroscopy in human cells, addressing long-standing challenges in sensitivity, specificity, and applicability to complex proteins and physiological models.

The first achievement was the establishment of cost-effective side-chain isotope labelling protocols based on the metabolic conversion of α-ketoacid precursors, which is an established practice in E. coli but was never applied to human cells. By exploiting the human transaminase activity, this approach enables efficient and residue-specific incorporation of isotopes without scrambling, resulting in high-quality in-cell and lysate NMR spectra of pharmacologically relevant proteins. The successful observation of ligand-induced chemical shift perturbations in living cells further validated the functional utility of this approach for probing protein-ligand interactions in their native environment.

The incorporation of a fluorinated tryptophan provided a complementary strategy, demonstrating the power of ^{19}F NMR combined with ^{13}C editing to resolve overlapping spectral signals. This advance expands the chemical space of in-cell NMR probes and highlights the potential of non-natural amino acids for dissecting structural and dynamic features of proteins in situ.

The thesis also introduced tailored selective labelling schemes for intrinsically disordered proteins (IDPs), exemplified by BRCA1. Optimized labelling, experimental protocols, and protein expression strategies enabled residue-specific observation of one of the largest IDP fragments ever studied by in-cell and in-lysate NMR. The creation of a monoclonal inducible BRCA1 cell line represents a pivotal resource for future investigations of BRCA1 dynamics and interactions, offering precise control over protein expression and labelling.

All these labelling schemes help overcome the challenges of low sensitivity, broad spectral lines, and signal overlap typically encountered when studying large and complex proteins within the crowded cellular environment. By reducing spectral complexity and enhancing sensitivity, they greatly simplify resonance assignment. These methodological advancements make it possible to investigate in-cell interactions of large globular proteins that were previously difficult to observe via NMR, thereby paving the way for exploring complex and unexplored molecular mechanisms. The ability to observe the interactions of intrinsically disordered proteins (IDPs) in their physiological environment allows the investigation of their highly dynamic and versatile interaction networks, characterized by structural plasticity and the ability to form diverse complexes with various partners. Overall, the opportunity to understand reaction mechanisms in their natural context lays the groundwork for the development of increasingly effective drugs against diseases that remain challenging to treat.

Finally, the work extended in-cell NMR applications to ligand-target interaction studies under increasing cell culture complexity. By time-resolved ^{19}F NMR in a flow NMR bioreactor it was possible to monitor drug binding and displacement across multiple cell culture models, including spheroids.

The ability to perform NMR experiments on complex biological models that mimic tissue-like or tumor-like environments opens the door to a wide range of investigations from drug screening and permeability kinetics to the study of dysfunctions in disease-related proteins,

particularly those implicated in cancer. These experiments underscore the potential of in-cell NMR to bridge the gap between molecular biophysics and physiologically relevant environments.

8. Bibliography

(1) 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.

(2) Luchinat, E.; Banci, L. In-Cell NMR in Human Cells: Direct Protein Expression Allows Structural Studies of Protein Folding and Maturation. *Acc. Chem. Res.* **2018**, *51* (6), 1550–1557. https://doi.org/10.1021/acs.accounts.8b00147.

(3) Luchinat, E.; Banci, L. In-Cell NMR: Recent Progresses and Future Challenges. *Rendiconti Lincei Sci. Fis. E Nat.* **2023**, *34* (3), 653–661. https://doi.org/10.1007/s12210-023-01168-y.

(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. https://doi.org/10.1021/acs.chemrev.1c00790.

(5) Banci, L.; Barbieri, L.; Bertini, I.; Cantini, F.; Luchinat, E. In-Cell NMR in E. Coli to Monitor Maturation Steps of hSOD1. *PLoS ONE* **2011**, *6* (8), e23561. https://doi.org/10.1371/journal.pone.0023561.

(6) Barbieri, L.; Luchinat, E.; Banci, L. Protein Interaction Patterns in Different Cellular Environments Are Revealed by In-Cell NMR. *Sci. Rep.* **2015**, *5* (1), 14456. https://doi.org/10.1038/srep14456.

(7) 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.

(8) Li, C.; Zhao, J.; Cheng, K.; Ge, Y.; Wu, Q.; Ye, Y.; Xu, G.; Zhang, Z.; Zheng, W.; Zhang, X.; Zhou, X.; Pielak, G.; Liu, M. Magnetic Resonance Spectroscopy as a Tool for Assessing Macromolecular Structure and Function in Living Cells. *Annu. Rev. Anal. Chem.* **2017**, *10* (1), 157–182. https://doi.org/10.1146/annurev-anchem-061516-045237.

(9) 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.

(10) Serber, Z.; Ledwidge, R.; Miller, S. M.; Dötsch, V. Evaluation of Parameters Critical to Observing Proteins Inside Living *Escherichia c Oli* by In-Cell NMR Spectroscopy. *J. Am. Chem. Soc.* **2001**, *123* (37), 8895–8901. https://doi.org/10.1021/ja0112846.

(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.* **2006**, *103* (32), 11904–11909. https://doi.org/10.1073/pnas.0604667103.

(12) 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. https://doi.org/10.1038/nature07839.

(13) 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. https://doi.org/10.1021/ja904407w.

(14) Theillet, F.-X.; Binolfi, A.; Bekei, B.; Martorana, A.; Rose, H. M.; Stuiver, M.; Verzini, S.; Lorenz, D.; Van Rossum, M.; Goldfarb, D.; Selenko, P. Structural Disorder of Monomeric α-Synuclein Persists in Mammalian Cells. *Nature* **2016**, *530* (7588), 45–50. https://doi.org/10.1038/nature16531.

(15) Martorana, A.; Bellapadrona, G.; Feintuch, A.; Di Gregorio, E.; Aime, S.; Goldfarb, D. Probing Protein Conformation in Cells by EPR Distance Measurements Using Gd^{3+} Spin

Labeling. *J. Am. Chem. Soc.* **2014**, *136* (38), 13458–13465. https://doi.org/10.1021/ja5079392.
(16) Banci, L.; Barbieri, L.; Bertini, I.; Luchinat, E.; Secci, E.; Zhao, Y.; Aricescu, A. R. Atomic-Resolution Monitoring of Protein Maturation in Live Human Cells by NMR. *Nat. Chem. Biol.* **2013**, *9* (5), 297–299. https://doi.org/10.1038/nchembio.1202.
(17) Annecke, H. T. P.; Eidelpes, R.; Feyrer, H.; Ilgen, J.; Gürdap, C. O.; Dasgupta, R.; Petzold, K. Optimising In-Cell NMR Acquisition for Nucleic Acids. *J. Biomol. NMR* **2024**, *78* (4), 249–264. https://doi.org/10.1007/s10858-024-00448-5.
(18) 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.
(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.; 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.
(21) Mu, X.; Choi, S.; Lang, L.; Mowray, D.; Dokholyan, N. V.; Danielsson, J.; Oliveberg, M. Physicochemical Code for Quinary Protein Interactions in *Escherichia Coli*. *Proc. Natl. Acad. Sci.* **2017**, *114* (23). https://doi.org/10.1073/pnas.1621227114.
(22) Smith, A. E.; Zhou, L. Z.; Gorensek, A. H.; Senske, M.; Pielak, G. J. In-Cell Thermodynamics and a New Role for Protein Surfaces. *Proc. Natl. Acad. Sci.* **2016**, *113* (7), 1725–1730. https://doi.org/10.1073/pnas.1518620113.
(23) Goodsell, D. S. *The Machinery of Life*, 2nd ed.; Springer, 2009.
(24) Luchinat, E.; Barbieri, L.; Cremonini, M.; Nocentini, A.; Supuran, C. T.; Banci, L. Intracellular Binding/Unbinding Kinetics of Approved Drugs to Carbonic Anhydrase II Observed by in-Cell NMR. *ACS Chem. Biol.* **2020**, *15* (10), 2792–2800. https://doi.org/10.1021/acschembio.0c00590.
(25) Hänsel, R.; Foldynová-Trantírková, S.; Löhr, F.; Buck, J.; Bongartz, E.; Bamberg, E.; Schwalbe, H.; Dötsch, V.; Trantírek, L. Evaluation of Parameters Critical for Observing Nucleic Acids Inside Living *Xenopus Laevis* Oocytes by In-Cell NMR Spectroscopy. *J. Am. Chem. Soc.* **2009**, *131* (43), 15761–15768. https://doi.org/10.1021/ja9052027.
(26) 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. https://doi.org/10.1002/anie.201207243.
(27) 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.
(28) Breindel, L.; Burz, D. S.; Shekhtman, A. Active Metabolism Unmasks Functional Protein–Protein Interactions in Real Time in-Cell NMR. *Commun. Biol.* **2020**, *3* (1), 249. https://doi.org/10.1038/s42003-020-0976-3.
(29) Kay, L. E. The Evolution of Solution State NMR Pulse Sequences through the 'Eyes' of Triple-Resonance Spectroscopy. *J. Magn. Reson.* **2019**, *306*, 48–54. https://doi.org/10.1016/j.jmr.2019.07.035.
(30) Amero, C.; Schanda, P.; Durá, M. A.; Ayala, I.; Marion, D.; Franzetti, B.; Brutscher, B.; Boisbouvier, J. Fast Two-Dimensional NMR Spectroscopy of High Molecular Weight Protein Assemblies. *J. Am. Chem. Soc.* **2009**, *131* (10), 3448–3449. https://doi.org/10.1021/ja809880p.
(31) Igumenova, T. I.; Palmer, A. G. Off-Resonance TROSY-Selected $R_{1\rho}$ Experiment with Improved Sensitivity for Medium- and High-Molecular-Weight Proteins. *J. Am. Chem. Soc.* **2006**, *128* (25), 8110–8111. https://doi.org/10.1021/ja061692f.

(32) 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. https://doi.org/10.1007/s10858-021-00358-w.

(33) Krishnarjuna, B.; Jaipuria, G.; Thakur, A.; D'Silva, P.; Atreya, H. S. Amino Acid Selective Unlabeling for Sequence Specific Resonance Assignments in Proteins. *J. Biomol. NMR* **2011**, *49* (1), 39–51. https://doi.org/10.1007/s10858-010-9459-z.

(34) Kasai, T.; Ono, S.; Koshiba, S.; Yamamoto, M.; Tanaka, T.; Ikeda, S.; Kigawa, T. Amino-Acid Selective Isotope Labeling Enables Simultaneous Overlapping Signal Decomposition and Information Extraction from NMR Spectra. *J. Biomol. NMR* **2020**, *74* (2–3), 125–137. https://doi.org/10.1007/s10858-019-00295-9.

(35) Proudfoot, A.; Frank, A. O.; Ruggiu, F.; Mamo, M.; Lingel, A. Facilitating Unambiguous NMR Assignments and Enabling Higher Probe Density through Selective Labeling of All Methyl Containing Amino Acids. *J. Biomol. NMR* **2016**, *65* (1), 15–27. https://doi.org/10.1007/s10858-016-0032-2.

(36) Wu, P. S. C.; Ozawa, K.; Jergic, S.; Su, X.-C.; Dixon, N. E.; Otting, G. Amino-Acid Type Identification in 15N-HSQC Spectra by Combinatorial Selective 15N-Labelling. *J. Biomol. NMR* **2006**, *34* (1), 13–21. https://doi.org/10.1007/s10858-005-5021-9.

(37) Rosati, M.; Barbieri, L.; Hlavac, M.; Kratzwald, S.; Lichtenecker, R. J.; Konrat, R.; Luchinat, E.; Banci, L. Towards Cost-Effective Side-Chain Isotope Labelling of Proteins Expressed in Human Cells. *J. Biomol. NMR* **2024**. https://doi.org/10.1007/s10858-024-00447-6.

(38) 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.

(39) Toscano, G.; Rosati, M.; Barbieri, L.; Maier, K.; Banci, L.; Luchinat, E.; Konrat, R.; Lichtenecker, R. J. The Synthesis of Specifically Isotope Labelled Fluorotryptophan and Its Use in Mammalian Cell-Based Protein Expression for^{19} F-NMR Applications. *Chem. Commun.* **2024**, *60* (96), 14188–14191. https://doi.org/10.1039/D4CC04789C.

(40) Damman, R.; Lucini Paioni, A.; Xenaki, K. T.; Beltrán Hernández, I.; Van Bergen En Henegouwen, P. M. P.; Baldus, M. Development of in Vitro-Grown Spheroids as a 3D Tumor Model System for Solid-State NMR Spectroscopy. *J. Biomol. NMR* **2020**, *74* (8–9), 401–412. https://doi.org/10.1007/s10858-020-00328-8.

(41) Pinto, B.; Henriques, A. C.; Silva, P. M. A.; Bousbaa, H. Three-Dimensional Spheroids as In Vitro Preclinical Models for Cancer Research. *Pharmaceutics* **2020**, *12* (12), 1186. https://doi.org/10.3390/pharmaceutics12121186.

(42) Tang, X.-Y.; Wu, S.; Wang, D.; Chu, C.; Hong, Y.; Tao, M.; Hu, H.; Xu, M.; Guo, X.; Liu, Y. Human Organoids in Basic Research and Clinical Applications. *Signal Transduct. Target. Ther.* **2022**, *7* (1), 168. https://doi.org/10.1038/s41392-022-01024-9.

(43) Bogacheva, M. S.; Harjumäki, R.; Flander, E.; Taalas, A.; Bystriakova, M. A.; Yliperttula, M.; Xiang, X.; Leung, A. W.; Lou, Y.-R. Differentiation of Human Pluripotent Stem Cells Into Definitive Endoderm Cells in Various Flexible Three-Dimensional Cell Culture Systems: Possibilities and Limitations. *Front. Cell Dev. Biol.* **2021**, *9*, 726499. https://doi.org/10.3389/fcell.2021.726499.

(44) Gunti, S.; Hoke, A. T. K.; Vu, K. P.; London, N. R. Organoid and Spheroid Tumor Models: Techniques and Applications. *Cancers* **2021**, *13* (4), 874. https://doi.org/10.3390/cancers13040874.

(45) Tevis, K. M.; Colson, Y. L.; Grinstaff, M. W. Embedded Spheroids as Models of the Cancer Microenvironment. *Adv. Biosyst.* **2017**, *1* (10), 1700083. https://doi.org/10.1002/adbi.201700083.

(46) Zanoni, M.; Cortesi, M.; Zamagni, A.; Arienti, C.; Pignatta, S.; Tesei, A. Modeling Neoplastic Disease with Spheroids and Organoids. *J. Hematol. Oncol.J Hematol Oncol* **2020**, *13* (1), 97. https://doi.org/10.1186/s13045-020-00931-0.

(47) Browning, A. P.; Sharp, J. A.; Murphy, R. J.; Gunasingh, G.; Lawson, B.; Burrage, K.; Haass, N. K.; Simpson, M. J. Quantitative Analysis of Tumour Spheroid Structure.
(48) Murphy, R. J.; Browning, A. P.; Gunasingh, G.; Haass, N. K.; Simpson, M. J. Designing and Interpreting 4D Tumour Spheroid Experiments. *Commun. Biol.* **2022**, *5* (1), 91. https://doi.org/10.1038/s42003-022-03018-3.
(49) Gilazieva, Z.; Ponomarev, A.; Rutland, C.; Rizvanov, A.; Solovyeva, V. Promising Applications of Tumor Spheroids and Organoids for Personalized Medicine. *Cancers* **2020**, *12* (10), 2727. https://doi.org/10.3390/cancers12102727.
(50) Vu, N. B.; Nguyen, M. T.-N. A Simple and Scalable Method to Generate Spheroids from Human Mesenchymal Stem Cells for Use in Tissue Engineering. *Biomed. Res. Ther.* **2020**, *7* (12), 4139–4151. https://doi.org/10.15419/bmrat.v7i12.652.
(51) Vadivelu, R.; Kamble, H.; Shiddiky, M.; Nguyen, N.-T. Microfluidic Technology for the Generation of Cell Spheroids and Their Applications. *Micromachines* **2017**, *8* (4), 94. https://doi.org/10.3390/mi8040094.
(52) Rynes, J.; Istvankova, E.; Dzurov Krafcikova, M.; Luchinat, E.; Barbieri, L.; Banci, L.; Kamarytova, K.; Loja, T.; Fafilek, B.; Rico-Llanos, G.; Krejci, P.; Macurek, L.; Foldynova-Trantirkova, S.; Trantirek, L. Protein Structure and Interactions Elucidated with In-Cell NMR for Different Cell Cycle Phases and in 3D Human Tissue Models. *Commun. Biol.* **2025**, *8* (1), 194. https://doi.org/10.1038/s42003-025-07607-w.
(53) Yao, Q.; Cheng, S.; Pan, Q.; Yu, J.; Cao, G.; Li, L.; Cao, H. Organoids: Development and Applications in Disease Models, Drug Discovery, Precision Medicine, and Regenerative Medicine. *MedComm* **2024**, *5* (10), e735. https://doi.org/10.1002/mco2.735.
(54) Yang, S.; Hu, H.; Kung, H.; Zou, R.; Dai, Y.; Hu, Y.; Wang, T.; Lv, T.; Yu, J.; Li, F. Organoids: The Current Status and Biomedical Applications. *MedComm* **2023**, *4* (3), e274. https://doi.org/10.1002/mco2.274.
(55) Zhao, Z.; Chen, X.; Dowbaj, A. M.; Sljukic, A.; Bratlie, K.; Lin, L.; Li Shan Fong, E.; Balachander, G. M.; Chen, Z.; Soragni, A.; Huch, M.; Zeng, Y. A.; Wang, Q.; Yu, H. Organoids. *Nat. Rev. Methods Primer* **2022**, *2* (1), 94. https://doi.org/10.1038/s43586-022-00174-y.
(56) Moss, S. P.; Bakirci, E.; Feinberg, A. W. Engineering the 3D Structure of Organoids. *Stem Cell Rep.* **2025**, *20* (1), 102379. https://doi.org/10.1016/j.stemcr.2024.11.009.
(57) Rossi, G.; Manfrin, A.; Lutolf, M. P. Progress and Potential in Organoid Research. *Nat. Rev. Genet.* **2018**, *19* (11), 671–687. https://doi.org/10.1038/s41576-018-0051-9.
(58) Mallis, R. J.; Lee, J. J.; Berg, A. V. D.; Brazin, K. N.; Viennet, T.; Zmuda, J.; Cross, M.; Radeva, D.; Rodriguez-Mias, R.; Villén, J.; Gelev, V.; Reinherz, E. L.; Arthanari, H. Efficient and Economic Protein Labeling for NMR in Mammalian Expression Systems: Application to a preT-cell and T-cell Receptor Protein. *Protein Sci.* **2024**, *33* (4), e4950. https://doi.org/10.1002/pro.4950.
(59) Rowlinson, B.; Crublet, E.; Kerfah, R.; Plevin, M. J. Specific Isotopic Labelling and Reverse Labelling for Protein NMR Spectroscopy: Using Metabolic Precursors in Sample Preparation. *Biochem. Soc. Trans.* **2022**, *50* (6), 1555–1567. https://doi.org/10.1042/BST20210586.
(60) Boisbouvier, J.; Kay, L. E. Advanced Isotopic Labeling for the NMR Investigation of Challenging Proteins and Nucleic Acids. *J. Biomol. NMR* **2018**, *71* (3), 115–117. https://doi.org/10.1007/s10858-018-0199-9.
(61) Lacabanne, D.; Meier, B. H.; Böckmann, A. Selective Labeling and Unlabeling Strategies in Protein Solid-State NMR Spectroscopy. *J. Biomol. NMR* **2018**, *71* (3), 141–150. https://doi.org/10.1007/s10858-017-0156-z.
(62) Kerfah, R.; Plevin, M. J.; Sounier, R.; Gans, P.; Boisbouvier, J. Methyl-Specific Isotopic Labeling: A Molecular Tool Box for Solution NMR Studies of Large Proteins. *Curr. Opin. Struct. Biol.* **2015**, *32*, 113–122. https://doi.org/10.1016/j.sbi.2015.03.009.

(63) Sugiki, T.; Furuita, K.; Fujiwara, T.; Kojima, C. Amino Acid Selective 13C Labeling and 13C Scrambling Profile Analysis of Protein α and Side-Chain Carbons in Escherichia Coli Utilized for Protein Nuclear Magnetic Resonance. *Biochemistry* **2018**, *57* (26), 3576–3589. https://doi.org/10.1021/acs.biochem.8b00182.
(64) Bhinderwala, F.; Gronenborn, A. M. Exploiting^{19} F NMR in a Multiplexed Assay for Small GTPase Activity. *J. Am. Chem. Soc.* **2025**, *147* (1), 1028–1033. https://doi.org/10.1021/jacs.4c14294.
(65) Kigawa, T.; Muto, Y.; Yokoyama, S. Cell-Free Synthesis and Amino Acid-Selective Stable Isotope Labeling of Proteins for NMR Analysis. *J. Biomol. NMR* **1995**, *6* (2), 129–134. https://doi.org/10.1007/BF00211776.
(66) Lazarova, M.; Löhr, F.; Rues, R.-B.; Kleebach, R.; Dötsch, V.; Bernhard, F. Precursor-Based Selective Methyl Labeling of Cell-Free Synthesized Proteins. *ACS Chem. Biol.* **2018**, *13* (8), 2170–2178. https://doi.org/10.1021/acschembio.8b00338.
(67) Strauss, A.; Bitsch, F.; Cutting, B.; Fendrich, G.; Graff, P.; Zurini, M.; Jahnke, W. Amino–Acid-Type Selective Isotope Labeling of Proteins Expressed in Baculovirus-Infected Insect Cells Useful for NMR Studies.
(68) Yabuki, T.; Kigawa, T.; Dohmae, N.; Takio, K.; Terada, T.; Ito, Y.; Laue, E. D.; Cooper, J. A.; Kainosho, M. Dual Amino Acid-Selective and Site-Directed Stable-Isotope Labeling of the Human c-Ha-Ras Protein by Cell-Free Synthesis.
(69) Lin, M. T.; Sperling, L. J.; Frericks Schmidt, H. L.; Tang, M.; Samoilova, R. I.; Kumasaka, T.; Iwasaki, T.; Dikanov, S. A.; Rienstra, C. M.; Gennis, R. B. A Rapid and Robust Method for Selective Isotope Labeling of Proteins. *Methods* **2011**, *55* (4), 370–378. https://doi.org/10.1016/j.ymeth.2011.08.019.
(70) Marley, J.; Lu, M.; Bracken, C. A Method for Efficient Isotopic Labeling of Recombinant Proteins.
(71) Lee, K. M.; Androphy, E. J.; Baleja, J. D. A Novel Method for Selective Isotope Labeling of Bacterially Expressed Proteins. *J. Biomol. NMR* **1995**, *5* (1), 93–96. https://doi.org/10.1007/BF00227474.
(72) Reilly, D.; Fairbrother, WayneJ. A Novel Isotope Labeling Protocol for Bacterially Expressed Proteins. *J. Biomol. NMR* **1994**, *4* (3). https://doi.org/10.1007/BF00179352.
(73) Giraud, A.; Imbert, L.; Favier, A.; Henot, F.; Duffieux, F.; Samson, C.; Frances, O.; Crublet, E.; Boisbouvier, J. Enabling Site-Specific NMR Investigations of Therapeutic Fab Using a Cell-Free Based Isotopic Labeling Approach: Application to Anti-LAMP1 Fab. *J. Biomol. NMR* **2024**, *78* (2), 73–86. https://doi.org/10.1007/s10858-023-00433-4.
(74) Pham, T. T.; Kim, J.-Y.; Tuomivaara, S. T.; Lee, Y.-I.; Kim, S.; Wells, L.; Lim, J.-M. Triplex Glycan Quantification by Metabolic Labeling with Isotopically Labeled Glucose in Yeast. *Anal. Chim. Acta* **2024**, *1288*, 342114. https://doi.org/10.1016/j.aca.2023.342114.
(75) Kim, J.-Y.; Joo, W. H.; Shin, D.-S.; Lee, Y.-I.; Teo, C. F.; Lim, J.-M. Metabolic Labeling of Glycans with Isotopic Glucose for Quantitative Glycomics in Yeast. *Anal. Biochem.* **2021**, *621*, 114152. https://doi.org/10.1016/j.ab.2021.114152.
(76) Gossert, A. D.; Hinniger, A.; Gutmann, S.; Jahnke, W.; Strauss, A.; Fernández, C. A Simple Protocol for Amino Acid Type Selective Isotope Labeling in Insect Cells with Improved Yields and High Reproducibility. *J. Biomol. NMR* **2011**, *51* (4), 449–456. https://doi.org/10.1007/s10858-011-9570-9.
(77) Staunton, D.; Schlinkert, R.; Zanetti, G.; Colebrook, S. A.; Campbell, I. D. Cell-Free Expression and Selective Isotope Labelling in Protein NMR. *Magn. Reson. Chem.* **2006**, *44* (S1), S2–S9. https://doi.org/10.1002/mrc.1835.
(78) Morató, A.; Elena-Real, C. A.; Popovic, M.; Fournet, A.; Zhang, K.; Allemand, F.; Sibille, N.; Urbanek, A.; Bernadó, P. Robust Cell-Free Expression of Sub-Pathological and Pathological Huntingtin Exon-1 for NMR Studies. General Approaches for the Isotopic Labeling of Low-Complexity Proteins. *Biomolecules* **2020**, *10* (10), 1458. https://doi.org/10.3390/biom10101458.

(79) Sobhanifar, S.; Reckel, S.; Junge, F.; Schwarz, D.; Kai, L.; Karbyshev, M.; Löhr, F.; Bernhard, F.; Dötsch, V. Cell-Free Expression and Stable Isotope Labelling Strategies for Membrane Proteins. *J. Biomol. NMR* **2010**, *46* (1), 33–43. https://doi.org/10.1007/s10858-009-9364-5.
(80) Schwarz, D.; Junge, F.; Durst, F.; Frölich, N.; Schneider, B.; Reckel, S.; Sobhanifar, S.; Dötsch, V.; Bernhard, F. Preparative Scale Expression of Membrane Proteins in Escherichia Coli-Based Continuous Exchange Cell-Free Systems. *Nat. Protoc.* **2007**, *2* (11), 2945–2957. https://doi.org/10.1038/nprot.2007.426.
(81) Torizawa, T.; Shimizu, M.; Taoka, M.; Miyano, H.; Kainosho, M. Efficient Production of Isotopically Labeled Proteins by Cell-Free Synthesis: A Practical Protocol. *J. Biomol. NMR* **2004**, *30* (3), 311–325. https://doi.org/10.1007/s10858-004-3534-2.
(82) Klammt, C.; Löhr, F.; Schäfer, B.; Haase, W.; Dötsch, V.; Rüterjans, H.; Glaubitz, C.; Bernhard, F. High Level Cell-Free Expression and Specific Labeling of Integral Membrane Proteins. *Eur. J. Biochem.* **2004**, *271* (3), 568–580. https://doi.org/10.1111/j.1432-1033.2003.03959.x.
(83) Subedi, G. P.; Roberts, E. T.; Davis, A. R.; Kremer, P. G.; Amster, I. J.; Barb, A. W. A Comprehensive Assessment of Selective Amino Acid 15N-Labeling in Human Embryonic Kidney 293 Cells for NMR Spectroscopy. *J. Biomol. NMR* **2024**, *78* (2), 125–132. https://doi.org/10.1007/s10858-023-00434-3.
(84) Wang, Q.; Parrish, A. R.; Wang, L. Expanding the Genetic Code for Biological Studies. *Chem. Biol.* **2009**, *16* (3), 323–336. https://doi.org/10.1016/j.chembiol.2009.03.001.
(85) Meineke, B.; Heimgärtner, J.; Eirich, J.; Landreh, M.; Elsässer, S. J. Site-Specific Incorporation of Two ncAAs for Two-Color Bioorthogonal Labeling and Crosslinking of Proteins on Live Mammalian Cells. *Cell Rep.* **2020**, *31* (12), 107811. https://doi.org/10.1016/j.celrep.2020.107811.
(86) Nikić, I.; Kang, J. H.; Girona, G. E.; Aramburu, I. V.; Lemke, E. A. Labeling Proteins on Live Mammalian Cells Using Click Chemistry. *Nat. Protoc.* **2015**, *10* (5), 780–791. https://doi.org/10.1038/nprot.2015.045.
(87) Somogyi, M.; Szimler, T.; Baksa, A.; Végh, B. M.; Bakos, T.; Paréj, K.; Ádám, C.; Zsigmond, Á.; Megyeri, M.; Flachner, B.; Sajó, R.; Gráczer, É.; Závodszky, P.; Hajdú, I.; Beinrohr, L. A Versatile Modular Vector Set for Optimizing Protein Expression among Bacterial, Yeast, Insect and Mammalian Hosts. *PLOS ONE* **2019**, *14* (12), e0227110. https://doi.org/10.1371/journal.pone.0227110.
(88) Kesidis, A.; Depping, P.; Lodé, A.; Vaitsopoulou, A.; Bill, R. M.; Goddard, A. D.; Rothnie, A. J. Expression of Eukaryotic Membrane Proteins in Eukaryotic and Prokaryotic Hosts. *Methods* **2020**, *180*, 3–18. https://doi.org/10.1016/j.ymeth.2020.06.006.
(89) Schörghuber, J.; Geist, L.; Platzer, G.; Feichtinger, M.; Bisaccia, M.; Scheibelberger, L.; Weber, F.; Konrat, R.; Lichtenecker, R. J. Late Metabolic Precursors for Selective Aromatic Residue Labeling. *J. Biomol. NMR* **2018**, *71* (3), 129–140. https://doi.org/10.1007/s10858-018-0188-z.
(90) Schörghuber, J.; Sára, T.; Bisaccia, M.; Schmid, W.; Konrat, R.; Lichtenecker, R. J. Novel Approaches in Selective Tryptophan Isotope Labeling by Using *Escherichia Coli* Overexpression Media. *ChemBioChem* **2015**, *16* (5), 746–751. https://doi.org/10.1002/cbic.201402677.
(91) Miyanoiri, Y.; Takeda, M.; Okuma, K.; Ono, A. M.; Terauchi, T.; Kainosho, M. Differential Isotope-Labeling for Leu and Val Residues in a Protein by E. Coli Cellular Expression Using Stereo-Specifically Methyl Labeled Amino Acids. *J. Biomol. NMR* **2013**, *57* (3), 237–249. https://doi.org/10.1007/s10858-013-9784-0.
(92) Caligiore, F.; Zangelmi, E.; Vetro, C.; Kentache, T.; Dewulf, J. P.; Veiga-da-Cunha, M.; Van Schaftingen, E.; Bommer, G.; Peracchi, A. Human Cytosolic Transaminases: Side Activities and Patterns of Discrimination towards Physiologically Available Alternative

Substrates. *Cell. Mol. Life Sci.* **2022**, *79* (8), 421. https://doi.org/10.1007/s00018-022-04439-3.
(93) McShan, A. C. Utility of Methyl Side Chain Probes for Solution NMR Studies of Large Proteins. *J. Magn. Reson. Open* **2023**, *14–15*, 100087. https://doi.org/10.1016/j.jmro.2022.100087.
(94) Danmaliki, G. I.; Yu, S.; Braun, S.; Zhao, Y. Y.; Moore, J.; Fahlman, R. P.; West, F. G.; Hwang, P. M. Cost-Effective Selective Deuteration of Aromatic Amino Acid Residues Produces Long-Lived Solution 1H NMR Magnetization in Proteins. *J. Magn. Reson.* **2023**, *353*, 107499. https://doi.org/10.1016/j.jmr.2023.107499.
(95) Miyanoiri, Y.; Takeda, M.; Okuma, K.; Terauchi, T.; Kainosho, M. Enhancing Solution Structural Analysis of Large Molecular Proteins through Optimal Stereo Array Isotope Labeling of Aromatic Amino Acids. *Biophys. Chem.* **2024**, *315*, 107328. https://doi.org/10.1016/j.bpc.2024.107328.
(96) Levorin, L.; Becker, N.; Uluca-Yazgi, B.; Gardon, L.; Kraus, M.; Sevenich, M.; Apostolidis, A.; Schmitz, K.; Rüter, N.; Apanasenko, I.; Willbold, D.; Hoyer, W.; Neudecker, P.; Gremer, L.; Heise, H. Isoleucine Side Chains as Reporters of Conformational Freedom in Protein Folding Studied by DNP-Enhanced NMR. *J. Am. Chem. Soc.* **2025**, *147* (18), 15867–15879. https://doi.org/10.1021/jacs.5c04159.
(97) Toscano, G.; Höfurthner, T.; Nagl, B.; Beier, A.; Mayer, M.; Geist, L.; McConnell, D. B.; Weinstabl, H.; Konrat, R.; Lichtenecker, R. 13Cβ-Valine and 13Cγ-Leucine Methine Labeling ToProbe Protein Ligand Interaction. *Eur. J. Chem. Biol.* **2024**, *25* (6). https://doi.org/10.1002/cbic.202300762.
(98) Rodriguez-Mias, R. A.; Pellecchia, M. Use of Selective Trp Side Chain Labeling To Characterize Protein−Protein and Protein−Ligand Interactions by NMR Spectroscopy. *J. Am. Chem. Soc.* **2003**, *125* (10), 2892–2893. https://doi.org/10.1021/ja029221q.
(99) Schütz, S.; Sprangers, R. Methyl TROSY Spectroscopy: A Versatile NMR Approach to Study Challenging Biological Systems. *Prog. Nucl. Magn. Reson. Spectrosc.* **2020**, *116*, 56–84. https://doi.org/10.1016/j.pnmrs.2019.09.004.
(100) Ollerenshaw, J. E.; Tugarinov, V.; Kay, L. E. Methyl TROSY: Explanation and Experimental Verification. *Magn. Reson. Chem.* **2003**, *41* (10), 843–852. https://doi.org/10.1002/mrc.1256.
(101) Korzhnev, D. M.; Kloiber, K.; Kanelis, V.; Tugarinov, V.; Kay, L. E. Probing Slow Dynamics in High Molecular Weight Proteins by Methyl-TROSY NMR Spectroscopy: Application to a 723-Residue Enzyme. *J. Am. Chem. Soc.* **2004**, *126* (12), 3964–3973. https://doi.org/10.1021/ja039587i.
(102) Tugarinov, V.; Kay, L. E. An Isotope Labeling Strategy for Methyl TROSY Spectroscopy. *J. Biomol. NMR* **2004**, *28* (2), 165–172. https://doi.org/10.1023/B:JNMR.0000013824.93994.1f.
(103) Bolik-Coulon, N.; Cousin, S. F.; Kadeřávek, P.; Dumez, J.-N.; Ferrage, F. Understanding the Methyl-TROSY Effect over a Wide Range of Magnetic Fields. *J. Chem. Phys.* **2019**, *150* (22), 224202. https://doi.org/10.1063/1.5095757.
(104) Simplaceanu, V.; Lukin, J. A.; Fang, T.-Y.; Zou, M.; Ho, N. T.; Ho, C. Chain-Selective Isotopic Labeling for NMR Studies of Large Multimeric Proteins: Application to Hemoglobin. *Biophys. J.* **2000**, *79* (2), 1146–1154. https://doi.org/10.1016/S0006-3495(00)76368-5.
(105) Shukla, V. K.; Karunanithy, G.; Vallurupalli, P.; Hansen, D. F. A Combined NMR and Deep Neural Network Approach for Enhancing the Spectral Resolution of Aromatic Side Chains in Proteins. *Sci. Adv.* **2024**, *10* (51), eadr2155. https://doi.org/10.1126/sciadv.adr2155.
(106) Ninković, D. B.; Blagojević Filipović, J. P.; Hall, M. B.; Brothers, E. N.; Zarić, S. D. What Is Special about Aromatic–Aromatic Interactions? Significant Attraction at Large

Horizontal Displacement. *ACS Cent. Sci.* **2020**, *6* (3), 420–425. https://doi.org/10.1021/acscentsci.0c00005.
(107) Serrano, L.; Bycroft, M.; Fersht, A. R. Aromatic-Aromatic Interactions and Protein Stability. *J. Mol. Biol.* **1991**, *218* (2), 465–475. https://doi.org/10.1016/0022-2836(91)90725-L.
(108) Sattler, M.; Fesik, S. W. Use of Deuterium Labeling in NMR: Overcoming a Sizeable Problem. *Structure* **1996**, *4* (11), 1245–1249. https://doi.org/10.1016/S0969-2126(96)00133-5.
(109) Gardner, K. H.; Rosen, M. K.; Kay, L. E. Global Folds of Highly Deuterated, Methyl-Protonated Proteins by Multidimensional NMR. *Biochemistry* **1997**, *36* (6), 1389–1401. https://doi.org/10.1021/bi9624806.
(110) Tugarinov, V.; Kay, L. E. Ile, Leu, and Val Methyl Assignments of the 723-Residue Malate Synthase G Using a New Labeling Strategy and Novel NMR Methods. *J. Am. Chem. Soc.* **2003**, *125* (45), 13868–13878. https://doi.org/10.1021/ja030345s.
(111) Mandal, P. K.; Majumdar, A. A Comprehensive Discussion of HSQC and HMQC Pulse Sequences. *Concepts Magn. Reson. Part A* **2004**, *20A* (1), 1–23. https://doi.org/10.1002/cmr.a.10095.
(112) Vuister, G. W.; Boelens, R.; Kaptein, R.; Hurd, R. E.; John, B.; Van Zijl, P. C. M. Gradient-Enhanced HMQC and HSQC Spectroscopy. Applications to 15N-Labeled Mnt Repressor. *J. Am. Chem. Soc.* **1991**, *113* (25), 9688–9690. https://doi.org/10.1021/ja00025a053.
(113) Bax, A. Two-Dimensional Heteronuclear Relayed Coherence Transfer Spectroscopy. *J. Magn. Reson. 1969* **1983**, *53* (1), 149–153. https://doi.org/10.1016/0022-2364(83)90086-0.
(114) Gal, M.; Schanda, P.; Brutscher, B.; Frydman, L. UltraSOFAST HMQC NMR and the Repetitive Acquisition of 2D Protein Spectra at Hz Rates. *J. Am. Chem. Soc.* **2007**, *129* (5), 1372–1377. https://doi.org/10.1021/ja066915g.
(115) Schanda, P.; Kupče, Ē.; 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. https://doi.org/10.1007/s10858-005-4425-x.
(116) Wang, S.; Parthasarathy, S.; Nishiyama, Y.; Endo, Y.; Nemoto, T.; Yamauchi, K.; Asakura, T.; Takeda, M.; Terauchi, T.; Kainosho, M.; Ishii, Y. Nano-Mole Scale Side-Chain Signal Assignment by 1H-Detected Protein Solid-State NMR by Ultra-Fast Magic-Angle Spinning and Stereo-Array Isotope Labeling. *PLOS ONE* **2015**, *10* (4), e0122714. https://doi.org/10.1371/journal.pone.0122714.
(117) Maiti, S.; Singh, A.; Maji, T.; Saibo, N. V.; De, S. Experimental Methods to Study the Structure and Dynamics of Intrinsically Disordered Regions in Proteins. *Curr. Res. Struct. Biol.* **2024**, *7*, 100138. https://doi.org/10.1016/j.crstbi.2024.100138.
(118) Chhabra, S.; Fischer, P.; Takeuchi, K.; Dubey, A.; Ziarek, J. J.; Boeszoermenyi, A.; Mathieu, D.; Bermel, W.; Davey, N. E.; Wagner, G.; Arthanari, H. ^{15}N Detection Harnesses the Slow Relaxation Property of Nitrogen: Delivering Enhanced Resolution for Intrinsically Disordered Proteins. *Proc. Natl. Acad. Sci.* **2018**, *115* (8). https://doi.org/10.1073/pnas.1717560115.
(119) Bermel, W.; Bruix, M.; Felli, I. C.; Kumar M. V., V.; Pierattelli, R.; Serrano, S. Improving the Chemical Shift Dispersion of Multidimensional NMR Spectra of Intrinsically Disordered Proteins. *J. Biomol. NMR* **2013**, *55* (3), 231–237. https://doi.org/10.1007/s10858-013-9704-3.
(120) Tong, K. I.; Yamamoto, M.; Tanaka, T. A Simple Method for Amino Acid Selective Isotope Labeling of Recombinant Proteins in E. Coli. *J. Biomol. NMR* **2008**, *42* (1), 59–67. https://doi.org/10.1007/s10858-008-9264-0.

(121) Löhr, F.; Gebel, J.; Henrich, E.; Hein, C.; Dötsch, V. Towards Complete Polypeptide Backbone NH Assignment via Combinatorial Labeling. *J. Magn. Reson.* **2019**, *302*, 50–63. https://doi.org/10.1016/j.jmr.2019.03.010.
(122) Takeuchi, K.; Ng, E.; Malia, T. J.; Wagner, G. 1-13C Amino Acid Selective Labeling in a 2H15N Background for NMR Studies of Large Proteins. *J. Biomol. NMR* **2007**, *38* (1), 89–98. https://doi.org/10.1007/s10858-007-9152-z.
(123) Yamazaki, T.; Lee, W.; Revington, M.; Mattiello, D. L.; Dahlquist, F. W.; Arrowsmith, C. H.; Kay, L. E. An HNCA Pulse Scheme for the Backbone Assignment of 15N,13C,2H-Labeled Proteins: Application to a 37-kDa Trp Repressor-DNA Complex. *J. Am. Chem. Soc.* **1994**, *116* (14), 6464–6465. https://doi.org/10.1021/ja00093a069.
(124) Salzmann, M.; Wider, G.; Pervushin, K.; Senn, H.; Wüthrich, K. TROSY-Type Triple-Resonance Experiments for Sequential NMR Assignments of Large Proteins. *J. Am. Chem. Soc.* **1999**, *121* (4), 844–848. https://doi.org/10.1021/ja9834226.
(125) Farmer, B. T.; Venters, R. A.; Spicer, L. D.; Wittekind, M. G.; Müller, L. A Refocused and Optimized HNCA: Increased Sensitivity and Resolution in Large Macromolecules. *J. Biomol. NMR* **1992**, *2* (2), 195–202. https://doi.org/10.1007/BF01875530.
(126) Meissner, A.; Sørensen, O. W. A Sequential HNCA NMR Pulse Sequence for Protein Backbone Assignment. *J. Magn. Reson.* **2001**, *150* (1), 100–104. https://doi.org/10.1006/jmre.2001.2317.
(127) Mäntylahti, S.; Koskela, O.; Jiang, P.; Permi, P. MQ-HNCO-TROSY for the Measurement of Scalar and Residual Dipolar Couplings in Larger Proteins: Application to a 557-Residue IgFLNa16-21. *J. Biomol. NMR* **2010**, *47* (3), 183–194. https://doi.org/10.1007/s10858-010-9422-z.
(128) Hu, K.; Vögeli, B.; Clore, G. M. Spin-State Selective Carbon-Detected HNCO with TROSY Optimization in All Dimensions and Double Echo−Antiecho Sensitivity Enhancement in Both Indirect Dimensions. *J. Am. Chem. Soc.* **2007**, *129* (17), 5484–5491. https://doi.org/10.1021/ja0679811.
(129) Parker, M. J.; Aulton-Jones, M.; Hounslow, A. M.; Craven, C. J. A Combinatorial Selective Labeling Method for the Assignment of Backbone Amide NMR Resonances. *J. Am. Chem. Soc.* **2004**, *126* (16), 5020–5021. https://doi.org/10.1021/ja039601r.
(130) Myshkin, M. Yu.; Dubinnyi, M. A.; Kulbatskii, D. S.; Lyukmanova, E. N.; Kirpichnikov, M. P.; Shenkarev, Z. O. CombLabel: Rational Design of Optimized Sequence-Specific Combinatorial Labeling Schemes. Application to Backbone Assignment of Membrane Proteins with Low Stability. *J. Biomol. NMR* **2019**, *73* (10–11), 531–544. https://doi.org/10.1007/s10858-019-00259-z.
(131) Dey, S.; MacAinsh, M.; Zhou, H.-X. Sequence-Dependent Backbone Dynamics of Intrinsically Disordered Proteins.
(132) Salmon, L.; Nodet, G.; Ozenne, V.; Yin, G.; Jensen, M. R.; Zweckstetter, M.; Blackledge, M. NMR Characterization of Long-Range Order in Intrinsically Disordered Proteins. *J. Am. Chem. Soc.* **2010**, *132* (24), 8407–8418. https://doi.org/10.1021/ja101645g.
(133) 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), e4910. https://doi.org/10.1002/pro.4910.
(134) Hanson, G. S. M.; Coxon, C. R. Fluorinated Tags to Study Protein Conformation and Interactions Using^{19} F NMR. *ChemBioChem* **2024**, *25* (15), e202400195. https://doi.org/10.1002/cbic.202400195.
(135) Werle, Y.; Kovermann, M. Fluorine Labeling and^{19} F NMR Spectroscopy to Study Biological Molecules and Molecular Complexes. *Chem. – Eur. J.* **2025**, *31* (2), e202402820. https://doi.org/10.1002/chem.202402820.
(136) Galles, G. D.; Infield, D. T.; Clark, C. J.; Hemshorn, M. L.; Manikandan, S.; Fazan, F.; Rasouli, A.; Tajkhorshid, E.; Galpin, J. D.; Cooley, R. B.; Mehl, R. A.; Ahern, C. A. Tuning

Phenylalanine Fluorination to Assess Aromatic Contributions to Protein Function and Stability in Cells. *Nat. Commun.* **2023**, *14* (1), 59. https://doi.org/10.1038/s41467-022-35761-w.

(137) Mishra, N. K.; Urick, A. K.; Ember, S. W. J.; Schönbrunn, E.; Pomerantz, W. C. Fluorinated Aromatic Amino Acids Are Sensitive^{19} F NMR Probes for Bromodomain-Ligand Interactions. *ACS Chem. Biol.* **2014**, *9* (12), 2755–2760. https://doi.org/10.1021/cb5007344.

(138) Hammill, J. T.; Miyake-Stoner, S.; Hazen, J. L.; Jackson, J. C.; Mehl, R. A. Preparation of Site-Specifically Labeled Fluorinated Proteins for 19F-NMR Structural Characterization. *Nat. Protoc.* **2007**, *2* (10), 2601–2607. https://doi.org/10.1038/nprot.2007.379.

(139) Costello, A.; Peterson, A. A.; Chen, P.-H.; Bagirzadeh, R.; Lanster, D. L.; Badran, A. H. Genetic Code Expansion History and Modern Innovations. *Chem. Rev.* **2024**, *124* (21), 11962–12005. https://doi.org/10.1021/acs.chemrev.4c00275.

(140) Qianzhu, H.; Tan, Y. J.; Abdelkader, E. H.; Huber, T.; Otting, G. Genetic Encoding of Fluorinated Analogues of Phenylalanine for^{19} F NMR Spectroscopy: Detection of Conformational Heterogeneity in Flaviviral NS2B-NS3 Proteases. *ACS Sens.* **2025**, *10* (4), 3152–3161. https://doi.org/10.1021/acssensors.5c00432.

(141) Orton, H. W.; Qianzhu, H.; Abdelkader, E. H.; Habel, E. I.; Tan, Y. J.; Frkic, R. L.; Huber, T.; Otting, G. Through-Space Scalar 19F–19F Couplings between Fluorinated Non-Canonical Amino Acids for the Detection of Specific Contacts in Proteins.

(142) Maleckis, A.; Abdelkader, E. H.; Herath, I. D.; Otting, G. Synthesis of Fluorinated Leucines, Valines and Alanines for Use in Protein NMR. *Org. Biomol. Chem.* **2022**, *20* (12), 2424–2432. https://doi.org/10.1039/D2OB00145D.

(143) Adan, A.; Alizada, G.; Kiraz, Y.; Baran, Y.; Nalbant, A. Flow Cytometry: Basic Principles and Applications. *Crit. Rev. Biotechnol.* **2017**, *37* (2), 163–176. https://doi.org/10.3109/07388551.2015.1128876.

(144) Brown, M.; Wittwer, C. Flow Cytometry: Principles and Clinical Applications in Hematology. *Clin. Chem.* **2000**, *46* (8), 1221–1229. https://doi.org/10.1093/clinchem/46.8.1221.

(145) McKinnon, K. M. Flow Cytometry: An Overview. *Curr. Protoc. Immunol.* **2018**, *120* (1). https://doi.org/10.1002/cpim.40.

(146) Song, Y.; Lee, Y. Brief Guide to Flow Cytometry. *Mol. Cells* **2024**, *47* (11), 100129. https://doi.org/10.1016/j.mocell.2024.100129.

(147) Cai, B.; Ding, J.; Yalikun, Y.; Ren, D. Progress of Cell Sorting in Flow Cytometry. *iLABMED* **2025**, *3* (1), 96–105. https://doi.org/10.1002/ila2.70000.

(148) Schraivogel, D.; Steinmetz, L. M. Cell Sorters See Things More Clearly Now. *Mol. Syst. Biol.* **2023**, *19* (3), e11254. https://doi.org/10.15252/msb.202211254.

(149) Telford, W. G. Flow Cytometry and Cell Sorting. *Front. Med.* **2023**, *10*, 1287884. https://doi.org/10.3389/fmed.2023.1287884.

(150) Herzenberg, L. A.; Parks, D.; Sahaf, B.; Perez, O.; Roederer, M.; Herzenberg, L. A. The History and Future of the Fluorescence Activated Cell Sorter and Flow Cytometry: A View from Stanford. *Clin. Chem.* **2002**, *48* (10), 1819–1827. https://doi.org/10.1093/clinchem/48.10.1819.

(151) Datta, R.; Heaster, T. M.; Sharick, J. T.; Gillette, A. A.; Skala, M. C. Fluorescence Lifetime Imaging Microscopy: Fundamentals and Advances in Instrumentation, Analysis, and Applications. *J. Biomed. Opt.* **2020**, *25* (07), 1. https://doi.org/10.1117/1.JBO.25.7.071203.

(152) Hickey, S. M.; Ung, B.; Bader, C.; Brooks, R.; Lazniewska, J.; Johnson, I. R. D.; Sorvina, A.; Logan, J.; Martini, C.; Moore, C. R.; Karageorgos, L.; Sweetman, M. J.; Brooks, D. A. Fluorescence Microscopy—An Outline of Hardware, Biological Handling,

and Fluorophore Considerations. *Cells* **2021**, *11* (1), 35. https://doi.org/10.3390/cells11010035.
(153) Ishikawa-Ankerhold, H. C.; Ankerhold, R.; Drummen, G. P. C. Advanced Fluorescence Microscopy Techniques—FRAP, FLIP, FLAP, FRET and FLIM. *Molecules* **2012**, *17* (4), 4047–4132. https://doi.org/10.3390/molecules17044047.
(154) Lavrentovich, O. D. Confocal Fluorescence Microscopy. In *Characterization of Materials*; Kaufmann, E. N., Ed.; Wiley, 2012; pp 1–15. https://doi.org/10.1002/0471266965.com127.
(155) Fazel, M.; Grussmayer, K. S.; Ferdman, B.; Radenovic, A.; Shechtman, Y.; Enderlein, J.; Pressé, S. Fluorescence Microscopy: A Statistics-Optics Perspective. *Rev. Mod. Phys.* **2024**, *96* (2), 025003. https://doi.org/10.1103/RevModPhys.96.025003.
(156) Light Sheet Fluorescence Microscopy. *Nat. Rev. Methods Primer* **2021**, *1* (1), 74. https://doi.org/10.1038/s43586-021-00077-4.
(157) Thiele, J. C.; Helmerich, D. A.; Oleksiievets, N.; Tsukanov, R.; Butkevich, E.; Sauer, M.; Nevskyi, O.; Enderlein, J. Confocal Fluorescence-Lifetime Single-Molecule Localization Microscopy. *ACS Nano* **2020**, *14* (10), 14190–14200. https://doi.org/10.1021/acsnano.0c07322.
(158) Milles, S.; Salvi, N.; Blackledge, M.; Jensen, M. R. Characterization of Intrinsically Disordered Proteins and Their Dynamic Complexes: From in Vitro to Cell-like Environments. *Prog. Nucl. Magn. Reson. Spectrosc.* **2018**, *109*, 79–100. https://doi.org/10.1016/j.pnmrs.2018.07.001.
(159) Delaforge, E.; Kragelj, J.; Tengo, L.; Palencia, A.; Milles, S.; Bouvignies, G.; Salvi, N.; Blackledge, M.; Jensen, M. R. Deciphering the Dynamic Interaction Profile of an Intrinsically Disordered Protein by NMR Exchange Spectroscopy. *J. Am. Chem. Soc.* **2018**, *140* (3), 1148–1158. https://doi.org/10.1021/jacs.7b12407.
(160) Schneider, R.; Blackledge, M.; Jensen, M. R. Elucidating Binding Mechanisms and Dynamics of Intrinsically Disordered Protein Complexes Using NMR Spectroscopy. *Curr. Opin. Struct. Biol.* **2019**, *54*, 10–18. https://doi.org/10.1016/j.sbi.2018.09.007.
(161) Bhattacharya, S.; Lin, X. Recent Advances in Computational Protocols Addressing Intrinsically Disordered Proteins. *Biomolecules* **2019**, *9* (4), 146. https://doi.org/10.3390/biom9040146.
(162) Charlier, C.; Bouvignies, G.; Pelupessy, P.; Walrant, A.; Marquant, R.; Kozlov, M.; De Ioannes, P.; Bolik-Coulon, N.; Sagan, S.; Cortes, P.; Aggarwal, A. K.; Carlier, L.; Ferrage, F. Structure and Dynamics of an Intrinsically Disordered Protein Region That Partially Folds upon Binding by Chemical-Exchange NMR. *J. Am. Chem. Soc.* **2017**, *139* (35), 12219–12227. https://doi.org/10.1021/jacs.7b05823.
(163) Vagrys, D.; Davidson, J.; Chen, I.; Hubbard, R. E.; Davis, B. Exploring IDP–Ligand Interactions: Tau K18 as a Test Case. *Int. J. Mol. Sci.* **2020**, *21* (15), 5257. https://doi.org/10.3390/ijms21155257.
(164) Naudi-Fabra, S.; Blackledge, M.; Milles, S. Synergies of Single Molecule Fluorescence and NMR for the Study of Intrinsically Disordered Proteins. *Biomolecules* **2021**, *12* (1), 27. https://doi.org/10.3390/biom12010027.
(165) Narod, S. A.; Foulkes, W. D. BRCA1 and BRCA2: 1994 and Beyond. *Nat. Rev. Cancer* **2004**, *4* (9), 665–676. https://doi.org/10.1038/nrc1431.
(166) Mark, W.-Y.; Liao, J. C. C.; Lu, Y.; Ayed, A.; Laister, R.; Szymczyna, B.; Chakrabartty, A.; Arrowsmith, C. H. Characterization of Segments from the Central Region of BRCA1: An Intrinsically Disordered Scaffold for Multiple Protein–Protein and Protein–DNA Interactions? *J. Mol. Biol.* **2005**, *345* (2), 275–287. https://doi.org/10.1016/j.jmb.2004.10.045.
(167) Yadav, L. R.; Rai, S.; Hosur, M. V.; Varma, A. K. Functional Assessment of Intrinsic Disorder Central Domains of BRCA1. *J. Biomol. Struct. Dyn.* **2015**, *33* (11), 2469–2478. https://doi.org/10.1080/07391102.2014.1000973.

(168) Górka, M.; Żerko, S.; Konrat, R.; Koźmiński, W.; Kurzbach, D. 1H, 13C and 15N Backbone Resonance Assignment of BRCA1 Fragment 219–504. *Biomol. NMR Assign.* **2020**, *14* (2), 289–293. https://doi.org/10.1007/s12104-020-09963-6.
(169) Epasto, L.; Pötzl, C.; Peterlik, H.; Khalil, M.; Saint-Pierre, C.; Gasparutto, D.; Sicoli, G.; Kurzbach, D. Protein Science - 2023 - Epasto - NMR-identification of the Interaction between BRCA1 and the Intrinsically Disordered (1).Pdf. *Protein Sci.* **2024**, *33* (1). https://doi.org/10.1002/pro.4849.
(170) Che, K.; Kress, T.; Górka, M.; Żerko, S.; Kozminski, W.; Kurzbach, D. Coupled MD Simulations and NMR Reveal That the Intrinsically Disordered Domain of the Breast-Cancer Susceptibility 1 Protein (BRCA1) Binds Head-on to DNA Double-Strand Ends. *J. Magn. Reson. Open* **2022**, *12–13*, 100069. https://doi.org/10.1016/j.jmro.2022.100069.
(171) Witus, S. R.; Tuttle, L. M.; Li, W.; Zelter, A.; Wang, M.; Kermoade, K. E.; Wilburn, D. B.; Davis, T. N.; Brzovic, P. S.; Zhao, W.; Klevit, R. E. BRCA1 / BARD1 Intrinsically Disordered Regions Facilitate Chromatin Recruitment and Ubiquitylation. *EMBO J.* **2023**, *42* (15), e113565. https://doi.org/10.15252/embj.2023113565.
(172) Wang, Q.; Zhang, H.; Kajino, K.; Greene, M. I. BRCA1 Binds C-Myc and Inhibits Its Transcriptional and Transforming Activity in Cells. *Oncogene* **1998**, *17* (15), 1939–1948. https://doi.org/10.1038/sj.onc.1202403.
(173) Blackwood, E. M.; Eisenman, R. N. Max: A Helix-Loop-Helix Zipper Protein That Forms a Sequence-Specific DNA-Binding Complex with Myc. *Sci. New Ser.* **1991**, *251* (4998), 1211–1217.
(174) Savage, K. I.; Harkin, D. P. BRCA 1, a 'Complex' Protein Involved in the Maintenance of Genomic Stability. *FEBS J.* **2015**, *282* (4), 630–646. https://doi.org/10.1111/febs.13150.
(175) Livernois, A. M.; Hnatchuk, D. J.; Findlater, E. E.; Graether, S. P. Obtaining Highly Purified Intrinsically Disordered Protein by Boiling Lysis and Single Step Ion Exchange. *Anal. Biochem.* **2009**, *392* (1), 70–76. https://doi.org/10.1016/j.ab.2009.05.023.
(176) Romero-Pérez, S. P.; Covarrubias, A. A.; Campos, F. A Simple Method to Purify Intrinsically Disordered Proteins by Adjusting Trichloroacetic Acid Concentration. *Protein Expr. Purif.* **2023**, *202*, 106183. https://doi.org/10.1016/j.pep.2022.106183.
(177) Graether, S. P. Troubleshooting Guide to Expressing Intrinsically Disordered Proteins for Use in NMR Experiments. *Front. Mol. Biosci.* **2019**, *5*, 118. https://doi.org/10.3389/fmolb.2018.00118.
(178) Rößler, P.; Mathieu, D.; Gossert, A. D. Enabling NMR Studies of High Molecular Weight Systems Without the Need for Deuteration: The XL-ALSOFAST Experiment with Delayed Decoupling. *Angew. Chem. Int. Ed.* **2020**, *59* (43), 19329–19337. https://doi.org/10.1002/anie.202007715.
(179) Chérot, H.; Pred'homme, T.; Theillet, F.-X. In-Cell Residue-Resolved NMR of Micromolar α-Synuclein and Tau at 310K. October 8, 2024. https://doi.org/10.1101/2024.10.07.617093.
(180) Herrera-Nieto, P.; Pérez, A.; Fabritiis, G. D. Binding-and-Folding Recognition of an Intrinsically Disordered Protein Using Online Learning Molecular Dynamics. arXiv February 20, 2023. https://doi.org/10.48550/arXiv.2302.10348.
(181) Zavrtanik, U.; Medved, T.; Purič, S.; Vranken, W.; Lah, J.; Hadži, S. Leucine Motifs Stabilize Residual Helical Structure in Disordered Proteins. *J. Mol. Biol.* **2024**, *436* (4), 168444. https://doi.org/10.1016/j.jmb.2024.168444.
(182) Supuran, C. T. Emerging Role of Carbonic Anhydrase Inhibitors. *Clin. Sci.* **2021**, *135* (10), 1233–1249. https://doi.org/10.1042/CS20210040.
(183) Supuran, C. T. Structure and Function of Carbonic Anhydrases. *Biochem. J.* **2016**, *473* (14), 2023–2032. https://doi.org/10.1042/BCJ20160115.
(184) Supuran, C. T. Carbonic Anhydrases: Novel Therapeutic Applications for Inhibitors and Activators. *Nat. Rev. Drug Discov.* **2008**, *7* (2), 168–181. https://doi.org/10.1038/nrd2467.

(185) Nocentini, A.; Supuran, C. T. Advances in the Structural Annotation of Human Carbonic Anhydrases and Impact on Future Drug Discovery. *Expert Opin. Drug Discov.* **2019**, *14* (11), 1175–1197. https://doi.org/10.1080/17460441.2019.1651289.
(186) McDonald, P. C.; Chia, S.; Bedard, P. L.; Chu, Q.; Lyle, M.; Tang, L.; Singh, M.; Zhang, Z.; Supuran, C. T.; Renouf, D. J.; Dedhar, S. A Phase 1 Study of SLC-0111, a Novel Inhibitor of Carbonic Anhydrase IX, in Patients With Advanced Solid Tumors. *Am. J. Clin. Oncol.* **2020**, *43* (7), 484–490. https://doi.org/10.1097/COC.0000000000000691.
(187) 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.